Synthese Special Issue: PhilMethods manuscript No.
(will be inserted by the editor)

The Constituents of an Explication

Moritz Cordes

Received: date / Accepted: date

Abstract The method of explication has been somewhat of a hot topic in the last ten years. Despite the multifaceted research that has been directed at the issue, one may perceive a lack of step-by-step procedural or structural accounts of explication. This paper aims at providing a *structural* account of the method of explication in continuation of the works of Geo Siegwart. It is enhanced with a detailed terminology for the assessment and comparison of explications. The aim is to provide means to talk about explications including their criticisms and their interrelations. There is hope that this treatment will be able to serve as a foundation to a step-by-step guide to be established for explicators. At least it should help to frame and mediate explicative disputes. In closing the enterprise will be considered an explication of 'explication', though consecutive explications improving on this one are undoubtedly conceivable.

Keywords Explication · Criteria of Adequacy · Philosophical methodology · Metaphilosophy

1 Introduction

In the following I will provide a structural account of the method of explication, or one might say a formal framework that imposes some restraints on what kind of entity an explication could be. For this purpose I take explication to be concerned with giving meaning to expressions that are imprecise or otherwise problematic. I will avoid the talk about concepts since this sometimes dazzling expression may lead to complications in the discussions about explication.[1]

M. Cordes
University of Greifswald
Baderstraße 6-7, 17489 Greifswald, Germany
Tel.: +49-3834-4203458
E-mail: cordesm@uni-greifswald.de

[1] This is evidenced, I suppose, by some debates following Strawson (1963) and Schupbach (2015).

I consider this view to be compatible with Carnap, although he prefers to talk about concepts in his seminal work on explication.[2] More to the point – and not only concerning the concept/expression distinction –, I will follow Siegwart (1997b, ch. 24), (1997a) who provided a step-by-step account of explication improving on Carnap.[3] Other procedural approaches to explication are rare.[4] In fact, explication is often portrayed informally as an isolated act of introduction of a concept or an expression that, apparently, does not include much preparation or postprocessing.

The paper will run thus: First, I will give some characterizations (postulates and definitions) of the basic concepts of explication including the six constituents of an explication (2). In a short excursus I will consider the construction of a step-by-step guide of explication which may help (potential) explicators in their business (3). Nothing close to a vade mecum will be provided! Then I will facilitate means to talk about explication disputes, chains of explications, and other phenomena in a precise fashion allowing for some distinctions that are often ignored (4). In some parts of the sections 2 and 4 I will heavily rely on the work of Siegwart, although several modifications will be made which will be pointed out then. The final step will comprise some reflections on what I have done so far and whether this can be considered an explication of ‘explication’. In the light of these deliberations I will point to alternative explications of ‘explication’ (5).

What is being proposed is a theoretical and neutral account. (Here) I do not defend any theses on the usefulness of explication or on the preference of one kind of explication over another. There may be some implicit valuations on these matters in the last section, but they should not be read

[2] Carnap (1950, ch. I). Note especially the technical remarks on pages 7 and 8, but also his talk about terms on page 3. At any rate, one should take care not to interpret Carnap’s use of the expression ‘concept’ in any of the current thick senses. – (1950) is not Carnap’s first investigation into the method of explication: (1945) is an earlier one.

[3] Siegwart’s work seems to have fallen into oblivion with respect to current methodological and metaphilosophical debates. The earlier Hanna (1968) has suffered a similar fate. It seems, only Greimann (2007) and Brun (2016) refer to Siegwart. Quite remote from metaphilosophical debates Glatzer (2012), Hahn (2013), Paasch (2016), and Cordes (2016) employ Carnapian explication in its Siegwartian implementation.

[4] Greimann (2007) and Hanna (1968) have been mentioned already. Justus (2012), Schupbach (2015), and Shepherd and Justus (2015) make some contributions from experimental philosophy. Interestingly, Justus (2012, p. 162) observes a related methodological deficit in contemporary (meta)philosophy as I do: “Rather than become a staple of philosophical methodology, explication has been challenged on several grounds.” However, with respect to the works referred to at the end of the preceding note, the first part of the quote appears debatable. – As to the current discourse on explication, however, even such dedicated volumes like Wagner (2012) do not boast any systematic descriptions of the full procedure of explication, though it sheds light on many historical issues and includes several comparative investigations into aspects of methodology. To give one (maybe representative) example from this book: Carus’s essay (2012) presents an intuitively interesting distinction between local and global explication; but he describes neither in sufficient detail for the reader to get some clear directions as to how to perform a local or global explication or even what components must be stated when presenting any such explication. It would be desirable to have Carus’ local and global explication defined in an explicit fashion comparable to, for example, definitions 17, 18, 20 or 21 below.

as substantial claims – they only serve to test the ground in order to prepare further developments of the explicative method. For now, the aims are purely clarificatory. For that matter, some refinements of explication discussed in recent studies are neglected here. This is motivated by the focus on a broad account which leaves out issues that are still in dispute. Furthermore, I am not here involved in an exegetical investigation into Carnap's specific ideas about explication. I especially do not deal with any of his four requirements (similarity, exactness, fruitfulness, simplicity) in particular. Each of these requirements has been under scrutiny in the literature on explication. Many insights resulted from such investigations, but it is hard to condense them into specific conditions on how to take a specific step within an articulated procedure of explication. In fact, often it is hard to even relate what a given researcher says about explication because there is no theory of explication that is independent from the various readings of the Carnapian requirements. Thus, as long as philosophers keep referring to Carnap's outline of the explicative procedure including the four requirements while understanding the requirements differently, confusion is bound to ensue. For all these reasons, it seems important to first give a clear and very general understanding of what an explication is. Therefore the Carnapian requirements will not be discussed.

A note on the style: I employ a basic formal language for all clarificatory characterizations below. It is a first-order predicate language with a standard set theory. The reason for using such a language is clarity. There are strict regulations on how to correctly postulate axioms and define expressions in such a setting. Since informal talk adds to the complex state of discussion about explication, I think one is justified in trying more formal means. Perhaps some parts of the methodological debate are entangled in misconceptions caused by the intuitive use of language. At any rate, the goals pursued in this text are not dependent on the specific formulation of the definitions provided and other formulations could have been chosen as well. Furthermore, the ›spirit‹ of these definitions can always be carried over to the informal talk about explications, too, as long as the informal conception sticks to that ›spirit‹.

2 A Definition of Explication

As I mentioned in the beginning 'explication' will be understood as a procedure that deals with expressions.[5] The expression being explicated will be called *explicandum*. The explicandum is usually used within some formal or informal language, the *explicandum language*. When explicating, the explicandum and the explicandum language are ›replaced‹ by *explicatum* and *explicatum*

[5] In connection with explication Quine (1960, §§53-55) prefers to talk about expressions, too. That there may be a more or less vague *concept* behind the explicandum (or the explicatum) is not contested here. Theories about the explicandum *concept* and the explicatum *concept* can be built upon the explication theory that is developed here and that deals with explicandum and explicatum as *expressions*. It would be more difficult the other way around if we agree that concepts are abstractions from expressions between which some kind of semantic or syntactic equivalence relation holds.

language.[6] This language, too, can be either formal or informal.[7] The idea is, that the explicatum in the explicatum language is ›more precise‹ and does not have the problems of the explicandum in the explicandum language. In fact, the explicatum is taken to fulfill certain specific requirements with respect to the explicatum language, namely the *criteria of (explicative) adequacy*.[8] To this purpose there has to be an *explicative introduction* of the explicatum which ensures these criteria.

The six italic expressions in the preceding paragraph can be taken to refer to the six constituents of an explication.[9] Sometimes only the last item, the explicative introduction, is referred to as an explication (Cohnitz and Rossberg, 2006, p. 58). But if explication is taken serious as a procedure with considerable complexities – as is done here – then it is advisable to let 'explication' refer to more than just one item within the whole procedure. An explication will thus be defined below (definition 3) as the sextuple of the constituents.[10] Before 'explication' can be defined thus, it is necessary to ensure that the six constituents have generic features qualifying them as the kind of things that can take the respective part in an explication. Hence, we need to shed some light on what ›expressions‹ and ›languages‹ are in that context – ›criteria‹ (sets of formulas) and ›introductions‹ (formulas) can be given directly in the definition of 'explication' together with definitions 8 and 9. In order to keep the exposition short, I restrict myself to a brief postulation of those syntactic object language features that are of prime interest to explication. In that spirit I take *expressions* to be simples. Only the empty set is excluded from the set of all expressions (EXP) because it will serve as a substitute referent for later definitions 4 to 9:

Postulate 1
$\emptyset \notin \mathrm{EXP}$

Thus, basically, expressions can be anything but the empty set. This leaves enough room to see expressions as entities extraneous to pure set theory (and thus as urelements) or as set theoretic entities (possibly made up from

[6] Some authors prefer the term 'explicans' instead of 'explicatum' (e. g. Reichenbach (1951, p. 49); an anonymous reviewer drew my attention to this article). I stick with 'explicatum' for two reasons: First, the thing doing the explication (literally, the explicans) is not the expression but that what will be called 'explicative introduction' below. Second, 'explicatum' is the technical term employed in most of the literature.

[7] Usually, if the explicandum language is formal in some sense already, so will be the explicatum language.

[8] These *specific* requirements are to be distinguished from the *generic* requirements proposed by Carnap (similarity, exactness, fruitfulness, simplicity), though in some cases the specific requirements may be seen as appropriations of the generic requirements to the explicative scenario at hand. Toward the end of the current section the exclusion of Carnap's requirements will be briefly motivated.

[9] Siegwart (1997a, §17) lists the same six constituents of an explication ("Explikationsfaktoren").

[10] Additional constituents could be incorporated (cf. sect. 5). For now the six constituents named shall suffice. That is, they will at least suffice to make a point about what a structural account of explication could look like.

urelements or from the empty set). For example, one might identify expressions by their Gödel numbers or one might take any expression, be it formal or informal, as primitive or one may take some (atomic) expressions as primitive and others (molecular) as set theoretic constructs (sequences) of the former.

A *language* will be taken to consist of (i) a grammar including a vocabulary (VOC) of atomic expressions and a syntax of molecular expressions with a category of formulas (FORM)[11] having elements of the vocabulary as atomic subexpressions (ASE) and (ii) a set of rules setting up a (possibly empty) consequence relation ($\vdash$), though the rules do not have to be limited to logical or even inferential rules. Instead of defining all components of a language in detail, the following postulate just captures the relevant features. Their usual behavior will have to be ensured elsewhere by individually establishing or describing specific languages:

Postulate 2
For all L: if L is a language, then
$L \neq \emptyset$ and
$\mathrm{VOC}(L) \subseteq \mathrm{EXP}$ and $\mathrm{VOC}(L) \neq \emptyset$ and
$\mathrm{FORM}(L) \subseteq \mathrm{EXP}$ and $\mathrm{FORM}(L) \neq \emptyset$ and
$\mathrm{ASE}_L \subseteq \mathrm{VOC}(L) \times \mathrm{FORM}(L)$ and
$\vdash_L \subseteq \wp(\mathrm{FORM}(L)) \times \mathrm{FORM}(L)$

Note, definitions and axioms are not considered to be a separate constituent of a language. But there may be rules in a language regulating the use of formulas as axioms or definitions. Furthermore, this presentation of a language does not determine languages to be regimented in any way. This can be seen in the following possibilities that do not conflict with postulate 2: There need not be a full syntax; expressions may belong to more than one syntactical category; vagueness and ambiguity are not excluded; the consequence relation ($\vdash_L$) may be undecidable; inconsistency is not excluded. For the projected definition of explication all this is not necessary.

However, any formal language in an intuitive sense can be considered a language in accordance with this postulate. Thus, Carnap's languages I and II (1934) are examples for languages in that sense. ›Natural‹ languages like English, German, and Latin can be understood as satisfying postulate 2, too, if one decides on a list of expressions as the vocabulary (VOC) and on a grammar at least establishing the category of sentences/statements/formulas (FORM) and the relation of being an atomic subexpression (ASE). In this instance the consequence relation ($\vdash$) may or may not be kept empty.[12] A third kind of example for a language Z is the talk practiced in a given field of study, say, zoology. This would best be rendered as including a consequence relation which

[11] In accord with the decision not to dive too deep into the syntactic structure of the expressions of the object languages, I will not distinguish between closed and open formulas here. In a more detailed analysis one should be more precise.

[12] If such a language with an empty consequence relation is employed as explicatum language and not only as explicandum language, it will be impossible to prove the success of an explication (cf. definition 10).

is not empty but comprises any combination of premises and conclusion which would have to be accepted within the respective community. We may have, for example, {'specimen 13150225 is physeter macrocephalus'}$\vdash_Z$'the class mammalia includes specimen 13150225'. So, in this case, zoological taxonomy together with some informal logic and some grammatical equivalences is seen as setting up the consequence relation.[13]

A little step back may be in order. So far I have given two postulates. The purpose of these postulates is to establish some basic features of languages which we need if we take language(s) to play a central role in explication. For this purpose the features should not be seen to be too restrictive. In fact, they may be too liberal with regard to what Carnap would allow as explicandum language and explicatum language. I do not oppose introducing additional restrictions but currently this liberal understanding will suffice. Note, that the postulates and the following definitions do not call for a reconstruction of natural languages in order to provide an explication. Not even a rough imprecise outlining, as was done in the preceding paragraph, is necessarily needed – that was only done to give an impression of what may be considered a language for the purposes of systematic talk about explications. Postulate 2 just introduced the minimal properties a language must have to serve as explicandum language or explicatum language. Consequentially, now we can take a look at a structural definition of 'explication', which, for now, leaves this notion insensitive to ›success‹:[14]

Definition 3
For all E: E <u>is an explication</u> iff there are $L_1, L_2, \mu_1, \mu_2, X, \phi$ such that:
$E = \langle L_1, L_2, \mu_1, \mu_2, X, \phi\rangle$
L_1 and L_2 are languages and
$\mu_1 \in \mathrm{VOC}(L_1)$ and $\mu_2 \in \mathrm{VOC}(L_2)$ and
$X \subseteq \mathrm{FORM}(L_2)$ and $X \neq \emptyset$ and
$\phi \in \mathrm{FORM}(L_2)$ and μ_2 is an ASE_{L_2} of ϕ

In this and the following definitions the expressions being defined will be underscored. – One aim of definition 3 is to identify certain phenomena as explication, though some sextuples are included that are not explications in any intuitive sense. This will be tolerated in order to enable anyone to call a

[13] The conception of a language supposed here can be modified to accomodate more complex consequence conceptions. For example, as one reviewer suggests, Bayesian settings may require a consequence relation that admits of strengths.

[14] Siegwart's rather material definition of an ((in)adequate) explication (1997b, p. 262, 01-02) is not equivalent. Most notably, Siegwart demands the fifth constituent (the criteria of adequacy) to represent distinguished use patterns ("ausgezeichente Verwendungsmöglichkeiten repräsentieren"). In addition, the sixth constituent is just a formula here while, in fact, the explicative introduction may as well assume the form of, say, a metalanguage rule. This is possible in Siegwart's definition. (For further discussion see sect. 5!) Despite the differences, I take definition 3 and the following definitions to keep in the spirit of Siegwart's theory of explication.

given constellation of constituents an explication if that helps their respective purposes.[15]

Note that definition 3 does not amount to saying a lot. This is so because up to now the only ›substantial‹ terms (i. e. the non-set-theoretic terms) in this definition are 'language', 'VOC', 'FORM', and 'ASE' and the meaning of these terms is intentionally left thin as per postulates 1 and 2. It would be wrong to assume that with the two postulates and this definition of 'explication' one has a thorough grasp of the procedure of explication or one is automatically able to conduct an explication. All one can do with definition 3 is to call a certain constellation of linguistic entities that are in accord with the postulates an explication. It is not even possible to refer to the six constituents by specific terms. This will be possible only after definitions 4 to 9.

Having said that, it should be clear that calling something an explication and thus giving an example for an explication is now possible with the help of definition 3. Here is one example: ⟨English, I, 'successor', 'nf', {'nf(0) = 1'}, ⌜D1.D7-1⌝⟩ is an explication, more precisely a mathematical successor explication by Carnap.[16] After definitions 4 to 9 we will be able to refer to each component of this explication with ease. Of course, with the liberal postulates and the liberal definition of 'explication' there will also be unlikely examples for explications: ⟨French, English, 'chien', 'cat', {'love gains some inertia'}, 'cats eat grass'⟩ or ⟨I, I, 'nf', 'nf', {'$(x)(\mathrm{nf}(x) = \mathrm{nf}(x))$'}, D2⟩. I will come back to some examples for explications once suitable means to talk about them have been provided.

On the side of completeness one may fear that some explications (in an intuitive sense) are not explications in the sense just provided. For example, most explications are not stated in the same fashion as the ones in the preceding paragraph or one of the constituents might not be directly identifiable.[17] Two things should be said about that: (i) It is true that some things that one thinks should be explications will not turn out to be explications on this account. The remedy for this is to switch to a different theory of explication (which may or may not be a close variant of the one being proposed here). (ii) But the examples described at the beginning of this paragraph do not even necessarily fall out of the scope of the proposed sense of explication. It should be clear that by definition 3 explications are abstract entities and should not be confused with their (re)presentations in words and symbols. When explicators ›provide an explication‹ they just utter a *description* of an abstract entity. This is what happens in the preceding paragraph, too, albeit in a formal fashion. Thus, informal or unclear or

---

[15] If one reads definition 3 as (part of) an explication of 'explication' (cf. sect. 5), one may object with Carnap (1950, p. 479): "The result that a proposed explicatum is found too narrow constitutes a much less serious objection than the result that it is too wide." But the truth of this comparative statement largely depends on what the explicator aims for.

[16] This refers to Carnap's language I with its successor function 'nf'. The two definitions D1 and D7-1, here conjoined by conjunction, are the following: '$\mathrm{nf}(x) = x'$' and '$1 = 0'$', respectively (Carnap, 1934, pp. 15, 51, 52).

[17] An anonymous reviewer confronted me with such a scenario.

incomplete presentations (descriptions) of explications can still be associated with (one or several) abstract entities which are the explications (in the sense of definition 3). One purpose of a *procedural* (not a structural) account of explication is to lead explicators to give presentations of explications that do not leave the task of identifying the constituents of these explications to the reader (see sect. 3).

Next the six constituents of any explication will be defined relative to this explication in a straightforward fashion. Abbreviation function constants are given in brackets in the underscored part of each definition. The empty set serves as a substitute referent.[18]

Definition 4
For all L, E: $L =$ <u>the explicandum language (L1) of</u> E iff
E is an explication and there are $L_2, \mu_1, \mu_2, X, \phi$ such that
$E = \langle L, L_2, \mu_1, \mu_2, X, \phi\rangle$
or
E is not an explication and $L = \emptyset$

Definition 5
For all L, E: $L =$ <u>the explicatum language (L2) of</u> E iff
E is an explication and there are $L_1, \mu_1, \mu_2, X, \phi$ such that
$E = \langle L_1, L, \mu_1, \mu_2, X, \phi\rangle$
or
E is not an explication and $L = \emptyset$

Definition 6
For all μ, E: $\mu =$ <u>the explicandum (EX1) of</u> E iff
E is an explication and there are L_1, L_2, μ_2, X, ϕ such that
$E = \langle L_1, L_2, \mu, \mu_2, X, \phi\rangle$
or
E is not an explication and $\mu = \emptyset$

Definition 7
For all μ, E: $\mu =$ <u>the explicatum (EX2) of</u> E iff
E is an explication and there are L_1, L_2, μ_1, X, ϕ such that
$E = \langle L_1, L_2, \mu_1, \mu, X, \phi\rangle$
or
E is not an explication and $\mu = \emptyset$

[18] The definitions in (Siegwart, 1997b, p. 262) are different in two respects from the ones given here. (i) His definitions are conditional definitions. (ii) He defines by identity and not by biconditional. Each of the two techniques on its own allows him to avoid providing a substitute referent like the empty set. The latter option just moves the reference problem elsewhere. That is why it is avoided here. In addition, conditional definitions are avoided here in order to guarantee eliminability, which Siegwart sacrifices. The resulting disjunctive form of each definiens given here is not seen as overly displeasing and the second disjunct can be ignored for most purposes.

Definition 8
For all X, E: $X =$ the set of criteria of explicative adequacy (CEA) of E iff

> E is an explication and there are $L_1, L_2, \mu_1, \mu_2, \phi$ such that $E = \langle L_1, L_2, \mu_1, \mu_2, X, \phi \rangle$
> or
> E is not an explication and $X = \emptyset$

Definition 9
For all ϕ, E: $\phi =$ the explicative introduction (EI) of E iff

> E is an explication and there are $L_1, L_2, \mu_1, \mu_2, X$ such that $E = \langle L_1, L_2, \mu_1, \mu_2, X, \phi \rangle$
> or
> E is not an explication and $\phi = \emptyset$

To give an example, in the sequel 'L2(E)' refers to the explicatum language of E.[19] Note that it is not presupposed that E is an explication. If some E is not an explication, then L2(E) is just the empty set. In fact, E is not an explication iff L1(E) or ... or EI(E) is the empty set.[20]

The purpose of definitions 4 to 9 is quite modest: They allow one to refer to each of the six constituents of an explication by a suggestive term. Only with the further definitions below the appropriateness of the specific terms chosen will start to unfold. At least it will do so to the degree that the full theory suits descriptions of what one would intuitively consider explications.

In the remainder of the text I will often omit 'explicative' in 'criteria of explicative adequacy' (definition 8). – After identifying the six constituents of an explication it is now possible to introduce some helpful distinctions among explications. In this section I will only give the most immediate distinctions, starting with the definition of a successful explication as one that fulfills its criteria of adequacy.[21] This yields a binary definition (alternatively, success could be defined as having degrees).

[19] Note that these definitions are not circular. At any stage this is obvious from just one look at the preceding postulates and definitions. It is true that even before definition 3 reference was made to the six constituents. But it was made only in the motivating prose, not in the postulates or *in* definition 3. In fact, one reason for choosing a formal presentation for the theory of explication lies in the ease of monitoring and preventing such semantic troubles as circularity. (There is occasion for this remark: A commentator on an early version of this paper saw some "weird circularity" in these structural definitions.)

[20] To elaborate further: The explicandum of anything that is not an explication is the empty set and it is identical with its explicatum and even with its explicandum language and so on. The underlying mistake here is to speak about the explicandum (explicatum, ...) of a non-explication at all. One reviewer drew my attention to the fact that some find this way of talking problematic. I agree, but this way of talking is not avoided by a conditional definition – conditional definitions just move the matter to the realm of the undecidable. In order to prevent talk about the explicandum of a non-explication, one would have to resort to *syntactical* exclusion. But this kind of intervention seems disproportionate.

[21] Again, this definition excludes Carnap's requirements. This will be motivated toward the end of this section.

Definition 10
For all E: E is a successful explication iff
E is an explication and
for all ψ: if $\psi \in \mathrm{CEA}(E)$, then $\{\mathrm{EI}(E)\} \vdash_{\mathrm{L2}(E)} \psi$

Obviously, the criteria of adequacy (CEA), the explicative introduction (EI), and the explicatum language (L2) are the three relevant constituents. Let us reconsider the first example for an explication given above: $\langle$English, I, 'successor', 'nf', {'nf(0) = 1'}, $\ulcorner$D1.D7-1$\urcorner\rangle$. It turns out that this is a successful explication since in language I the only criterion of adequacy (namely 'nf(0) = 1' – that the successor of zero is one) follows from the explicative introduction, which is the conjunction of D1 and D7-1. In formal style: $\{\ulcorner$D1.D7-1$\urcorner\} \vdash_{\mathrm{I}}$'nf(0) = 1'. To be sure, in 1934 Carnap did not name 'nf(0) = 1' as an explicative criterion nor did he talk about explication at all. But with the terminology presented here we are able to talk about this as an explication.

Definition 10 points out the relevance of criteria of adequacy when explicating. They serve as a measure of success. Thus, if an author tries to provide an explication without in some fashion providing criteria of adequacy readers are bound to ascribe unintended criteria of adequacy to it and assessments of success will probably vary. Taking 'nf(0) = 1' as such a criterion was one example for an unwarranted, though presumably unproblematic ascription in the above paragraph. Some explicators miss out on providing criteria of adequacy, but others explicitly provide them.[22]

In general, the easiest way to produce a successful explication for an explication with only finitely many criteria of adequacy is to choose as the explicative introduction just the conjunction of these criteria:

Theorem 11
For all E:
If E is an explication and
$\vdash_{\mathrm{L2}(E)}$ is closed under $\wedge$E and
$\mathrm{CEA}(E) = \{\phi_1, ..., \phi_n\}$ for some finite n and
$\mathrm{EI}(E) = \ulcorner \phi_1 \wedge ... \wedge \phi_n \urcorner$,
then E is a successful explication

Two things speak against generally synchronizing the criteria of adequacy and the explicative introduction in that fashion: (i) The proposing of such an explication may run counter to the rules of introduction[23] in the explicatum language, effectively preventing one from correctly introducing the explicatum by means of the explicative introduction within the explicatum language. (ii)

[22] D7-B in (Carnap, 1942, pp. 27-28) is an explicit example for a criterion of adequacy of a truth predicate (in a book that is not explicitly concerned with explication methodology).

[23] By 'rule of introduction' I mean the same as by 'rule of inference' except what is being licensed is an act of definition or of setting an axiom. Here is one example rule of introduction: If ϕ is a closed formula consistent with any axioms and definitions already set, then one may set ϕ as an axiom. For a full treatment see Reinmuth (2013).

Explicators will often want to decide on criteria independently from the specific form of the explicative introduction, forgoing direct synchronizing. Having said that, of course it is possible to trim the criteria or to tweak the introduction, if a successful explication is not reached on first attempt. But then the trimming and tweaking will need justification.

Often an explicative introduction can only unfold its power against a background theory. If so, success should be measured with respect to a suitable set of explicatum language formulas representing the background theory. Let us first define what a theory in a language and what a background theory for an explication are:

Definition 12
For all T, L: T <u>is a theory in</u> L iff
  L is a language and $T \subseteq \mathrm{FORM}(L)$

Definition 13
For all T, E: T <u>is a background theory to</u> E iff
  E is an explication and T is a theory in $\mathrm{L2}(E)$ and
  for all ψ: if $\psi \in T$, then $\mathrm{EX2}(E)$ is not an $\mathrm{ASE}_{\mathrm{L2}(E)}$ of ψ

The last line of definition 13 prohibits the explicatum from occurring in the background theory. Keeping in the liberal spirit of the preceding postulates and definitions no constraints of consistency or the like are imposed on these theories. It is only required that the explicatum does not already appear (is not an ASE of any formula) in the background theory. Otherwise the theory would provide more than a background. Now success can be defined in relation to a given background theory:

Definition 14
For all E, T: E <u>is a successful explication for</u> T iff
  T is a background theory to E and
  for all ψ: if $\psi \in \mathrm{CEA}(E)$, then $T \cup \{\mathrm{EI}(E)\} \vdash_{\mathrm{L2}(E)} \psi$

Definition 15
For all E, T: E <u>is a non-trivially successful explication for</u> T iff
  E is a successful explication for T and
  there is ψ such that: $\psi \in \mathrm{CEA}(E)$ and $T \nvdash_{\mathrm{L2}(E)} \psi$

With definition 14 we have that ⟨I, I, ‘nf’, ‘nf’, {‘$(x)(\mathrm{nf}(x) = \mathrm{nf}(x))$’}, D2⟩ is a trivially successful explication for any background theory since the sole criterion of adequacy (namely ‘$(x)(\mathrm{nf}(x) = \mathrm{nf}(x))$’) is logically true in language I; the explicative introduction D2 is redundant. D2 still is the explicative introduction of that explication, but it is a bogus one. – Here is an example for a non-trivially successful explication for a given background theory: ⟨English, RL, ‘mother’, ‘M’, {‘$\forall x \forall y \forall z(\mathrm{M}(x, y) \wedge \mathrm{M}(z, y) \rightarrow x = z)$’}, ‘$\forall x \forall y(\mathrm{M}(x, y) \leftrightarrow \mathrm{P}(x, y) \wedge \mathrm{F}(x))$’⟩. Read ‘P’ as ‘is parent of’ and ‘F’ as ‘is female’. The explicative introduction is a straightforward definition of motherhood. The sole criterion of adequacy states left-uniqueness for the motherhood relation. Thus, RL shall be a formal first-order language for the talk about family relations. It shall have a standard consequence relation with no unexpected results that would amount to substantial axioms or definitions involving the material expressions for family relations. Axioms shall be included in the theories RT1 and RT2. Let us say, the axioms of RT1 are consistent and lay down that everybody has exactly two parents and only one of them is female. Let us say, the axioms of RT2 do not subscribe to that or any equivalent proposition; but RT2 shall be consistent, too. Then said explication is non-trivially successful for RT1 and it is not successful at all for RT2.

Definition 10 and definition 14 are related in the sense that any explication is successful iff it is successful for $\emptyset$. This fact could be taken as an alternative definition of the unary predicate 'successful explication', in which case Definition 10 would become a theorem. Furthermore, from definition 14 springs a theorem akin to theorem 11 but with a weaker EI.[24]

Theorem 16
For all E, T:
If T is a background theory to E and
$\vdash_{\mathrm{L2}(E)}$ is closed under $\wedge$I, $\wedge$E, $\to$E and
$\{\psi_1, ..., \psi_m\} \subseteq T$ for some finite m and
CEA(E) $=\{\phi_1, ..., \phi_n\}$ for some finite n and
EI(E) $= \ulcorner(\psi_1 \wedge ... \wedge \psi_m) \to (\phi_1 \wedge ... \wedge \phi_n)\urcorner$,
then E is a successful explication for T

The limited practicality of theorem 11, as explained there, holds for theorem 16, too.

A remark on Carnap's four requirements of adequacy is now in order. A full discussion of similarity, exactness, fruitfulness and simplicity or their appropriations to specific explications would be misplaced here, because what is offered is not a content discussion for specific explications but a structural account of explication. For that purpose definitions are proposed that relate six arbitrary constituents to one another. None of them is filled with definite content (except for illustration purposes). Thus, if one wants to use the present account of explication and wants all explications to be fruitful in any Carnapian sense relevant to the field within which that explication is situated, then it would be wise to try to codify that requirement somehow – for example as a condition which the fifth constituent (the CEA) may satisfy – and define fruitfulness according to the pattern suggested by definitions 10, 17, 18: For all E: E is a fruitful explication iff ... . Said condition takes the place of the the ellipsis.[25]

[24] An anonymous reviewer drew my attention to theorem 16 and requested the inclusion of more theorems like, for example, theorem 11.

[25] In instances of explication the Carnapian requirements are sometimes refined into more specific criteria that would fit well into what is here called the criteria of explicative adequacy (Pinder, 2016, sect. 4.2). This happens particularly often with the requirement of similarity. Here Siegwart (1997b, pp. 270-272) sees a transformation of Carnap's similarity requirement via Quine into what is now known as the criteria of explicative adequacy.

Similar avenues are available for simplicity and exactness. With such an approach to the Carnapian requirements the problem of what an explication is breaks down into two problems that could be tackled separately: (i) What is an explication, regardless of its cognitive value? One possible answer: A sextuple as described in definition 3. (ii) What makes an explication worthwhile? One possible answer: If it is successful as described in definition 10 and fruitful, simple, exact as described elsewhere.[26]

To conclude this section we will make a fundamental distinction between two kinds of explications. Telling intra-language and inter-language explications[27] apart will come in handy:

Definition 17
For all E: E is an intra-language explication iff
E is an explication and $\mathrm{L1}(E) = \mathrm{L2}(E)$

Definition 18
For all E: E is an inter-language explication iff
E is an explication and $\mathrm{L1}(E) \neq \mathrm{L2}(E)$

Of the last two example explications ⟨English, RL, ‘mother’, ‘M’, {‘$\forall x \forall y \forall z(\mathrm{M}(x,y) \wedge \mathrm{M}(z,y) \to x = y)$’}, ‘$\forall x \forall y(\mathrm{M}(x,y) \leftrightarrow \mathrm{P}(x,y) \wedge \mathrm{F}(x))$’⟩ is an inter-language explication since English and RL are distinct. ⟨I, I, ‘nf’, ‘nf’, {‘$(x)(\mathrm{nf}(x) = \mathrm{nf}(x))$’}, D2⟩ is an intra-language explication since both explicandum language and explicatum language are Carnap’s language I. Whether some given explication is inter-language or intra-language will often depend on what one takes to be the two relevant languages. Some authors do not say anything about that matter, but with others at least the target language (i. e. the explicatum language) is clearly identified.

3 A Procedure of Explication

An explication in the sense here proposed is a sextuple. That does not mean that every single explication has to be presented in a formal fashion displaying this sextuple structure at the surface. But the explication definition from the preceding section makes clear what the constituents are and those should at least be named somewhere when an explicator submits an explication. Thus, we arrive at a first rough procedural imperative for explicators: *Name each of the constituents of the explication!* Presenting an explication according to this maxim can be quite easy. It may look like this (the naming of each constituent is marked by the abbreviations used earlier):

> Let us get clear about what we mean by beauty in our everyday aesthetic assessments (L1) so that we may establish a philosophical theory of beauty (L2). This theory should reflect what we mean when

[26] Additional distinctions of success in explications are given in (Cordes, 2016, p. 38).
[27] Cf. Siegwart (1997a, §17): “Binnenexplikation” and “Brückenexplikation”.

> we say that something is beautiful (EX1). As aesthetic assessments are generally taken to be subjective let us settle on a relative predicate: x is beautiful to p (EX2). We all agree that only perceptible things are beautiful to anyone but not everything that is perceptible is beautiful to everyone. Additionally, it seems important that beautiful things are pleasing to the senses of the relevant beholder. (CEA) Thus let us settle on the following characterization: Something is beautiful to somebody if it has been perceived by that person and that person is pleased by that perceptive experience (EI).[28]

To be sure, much is left open in this example which is quite naïve in both methodic and systematic regards. But it has the advantage that it is easy to see all six components of the explication as indicated. Each of the components plays an essential role in the explication, so the author of above paragraph did right in not leaving anything out. From such a lucid presentation one could now, with the help of a given background theory, begin to argue for the criteria of adequacy as they were stated in two sentences. This would conclude the explicative procedure in a fuller sense.[29]

In the example the naming of the constituents happens rather ad hoc. It does not take a seasoned philosopher of aesthetics to see that this explication needs a lot more motivation for some of the constituents. This is especially clear with the choice of the explicatum and the criteria of adequacy. In both cases one may rightfully ask whether other alternatives would be more interesting, worthwhile, or suitable. To a certain degree all six constituents need some motivation. With each naming of a constituent a number of choices have been made. How are those choices justified? What would be a clever way of making those choices to produce successful explications? This leads to the inevitable question that is not conclusively answered here: How does the meaning of the explicandum figure within the explication?

These questions indicate the need for a comprehensive guide to explication. It is important to see that Carnap's contribution is only a beginning and does not provide such a guide. Maybe Hanna (1968, p. 29) was the first to explicitly name steps (in his case: five steps) that together make up the procedure of explication. Siegwart (1997a, §§11-16) gave a procedure that is much more detailed in most respects. He complemented his approach in 2007b by drawing on Lambert (1771). A similar procedural outline is given by Brun (2016, sect. 3). Even before Hanna, Tillman (1965) made a proposal to turn part of Strawson's critique of Carnap's explication ("linguistic portrayal") into a part of the method of explication – but he did not really come up with a full procedure. More recently, Hahn (2013, sect. 2.3) and Cordes (2016, sect. 1.2) contributed further with regard to the assessment of the need for an explication and to the refinement of criteria of adequacy, respectively.

[28] The reader may consider it an exercise to formulate a similar paragraph for another picturebook explication, for example for Carnap's fish/piscis explication.

[29] Siegwart (1997b, p. 256), (1997a, p. 29) explicitly includes this as the final step in the explicative procedure.

In another tradition Shepherd and Justus (2015, sect. 3) contributed some improvements concerning the explicandum individuation by experimental philosophy.[30]

In this article I am quite far from putting together a full guide from all these sources. Instead I am concerned with giving a theoretical account or framework of explication. But I would like to stress the fact that both enterprises, providing a theoretical account of explication and providing a guide for potential explicators, are closely dependent on one another. Any refinement of how explication is done may bear on how we should theoretically conceive of explication. On the other hand, some degree of theoretical description of explication is always needed to tell explicators what they should do when explicating. In that spirit, the theoretical account of explication given in the preceding section may serve as a foundation for developing a procedural account guiding the explicator. The maxim at the beginning of this section exemplifies this approach: Name each of the constituents of the explication!

Instead of giving a full procedural outline, I restrict myself to illustrating how one may draw on insights from other fields of study to refine the procedure. There is, for example, the question as to what regulations should apply to the explicative introduction. Often the explicator wants a definition. How this is correctly provided is an issue treated in formal and informal definition theory. Thus, if the explicatum is a predicate one may demand of the explicative introduction to be a universal biconditional (which would have to satisfy additional criteria). But under the current account of explication the explicative introduction could also be a conditional definition, an axiom, conjunctions of axioms and definitions, or some other kind of formula.[31] The most pertinent measure of what is a formally admissible explicative introduction can be provided by the explicatum language. For that purpose languages have to be construed in a way that they provide regulations for object-language means of conceptual introduction (like defining). This could be postulated in the following fashion:

Postulate 19
For all L: if L is a definitional language, then
L is a language and
$\mathrm{DEF}_L \subseteq \wp(\mathrm{FORM}(L)) \times \mathrm{FORM}(L)$ and $\mathrm{DEF}_L \neq \emptyset$

Just like the consequence relation the definition relation relates sets of formulas to formulas. That way it is possible to talk about some formula being a definition relative to a theory (which, for example, includes previous definitions and axioms). Definitional explications can then be easily defined:

[30] A paper providing further methodological references is in preparation.

[31] In the proposed conception the introduction of the explicatum cannot be facilitated by the setting of rules for inferring or for constating formulas – at least if the explicatum language does not include its own metalanguage. In this respect the account differs from Siegwart's.

Definition 20
For all E, T: E is a definitional explication for T iff
T is a background theory to E and
$(T,\, \mathrm{EI}(E)) \in \mathrm{DEF}_{\mathrm{L2}(E)}$

Definition 21
For all E: E is a definitional explication iff
there is a T such that E is a definitional explication for T

The explication of beauty above concluded thus: ‘something is beautiful to somebody if it has been perceived by that person and that person is pleased by that perceptive experience’. The ‘if’ can plausibly be read as ‘iff’. Then with some acceptable provisions for the definition rules of the philosophical explicatum language we can say that that explication is a definitional explication. Definition 20 allows to relativize the property of being a definitional explication to some background theory. It presupposes that in a language something is a definition only with respect to some theory. This makes it possible to require of a definition to have a definiens in which all material expressions have been introduced beforehand (namely in that background theory).

This only gives a rough first impression of what one needs to think about when trying to provide a guide to explication which respects customs like standard definition theory. And this example deals with just two of the constituents of an explication. The processes involved in the choice of the other constituents have to be described with recourse to fields of study other than definition theory. – So far it should have become clear that explication “is certainly not a mechanical method or decision procedure. But neither is it arbitrary.”[32] I will leave this issue at this point and, in the next section, give some distinctions which can be made without direct reference to features of an explication procedure or an explication guide.

4 A Way to Talk about Explications

Some means to talk about explication have been presented in the preceding sections. This will not be enough to facilitate a comprehensive discourse about explications. When, for example, one wants to assess the quality of an explication there are only the internal and theory relative assessment options (definitions 10, 14, and 15) available for now. A way to comparatively assess different explications has not yet been provided. This deficit will be amended now.

Note that the following and preceding definitions are all founded on a formal understanding of explications as sextuples. But the ›spirit‹ of the

[32] Creath (2012, p. 162) describes the related but distinct method of logical analysis in these words.

definitions can always be carried over to the informal talk about explications, as stated in the beginning. – The first batch of definitions largely follows Siegwart[33] and helps to specify at which various points two explications may differ and still be called alternatives to one another. This is the case if two explications explicate the same explicandum from the same explicandum language:

Definition 22
For all E, E^*: E is an explication alternative to E^* iff
E and E^* are explications and $E \neq E^*$ and
$\mathrm{L1}(E) = \mathrm{L1}(E^*)$ and $\mathrm{EX1}(E) = \mathrm{EX1}(E^*)$

The realm of explication alternatives is limited to those pairs of non-identical explications that have at least the explicandum and the explicandum language in common. Thus, if an explication of 'meaningless' and an explication of 'pseudo-problem' from everyday philosophy talk are provided, they should not be seen as explication alternatives vying for interpretive authority over what philosophers say. What is being explicated in the two explications is just different and both explicanda may have any right to an explication. Definition 22, thus, suggests that there is no (explicative) conflict if two explicators explicate different explicanda from different explicandum languages.

If, on the other hand, two explicators both try to explicate 'meaningless' from everyday philosophy talk and do not reach the same results in all respects, then they present *explication alternatives* which may give rise to discussion and dispute. Thus, once two explications have the same explicandum and explicandum language they will be considered explication alternatives with respect to one another. Then, of course, the question is, in what respects may explication alternatives differ? There are four remaining constituents. It is sensible to classify the alternatives with regard to these constituents:

Definition 23
For all E, E^*: E is a linguistic explication alternative to E^* iff
E is an explication alternative to E^* and
$\mathrm{L2}(E) \neq \mathrm{L2}(E^*)$

Definition 24
For all E, E^*: E is a lexical explication alternative to E^* iff
E is an explication alternative to E^* and
$\mathrm{EX2}(E) \neq \mathrm{EX2}(E^*)$

Definition 25
For all E, E^*: E is a criterial explication alternative to E^* iff
E is an explication alternative to E^* and
$\mathrm{CEA}(E) \neq \mathrm{CEA}(E^*)$

[33] See (1997b, pp. 263-264). Some adjustments to the current framework are made. Definitions 27 and 30 are entirely new.

Definition 26

For all E, E^*: E <u>is an introductory explication alternative to</u> E^* iff
E is an explication alternative to E^* and
$\mathrm{EI}(E) \neq \mathrm{EI}(E^*)$

There is nothing spectacular hidden in these definitions. Basically, they just nail down four adjectives in their application to the term ‘explication alternative’, namely ‘linguistic’, ‘lexical’, ‘criterial’, and ‘introductory’.[34]

Examples for the various kinds of explication alternatives are easily found: Maybe the best known *introductory alternatives* are those that are associated with the various definitions of an ordered pair. Insofar as they all have the characteristic property of ordered pairs as the (only) criterion of adequacy, none of them is a criterial explication alternative to any other. If one frames them in a way that they are neither lexical nor linguistic explication alternatives they still turn out to be *introductory explication alternatives*.

Criterial explication alternatives can be seen in various social explications of gender. Most of them can be thought of as starting with the explicandum ‘gender’ from everyday English. As an explicatum we can choose ‘g’ from a given sociological language as a one-place expression that works like a function constant. (Thus the explication alternatives considered are not linguistic or lexical explication alternatives.) But one explication may try to provide ‘g’ as an expression which attributes to someone a function value that reflects that person’s gender self-ascription, while another explication may try to provide that expression as a device for categorizing people into two groups (for example for the purpose of studying how people of those groups are treated differently). These stipulations can be framed as different criteria of adequacy[35] which, of course, ask for different explicative introductions.

Lexical explication alternatives that are not linguistic explication alternatives may not occur very frequently, but they can be constructed easily. The word ‘successor’ from informal mathematics is usually taken to mean ‘direct successor’. The occasional need for the word ‘direct’ is due to the alternative reading as ‘any number that comes later in the natural numbers’. While the former reading usually takes the function constant ‘S’ as the formal language explicatum, the latter reading should take the predicate constant ‘$<$’. Thus we have two lexical explication alternatives which – if successful and

[34] Instead of ‘introductory’ some may prefer ‘semantical’ for the fourth kind of explication alternative. This would recognize the explicative introduction as a semantical entity or as related to semantics. But it may also be perceived as inscribing specific meaning-theoretic views into the terminology. I would like to avoid this for now. – With regard to lexical explication alternatives theorists may further want to distinguish between whether two explicata differ in their grammatical category or in the arity or in another respect. See (Siegwart, 1997b, p. 263).

[35] The CEA for the first explication could be {‘for any person p it should hold that when asked for their gender in most cases p names $\mathrm{g}(p)$’} and for the second {‘the range of values of g is male and female’}. (There may be explications that satisfy both criteria, coincidentally or on purpose.)

within usual use of the formal expressions – are also criterial and introductory explication alternatives.[36]

With some background assumptions one will have that, usually, linguistic explication alternatives are lexical explication alternatives, because if the explicatum language is altered then often the new explicatum language does not have the same vocabulary and a different expression is chosen as the explicatum. Here is an example: Suppose someone has an explication for 'belongs' (explicandum) with a set theoretic explicatum language. The explicatum may be the element predicate '$\in$'. If, then, a step is made to a different explicatum language, say, mereology, we arrive at a linguistic explication alternative. The mereological language may not contain '$\in$'. Thus one has to choose a different explicatum. This then yields a lexical explication alternative. Furthermore, probably both the criteria of adequacy and the explicative introduction will change, too. In other words, the explication alternatives are also criterial and introductory explication alternatives. But this need not be so. Starting from one explication one may switch to a different explicatum language which is possibly weaker in a logical sense. In that case one may have two linguistic explication alternatives, say, a classical one and an intuitionistic one, which may share the explicatum, the criteria of adequacy and the explicative introduction.

Coming back to the example for introductory explication alternatives, namely the different definitions of the ordered pair, one may feel the need to point out that they are not different in the sense that they choose different forms of introduction. After all, they all are definitons. In addition, they are all intended to be employed against the same set-theoretic background theory in the same language and all of them are associated with the same criterion of adequacy. This suggests a specific refinement of the concept of introductory explication alternatives:

Definition 27
For all E, E^*: E <u>is a just definitional explication alternative to</u> E^* iff
E is an explication alternative to E^* and
$\mathrm{L2}(E) = \mathrm{L2}(E^*)$ and
$\mathrm{EX2}(E) = \mathrm{EX2}(E^*)$ and
$\mathrm{CEA}(E) = \mathrm{CEA}(E^*)$ and
E and E^* are definitional explications

The word 'just' in the definiendum signals that all constituents except the explicative introduction are in accord.[37] – Note, that from the first four conditions in the definiens it follows that $\mathrm{EI}(E) \neq \mathrm{EI}(E^*)$. Thus, the relation defined is irreflexive.

[36] Justus (2012, p. 168) draws attention to the fact that lexical explication alternatives are sometimes advisable even if the explicandum is not ambiguous in the explicandum language.
[37] Defining 'is a definitional explication alternative' (without 'just') would be more adequately achieved by accepting any pair of introductory explication alternatives which are both definitional explications.

Now, if suitably portrayed as explications, the different definitions of ‘ordered pair’ are definitional explication alternatives. This rather superficial observation can be more closely studied by paying attention to whether the different definitions ›amount to the same thing‹:[38]

Definition 28
For all E, E^*, T: E <u>is a consonant explication to</u> E^* <u>under</u> T iff
E and E^* are both definitional explications for T and
$E = E^*$
or
E is a just definitional explication alternative to E^* and
$T \cup \{\mathrm{EI}(E)\} \vdash_{\mathrm{L2}(E)} \mathrm{EI}(E^*)$ and $T \cup \{\mathrm{EI}(E^*)\} \vdash_{\mathrm{L2}(E^*)} \mathrm{EI}(E)$

Definition 29
For all E, E^*: E <u>is a consonant explication alternative to</u> E^* iff
E is a consonant explication to E^* under $\emptyset$

Now, the various definitions of ‘ordered pair’ can be qualified as just definitional explication alternatives that are not consonant under the usual set theories, since their definientia are not equivalent under these theories. Two explications that *are* consonant to one another under the usual set theories can be associated with one of the usual definitions, say, Kuratowski's ‘$\langle x, y\rangle = \{\{x\}, \{x, y\}\}$’ and a set-theoretically inert variant of it, namely ‘$\langle x, y\rangle = \{\{x, y\}, \{x\}\}$’. Here, both explicative introductions are equivalent with regard to the suggested background theories.

The preceding definitions mostly dealt with explicative introductions. In principle, explication alternatives can be related to one another with respect to the other constituents, too. For example, the criteria of adequacy even allow for a comparative assessment of two explications:

Definition 30
For all E, E^*: E <u>is a more demanding explication than</u> E^* iff
E is an explication alternative to E^* and
for all ψ: if $\psi \in \mathrm{CEA}(E^*)$, then $\mathrm{CEA}(E) \vdash_{\mathrm{L2}(E)} \psi$ and
there is ψ such that: $\psi \in \mathrm{CEA}(E)$ and $\mathrm{CEA}(E^*) \nvdash_{\mathrm{L2}(E^*)} \psi$

This definition suggests a point of departure for further elaborations once an explication has proven successful. For example, if an explication of ‘thinking’ in the field of A.I. is successful in that it excludes all computers from the realm of thinking things and qualifies at least some humans as thinking (two criteria of adequacy) one may want to strengthen these criteria by adding another one which requires all humans between 18 and 65 who are in good health to be thinking. Such an explicative endeavour would be more demanding, intuitively and also according to definition 30.

[38] Here, in distinction to Siegwart (1997b, p. 264), the binary consonance relation is reflexive for definitional explications.

What is the use of the definitions presented in this section so far with regard to the bigger picture? First and foremost they help to localize dissent between two disputing explicators. Definitions 23 to 26 suggest that a discussion between vying explicators may be streamlined by consciously focusing on one constituent. In reality, many disputing explicators speak of the differences in the explicative definition (i. e. the explicative introduction) only. Explicators should always consider that they may have different criteria of adequacy, which prove different explicative introductions successful. In everyday situations people tend to overlook this fact as one can easily see from semantic disputes that fall short of taking into account the different goals associated with different definitions of one term. But two explicators proposing criterial explication alternatives may not have the need to dispute the other's explication at all. – Definitions 27 to 30 allow for more fine-grained distinctions with respect to the criteria of adequacy and the explicative introduction. It might be interesting to develop similar or even subtler distinctions with respect to all six constituents of explications.

On a less harmonious view, this terminology of explication alternatives does not only relate explications to one another, it is a starting point for criticism directed at explications, too. Somebody who wants to criticize an explication may do so by just naming a modification of any of the six constituents and demanding a justification for the unmodified constituent in light of the modified one. Thus, if I am dissatisfied with an explication of 'gender' as a one-place function constant, I may direct the explicators attention to an alternative explicatum, say, a two-place function which is capable of relating a person to its gender dependent on a time index. The explicator and I can now engage in a dispute which may or may not involve consideration of the other explication constituents.

Explication alternatives could be a vast field of study but there are several other interesting ways to relate explications to one another. For example, convergent explications are something like an opposite to explication alternatives in that they may differ with respect to explicandum and explicandum language:

Definition 31
For all E, E^*, T: E <u>and</u> E^* <u>are convergent explications under</u> T iff
E and E^* are explications and T is a theory in $\mathrm{L2}(E)$ and
$\mathrm{L2}(E) = \mathrm{L2}(E^*)$ and $\mathrm{EX2}(E) = \mathrm{EX2}(E^*)$ and
$T \cup \{\mathrm{EI}(E)\} \vdash_{\mathrm{L2}(E)} \mathrm{EI}(E^*)$ and $T \cup \{\mathrm{EI}(E^*)\} \vdash_{\mathrm{L2}(E)} \mathrm{EI}(E)$

This is quite similar to definition 28, but the intended area of application for the consonance relation is among explication alternatives, although it is not limited to them. Definition 31 is intended for explications that are not explication alternatives. The latter of the two following definitions provides the suitable terminology for such non-alternatives.

Definition 32
For all E, E^*: E and E^* are convergent explications iff
E and E^* are convergent explications under $\emptyset$

Definition 33
For all E, E^*: E and E^* are genetically different explications iff
$\mathrm{EX1}(E) \neq \mathrm{EX1}(E^*)$ or $\mathrm{L1}(E) \neq \mathrm{L1}(E^*)$

These definitions suggest that sometimes one and the same expression in a language may serve as explicatum of different explications that started with different explicanda and/or in different explicandum languages. If such a constellation occurs one may consider both explications as cumulatively fruitful or illuminating depending on how far ›apart‹ the two explicanda or explicandum languages are. For example, 'pseudo-sentence' from the polemics of the Vienna Circle and 'fashionable nonsense' from Sokal and Bricmont's criticism of postmodernism (1998) can both be explicated. It is conceivable that the two explications lead to a common language where one explicatum with the same definition represents both explicanda.[39] Then we can speak of convergent explications that are genetically different.

Another interesting constellation are explications whose explicative introductions together constitute a theory in a common language:

Definition 34
For all A, T: A is an explicative constitution of T iff
A is a non-empty set of explications and
for all E, E^*: if $\{E, E^*\} \subseteq A$ and $E \neq E^*$, then $\mathrm{L2}(E) = \mathrm{L2}(E^*)$ and $\mathrm{EX2}(E) \neq \mathrm{EX2}(E^*)$ and
$T = \{\phi \mid$ there is E such that: $E \in A$ and $\phi = \mathrm{EI}(E)\}$

An explicative constitution of a theory is not that spectacular if one disregards all the work that went into each of the explications that are an element of the whole constellation. It is just a compilation of explications with different explicata that happen to have a common explicatum language. Thus, even the set of all explications ever proposed in, say, the language of some specific set theory (e. g. NBGU) explicatively constitutes a certain theory – probably an inconsistent one. It gets more interesting if the explications build up on one another by reusing the results of one explication in the explicative introduction of the next explication. These constellations are chains of explications (cf. Siegwart (1997a, §13), Lutz (2012, p. 20)):

Definition 35
For all E: E is a chain of explications iff
E is a non-empty, finite sequence of explications E_1, ..., E_k and
for all m, n: if $0 < m < n \leq k$, then $\mathrm{EX2}(E_n)$ is not an atomic subexpression of $\mathrm{EI}(E_m)$ and
for all n: if $0 < n < k$, then $\mathrm{L2}(E_k) = \mathrm{L2}(E_n)$

[39] For an explicative study of 'pseudo-sentence' and similar terms see (Cordes, 2016).

The language conception stipulated in postulate 2 at the beginning of this article is fairly broad. With a more specific conception definition 35 could be tailored to better suit scenarios in which consecutive definitions or axioms cumulatively create larger and larger languages. Depending on the specific proposal, relations between these languages could then be made explicit, either in set-theoretic terms (e. g. $\mathrm{L2}(E_1) \subseteq \mathrm{L2}(E_2)$) or in other terms (e. g. $\mathrm{L2}(E_2)$ is an extension of $\mathrm{L2}(E_1)$). For now the identity condition at the end of definition 35 simply requires a chain of explications to have a constant overarching language $\mathrm{L2}(E_k)$. – The following definition defines a special case of chains of explications:

Definition 36
For all E: E <u>is a single-source chain of explications</u> iff
E is a chain of explications E_1, ..., E_k and
for all n: if $0 < n < k$, then $\mathrm{L1}(E_k) = \mathrm{L1}(E_n)$

The single-source chain of explications seems to be the standard case for chains of explications because it seems sensible to relate explicata in an explicatum language whose explicanda were already used within the same context or language before the explication. In fact, the chain of explications may introduce each explicatum in an order that was inspired by some kind of order of the explicanda in the explicandum language. But, in general, no demands can be made about the kind of order between the explicanda, since, usually, the explicanda and the explicandum languages are not sufficiently structured in any serviceable sense. This ›disorder‹ may even be the motive for the explication. On the side of the explicata there is a constraint on the order that prevents definitional circularity.

Another phenomenon has been dubbed ‘chain of explication’, too (Siegwart, 1997b, p. 268). This is more appropriately called a ‘history of explications’. It refers to a number of several explication alternatives (i. e. same explicandum and same explicandum language) in a chronological order:

Definition 37
For all E: E <u>is a history of explications</u> iff
E is a non-empty sequence of explications that are mutual explication alternatives

This relates back to definition 22. It could also be expressed via definition 33 since explication alternatives are explications that are not genetically different. The *correct* chronological order is not included in definition 37 because that would require explications to have a temporal date which runs counter to the abstract character given to explications in this paper. Note that a history of explications is not defined as being finite which suggests that the business of explication can be perceived as open ended. – In part Frege’s *Foundations of Arithmetic* (1953) can be read as a presentation of a history of explications of the term ‘number’ (‘Anzahl’) supplemented with discussion and criticism.

Quite unrelated to chains or histories of explications it sometimes occurs that a given explication is improved upon because certain problems surface in the use of the explicatum within the explicatum language calling for yet another explication. A definition may run like this:

Definition 38
For all E, E^*: E is a consecutive explication to E^* iff
E and E^* are explications and
$\mathrm{L2}(E^*) = \mathrm{L1}(E)$ and $\mathrm{EX2}(E^*) = \mathrm{EX1}(E)$

Definition 39
For all E, E^*: E is a close consecutive explication to E^* iff
E is a consecutive explication to E^* and
$\mathrm{CEA}(E^*) \subseteq \mathrm{CEA}(E)$

The development of set theory can be seen as several consecutive explications. Given by the name of the explicators for the sake of brevity Cantor, Frege, and Russell/Whitehead are part of that progression. – Definition 39 mirrors the fact that sometimes the goal of an explication (formulated as the criteria of explicative adequacy) does not diminish in a consecutive explication but, instead, is enhanced by further requirements. However, this is not always so. The problems with an explicatum in an explicatum language which motivate the consecutive explication may be related to defects of the criteria of adequacy of the first explication. For instance, the criteria may be inconsistent. If so, then for the consecutive explication some of the criteria should be deleted.

5 An Explication of 'explication'

What has been done in the previous sections can be seen as presenting an explication of 'explication'[40], though it was presented without adherence to an explication manual (cf. sect. 3). Nonetheless it is possible to identify the constituents of this explication. The explicandum is 'explication'. The explicandum language is something we may call the contemporary philosopher's English (CPE). The explicatum according to definition 3 is 'is an explication'. I would like to consider the explicatum language to be CPE, too. Of course, not every contemporary English speaking philosopher will employ all the expressions I employed here, but that just means each contemporary philosopher employs only a part of CPE. Thus, CPE should be considered a language, in part formal and in part informal, that is shared by many speakers with none of them employing all parts of it. This arrangement

[40] The title of this section has been used before: (Hanna, 1968), (Wilson, 2012, p. 205). Greimann's employment of this construction (2007, sect. 2) is problematic: According to his own characterization of 'explication' he does not provide an explication of 'explication'. Kitcher (2012, p. 202) applies this term to Carnap's exposition.

renders the explication of ‘explication’ an intra-language explication according to definition 15.[41]

Obviously, the explicative introduction of the explication of ‘explication’ should at least include definition 3. In fact, it may be more sensible to include definitions 4 to 9, too, since they directly define the six constituents of any explication. Thus, the explicative introduction in this explication of ‘explication’ is the conjunction of definitions 3 to 9. Please note: It would be misleading to call definition 3 (or for that matter, the conjunction of definitions 3 to 9) the explication – it is just one constituent of the explication or even just one part of one constituent!

The fifth constituent, the criteria of explicative adequacy, are harder to come by. They are not explicitly given anywhere in the previous sections, but I can give some sample criteria here. For instance, it was an implicit aim to explicate ‘explication’ in a way that (i) identifies explicandum, explicatum, their respective languages, a specific explicative introduction, and the criteria of explicative adequacy in an explication. (ii) Furthermore, explications were supposed to be not confined to definitions. (iii) It must be possible for there to be explications with informal explication languages, too. One rather formal criterion was, that (iv) explications were supposed to be sets of some kind. This criterion was not motivated materially but methodologically: If explications are set-theoretic entities one has many means available to refine further talk about them.

In the preceding three paragraphs, a full explication is outlined. Note that the success of the explication (definition 10) cannot be assessed as long as there is no consequence relation spelled out for CPE. If we attribute to CPE a consequence relation along the lines of an informal classical first-order logic, then the explication is not successful on its own. More specifically, we would have to presuppose some set theory, some synonymies between informal expressions, and the existence of languages with suitable properties. Otherwise the third criterion of adequacy given in the previous paragraph would not turn out true. Hence, even if we accept these rather harmless criteria of adequacy, the explication is only successful for some background theory T. But not any set-theoretic or other formula will do as part of a background theory. For example, many definitions given here cannot be part of a background theory (definition 13) because they involve the explicatum and, thus, they are not really ›background‹ with respect to the explication of ‘explication’.

Up to now considerations in this section were a presentation of an explication of ‘explication’. But there are several unfinished construction sites here. One was pointed out in the preceding paragraph. Some others have been hinted at in the course of the paper. A consecutive explication of ‘explication’ may be in order. This will not be given here, but we may jot down some ideas that would help such an effort. It is important to realize that this need for another explication is relative. For many purposes the theory of explication

[41] To be sure, staying within the boundaries of CPE still allows us to transition from a less exact part of CPE to “a more exact part of it” (Carnap, 1963, p. 935).

presented here may work just fine, the same way as Carnap's informal theory of explication seems to have worked fine for him for many years. Maybe with Strawson things got problematic. After his contribution one may have several issues with Carnapian explication that call for more precision. But it is hard to find something that amounts to an explication of 'explication' (in the sense of either (i) a formal account of a theory of explication or (ii) a procedural outline of explication, which can go a long way toward a theory of explication). Most investigations into explication avoid these issues and focus on, for instance, justifying Carnapian criteria of adequacy without explaining what ›exact‹ conception of explication they presuppose.[42]

To get to some points where enhancement is conceivable: First, it would be convenient to have a whole explicatum theory incorporated into the explication as another constituent. This theory may comprise additional definitions building on the explicative introduction or definitions and postulates that serve as a foundation for the explicative introduction.[43] Thus, for example, all the postulates and definitions given in this text may together form the explicatum theory for the explication of 'explication'. Non-essential material may be included into an explicatum theory, too.

Second, one may want a stronger notion of languages, or at least of explicatum languages in order to generally distinguish between various kinds of introduction formats: axioms/postulates, unconditional definitions, conditional definitions etc.. This enhancement can be achieved without demanding that all explicatum languages be formal.

Third, in order to accommodate metalanguage introduction procedures a major modification in the sixth constituent, the explicative introduction, would be in order. It is obvious that in formal and informal languages we regularly turn to metalanguage support in order to regulate some expressions, especially the basic ones like logical operators, but also the empirical ones that have to be introduced operationally. Thus metalanguage rules could be considered admissible explicative introductions. A weaker version of this modification may include changing the explicative introduction from a single formula to a possibly infinite set of formulas.

Fourth, the explicandum and the explicatum rarely are single expressions. When explicating an expression we often also explicate other expressions that are taken to be synonymous in some intuitive sense (cf. definition 31). On the explicatum side, we often find out that we want a more diverse squad of expressions serving our post-explicative purposes.[44] This can be seen in

[42] This does not mean that each such investigation is problematic. This is most decidedly not so. Carnap's criteria can be discussed without setting up a whole theory of explication. But one has to accept, that this way a particular systematic shortcoming in the overall explication debate will always remain.

[43] One may compare this enhanced approach to the crutch employed earlier in this section in the identification of the explicative introduction as a conjunction of multiple definitions.

[44] Greimann (2007, p. 266) misleadingly talks about explications without explicata.

the current paper, too, as there are many definitions of expressions without a similar multitude of expressions in the pre-explicative discourse.[45]

Fifth and finally, the criteria of explicative adequacy are currently all object language formulas although sometimes explicators may want some formula to not follow from the explicative introduction (and its negation following from it neither). If the explicatum language does not include its own metalanguage, as is often the case, the desideratum of not being a consequence cannot be expressed in the explicatum language.[46] Hence, the criteria of adequacy could be re-imagined as a set of either object- *and* metalanguage formulas or as metalanguage formulas exclusively (Cordes, 2016, p. 31). The switch to the metalanguage would accommodate explications that have ›superpropositional‹ explicata, for example illocutionary force indicators.[47]

6 Conclusion

In this paper I proposed a structural account of explication. I tried to stay true to a tradition that prominently started with Carnap and – less prominently – was continued by Greimann, Hanna, and Siegwart. One characteristic of this approach was to disregard any concept ontology and to relate explication to expressions exclusively. Another one was to provide some measure of independence from any qualificatory desiderata like the Carnapian requirements. The approach included a demand for a step-by-step procedure that can guide any philosopher in the practice of explication. Some terminology was established in order to relate explications to one another. The merit of all this was hinted at by means of some textbook style examples. I hope to have provided a serviceable theory of explication, but I also admitted that there are several aspects where improvement is possible.

Acknowledgements For very valuable comments I would like to thank all the participants of the advanced seminar on Theoretical Philosophy at the University of Greifswald and several anonymous reviewers. This considerably improved the text.

[45] This is true at least if the pre-explicative use refers to single authors writing on explication. Presumably the cumulative pre-explicative talk about explication by all authors writing on that topic is much more diverse than the expressions defined here. – Glatzer (2012) is an example for disambiguation by means of explication.

[46] This goes for some other desiderata, too. One reviewer drew my attention to the Carnapian desideratum of fruitfulness, specifically.

[47] One example: In (Reinmuth and Cordes, 2017, sect. 4-5) a formal language is introduced which includes 'Therefore' as an illocutionary operator. Its application to a formula does not yield a formula. So criteria of adequacy limited to object language formulas will not involve 'Therefore'. Still, this expression may be seen as an explicatum. Suitable explicanda are any natural language terms expressing acts of inference. Cf. (Hinst, 1982), (Siegwart, 2007a).

References

Brun G (2016) Explication as a method of conceptual re-engineering. Erkenntnis 81(6):1211–1241, DOI: 10.1007/s10670-015-9791-5

Carnap R (1934) Logische Syntax der Sprache. Springer, Wien

Carnap R (1942) Introduction to Semantics. Harvard University Press, Cambridge, Mass.

Carnap R (1945) The two concepts of probability. Philosophy and Phenomenological Research 5(4):513–532

Carnap R (1950) Logical Foundations of Probability. University of Chicago Press, Chicago

Carnap R (1963) The philosopher replies. In: Schilpp PA (ed) The Philosophy of Rudolf Carnap, Open Court, La Salle, Ill., pp 857–1013

Carus AW (2012) Engineers and drifters: The ideal of explication and its critics. In: Wagner P (ed) Carnap's Ideal of Explication and Naturalism, Palgrave Macmillan, Houndsmill, Basingstoke, Hampshire, pp 225–239

Cohnitz D, Rossberg M (2006) Nelson Goodman. Acumen, Chesham, Bucks

Cordes M (2016) Scheinprobleme. ein explikativer Versuch. Dissertation, Available online

Creath R (2012) Before explication. In: Wagner P (ed) Carnap's Ideal of Explication and Naturalism, Palgrave Macmillan, Houndsmill, Basingstoke, Hampshire, pp 161–174

Frege G (1953) The Foundations of Arithmetic. A logico-mathematical enquiry into the concept of number. Basil Blackwell, Oxford

Glatzer J (2012) Schönheit. Ein Klärungsversuch. Ontos, Frankfurt et al.

Greimann D (2007) Regeln für das korrekte Explizieren von Begriffen. Zeitschrift für philosophische Forschung 61(3):261–282

Hahn S (2013) Rationalität. Eine Kartierung. Mentis, Münster

Hanna JF (1968) An explication of 'explication'. Philosophy of Science 35(1):28–44

Hinst P (1982) Pragmatische Regeln des logischen Argumentierens. In: Gethmann CF (ed) Logik und Pragmatik. Zum Rechtfertigungsproblem logischer Sprachregeln, Suhrkamp, Frankfurt am Main, pp 199–215

Justus J (2012) Carnap on concept determination: Methodology for philosophy of science. European Journal for Philosophy of Science 2:161–179

Kitcher P (2012) Carnap and the caterpillar. In: Preludes to Pragmatism. Toward a Reconstruction of Philosophy, OUP, Oxford, New York, pp 192–209

Lambert JH (1771) Anlage zur Architectonic oder Theorie des Einfachen und des Ersten in der philosophischen und mathematischen Erkenntniß. Johann Friedrich Hartknoch, Riga, 2 volumes

Lutz S (2012) Criteria of Empirical Significance. Foundations, Relations, Applications, Quaestiones Infinitae. Publications of the Zeno Institute of Philosophy, vol LXX. University Utrecht, Utrecht

Paasch S (2016) Mehr vom Sinn als nichts: Eine Abhandlung über das Sinnvollsein von Ausdrücken. Dissertation

Pinder M (2016) The explication defence of arguments from reference. Erkenntnis Online First, DOI: 10.1007/s10670-016-9868-9

Quine WVO (1960) Word and Object. The Technology Press (MIT), John Wiley & Sons, New York, London

Reichenbach H (1951) The verifiability theory of meaning. Proceedings of the American Academy of Arts and Sciences 80(1):46–60

Reinmuth F (2013) Definitionslehre. Lecture notes for the seminar "Logische Propädeutik und Methodische Begriffsbildung B: Definitionslehre", Available online

Reinmuth F, Cordes M (2017) Commentary and illocutionary expressions in linear calculi of natural deduction. Logic and Logical Philosophy 26(2):163–196, DOI: 10.12775/LLP.2017.002

Schupbach JN (2015) Experimental explication. Philosophy and Phenomenological Research 94(3):672–710, DOI: 10.1111/phpr.12207

Shepherd J, Justus J (2015) X-Phi and Carnapian explication. Erkenntnis 80(2):381–402

Siegwart G (1997a) Explikation. Ein methodologischer Versuch. In: Löffler W, Runggaldier E (eds) Dialog und System. Otto Muck zum 65. Geburtstag, Academia, Sankt Augustin, pp 15–45

Siegwart G (1997b) Vorfragen zur Wahrheit. Ein Traktat über kognitive Sprachen. Oldenbourg, München

Siegwart G (2007a) Alethic acts and alethiological reflection. an outline of a constructive philosophy of truth. In: Greimann D, Siegwart G (eds) Truth and Speech Acts. Studies in the Philosophy of Language, Routledge, New York, pp 41–58

Siegwart G (2007b) Johann Heinrich Lambert und die präexplikativen Methoden. Philosophisches Jahrbuch 114:95–116

Sokal A, Bricmont J (1998) Fashionable Nonsense. Postmodern Intellectuals' Abuse of Science. Picador, New York

Strawson PF (1963) Carnap's views on constructed systems versus natural languages in analytic philosophy. In: Schilpp PA (ed) The Philosophy of Rudolf Carnap, Open Court, La Salle, Ill., pp 503–518

Tillman FA (1965) Explication and ordinary language analysis. Philosophy and Phenomenological Research 25(3):375–383

Wagner P (ed) (2012) Carnap's Ideal of Explication and Naturalism. Palgrave Macmillan, Houndsmill, Basingstoke, Hampshire

Wilson M (2012) The perils of Pollyanna. In: Wagner P (ed) Carnap's Ideal of Explication and Naturalism, Palgrave Macmillan, Houndsmill, Basingstoke, Hampshire, pp 205–224