Erkenn (2016) 81:1211–1241
DOI 10.1007/s10670-015-9791-5

ORIGINAL ARTICLE

Explication as a Method of Conceptual Re-engineering

Georg Brun[1]

Received: 23 September 2014 / Accepted: 11 November 2015 / Published online: 14 December 2015

Abstract Taking Carnap's classic exposition as a starting point, this paper develops a pragmatic account of the method of explication, defends it against a range of challenges and proposes a detailed recipe for the practice of explicating. It is then argued that confusions are involved in characterizing explications as definitions, and in advocating precising definitions as an alternative to explications. Explication is better characterized as conceptual re-engineering for theoretical purposes, in contrast to conceptual re-engineering for other purposes and improving exactness for purely practical reasons. Finally, three limitations which call for further development of the method of explication are discussed.

1 Introduction

Explication in Carnap's sense is a method of re-engineering concepts with the aim of advancing theory: a concept is replaced by an explicitly characterized 'new' concept which can be used in place of the 'old' concept in relevant contexts but proves advantageous in respects such as being more exact, fruitful, simple or precise.[1] Examples of explications abound in science and philosophy. Mineralogists follow Mohs in explicating *hardness* with reference to scratching tests, economists explicate familiar notions such as *work* or *poverty,* epistemologists develop justified-true-belief accounts that explicate *knowledge,* logical theories offer

[1] Many other uses of "explication" can be found in the literature. In the tradition of Kant, for example, explication is frequently understood as a less ambitious alternative to a proper definition. I will not discuss such uses here. See Beaney (2004) for a historical study of Carnap's notion of explication in relation to Frege's, Kant's and Husserl's.

✉ Georg Brun
Georg.Brun@philo.unibe.ch

[1] Institute for Philosophy, University of Bern, Länggassstrasse 49a, 3012 Bern, Switzerland

equivalent as well as alternative explications of the concept *valid inference*, and in ethics, concepts such as *justice, blame, risk* and *promise* are explicated.[2]

Interest in explication has recently seen a revival. Discussions range from Maher's (2006, 2007, 2010) resurrection of Carnap's project of explicating probability to Kitcher's (2008) appeal to explication in the philosophy of biology and Justus and Shepherd's (Justus 2012; Shepherd and Justus 2015) defence of explication as a philosophical method. Many studies (see, e.g. Wagner 2012; Richardson 2013) have been sparked by Carus's (2007a) thesis that explication is not only the key to understanding Carnap's philosophical projects after *Logical Syntax of Language,* but that the "the ideal of explication" is the heart of an attractive philosophical programme in the spirit of the Enlightenment. Most contributions to the recent discussion about explication focus on Carnap's philosophical programme, specifically the principle of tolerance, the distinction between internal and external questions, and pragmatic tendencies in Carnap's later philosophy. Surprisingly little effort, however, has been put into analysing the details of Carnap's explicit accounts of the method of explication, and attempts at developing the method further have been scarce. Besides Hanna (1968), who is critical of Carnap's ideas, Siegwart (1997a, b) and Greimann (2007) are most notable.[3] The lack of attention is surprising, for one thing, because it is widely known that explication plays a central role for Carnap and other philosophers. Hempel, for example, conceived of his theory of explanation as a project of explication (see 1970: 489), and Quine defended explication as a methodological cornerstone of philosophy in *Word and Object* (1960: § 53). However, philosophers who explicitly appeal to explication are only the tip of the iceberg. Explications are simply omnipresent in science and philosophy. One reason why explications often go unnoticed might be that they are frequently identified not as explications but as so-called precising definitions; that is, definitions which simultaneously aim at capturing core features of the established use of a term and at giving it a more precise meaning.

The two main goals of this paper are to work out a sufficiently detailed account of the method of explication and to analyse its relation to definitions and, more generally, to conceptual re-engineering, by which I understand methods of concept formation that introduce a concept which is both similar to an existing concept and useful for some given purpose. More specifically, I aim to suggest an account of the method of explication which is Carnapian in spirit but not tied to his specific philosophical projects of semantics and inductive logic. Rather, it should be, as Carnap himself thought (1963: 933–939), available to a broad range of philosophical programmes. This goal is a motivation for focusing on Carnap's explicit discussions of explication rather than analysing his overarching philosophical programme or his specific projects. I will therefore not try to 'reverse engineer' the

[2] It might be argued that these examples should be interpreted as instances of some other method such as conceptual analysis or definition. In Sect. 4, I will discuss some arguments which tell against interpreting them as definitions, but defending the view that they are best interpreted as explications would require case studies, which cannot be undertaken here.

[3] Developments less directly relevant to the project of this paper include Martin's (1973) proposal to extend the method from concepts to theories, and Justus and Shepherd's methodological work (Justus 2012; Shepherd and Justus 2015), which focuses on combining explication with experimental philosophy, and on the potential uses of these methods.

method of explication from Carnap's actual practice of explication. Such an approach would seek to extract general features of the method of explication from, for example, Carnap's work on inductive logic. Doing so would, firstly, run the risk of incorporating into the resulting method specific features of Carnap's work which he did not consider to be general characteristics of explication (e.g. that explications *always* use formal methods; see Sect. 2.2). Secondly, it may also easily give the misleading impression that Carnap had given explicit methodological explanations he in fact did not provide (e.g. on how to explicate entire systems of concepts in contrast to individual concepts; see Sect. 5.2).

As a starting point for developing an account of explication, Sect. 2.1 presents the classic exposition from *Logical Foundations of Probability* (Carnap 1962). The way Carnap explains his ideas, however, also raises a number of questions of interpretation and systematic problems. In my analysis, I emphasize pragmatic aspects of Carnap's thinking and show how his later writings (1963) can be used as a guide for developing an account of explication that resolves these difficulties. In Sects. 2.2–2.3, this interpretation of Carnap's ideas is developed and contrasted with alternative readings of Carnap; in Sect. 2.4 my interpretative strategy is briefly discussed. Section 3 then suggests a recipe for developing explications, which summarizes, completes and presents the results of the preceding sections in a systematic way.

Section 4 contrasts explications with definitions and argues two points. First, although explications often include a definition, there are no explications which *are* definitions (Sect. 4.1). Second, the strategy of using precising definitions as an alternative to explications fails because the idea of definitions which are a compromise between reportive and stipulative definitions is not viable (Sect. 4.2). Yet the comparison of explications with definitions also shows that there are a number of related forms of conceptual re-engineering, which share explication's structure but pursue different aims and hence are subject to different criteria of adequacy (Sect. 4.3).

Section 5, finally, discusses some of the method's limitations. The method presupposes an account of the relation between different systems of concepts, which may in turn require a theory of formalization (Sect. 5.1); explication focuses on individual concepts although it has aspects which can only be described systematically in a wider context of theory development (Sect. 5.2); and the method has basically a linear structure although the process of explicating may involve various feedback effects (Sect. 5.3).

2 A Carnapian Account of Explication

The two most important places where Carnap explicitly explains the method of explication are his classic exposition in §§ 2–3 of *Logical Foundations of Probability* (1962; henceforth "LFP")[4] and the explanations and clarifications he

[4] Other relevant passages in LFP are referenced in what follows. The remarks on explication which can be found in Carnap's other writings of the 1940s and 50s (esp. 1945; 1947; 1953; 1956; Creath 1990) are in line with the account in LFP. "Explication" first appears in Carnap (1945: 513); earlier, Carnap used

later included in his replies to Strawson and Goodman in the Schilpp-volume (Carnap 1963; henceforth "RSE").[5] In this section, I roughly follow this historical development. Section 2.1 outlines Carnap's classic account from LFP and gives some examples of explications in philosophy. Unclear and controversial points are then discussed in detail in Sects. 2.2 and 2.3. On the surface, Carnap treatment of explication in RSE differs in many respects from the one he gives in LFP. As I will argue, however, Carnap's later explanations in RSE are best read as a guide to a charitable (re)interpretation of his exposition in LFP. Section 2.4 briefly reflects on this interpretative strategy.

2.1 Carnap's Classic Exposition of Explication

Carnap's basic idea is that explication is a process which replaces an inexact concept (the explicandum) with a more exact concept (the explicatum); this process serves some theoretical purpose and explicitly introduces the explicatum into the system of concepts of a target theory.

One of Carnap's standard examples is the everyday concept *fish*[6] (the explicandum), which is perfectly appropriate for fishmongers, but not for the purpose of biological theory, where it was replaced by the concept *piscis* (the explicatum), which is characterized in terms of biological concepts as (let us assume for the sake of simplicity) "cold-blooded aquatic vertebrate" (LFP §3). Further examples are the everyday concepts *cold, warm, hot* etc., which for scientific purposes are replaced by a quantitative concept of temperature (LFP §§4–5). It is important to note that explications are frequently used not only in science, but also in philosophy. Carnap specifically refers to Tarski's definition of truth, Frege's explications of arithmetical terms, and his own explications of, for example, *probability*, *estimate, logical truth* and *entailment.* These examples might suggest that explication is a method of formal philosophy. But in fact, explications can also be found in great numbers in canonical texts and in entirely non-technical writings in, for example, ethics[7]:

> (1) "*Having an opinion* is taking something to be true with the consciousness that it is subjectively *as well as* objectively insufficient. If taking something to be true is only subjectively sufficient and is at the same time held to be objectively insufficient, then it is called *believing*. Finally, when taking something to be true is both subjectively and objectively sufficient it is called *knowing*. Subjective sufficiency is called *conviction* (for myself), objective sufficiency, *certainty* (for everyone). I will not pause for the exposition of such

Footnote 4 continued
other terms such as "systematization" (in a letter to Neurath, 29.1.1943; see Hegselmann 1985: 283); see also footnote 47.

[5] RSE was written in the mid fifties and completed by October 1958 (Creath 1990: 449).

[6] Italics signal that "fish" does not refer to the animals labelled by the word "fish" but to the concept of being a fish.

[7] Carnap alludes to this in when he points out that philosophers often neglect the clarification of explicanda such as *causality, life, mind* and *justice* (LFP 4).

readily grasped concepts." (Kant 1998: A 822/B 850) This quote presents five explications.

(2) "To claim that a person is *blameworthy* for an action is to claim that the action shows something about the agent's attitudes toward others that impairs the relations that others can have with him or her. […] To *blame* a person is to judge him or her to be blameworthy and to take your relationship with him or her to be modified in a way that this judgment of impaired relations holds to be appropriate." (Scanlon 2008: 128–129) In this example, the explication of *blameworthy* is the basis for the explication of *blame*.

Carnap discusses the method of explication from two perspectives. One deals with the method as a procedure; the other focuses on explication as a result and specifically on the characteristics of adequate explications.

The process of explicating has two phases. Firstly, the explicandum must be identified as clearly as possible. If the explicandum-term is ambiguous, it must be disambiguated. As the explicandum is inexact, it cannot be defined exactly, but has to be characterized informally, for example, by describing cases in which the explicandum clearly does or does not apply. This clarification of the explicandum will not yet use the resources of the target system of concepts.[8] Secondly, an explicatum must be introduced. This requires explicitly specifying rules for using the explicatum in terms of the target system of concepts. Ideally, the explicatum can be defined, but less strict methods of concept introduction are permissible as well.

Since the explicandum is inexact, there is not one correct explicatum, but several more or less adequate explicata are possible. Adequacy must be assessed in light of the role the explicatum is expected to play in the target theory. Carnap specifies four criteria of adequacy. They are all a matter of degree, and an explicatum counts as adequate just in case it meets these criteria to a sufficient degree. The first criterion requires that the explicatum has a certain similarity to the explicandum; that is, it must be possible to use the explicatum instead of the explicandum in relevant contexts, but differences are permitted. For example, many explicata have a wider range of application because they cover extreme cases (e.g. zero velocity) or provide additional discriminations (e.g. 0.04 °F warmer). Reclassification of pre-theoretically clear cases is also possible (e.g. *piscis* excludes whales). Second, the explicatum must be more exact than the explicandum. Third, the explicatum should be as fruitful as possible; that is, it should be possible to formulate many laws or generalizations featuring the explicatum. If two explications both meet these criteria satisfactorily, simplicity can be used to select one. This forth criterion includes simple rules for using the explicatum as well as the simplicity of the laws which include the explicatum.

[8] That this is Carnap's view is confirmed by his emphasis that "the explicandum cannot be given in exact terms" (LFP 4) and that "Naturally, such an elucidation [of the explicandum] can be rendered only in terms that are themselves not yet exact." (1990: 430; with reference to LFP §2). I pick up on this point in Sect. 5.3.

2.2 The Structure of Explications

Several aspects of Carnap's characterization of the method of explication raise questions of interpretation or have sparked criticism. Some of these issues have their roots in tensions or perhaps even in inconsistencies in Carnap's writings, which I propose to resolve with the help of Carnap's own clarifications in RSE. This section provides a systematic discussion of questions directly relating to the basic structure of explications; Sect. 2.3 addresses issues concerning the criteria of adequacy. Both sections are organized by focus questions which have been controversially discussed in the literature.

Does explication deal with terms or with concepts? Carnap seems to say that we can have it both ways: "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.*" (LFP 3) But later on, he emphasizes that a "genuine explication" must give us not only a term but also specify its meaning (LFP 16–17). Hence, genuine explicata cannot simply be terms but must involve a concept since in Carnap's terminology, terms are words or phrases that have as their meaning a concept, which is a non-linguistic, abstract entity, namely a property, relation or function (LFP 7–8).[9]

With respect to the explicandum, however, there seem to be problems with both options. Carnapian meanings are not ambiguous; terms are ambiguous if they are used with more than one meaning. Therefore, the first step of an explication, the clarification of the explicandum, must be understood as dealing with terms. But explicanda cannot simply be terms either since clarifying the explicandum calls for selecting one meaning of a given term and this implies that a concept is identified, not merely a term.[10] These arguments show that "explicandum" in fact covers two different items. The potentially ambiguous starting point of an explication is a term, but the unambiguous result of the first step, which the explicatum replaces, is a concept.

Independently, we face the question of whether the method of explication should really be tied to Carnap's theory of concepts. After all, it seems possible and desirable that the method of explication be available for those who do not accept Carnap's semantics. But there is no real problem here: although Carnap insists that explications need to specify meanings, his discussion of explication nowhere substantially relies on his particular theory of meaning. Instead, he often simply speaks of using a term in a particular way.[11] Taking up this clue, I suggest to free the method of explication from its association with Carnap's theory of concepts by adopting an alternative terminology. In what follows, I use "concept" to refer to an

[9] To simplify, I limit the discussion to explications of concepts that involve *n*-ary predicates. I leave open how the method can be extended to other categories of expressions [Carnap (1956: 8) mentions, e.g. the definite descriptor *the*; cf. Siegwart 1997a: 27].

[10] Maher (2010) argues this point against Quine (1960) and Hanna (1968).

[11] Maher (2010: 17) comments: "to talk of a term 'as used in' a particular way is just a misleading way of talking about a concept". As an argument against Carnap's position in LFP, this misses the point since there Carnap holds that concepts are abstract entities, not terms used in a certain way. The latter, however, reflects Carnap's use of "concept" in some earlier writings (e.g. Carnap 1966: 3–4).

elementary linguistic entity, a "term", together with rules for its use. So identity of concepts requires identity of the term and identity of the rules for its use. I take a neutral stance on the nature of such rules; they may specify a term's intension or extension; they may be stated explicitly or be given implicitly in usage, fairly clearly or rather turbidly. I also leave open how such rules are related to mental or abstract entities which are often called "concepts". We now can say that the clarification of an explicandum (typically) deals with an ambiguous term, the explicandum-term, identifies one way of using it and hence identifies a concept, the explicandum. The explicatum is a concept as well, although a different one, even if the same term is employed for both.

Does explication aim at clarifying everyday concepts for scientific purposes? Many explications Carnap mentions as examples have this aim, not only *fish* and *warm*, but also Frege's definition of numbers and Tarski's definition of truth. Furthermore, Carnap often characterizes the explicandum as "pre-scientific" and the explicatum "scientific". All that might be taken as evidence that explication is tailored to the conceptual needs of science, logics and mathematics in contrast to philosophy (see, e.g., Strawson 1963; Leitgeb 2013). Such an interpretation, however, misconstrues Carnap's position. Three points need to be noted about Carnap's use of "(pre-)scientific". Firstly, not only everyday concepts may need explication but also concepts already in scientific use (LFP 3). Secondly, there is no clear-cut distinction between scientific and pre-scientific concepts but a more or less continuous transition (RSE 634). Thirdly, "scientific" is not meant to draw a contrast to philosophy. On the contrary, Carnap holds that explication plays a major role in philosophy (RSE 917, 933–939). To avoid misunderstandings, it may be better to say that explicanda belong to a "pre-theoretical" system of concepts, where "pre-" is to be understood not in any absolute sense, but as relative to the target theory into which the explicata are supposed to be introduced.

Must the target system of concepts be framed in a formal or ideal language? A positive answer is given or assumed by a large number of commentators, both approving and critical ones (to name a few: Greimann 2007; Putnam 1997; Reck 2012; Siegwart 1997b[12]; Strawson 1963). However, as Carnap explicitly says (RSE 936, 937) and as his example of *fish* clearly shows, he does not require the target system of concepts to be framed in a formalism or some kind of ideal language.

Explications which draw on formal languages are an important paradigm nonetheless for two reasons. They exemplify the possibility of introducing explicata which exhibit a maximum of exactness, and they show that we generally need to distinguish between two systems of concepts involved in an explication. This distinction is obvious for theories which use terms or symbols not to be found in the vernacular (e.g. if "$\models$" explicates "implies"). But if a project of explication is carried out in ordinary English, it may be tempting to speak of expanding English rather than to distinguish two systems of concepts. However, drawing such a distinction is advisable nonetheless. Even philosophical theories entirely framed in terms of ordinary English typically develop a complex network of concepts which are the result of explicating ordinary language concepts. Scanlon's *Moral*

[12] Siegwart notes that his interpretation targets LFP but not RSE.

Dimensions (2008), for example, is devoted (inter alia) to developing specific concepts of *blame, meaning,* and of what is permissible (cf. quote 2). Of course, a great deal of the claims defended in such a philosophical work are to be understood as framed not in plain everyday-English concepts, but in the concepts introduced by explications. Hence there is a tacit distinction between two systems of concepts at work.[13] In any case, the risk of running into conceptual muddles and misunderstandings is much higher for those who insist that the same system of concepts includes a more and a less exact version of the same concept, which, in the terminology introduced above, amounts to using the same term for two concepts, one more, one less exact.

What is the nature of the transition from explicandum to explicatum? Carnap points out that explication is more precisely characterized as a process of "replacing" rather than "transforming" or "making more exact" an explicandum (LFP 3, 5; 1956: 7–8). Two reasons speak in favour of speaking of replacing. Firstly, explicandum and explicatum are different concepts, even if the same term is employed for both, since the rules for using the term(s) are different. Hence, speaking of "transforming" or "making more exact" a concept would be misleading since it would suggest that although the explicatum is more exact than the explicandum, it is still the same concept. Secondly, Carnap writes: "The explicatum is intended to take the place of the explicandum, and that means, of course, that it is to be used for the same purpose as the explicandum." (RSE 936) So speaking of "replacing" is appropriate since, for those purposes which motivate the explication, the explicatum literally takes the place of the explicandum in the sense that the explicatum rather than the explicandum is used.

Nevertheless, speaking of replacing may lead to misunderstandings and objections. Firstly, replacing an explicandum by an explicatum does not mean that the explicatum-term will be plugged into sentences in which the explicandum-term occurred so far. That would often result in a 'hybrid' of two systems of concepts which may be hard to interpret (cf. Siegwart 1997b: 271). Secondly, that explications replace concepts does not imply that every explication has to coin a new term. Most explications use the explicandum-term for the explicatum as well. This creates potential ambiguities, but it is a problem only in instances in which we have no explicit or implicit clue whether the term belongs to the explicandum or to the explicatum. Thirdly, speaking of replacing does not imply that the explicatum can replace the explicandum in all contexts. This is clearly not the case since the two concepts are different (LFP 6). Neither does it imply that an explication renders the explicandum completely useless. Explications are undertaken for specific purposes (RSE 935–938). The biological concept *piscis* is not intended to replace the everyday concept *fish* at the fish market.

Finally, there is the objection that replacing concepts results in changing the subject.[14] The worry is that instead of suggesting a suitable modification,

[13] This is not to say that there is a tacit distinction between two languages. I will not discuss criteria of identity for languages here, but simply assume that in general the target system of concepts can, but need not belong to the same language as the explicandum.

[14] See Justus (2012) for a more extensive discussion.

explications simply introduce a new, different, concept and in this sense change the subject. A harmless version of this objection adduces examples of a stark difference in extension and then points to contexts in which we want to insist on much closer similarity. As an example we may take biological explications of *berry* which exclude blackberries and strawberries, but include aubergines and bananas.[15] In a culinary context, such explications seem inappropriate because they simply do not even approximately deal with what we count as berries in this context; from the culinary perspective, the biologist has just changed the subject. However, such examples do not raise a fundamental problem. They just show that an explication is inadequate if the explicatum is not similar enough to the explicandum in order to take its place for the purpose at hand. To prevent that, explications should be guided by conditions of similarity which specify what deviations from the use of the explicandum are acceptable for the task of explication at hand. This takes into account that "changing the subject" is a notion which has to be detailed on a case-by-case basis.

The subject-change challenge has a valid point, however, in contexts in which the explicandum itself is the subject of enquiry, for example, in Strawson's philosophical programme (1963). Here, the aim is to describe how the explicandum-term is actually used and to analyse the problems that result from certain ways of using it. Such an analysis, in turn, may be the basis for avoiding the identified problems by refraining from certain uses of the explicandum. Explication is clearly not designed to implement such a project. It is rather an alternative which, once a concept is identified as creating problems, aims at solving those problems by developing a concept which no longer gives rise to them. As Maher (2007) has argued, explications can be helpful to Strawsonian projects nonetheless by uncovering ambiguities and suggesting ways of arguing for or against claims involving the explicandum.

Must the explicatum be introduced by a definition? Carnap explicitly denies this (RSE 935; cf. LFP 3;) and allows introducing explicata by other methods such as postulates or reduction sentences (cf. Carnap 2003: viii–ix). Hempel (1970: 489), for example, points out that he explicates *explanation* without giving a full definition. Methods other than definitions are needed, at any rate, for explicating concepts which are basic in the target theory.

The results of this section are summarized in Fig. 1, which visualizes in a simplified manner the explication of *fish* by *piscis* with the help of a definition. Note that in other cases the explicatum may be characterized not by a definition, but by some other method.

2.3 The Adequacy of Explications

The discussion of Carnap's criteria of adequacy in this section begins with similarity and exactness and then turns to additional virtues such as fruitfulness and simplicity.

What does Carnap mean by "similarity"? In LFP, the criterion of similarity is: "in most cases in which the explicandum has so far been used, the explicatum can

[15] See, e.g., Singh (2010: 82).

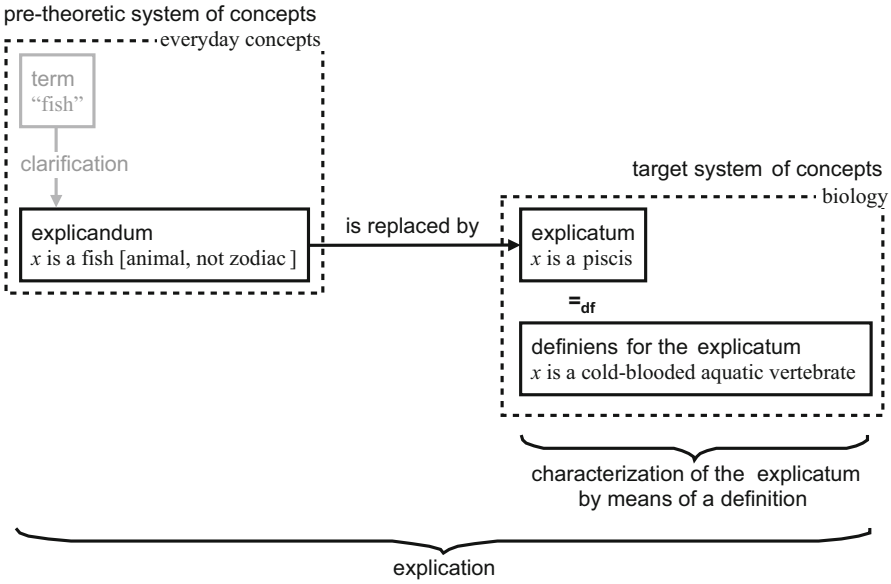

Fig. 1 Introducing an explicatum by giving a definition

be used; however close similarity is not required, and considerable differences are permitted" (LFP 7). "Most" makes clear that similarity does not require that the explicatum can always replace the explicandum. But there are two problems here. "Most cases" may also be misleading since what counts for Carnap is obviously not the sheer number of cases but that the explicatum can be used in *relevant* cases. More importantly, Carnap's criterion has been interpreted in different ways, for example, by Quine, who offered two influential readings. In *Two Dogmas* (1980: 25), he championed a semantic interpretation and claimed that Carnap's criterion requires synonymy in "favoured contexts". In *Word and Object* (1960: 258–259), he switched to a more pragmatic reading and argued that similarity requires only that in relevant contexts the explicatum can be used in place of the explicandum. This second reading leaves room for interpreting the similarity criterion differently in different situations; which contexts are relevant depends on the purpose the explicatum is supposed to serve as part of the target system of concepts. Clearly, Carnap has the second reading in mind, not the first, and in RSE (936, 938), he uses formulations closely resembling Quine's. This interpretation is vindicated by the following more specific considerations (see also Kitcher 2008).

Can the similarity condition be stated in terms of extensions? The strongest contention which still allows for reducing vagueness is that similarity requires extensional identity for all non-borderline cases; that is, the explicatum must clearly (not) apply to everything the explicandum clearly does (not) apply to (e.g. Hanna

1968).[16] However, Carnap explicitly denies that adjustments of the extension are *always* limited to borderline cases (LFP 5). If this were so, *piscis* would not qualify as an adequate explication of *fish* since whales are clear-cut cases of fish (cf. Lutz 2012).

In his discussion of Hempel's explication of *confirming evidence*, Carnap argues: "The result that a proposed explicatum is found too narrow constitutes a much less serious objection than the result that it is too wide." (LFP 479) This is in line with the example of *piscis.* So the condition of similarity might be interpreted as a negative extensional constraint, requiring that the explicatum should not cover the cases to which the explicandum does not apply (e.g. Hempel 2000: 207). But Carnap explicitly rejects this as well and points out that the explicatum is often wider than the explicandum (LFP 11, 13). Explicata which cover extreme cases (e.g. zero velocity) are a case in point, and if a comparative concept is explicated by a quantitative concept that allows more fine-grained distinctions, the explicatum will cover more cases than the explicandum. As these examples show, Carnap appeals to the merits of both narrower and wider explicata.[17] There is thus no general rule and we need to decide on a case-by-case basis what similarity requires.

The points discussed so far still leave room for a minimal extensional requirement: whatever the criterion of similarity exactly amounts to in the explication-task at hand, the extensions of the explicandum and the explicatum must at least overlap. Though even this claim has to be rejected. In the following passage, for example, Stalnaker's offers two alternative explicata for *proposition:*

> (3) "the *proposition* [a sentence expresses] will be a function taking possible worlds into truth values. Equivalently, a proposition may be thought of as a set of possible worlds […]". (Stalnaker 1976: 80)

These two explicata have disjoint extensions because no function *is* a set of possible worlds. Assuming that it is not the case that some propositions are functions while others are sets of possible worlds, this also means that at least one of the explicata is extensionally disjoint with the explicandum. A similar point can be made with respect to different explications of the number *two,* which Carnap (1956: 8) mentions as examples. Whatever the extension of *two* may be, it cannot be both Zermelo's {{Ø}} and von Neumann's {Ø, {Ø}}. Similarity therefore cannot, in general, be specified with reference to extensions[18] but must be characterized by a

[16] Carnap (LFP 7) reports that Menger endorses such a criterion of similarity for definitions. It is clear from the context, that Carnap does not approve of Menger's criterion but rather of the fact that Menger states an explicit criterion (contra Hanna 1968: 33).

Hanna's criterion is proposed as an alternative, not as an interpretation of Carnap. Bizarrely, Hanna's official definition (1968: 37–38) does not require that everything clearly (not) in the explicandum's extension is (not) in the extension of the explicatum because he allows for the explicandum and the explicatum to have different domains linked by an "effective mapping" (which need not be 1–1). Hence explicandum and explicatum may even have disjoint extensions.

[17] Hanna (1968: 41) concludes that Carnap's position in LFP is inconsistent. I suggest a more charitable reading which resolves the tensions in line with Carnap's explanations in RSE (which Hanna does not mention).

[18] Another argument is due to Goodman, who was the first to note explicitly that similarity does not require overlapping extensions (Goodman 1977: 5–7). Carnap accepts this diagnosis in RSE (945) and

specification of the contexts in which and the purposes for which the explicatum can replace the explicandum; that is, perform the explicandum's function.

Is synonymy admissible as a criterion of similarity? In RSE (945), Carnap emphasizes that different situations may call for different standards of similarity. As examples, he mentions synonymy, extensional equivalence and a weaker requirement of extensional isomorphism proposed by Goodman. That extensional equivalence or even synonymy can be required may come as a surprise because these requirements exclude reclassifications (such as excluding the whales when explicating *fish* by *piscis*) and imply that the explicatum will be just as vague or exact as the explicandum. This raises the question of what can be accomplished by such an explication. Several points can be argued for. In addition to the clarification of the explicandum, we will have an explicit characterization of the explicatum which incorporates it into the target system of concepts.[19] Thereby, relations to other concepts of the target system are established and the explicatum becomes available for the formulation of generalizations. In this way, the explicatum may prove fruitful. Simplicity can be enhanced as well, for example, if an explication can be found which reduces the conceptual basis of the theory (e.g. when an explication is given for a concept that hitherto was taken as basic). So it seems reasonable to consider extensional equivalence and synonymy as extreme, but still legitimate forms of similarity requirements which lead to limiting cases of explication.

What does Carnap mean by "exactness"? Carnap's use of "exact" and "precise" in fact covers a whole range of qualities of an explication. The rules for using the explicatum-term must be formulated explicitly in terms of the target system (LFP 7), they must eliminate ambiguity (LFP 4), they must not lead to paradoxes or contradictions (RSE 935), and they should allow in as many cases as possible a clear decision on whether the explicatum does or does not apply; that is, the explicatum should be less—or at least not more—vague than the explicandum (LFP 5; more on that below). Additionally, a concept may be considered to be more exact if it allows more precise descriptions and finer discriminations, as quantitative concepts do in comparison with qualitative and comparative concepts (LFP 13).

I suggest disentangling these points as follows. Formulating the rules for using the explicatum-term explicitly in terms of the target system is just a necessary condition for giving an explication. That the rules are unambiguous[20] and do not lead to paradoxes or contradictions are then necessary conditions of adequacy.

Footnote 18 continued
acknowledges that in some situations similarity may be captured, as Goodman suggests, by a certain kind of isomorphism, which admits that explicandum and explicatum may have disjoint extensions. However, Carnap's acceptance of Goodman's isomorphism does not withstand closer analysis because Goodman's proposal cannot be interpreted as a similarity condition fitting into Carnap's method of explication. Goodman's criterion is not meant to be applied to individual concepts but to entire systems of concepts (Goodman 1963: 556; Goodman 1977: 16).

[19] Carnap emphasizes this point: "The new definitions should be superior to the old in clarity and exactness, and, *above all,* should fit into a systematic structure of concepts." (2003: v; italics GB). His account of Frege's explication of the number *two* points in the same direction. Frege's achievement was not so much reducing the vagueness of *two,* but giving a definition in logical terms (RSE 935; cf. LFP 17).

[20] This includes that the logical form of the explicatum needs to be determined (cf. Hempel 1952: 12–14).

Explicata meeting these conditions are called "unambiguous" and "consistent".[21] In contrast to Carnap's four officially listed criteria, these two conditions do not admit of trade-offs; they are not even a matter of degree. They are necessary conditions in the sense that no ambiguous or inconsistent explicatum counts as adequate. Another aspect of adequacy is what Carnap officially lists as his criterion number two (LFP 5, 7), which I interpret as requiring that the explicatum is not more vague than the explicandum.[22] For this, I use "exact", whereas "precise" is reserved for the precision and discriminating power of comparative and quantitative concepts. As I argue below, precision is best dealt with together with fruitfulness and simplicity.

Is being more exact a necessary condition for adequate explications? Carnap answers in the affirmative: "the explicandum is more or less vague and certainly more so than the explicatum" (LFP 5) and "The only essential requirement is that the explicatum be more precise than the explicandum." (RSE 936). His further explanations make it clear that how much reduction of vagueness is required also depends on the target system of concepts. Carnap's constructed languages are designed to be free of vagueness as much as possible, but less ambitious explications are acceptable as well (RSE 936–937). We need not ascribe to Carnap the dubious view that explicata must be 'absolutely' exact.

Virtually all of Carnap's interpreters hold that explications have to reduce vagueness (an exception is Sjögren 2011). However, there are two reasons why I think this interpretation should be given up even in the face of Carnap's explicit remarks just quoted. Firstly, Carnap's requirement of reducing vagueness is incompatible with his claim (discussed above) that similarity sometimes calls for extensional equivalence or even synonymy. More importantly, Carnap's example *piscis* is incompatible with reducing vagueness as a necessary condition of adequacy. What recommends *piscis* as an explicatum is its adjusted extension, which makes it a much more fruitful concept than *fish,* even if we assume that *piscis* is exactly as vague or exact as *fish*. So if we accept Carnap's example of *piscis,* but want to retain a criterion of exactness, the only plausible requirement is the minimal condition that the explicatum must not be more vague than the explicandum.[23]

This requirement of exactness is a non-trivial feature of explication. It distinguishes explication from other forms of concept change and conceptual re-engineering (see Sect. 4.3). Exactness is also not implied by other aspects of the method. Specifically, an explicatum can be simpler and more fruitful in the sense of admitting more generalizations without being more exact.

Obviously, *exact* and *vague* are themselves candidates for explication. Not only is the nature of vagueness debated, but the method of explication raises challenging questions about vagueness. It presupposes that we can decide whether a proposed explicatum is less exact than the explicandum even if the explication involves reclassification or the explicatum is considerably wider or narrower than the

[21] For theories of inconsistent concepts, see Eklund (2014).

[22] The distinction between consistency and exactness in the context of explications (though not the terminology) is due to Hanna (1968).

[23] Contra LFP 5, this leaves open the possibility that the explicandum is not vague at all.

explicandum. In short, a theory of vagueness is needed for a more elaborate account of explication (see Sorensen 2013 for an overview).

Which further virtues should an adequate explication have? Carnap officially lists fruitfulness and simplicity. "Fruitful" is explained as "useful for the formulation of many universal statements" (LFP 7), which include "empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept" (LFP 7). Since Carnap emphasizes that explications play an important role in philosophy and need not employ a formal language (RSE 936), universal statements adding to fruitfulness may also be, for example, normative statements (e.g. an explication of *valid argument*) and may include informally specified explicata.

On simplicity, Carnap says very little, but two points are clear. He refers not to ontological parsimony, but to what is sometimes called "syntactic simplicity" (Baker 2013), specifically to "the simplicity of the form of [the explicatum's] definition" and "the simplicity of the forms of the laws connecting it with other concepts" (LFP 7). He considers simplicity to be relatively unimportant, but relevant if two explicata fare approximately equally well with respect to similarity, exactness and fruitfulness. Meanwhile, there has been an extensive discussion of various notions of simplicity and their measurement (see, e.g. Scorzato 2013 and the references given there), and some have also argued that simplicity plays not a secondary but a pivotal role in theory assessment (e.g. Goodman 1972).

However, in assessing the adequacy of explications we certainly need to account for more than fruitfulness and simplicity (besides similarity, consistency and exactness). Carnap considers comparative and quantitative concepts to be frequently "superior" or "more powerful" because "they enable us to give a more precise description of a concrete situation and, more important, to formulate more comprehensive general laws" (LFP 12, 13). Hence Carnap considers increased precision and wider scope of the resulting theory to be desirable results of explications. Frequently, quantitative concepts are also more fruitful, but this need not be the case. It might not be possible to formulate more laws with the quantitative concept (LFP 13–15). In light of subsequent discussions about 'virtues' of theories and theory choice (mainly sparked by Kuhn 1977), more desiderata come to mind (see Douglas 2013 for a short survey). Explications may also contribute to the explanatory power of a theory, to its ability to be used for predicting novel phenomena or to more specific virtues such as visualizability (see de Regt and Dieks 2005) or the possibility to use it as an effective means to decide on practical problems.

The discussions about theory choice and various theoretical virtues cannot be rehearsed here, but a few consequences can be drawn immediately. Firstly, Carnap's four criteria need to be supplemented with additional aspects of theoretical usefulness and we cannot simply assume that fruitfulness trumps all other such aspects. Secondly, the adequacy of an explicatum depends to a large extent on criteria which primarily apply not to the explicatum taken in isolation, but to the theory the explicatum is part of (see Sect. 5.2). And we have to accept that we have neither exact formulations for many of these criteria nor exact rules for balancing the various criteria against each other or for determining whether a criterion is met

to a sufficient degree. Assessing the overall adequacy of an explicatum calls for judgement in light of its role in the target theory. Such judgements are context-dependent and, in general, it cannot be expected that one explicatum can be singled out as outperforming all competitors (see, e.g. the case-studies in Justus 2012). This situation may seem unsatisfactory, but it is a general problem of theory assessment, not a specific weakness of a Carnapian account of explication.

2.4 The Development of Carnap's Views on Explication

A more general pattern can be discerned in many of the points discussed in the preceding two sections. In his later explanations in RSE, Carnap takes a decidedly more pragmatic perspective on explication and underlines that the method should not be interpreted too restrictively. He holds that consistency and exactness are the only indispensable requirements an explicatum must meet, apart from similarity in the sense of serving some purpose the explicandum was used for; and he emphasizes that explicating primarily calls for giving explicit rules for the use of the explicatum-term in a way that eliminates some problems of using the explicandum (RSE 936–937). The issue at hand decides how much weight is to be given to criteria such as exactness, fruitfulness and simplicity, though Carnap hardly mentions specific criteria any more and rather speaks more generally of clarity and aiming to solve philosophical or scientific problems. This is a more pragmatic conception of explication than the one we find in LFP because Carnap emphasizes much more the following two points. Firstly, the requirements an adequate explication has to meet cannot be specified in general but only with respect to a specific task of explication. Secondly, choosing an adequate explicatum is a practical decision which has to be taken in view of the specific problems the explicatum is expected to solve and in view of the role it is expected to play in the target theory.

In effect, Carnap's position in RSE comes very close to Quine's account in *Word and Object* (1960: § 53): explication is a method of supplying theories with concepts that take over certain useful functions of the explicandum without having some of its defects. What counts as useful and what as a defect is a pragmatic question insofar as it has to be answered in the context of specific problems and with reference to a target theory. All that should not be read as claiming that explication is a method that serves whatever goals a scientist or philosopher happens to have. Rather, Carnap and Quine acknowledge that paradox, vagueness, ambiguity, sterility and being too complicated are just some of the most common problems standing in the way of scientific and philosophical understanding. Other problems, such as appealing to suspicious entities, may be more prominent in other cases (as in Quine's (1960: § 53) discussion of explications of *the ordered pair* a *and* b, which do not appeal to a third entity besides *a* and *b*).[24]

[24] Carus (2007a: 265–267) argues for a profound difference between Quine's and Carnap's account of explication. However, he "takes a broader view of explication than is usual in the literature" (2007a: 21n23) and uses "explication" to refer to the philosophical programme of Carnap after the introduction of the principle of tolerance. From this perspective, he argues for example that explication is an "external matter" for Carnap, whereas Quine restricts it to internal questions and uses it as a tool for his eliminativist ontological programme. Other writers argue that the similarity between Carnap's and

An interpretation of Carnap which emphasizes pragmatic aspects of explicating has a firm basis in RSE. But since there are also apparent differences to the account of explication in LFP, we face the question of whether we should interpret Carnap as advocating a new position in RSE or merely as clarifying what he wrote in LFP. This paper adopts the second interpretation, which is more charitable because there are, independently of what Carnap writes in RSE, good reasons to (re)interpret the ideas of LFP in line with his explanations in RSE. As Sects. 2.2 and 2.3 have shown, there are tensions within LFP, especially between Carnap's explicit account of explication and the key examples he uses to explain this account. Carnap's discussion in RSE can be read as resolving these tensions. He relaxes his official account and instead takes examples more seriously, emphasizes philosophical uses of explication and frees the method of explication from a too close alliance with formal methods. Of course, the tensions in LFP could also be resolved in the other direction, but doing so has a price: one needs to reject Carnap's example of *fish* and discard his later explanations in RSE as misguided.[25] This makes for a rather uncharitable reading of Carnap and the resulting notion of explication loses much of its attraction since its range of applicability would be relatively narrow.

3 A Recipe for Explicating

If we accept the interpretation of Carnap's method of explication developed in the preceding sections, we cannot simply stick to Carnap's explicit account in LFP §§ 2–3. As an alternative description, I propose a general 'recipe' for explicating, which captures the thrust of Carnap's ideas read from the perspective of RSE.[26] The recipe distinguishes four main steps. Although they may be followed roughly in this order, some steps are best dealt with together and quite often, going back and forth will be needed as well (see Sect. 5.3).

1. *Identification of the task.*
   (a) *Clarification of the explicandum.* Which way of using the explicandum-term is relevant for the explication? This question calls for disambiguating the explicandum-term, as well as specifying its relevant use, logical

Footnote 24 continued
Quine's explanations of the method of explication is misleading since the two *in fact* put the method to different uses. Gustafsson (2014), e.g. argues that Quine's explications aim at eliminating ontological commitments, whereas Carnap's aim at incorporating a concept into a formal framework. However, these points do not directly bear on the approach of this paper, which takes a narrower perspective and discusses explication as a method of conceptual re-engineering available to different philosophical programmes. That Carnap and Quine had similar views on the workings of method of explication is compatible with their making different use of it.

[25] Somewhat surprisingly, the literature shows little awareness of the tensions within LFP and between LFP and RSE (exceptions include Hanna 1968; Justus 2012; Siegwart 1997b; Sjögren 2011). It is, for example, hardly ever mentioned that requiring explications to reduce vagueness or to use a formal target language is incompatible with the example of *fish* and *piscis*.

[26] The recipe also incorporates ideas from Greimann (2007) and Siegwart (1997a). For a simplified version see Brun and Hirsch Hadorn (2014).

form and range of meaningful application. Often one finds that an entire family of concepts are in need of explication (e.g. predicates with different numbers of arguments).

(b) *Theoretical framework.* Into which system of concepts is the explicatum to be integrated?

(c) *Theoretical purpose the concept is expected to serve.* Which functions should the explicatum fulfil in the target theory? This includes the functions of the explicandum that are to be preserved as well as the problems the explicatum is expected to solve.

2. *Specification of conditions of adequacy.*

(a) *Conditions of similarity.* In which contexts should it be possible to use the explicatum in place of the explicandum? Similarity conditions serve to ensure that certain functions of the explicandum are preserved (see (1.c) above) and that the explication does not change the subject.[27] Three common types of similarity conditions are:

(i) *Positive and negative examples.* Are there (actual or hypothetical) cases where the explicatum is clearly expected (not) to apply?

(ii) *Structural properties of the explicatum.* What logical form should the explicatum have?[28] Should it have logical properties such as being reflexive, transitive or symmetric? Should it be classificatory, comparative or quantitative?

(iii) *Relations to other concepts.* Is the explicatum expected to be, for example, a hyponym of some other concept?

(b) *Conditions of theoretical usefulness.* In addition to unambiguity and consistency, Carnap's exactness (the explicatum is not more vague than the explicandum), fruitfulness (the explicatum allows for generalizations) and simplicity are criteria relevant to all theories. Depending on the purpose the theory is supposed to serve, an explicatum's theoretical usefulness also depends on further virtues such as precision, measurability, scope of application, explanatory power and other properties of the target theory. Finally, the conditions of theoretical usefulness include the functions the explicatum is expected to fulfil and the problems it should help to solve, as specified under (1.c).

[27] In its most general form, a similarity condition is a statement involving the explicatum that we expect to have a certain truth value. Contra Siegwart (1997a: 32), similarity conditions need not be statements which are framed in terms of the target system and use the explicatum; they can also be metalanguage statements which mention the explicatum (e.g. stating that the explicatum should be an n-place predicate).

[28] The explicatum need not have the same logical form as the explicandum. It may have more argument places if, for example, a comparative concept is explicated by a quantitative one. However, I cannot think of good reasons for introducing an explicatum with less argument places than the explicandum.

Some conditions of adequacy may be understood as absolute requirements for any adequate explication (e.g. unambiguity, consistency and possibly some of the conditions of similarity[29]). Of the remaining conditions, some may be expected to be met to a satisfactory degree, others are taken to specify a property that we are trying to maximize or that we will only use to select one from a set of otherwise equally adequate explicata.

3. *Introduction of the explicatum.* A term must be selected and the rules for its use need to be laid down explicitly in terms of the system of concepts identified in (1.b), either by a definition or by another method of introducing a concept.
4. *Assessment of adequacy.* Checking whether the explicatum introduced in (3) meets the conditions specified in (2) cannot be done in a mechanical way, but is subject to informal evaluation and judgement. There are various reasons for that. Some criteria (e.g. simplicity) are open to interpretation and not defined exactly (they are the subject of further projects of explication). Many criteria admit for degrees, but there is no effective procedure for assessing how well they are met. And it is usually not possible to specify in an exact yet non-arbitrary way the relative weight of the various criteria, the admissible trade-offs between them and what is required for them to be met to a sufficient degree (e.g. which price in terms of similarity is acceptable in exchange for improved fruitfulness?). In any case, the adequacy of an explicatum must be assessed in the context of the target theory and with respect to the purpose the explicatum is supposed to serve (1.c). This perspective usually helps to provide at least some boundary conditions for the decisions that need to be taken.[30]

As an example, let us see how Mohs mineralogical hardness scale be interpreted as providing an explication of *harder*.[31]

1. *Identification of the task.* Mohs hardness scale deals with *hardness*—more exactly, with x *is harder than* y—as a property of minerals[32] that makes them resistant to scratching but not necessarily to other forms of deformation such as bending or breaking. The explicatum is to be defined in mineralogical terms. Its purpose is to yield an ordered classification that is robust, easy to use and

[29] Contra Reck (2012: 103), this makes room for conclusive counterexamples. In the example given below: any explication of *hardness* which deems diamonds to be soft has failed.

[30] Applying this recipe to a specific problem of explication is a creative task, but it can be facilitated by various practical strategies. The most important is to learn from previous or rival attempts at explicating the same or related explicanda. Studying the history of explications may reveal, for example, important ambiguities in the explicandum-term, merits or demerits of different ways of introducing an explicatum, or that certain conditions of adequacy may not be satisfiable (see Carnap 1945: 516–521; Carus 2007a: 270; Hahn 2013: ch. 2.3.2–7; Siegwart 1997a: 33–34).

[31] See Schumann (2008: 20).

[32] Technically, the range of meaningful application (here: minerals) can be captured in two ways. Either the explicatum is introduced by a conditional definition limited to that range (e.g. If x and y are minerals, then: x is harder than y iff …). Or it is introduced by a non-conditional definition with a definiens which includes a condition ensuring that predicating the explicandum is trivially false for all objects outside the range of meaningful application (e.g. x is harder than y iff x and y are minerals and …).

relevant to the identification of minerals. Closely related concepts in need of explication include first of all x *is equally hard as* y, but also x *is softer than* y and maybe x *is hard* and x *is soft*.

2. *Conditions of adequacy.* (a) An adequate explicatum for *harder* must meet the following *conditions of similarity:* (i) *Positive and negative examples:* diamonds should come out as harder than most other minerals, talcum as not hard at all. (ii) *Structural properties:* the explicatum for *harder* should be a comparative two-place-predicate that is transitive and asymmetric. (iii) *Relations to other concepts:* if *x* is harder than *y*, then *x* scratches *y; x* is harder than *y* iff *y* is softer than *x*. (b) The *conditions of theoretical usefulness* are: Comprehensive scope of application (all minerals) takes precedence over precision, explanatory power is not expected, but the explicatum should allow an identification of minerals that is robust and easy to use (see 1).
3. *Introduction of the explicatum.* x *is harder than* y is characterized as x *scratches* y *leaving a visible mark*.
4. *Assessment of adequacy.* The explication fares very well with respect to the criteria mentioned under (2). Its minor problems with robustness (some minerals exhibit different resistance to scratching on different faces or in different directions) are offset by its excellent performance with respect to the other criteria (scope of application and ease of practical use).

The account of explication developed so far can now be compared with other forms of concept formation.

4 Explication, Definition and Conceptual Re-engineering

Explications are regularly presented as definitions (as in Tarski's definition of truth[33]), the methodological literature frequently classifies explication as a "mixture" or "compromise" between reportive and stipulative definitions (Sober 2000: 244; see also Quine 1980: 25–27; Belnap 1993: 116), and some writers seem to hold that many purposes are better served by so-called precising definitions than by explications. An analysis of similarities and differences between definitions and explications shows that these views are highly misleading or simply mistaken (Sects. 4.1–4.2). The comparison also shows that besides explication there are other forms of conceptual re-engineering tailored to different aims (Sect. 4.3).

4.1 Explication Versus Definition

Explications and definitions differ in structure. Definitions are given within one system of concepts, explications involve two systems of concepts (at least in paradigmatic cases). More decisively, where a definition has two main parts, definiendum and definiens, an explication has three main parts: explicandum,

[33] Even Carnap (1945: 513) spoke of "redefining an old [concept]" in his first explicit explanation of explication.

explicatum and a characterization of the explicatum.[34] The latter may give a definition, but it need not do so. If it gives a definition, the explicatum plays the role of the definiendum, not of the definiens (see Fig. 1). For these reasons, Carnap decides against calling the explicatum misleadingly "explicans" (LFP 3; others follow Reichenbach (1951) and Quine (1960) in using "explicans" nonetheless). In short, an explication can include a definition, but an explication is not a definition.

Definitions and explications also have different criteria of adequacy. For definitions, we need to distinguish (non-exclusively) between reportive and stipulative definitions. They have the same form but different goals and hence need to be evaluated in different ways. Reportive definitions aim at capturing the actual usage of the definiendum-term; they are true or false, depending on whether definiendum and definiens are (at least extensionally[35]) equivalent. Stipulative definitions express a decision to use a term in a certain way; they establish equivalence and may be judged as more or less useful.[36] Purely reportive definitions can be found in dictionaries. Purely stipulative definitions are used, for example, for coining new terms (e.g. "veganism"[37]) or for establishing an entirely new use for an established term (e.g. the mathematical use of "ring" introduced by Hilbert[38]). More frequently, purely stipulative definitions are part of some form of conceptual re-engineering. Before we turn to that below, it is worth noting that a definition is both reportive and stipulative if it is a decision to use a term in a specific, already established way, which it simultaneously characterizes (if, e.g. somebody defines "I follow Frege in using 'name' for all singular terms", this definition will be adequate only if it correctly reports Frege's use of "name" and proves useful for the purpose at hand).

The adequacy of explications, on the other hand, is a matter of the criteria discussed in Sect. 2.3. The condition of similarity is responsible for a crucial difference between explications and definitions. For purely stipulative definitions, there is no similarity requirement and in an adequate reportive definition, definiendum and definiens need to have (at least) the same extension. In any case, definiendum and definiens of an adequate definition are equivalent. But the similarity requirement for explications cannot in general be explained in terms of the extensions of explicandum and explicatum.

This raises the question of what the criteria of adequacy are for a definition that is part of an explication. Such a definition is purely stipulative; it is adequate just in case it is useful for the explication at hand, which in turn is the case iff the resulting explication is adequate. Consequently, similarity of explicandum and explicatum is necessary for the adequacy of the proposed definition. However, this does not turn

[34] I leave aside here the explicandum-term and the expression connecting definiendum and definiens (e.g. "$=_{df}$", or "is defined as") as additional parts of explications and definitions.

[35] The following arguments are independent of whether a stronger form of equivalence is required for true reportive definitions.

[36] A note on terminology: my use of "stipulative" neither implies that a new term is introduced nor that the definition is not aimed at capturing an established use of the definiendum-term. Definitions with the former property I call "inventive", those with the second "purely stipulative".

[37] See *The Oxford English Dictionary*. 2nd ed. 1986. Oxford: Oxford University Press, entry "vegan".

[38] Hilbert (1894/1895).

such definitions into quasi-reportive ones since the relevant similarity is not a relation between definiendum and definiens; in this respect equivalence is established by the definition that characterizes the explicatum. In the example of *fish*, a stipulative definition establishes the equivalence of x *is a piscis* with x *is a cold-blooded aquatic vertebrate*. This definition is adequate iff *piscis* is an adequate explicatum for *fish,* which requires *inter alia* that *piscis* is sufficiently similar to *fish.*

If a label is needed for definitions which are part of an explication, "explicative definition" is a natural choice, but it remains crucial to keep in mind that this does not turn explications into definitions and that explicative definitions are not an alternative to, but a subclass of stipulative definitions.

4.2 Precising Definitions

At this point, one might claim that once we have acknowledged the differences between definitions and explications, we realize that we often seek, especially in philosophy, not explications, but rather 'precising' definitions.[39] Such definitions seek to improve exactness; they are neither purely reportive, nor purely stipulative, nor both, but aim at a compromise between reporting and stipulating. This may seem an attractive option, firstly, if one aims at making a vague concept more exact by deciding on borderline cases without re-classifying any non-borderline cases. Secondly, one may have the following picture in mind: explication is at home in theory construction and science, but improving exactness can also serve the more ordinary task of improving communication by eliminating misunderstandings, or the philosopher's need to clear up conceptual problems by deciding to use a term in a specific way, or the practical requirement of clear and exact legal regulations. In such cases, 'precising' the concept is all we need; considerations of fruitfulness and theoretical usefulness are beside the point and cannot motivate further tampering with the concept at hand. Hence, we are better off with the more familiar and simple structure of definitions.

Since the last remark is the most problematic one, I look at it first and bracket the others for the moment. The crucial point is that the idea of 'precising' definitions as a structural alternative to explications is a chimera. If 'precising' definitions are definitions, they should have two parts, a definiendum and a definiens. But on closer analysis, we find that with respect to their stipulative aspect, the definiendum is the 'new', more exact concept, whereas with respect to the reportive aspect, the definiendum is the 'old', less exact concept. Consequently, we have three elements and are back to the structure of an explication: 'old' and 'new' concept play the role

[39] The expression "precising definition" is especially popular in textbooks (e.g. Sinnott-Armstrong and Fogelin 2010; Pfister 2013) and goes back at least to Copi 1953. Sometimes "precisifying definitions" is used (e.g. Sorensen 1991). In the terminology of this paper, 'precising' and 'precisifying' definitions improve exactness, not precision; hence the scare-quotes.

This paragraph describes views I have regularly been confronted with in discussions. In print, it is hard to find them defended explicitly, probably just because advocates of 'precising' definitions virtually never discuss Carnapian explication. In any case, the following analysis is relevant to all who advocate 'precising' definitions or commit themselves to such an idea in their methodological remarks.

of explicandum and explicatum respectively, and the definiens of the alleged definition corresponds to the characterization of the explicatum.

One might object that treating 'precising' definitions as explications makes it seem that the so called old and new concepts are two different concepts; but surely a 'precising' definition deals with one concept only, it just makes it more exact. Appearances to the contrary are due to the identity criteria for concepts adopted in Sect. 2.2. If, for example, we define

(4) A person is an adult iff she has attained the age of 18 years.

we do not introduce a new, less vague concept *adult,* but just make the concept *adult* more exact by eliminating borderline cases. Nevertheless, even if we assume for the sake of argument that there is only one concept *adult* involved, the 'precising' definition in (4) still needs two definienda. Insofar as (4) is stipulative, being 18 years old implies being adult, but insofar as it is reportive, being 18 years old may be compatible with not being adult. Hence "adult" cannot be used with the same meaning in its two occurrences in the preceding sentence. The very idea of 'precising' definitions as mixing stipulating and reporting implies that the definiendum introduced by stipulation is not identical to, but more exact than the definiendum the use of which is reported. Up to this point, the best sense we can make of the idea of 'precising' definitions is this: they are explications which place a premium on improving exactness and use a definition for introducing the explicatum, but they do not introduce a new term. However, calling such explications "definitions" just adds to confusion.

We now also have a preliminary answer to the first line of argument above. 'Precising' definitions can be treated as that subclass of explications for which the conditions of adequacy boil down to eliminating vagueness without re-classifying any non-borderline cases.[40] But with the second line of argument, the real challenge becomes apparent, namely that improving exactness may be undertaken for a variety of reasons, some of which have little to do with science and advancing theory. To this, the explication-theorist can answer that explication is neither wedded to formalisms nor to science as opposed to philosophy (Sect. 2.2), and that improving exactness frequently serves theoretical purposes in philosophy as well. This is true, but the fact remains that the account of explication developed in Sects. 2 and 3 is tailored to the conceptual needs of theories. Proponents of 'precising' definitions are right in pointing out that there are other forms of conceptual re-engineering, although we have just seen that they do not have the structure of definitions but rather that of explications. Nonetheless, these considerations motivate a generalization of the Carnapian account of explication developed so far.

4.3 Conceptual Re-engineering

To make room for various forms of conceptual re-engineering pursuing different goals, we can first identify some core features. Conceptual re-engineering

[40] This is roughly the kind of explication Hanna (1968) focused on.

introduces a concept explicitly, in contrast to other forms of concept change such as non-intentional language development; all forms of conceptual re-engineering have the same basic structure as explications (illustrated in Fig. 1); and their adequacy depends on both similarity to an 'old' concept and usefulness of the 'new' concept for some given purpose.[41] Similarity is meant to ensure two things: that useful functions of the 'old' concept remain available and that introducing a 'new' concept does not simply change the subject. If a proposal for re-engineering a concept is challenged, then these aspects of adequacy need to be argued for.

Against this background, we can distinguish various forms of conceptual re-engineering. Explications introduce a concept that is unambiguous, consistent, at least as exact as the explicandum, useful for some given theoretical purpose and hence subject to corresponding requirements of adequacy as discussed in Sect. 2.3. Characteristically, the similarity condition for explications has to be interpreted pragmatically as well; that is, what exactly is required depends on the target theory, the functions of the explicandum which the explicatum is expected to preserve, and the problems it is expected to solve. Explications can take various forms; specifically, they can but need not make use of a definition and introduce a new term.

Advocates of 'precising definitions' focus on introducing concepts by means of a definition which reduces vagueness without reclassifying non-borderline cases or coining a new term. In addition, the introduced concept is typically also required to be unambiguous and consistent. This manner of improving exactness is not tied to any specific kind of context or purpose. Consequently, the distinction between this form of conceptual re-engineering and explication is neither exclusive nor is one a special case of the other (contra, e.g. Pfister 2013: 55). If the improvement of exactness serves theoretical purposes, this counts as explication; and there are explications which do not improve exactness, as well as exactness improvements which have no theoretical, but purely practical goals.

Further types of conceptual re-engineering include, for example, those aiming more directly at social change.[42] Haslanger's (2000) revisionary "analytical" project seeks to fight sexism and racism by re-engineering concepts such as x *is a woman* or x *is racialized.* Conceptual re-engineering guided by political correctness introduces a concept by choosing a new term, say "Black" instead of "Negro", with the pivotal aim of eliminating negative connotations as a means of improving the social situation of a certain group of people; usually, the extension is expected to remain the same, and neither reducing vagueness nor eliminating inconsistency are on the agenda. The last three properties can also be found in conceptual re-

[41] The term "re-engineering" is chosen in analogy to "reconstruction" in order to underline that there are two dimensions of adequacy, similarity to an existing concept and usefulness of the new concept. Speaking of "re-engineering" is not meant to suggest that the original concept is already the product of some previous conceptual engineering. The literature usually follows Carnap, who speaks of "engineering" (e.g. 1956: 43; "re-engineering" can be found in Carus (1999); see Eklund (2015) for other uses of "conceptual engineering"). Carnap's idea of philosophy as a form of engineering has recently been widely discussed, esp. in Carus (2007a) and Richardson (2013).

[42] Conceptual re-engineering with socially relevant goals plays also a key role in Carus's (2007a) defence of the ideal of explication as a tool of enlightenment.

engineering that serves other goals, ranging from conventions that help to avoid blasphemy ("G–d") to political newspeak (e.g. "enhanced interrogation techniques").

This is obviously only a start for a systematic account of various types of conceptual re-engineering. Although this task cannot be undertaken here, a few remarks on the status of explication are in order. Explication is a specific form of conceptual re-engineering intended as an element of theory development. Other goals, say didactical simplification, political correctness or making an impression of profoundness, may be better achieved by other forms of conceptual change or re-engineering. But even in the context of theory development, there are many situations in which explication may not be the best method of concept formation. For one thing, it may sometimes be impossible to come up with an adequate explicatum, for example, because we cannot find a way of eliminating inconsistency without introducing additional vagueness. Moreover, explicating is not an end in itself, and the specific requirements an explication is supposed to meet as well as the theoretical enterprise that motivates them must eventually be argued for as well. Exactness need not be invariably helpful in science and philosophy. In some contexts, a theory based on relatively vague concepts may fare better overall than its more exact rivals (see, for example, the debate about Sen's capability approach; Chiappero-Martinetti 2008). As Feyerabend repeatedly reminded us, vague and inconsistent concepts may be instrumental in advancing science by stimulating creativity and opening up new perspectives (e.g. Feyerabend 1993).

5 Limitations and a Perspective for Further Developments

There are aspects of conceptual re-engineering for theoretical purposes which can only be systematically addressed if we substantially go beyond a Carnapian account of explication. I briefly discuss three related clusters of issues.

5.1 Relation Between Two Systems of Concepts

Since explicandum and explicatum in general belong to two systems of concepts, the following question arises: how exactly can we relate expressions in one system to expressions in the other, especially if the two are not framed in the same language? The explication of logical validity in standard first-order logic provides a vivid illustration. If we are looking for similarity conditions, stock examples of valid arguments are obvious candidates. Since "All men are mortal; Socrates is a man. Therefore: Socrates is mortal" is clearly in the extension of the explicandum *valid argument,* we may want to require that this verdict is respected by the explication of *logically valid.* But as a similarity condition we cannot simply require that the validity of the quoted argument can be proved in first-order logic because the example is an ordinary-language argument and not an expression of first-order logic. As our similarity condition we rather need to specify a corresponding expression of the logical target theory, say $\forall x(Fx \rightarrow Gx)$; $Fa \models Ga$, and require that it should be provable.

This situation, firstly, brings explication-like relations between more complex expressions into focus, in our example, between an argument and a sequence of formulas. Such relations share important features with explications, but they are not explications because explications deal with concepts (Brun 2004: ch 8.2; Leitgeb 2013: 272–273). Secondly, we are confronted with a host of questions about what counts as adequately relating expressions of one system of concepts to expressions of another such system. Using the example above: is the formula mentioned an adequate formalization of the ordinary language argument? (see Brun 2004, 2014). And even if the explicatum belongs to the same language as the explicandum, transforming expressions containing the explicandum into expressions containing the explicatum may be not trivial, for example, if the two concepts differ in logical form (as x *is warm* vs. x *has temperature* y). The literature on explication is typically not aware of these questions at all (Hanna 1968 is an exception[43]).

This lacuna may be difficult to address, but it does not directly affect the basic structure of the method of explication. In this respect, the two following points are more critical.

5.2 Focus on Individual Concepts

There is a tension between explaining explication as a method for replacing individual concepts and insisting that explications require the context of a theory. Strictly speaking, there is no such thing as an explication that deals with one concept only. To begin with, further concepts must be used for introducing an explicatum. In (2), for example, Scanlon explicates *blame* with the help of the previously introduced explicatum *blameworthy*. The same goes for some similarity conditions. For example, a condition of type (iii) for the explication of a concept *c* refers to a relation between *c* and another concept *d*, and requires the explicatum for *c* to stand in a corresponding relation to the explicatum of *d*. Such conditions become especially prominent, when a whole family of explicanda is replaced with the help of the same concept, as in Carnap's example of using *temperature* to explicate *warm, cold, warmer, colder* etc. (LFP §§ 4–5). In these ways, a 'chain reaction' of explications is started because we need to introduce additional concepts, so that they can be used in the explication at hand (Siegwart 1997a: 31; cf. LFP 173).

Secondly, criteria of theoretical usefulness operate primarily on the level of systems of concepts. The vagueness of the explicatum, for example, cannot be assessed without taking into consideration whether the concepts used to introduce the explicatum are vague (cf. Carnap 1990: 432). And new criteria, such as consistency, scope and explanatory power of the resulting theory, become prominent when we turn our attention from individual concepts to theories.

Finally, explicating a concept often calls for more than simply introducing an explicatum into a readily available theory; an appropriate theoretical framework must be selected, perhaps adapted or even developed during the process of explicating. Explicating concepts and developing the target theory are then

[43] Actually Hanna's way of dealing with the distinction between two systems of concepts leads him into serious problems; see footnote 16.

intertwined aspects of the same project. A clear example is, again, the explication of *logical validity,* which is typically not understood as the search for an explication of *valid* in a given logical theory, but as calling for formulating a logical theory that includes a definition of validity. Consequently, explication is best seen as a component of a more comprehensive process that deals not with replacing individual concepts but with developing systems of concepts and theories. The criteria of adequacy for explications are then subordinate to criteria for theory choice.[44] This, of course, branches a new subject, which is extensively discussed in the philosophy of science.[45]

When Carnap writes about explication, such a more comprehensive methodological perspective is not very prominent and the examples he gives focus on individual concepts. But of course, his primary goals in LFP and *Meaning and Necessity* are theories of logical probability and semantics, not individual explications.[46] And his actual work included the development of systems of concepts which provide interconnected explicata for a range of explicanda (e.g. in LFP, the explication of *estimate* is based on one for *probability*; see pp. 279, 514; in 1956: 8–12, *self-contradiction, entailment* and *factual truth* are explicated with the help of the explicatum for *necessary* or *analytic truth*). However, the fact remains that Carnap did not provide an explicit account of a method for explicating entire systems of concepts or theories. This is not to say that he was not aware of the difference between dealing with a single concept and with more comprehensive systems and theories. The end of LFP, for example, is a 'counterpart' to the opening chapter on explication, in which Carnap very briefly characterizes rational reconstruction as the method of replacing "a body of generally accepted but more-or-less vague beliefs" with a theory (LFP 576). This includes explicating concepts, but also representing those beliefs in a consistent, more exact and more systematic way. Reconstruction in this sense is clearly not reducible to giving explications since there are, in addition to concepts, other sources of inconsistency, vagueness and theoretical ineptness that need to be dealt with. The fluctuations of Carnap's focus are paralleled in terminology. Although he tends to use "explication" with respect to concepts and "rational reconstruction" when referring to theories (as in LFP ch. I and § 110.J), he occasionally equates explication with (rational) reconstruction (e.g. Carnap 1947: 147–148; LFP 453; RSE 945).[47]

[44] An extension of the method of explication to theories was suggested by Martin (1973), who understands explication exclusively as a matter of vagueness reduction. As far as I know, his idea has not been pursued further.

[45] Hempel (2000) provides a clear account of this shift from individual concepts to theories both from a historical and systematic perspective. See also the references given in Sect. 2.3.

[46] As he occasionally explicitly underlines (e.g. LFP 173; 1953: 190).

[47] In the foreword to the second edition of *The Logical Structure of the World* (Carnap 2003: v), Carnap declares "explication" to be a more recent replacement for "rational reconstruction" (*rationale Nachkonstruktion*), although the latter was in fact explained with respect to theories not individual concepts in (2003: § 100). Hempel (2000: 206; see also 1952: 11) reports that the logical empiricists used "explication", "logical analysis" and "rational reconstruction" in the same sense.

Carus (2007a) has recently argued for a deep-running distinction between explication and rational reconstruction, while underlining that both proceed by "piecemeal" replacement of concepts (e.g. 2007a: 15–16, 20, 278). The contrast Carus draws primarily relates to the development of Carnap's overall

5.3 Linear Structure

Descriptions of the method of explication typically present explicating as a process that involves a specific sequence of steps. But in practice, the process is non-linear and not rigidly structured for several reasons (cf. Stegmüller 1973: 25–26). Attempts at introducing an explicatum may prompt us to revise what we did in one of the 'previous' steps. It may turn out that the explicandum needs further disambiguation or we discover that we need more than one explicatum for different purposes. Or we find that the specified conditions of theoretical usefulness cannot be jointly satisfied and this may motivate us to stick to a certain explicatum and adapt the conditions of adequacy.[48] In some cases, the necessary clarification of the explicandum may call for identifying a subtle ambiguity which is most effectively identified indirectly by tentatively introducing explicata and comparing them with the help of the resources of the target system of concepts. Another reason is that explications give rise to feedback effects. A successful explication of a ordinary language concept can have the effect that the meaning of the explicandum-term changes or that the explicatum gets adopted into everyday language (as in the case of *fish;* see Laporte 2004: ch. 3.IV). A basically sequential structure of explicating has no room for these phenomena, but an adequate account of conceptual re-engineering needs to deal with them as well.

These aspects are not entirely absent in Carnap's account of explication. But they are mentioned only in passing and appear to be mere side-effects.[49] On the one hand, this underestimates the importance of feedback effects. Often, we want to introduce an explicatum for exactly those purposes for which the explicandum was used so far (RSE 936), and would not consider an explication to be fully successful if the explicatum did not replace the explicandum in its original usage. Biologists may not bother whether greengrocers sell bananas as "berries", but philosophical projects often aim at clarifying conceptual problems by means of explications. A logician may insist on replacing in non-technical discourse "that's only logical" by

Footnote 47 continued
philosophical programme. In his interpretation, rational reconstruction rests on the "hope that there could be a single, permanent logical framework for the whole of knowledge" (2007a: 20). Explication, on the other hand, rests on the principle of tolerance and a new position on internal and external questions, which admits of introducing alternative explicata and choosing between them on practical grounds (2007a: 263–265).

[48] An example can be found in Bar-Hillel and Carnap (1953: 150), where it is argued that we need two explicata for *amount of information* since no explicatum will meet the following two conditions: "the content of a conjunction should be equal to the sum of the contents of its components if and only if these components […] have no factual consequences in common" and "the content of the conjunction of two basic statements, say 'P_1a_1' and '$\sim P_2a_3$' [where P_1 and P_2 are different primitive predicates] should be equal to the sum of the contents of these statements since they are independent, and this not only in the weak deductive sense of this term, but even in the much stronger sense of initial irrelevance [i.e. inductive independence].".

[49] Carus emphasizes that Carnap's philosophical outlook recognizes "the dialectical relation between [ordinary language and constructed systems]" (2007a: xi; see also 2007b: 41–42). However, he also acknowledges that Carnap gives no explicit account of this dialectics (2007a: 19). His interpretation of Carnap's programme is therefore compatible with the diagnosis I give with respect to Carnap's explicit methodology.

a distinction of valid, sound and persuasive arguments; and Singer's ethical project includes arguments for replacing the traditional notion of moral standing to include non-human animals also in everyday language. On the other hand, the fact remains that Carnap's method of explication lacks a systematic treatment of such feedback effects. To accommodate for them calls for expanding and possibly adapting the basic structure of the method of explication.

From a historical perspective, we may note that this issue was picked up by Goodman, who worked on similar problems of explication at about the same time as Carnap. When he (Goodman 1983: 63–67) introduced what later became known as the method of reflective equilibrium, he was addressing exactly this kind of feedback structure in a further development of his theory of "constructional definitions" (the term for explication in Goodman 1977). Tracing this history and analysing the potential of the method of reflective equilibrium to overcome the problems mentioned in this and the preceding section must be left to further papers.

6 Conclusion

In this paper, I have developed a detailed account of the method of explication which is based on Carnap's classical exposition read through the lens of his later, more pragmatic interpretation of the method. Now the discussion in Sect. 5 has shown that the method of explication provides just a start for a systematic account of conceptional re-engineering for theoretical purposes. Overcoming its limitations requires substantial extensions. We need systematic accounts of explication-like relations between expressions more complex than individual concepts, of how explicating is embedded in theory development, and of the feed-back effects from explicatum to explicandum as well as between the various activities involved in explicating.

These limitations, however, neither invalidate the method of explication nor do they show that it is merely of historical interest. Even if extensions are needed, and neither theory development nor philosophical clarification reduces to giving explications, the analysis of the method of explication provides a number of important insights into the nature of conceptual re-engineering in general and specifically as a tool for the advancement of theories.

Firstly, conceptual re-engineering needs to have a clear target. Clarifying the explicandum plays a key role, but also trying to get as clear as possible about the problems the explication is supposed to solve and about the functions of the explicandum we expect to be preserved. Secondly, these pragmatic considerations as well as the whole process of explication have to be understood against a theoretical background to which the explication at hand is meant to contribute. Thirdly, conceptual re-engineering by explication uses a structure of three main elements: it replaces a concept in use, an explicandum, with a new concept, an explicatum, by means of an explicit characterization. Definitions are an important tool for characterizing concepts and an effective component of explications but since they are based on the two main elements of definiendum and definiens, they are structurally too simple for the tasks of conceptual re-engineering. Fourthly, we

can distinguish several forms of conceptual re-engineering tailored to different goals. Explicating characteristically serves some theoretical purpose whereas improving exactness may also be undertaken for other, often purely practical reasons, and still other forms of conceptual re-engineering are guided by social aims such as fighting stereotypes. Fifthly, explications can provide a number of benefits for theories—concepts are integrated in the target system of concepts, they are unambiguously, consistently and exactly characterized, and they are theoretically useful in various ways by enabling more generalizations, more precise statements, simpler theories and so on. Sixthly, judging the adequacy of an explicatum calls for balancing two basic considerations: whether the newly introduced concept is sufficiently similar to its predecessor to be used in its place in relevant contexts, and how the new concept performs with respect to the purposes it was introduced for, specifically with respect to the general theoretical desiderata and particular targets of the explication at hand. Doing sufficiently well on both counts—providing a theoretically useful concept that can take over important functions of an established concept—is what makes for successful conceptual re-engineering by explication.

Acknowledgments A first draft of this paper was written while I was a visiting scholar at Harvard University sponsored by the Swiss National Science Foundation Grant IZK0Z1-125823. For helpful discussions and critical comments, I would like to thank audiences in Zurich, Frankfurt am Main and Karlsruhe, and in particular Christoph Baumberger, Claus Beisbart, Gregor Betz, Catherine Elgin, Jonas Pfister, Geo Siegwart and Inga Vermeulen. Thanks also to two anonymous referees of this journal.

References

Baker, A. (2013). Simplicity. In *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/archives/fall2013/entries/simplicity/

Bar-Hillel, Y., & Carnap, R. (1953). Semantic information. *British Journal for the Philosophy of Science, 4*, 147–157.

Beaney, M. (2004). Carnap's conception of explication. In S. Awodey & C. Klein (Eds.), *Carnap brought home. The view from Jena* (pp. 117–150). Chicago/La Salle: Open Court.

Belnap, N. (1993). On rigorous definitions. *Philosophical Studies, 72*, 115–146.

Brun, G. (2004). *Die richtige Formel. Philosophische Probleme der logischen Formalisierung* (2nd ed.). Frankfurt a. M: Ontos.

Brun, G. (2014). Reconstructing arguments. Formalization and reflective equilibrium. *Logical Analysis and History of Philosophy, 17*, 94–129.

Brun, G., & Hirsch Hadorn, G. (2014). *Textanalyse in den Wissenschaften. Inhalte und Argumente analysieren und verstehen* (2nd ed.). Zürich: vdf.

Carnap, R. (1945). The two concepts of probability. *Philosophy and Phenomenological Research, 5*, 513–532.

Carnap, R. (1947). Probability as a guide in life. *Journal of Philosophy, 44*, 141–148.

Carnap, R. (1953). Inductive logic and science. *Proceedings of the American Academy of Arts and Sciences, 80*, 189–197.

Carnap, R. (1956) [1947]. *Meaning and necessity. A study in semantics and modal logic* (2nd ed.). Chicago: University of Chicago Press.

Carnap, R. (1962) [1950]. *Logical foundations of probability* (2nd ed.). Chicago/London: University of Chicago Press/Routledge and Kegan Paul. Referenced as LFP.

Carnap, R. (1963). Replies and systematic expositions. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 859–1013). La Salle: Open Court. Referenced as RSE.

Carnap, R. (1966) [1926]. *Physikalische Begriffsbildung*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Carnap, R. (1990) [written 1952]. Quine on analyticity. In Creath, R. (Ed.). *Dear Carnap. Dear Van. The Quine–Carnap correspondence and related work* (pp. 427–432). Berkeley: University of California Press.
Carnap, R. (2003) [1928/34]. *The logical structure of the world. Pseudoproblems in philosophy.* Chicago/La Salle: Open Court.
Carus, A. W. (1999). Carnap, syntax, and truth. In J. Peregrin (Ed.), *Truth and its nature (if any)* (pp. 15–35). Dordrecht: Kluwer.
Carus, A. W. (2007a). *Carnap and twentieth-century thought. Explication as enlightenment*. Cambridge: Cambridge University Press.
Carus, A. W. (2007b). Carnap's intellectual development. In M. Friedman & R. Creath (Eds.), *The Cambridge companion to Carnap* (pp. 19–42). Cambridge: Cambridge University Press.
Chiappero-Martinetti, E. (2008). Complexity and vagueness in the capability approach: Strengths or weaknesses? In F. Comim, M. Qizilbash, & S. Alkire (Eds.), *The capability approach. Concepts, measures and applications* (pp. 268–309). Cambridge: Cambridge University Press.
Copi, I. M. (1953). *Introduction to logic*. New York: MacMillan.
Creath, R. (Ed.). (1990). *Dear Carnap. Dear Van. The Quine-Carnap correspondence and related work*. Berkeley: University of California Press.
de Regt, H. W., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese, 144*, 137–170.
Douglas, H. (2013). The value of cognitive values. *Philosophy of Science, 80*, 796–806.
Eklund, M. (2014). Replacing truth? In A. Burgess & B. Sherman (Eds.), *Metasemantics* (pp. 293–310). Oxford: Oxford University Press.
Eklund, M. (2015). Intuitions, conceptual engineering, and conceptual fixed points. In C. Daly (Ed.), *Palgrave handbook of philosophical methods* (pp. 363–385). Basingstoke: Palgrave.
Feyerabend, P. (1993). *Against method* (3rd ed.). London/New York: Verso.
Goodman, N. (1963). The significance of Der logische Aufbau der Welt. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 545–558). La Salle: Open Court.
Goodman, N. (1972) [1958]. The test of simplicity. In *Problems and projects* (pp. 279–294). Indianapolis/New York: Bobbs-Merrill.
Goodman, N. (1977) [1951]. *The structure of appearance* (3rd ed.). Dordrecht: Reidel.
Goodman, N. (1983) [1954]. *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.
Greimann, D. (2007). Regeln für das korrekte Explizieren von Begriffen. *Zeitschrift für Philosophische Forschung, 61*, 261–282.
Gustafsson, M. (2014). Quine's conception of explication—And why it isn't Carnap's. In G. Harman & E. Lepore (Eds.), *A companion to W.V.O. Quine* (pp. 508–525). Chichester: Wiley Blackwell.
Hahn, S. (2013). *Rationalität. Eine Kartierung*. Münster: Mentis.
Hanna, J. F. (1968). An explication of 'explication'. *Philosophy of Science, 35*, 28–44.
Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs, 34*, 31–55.
Hegselmann, R. (1985). Die Korrespondenz zwischen Otto Neurath und Rudolf Carnap aus den Jahren 1934 bis 1945. Ein vorläufiger Bericht. In H.-J. Dahms (Ed.), *Philosophie, Wissenschaft, Aufklärung* (pp. 276–290). Berlin: de Gruyter.
Hempel, C. G. (1952). *Fundamentals of concept formation in empirical science*. Chicago: University of Chicago Press.
Hempel, C. G. (1970). Aspects of scientific explanation. In *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331–496). New York: Free Press.
Hempel, C. G. (2000) [1988]. On the cognitive status and the rationale of scientific methodology. In *Selected philosophical essays* (pp. 199–228). Cambridge: Cambridge University Press.
Hilbert, D. (1894/1895). Die Theorie der algebraischen Zahlkörper. *Jahresbericht der Deutschen Mathematiker-Vereinigung, 4*, 175–546.
Justus, J. (2012). Carnap on concept determination. Methodology for philosophy of science. *European Journal for the Philosophy of Science, 2*, 161–179.
Kant, I. (1998) [1781/1787]. *Critique of pure reason*. Cambridge: Cambridge University Press.
Kitcher, P. (2008). Carnap and the Caterpillar. *Philosophical Topics, 36*, 111–127.
Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension* (pp. 320–339). Chicago: University of Chicago Press.
Laporte, J. (2004). *Natural kinds and conceptual change*. Cambridge: Cambridge University Press.

Leitgeb, H. (2013). Scientific philosophy, mathematical philosophy, and all that. *Metaphilosophy, 44*, 267–275.
Lutz, S. (2012). On a straw man in the philosophy of science. A defense of the received view. *HOPOS, 2*, 77–119.
Maher, P. (2006). The concept of inductive probability. *Erkenntnis, 65*, 185–206.
Maher, P. (2007). Explication defended. *Studia Logica, 86*, 331–341.
Maher, P. (2010). *What is probability?* MS. http://patrick.maher1.net/preprints/pop.pdf. Accessed November 10, 2015.
Martin, M. (1973). The explication of a theory. *Philosophia, 3*, 179–199.
Pfister, J. (2013). *Werkzeuge des Philosophierens*. Stuttgart: Reclam.
Putnam, H. (1997). A half century of philosophy, viewed from within. *Daedalus, 126*, 175–208.
Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
Quine, W. V. O. (1980) [1951]. Two Dogmas of empiricism. In *From a logical point of view* (2nd ed., pp. 20–46). Cambridge, MA: Harvard University Press.
Reck, E. (2012). Carnapian explication. A case study and critique. In P. Wagner (Ed.), *Carnap's ideal of explication and naturalism* (pp. 96–116). Basingstoke: Palgrave Macmillan.
Reichenbach, H. (1951). The verifiability theory of meaning. *Proceedings of the American Academy of Arts and Sciences, 80*, 46–60.
Richardson, A. (2013). Taking the measure of Carnap's philosophical engineering. Metalogic as metrology. In E. Reck (Ed.), *The historical turn in analytic philosophy* (pp. 60–77). New York: Palgrave MacMillan.
Scanlon, T. M. (2008). *Moral dimensions. Permissibility, meaning, blame*. Cambridge, MA: Harvard University Press.
Schumann, W. (2008). *Minerals of the world* (2nd ed.). New York: Sterling.
Scorzato, L. (2013). On the role of simplicity in science. *Synthese, 190*, 2867–2895.
Shepherd, J., & Justus, J. (2015). X-Phi and Carnapian explication. *Erkenntnis, 80*, 381–402.
Siegwart, G. (1997a). Explikation. Ein methodologischer Versuch. In W. Löffler & E. Runggaldier (Eds.), *Dialog und System* (pp. 15–45). St. Augustin: Academia.
Siegwart, G. (1997b). *Vorfragen zur Wahrheit. Ein Traktat über kognitive Sprachen*. München: Oldenbourg.
Singh, G. (2010). *Plant systematics* (3rd ed.). Enfield, NH: Science Publishers.
Sinnott-Armstrong, W., & Fogelin, R. J. (2010). *Understanding arguments* (8th ed.). Belmont: Wadsworth.
Sjögren, J. (2011). *Concept formation in mathematics*. Gothenburg: University of Gothenburg.
Sober, E. (2000). Quine's two Dogmas. *Proceedings of the Aristotelian Society, 74*, 237–280.
Sorensen, R. (1991). Vagueness and desiderata for definition. In J. H. Fetzer, D. Shatz, & G. N. Schlesinger (Eds.), *Definitions and definability* (pp. 71–109). Dordrecht: Kluwer.
Sorensen, R. (2013). Vagueness. In *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/archives/win2013/entries/vagueness/
Stalnaker, R. (1976). Propositions. In A. F. MacKay & D. D. Merrill (Eds.), *Issues in the philosophy of language*. New Haven: Yale University Press.
Stegmüller, W. (1973). *Jenseits von Popper und Carnap. Die logischen Grundlagen des statistischen Schließens.* Berlin: Springer. (= *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie; Bd. 4: Personelle und statistische Wahrscheinlichkeit. Studienausgabe, Teil D*).
Strawson, P. F. (1963). Carnap's views on constructed systems versus natural languages in analytic philosophy. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 503–518). La Salle: Open Court.
Wagner, P. (Ed.). (2012). *Carnap's ideal of explication and naturalism*. Basingstoke: Palgrave Macmillan.