



LUND UNIVERSITY

Broadening Explication

An interpretation and development of Carnap's method of explication

Österblom, Fredrik

2025

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Österblom, F. (2025). *Broadening Explication: An interpretation and development of Carnap's method of explication*. [Doctoral Thesis (monograph), Department of Philosophy]. Department of Philosophy, Lund University.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Broadening Explication

An interpretation and development of Carnap's
method of explication

FREDRIK ÖSTERBLOM

DEPARTMENT OF PHILOSOPHY | LUND UNIVERSITY





LUND
UNIVERSITY

Faculty of Humanities and Theology
Department of Philosophy

Serie
ISBN 978-91-90055-10-6



Broadening Explication

Broadening Explication

An interpretation and development of Carnap's
method of explication

Fredrik Österblom



LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Humanities and Theology at Lund University to be publicly defended on 13th of September at 10:15 in LUX C126, Helgonavägen 3, Lund

Faculty opponent
Professor Georg Brun

Organization: LUND UNIVERSITY

Document name: DOCTORAL DISSERTATION

Date of issue: 2025-08-15

Author: Fredrik Österblom

Title and subtitle: Broadening Explication: An interpretation and development of Carnap's method of explication

Abstract:

In this thesis I critically evaluate and develop the method of explication. To explicate a concept is, roughly, to replace it with a similar but more theoretically useful concept. The method was first articulated in the 1940s by Rudolf Carnap (1891–1970) but had been pursued in science and philosophy long before that. I begin the thesis with a critical evaluation of Carnap's own writings on explication, and comment on contemporary interpretations of Carnapian explication. I discuss the relative importance of the criteria of adequacy for explication, and the internal structure among them. I also give an overview of various versions and conceptions of explication beside Carnap's. As part of this task I will include accounts of concept formation in the social sciences which previously have not been treated as explications, and I argue that they may plausibly be understood as versions of explication. Most importantly, I will propose new ways to develop and modify Carnap's method. The modifications are partly based on insights from the literature on concept formation in the social sciences. In this sense I am broadening explication. Although I adapt Carnap's method for purposes beyond those intended by him, my project is in certain aspects in the spirit of Carnap. While I modify and supplement his criteria of adequacy, I retain the ideal of a common standard of evaluation for concepts in distinct fields of inquiry. The overall aim is to shed light on the question of how we should deal with concepts that are defective (at least in some contexts) and yet indispensable to our cognitive lives.

Key words: Explication, Conceptual Engineering, Carnap, Philosophical Methodology, Concept Formation

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

Number of pages: 222

ISSN and key title:

ISBN (print) 978-91-90055-10-6

ISBN (digital) 978-91-90055-11-3

Recipient's notes

Price

Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date 2025-07-22

Broadening Explication

An interpretation and development of Carnap's
method of explication

Fredrik Österblom



LUND
UNIVERSITY

Coverphoto by: Kazimir Malevich, *Suprematist Composition: White on White*, 1918, Museum of Modern Art, New York, Public Domain

© Fredrik Österblom

Faculty of Humanities and Theology

Department of Philosophy

Lund University

ISBN (print) 978-91-90055-10-6

ISBN (digital) 978-91-90055-11-3

Printed in Sweden by Media-Tryck, Lund University

Lund 2025



Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN 

Till Mia och Henrik

Table of Contents

| | |
|---|-----------|
| Acknowledgements..... | 13 |
| Abstract | 15 |
| Abbreviations of Carnap’s works | 16 |
| 1 Introduction | 17 |
| 1.1 Opening remarks..... | 17 |
| 1.2 Carnap’s views on explication in brief summary | 20 |
| 1.3 How I use the term ‘explication’ | 22 |
| 1.4 Examples of explication..... | 24 |
| 1.4.1 Examples of explication in empirical science..... | 24 |
| 1.4.2 Examples of explications in philosophy and mathematics | 25 |
| 1.5 Why spend time on interpretations of Carnap? | 27 |
| 1.6 Explication beyond Carnap | 29 |
| 1.7 The objectives of the thesis | 30 |
| 1.8 My position in relation to extant literature on explication | 31 |
| 1.9 The plan of the thesis | 32 |
| 2 Carnapian explication..... | 34 |
| 2.1 Overview | 34 |
| 2.2 Before explication: rational reconstruction | 34 |
| 2.3 Carnap’s writings on explication | 37 |
| 2.4 The structure and motivation of explication..... | 40 |
| 2.5 The two steps..... | 42 |
| 2.5.1 The first step: clarification of the explicandum | 42 |
| 2.5.2 The second step: the specification of the explicatum | 43 |
| 2.6 Three kinds of concepts | 45 |
| 2.7 The four criteria of adequacy | 49 |
| 2.7.1 The similarity criterion | 49 |

| | | |
|----------|---|-----------|
| 2.7.2 | The exactness criterion | 53 |
| 2.7.3 | The fruitfulness criterion | 57 |
| 2.7.4 | The simplicity criterion | 59 |
| 2.8 | Concepts versus terms | 60 |
| 2.9 | The ideal of scientific philosophy in Carnap's conception of explication | 62 |
| 2.9.1 | Explication as the method of scientific philosophy | 62 |
| 2.10 | Concluding remarks | 66 |
| 3 | Early commentary and debate on Carnap's method | 67 |
| 3.1 | Overview | 67 |
| 3.2 | The exchange between Strawson and Carnap | 67 |
| 3.3 | Early writing on explication | 72 |
| 3.4 | Renewed interest in explication | 74 |
| 3.5 | Concluding remarks | 76 |
| 4 | The purpose of explication | 77 |
| 4.1 | Overview | 77 |
| 4.2 | Immediate and ultimate purpose | 78 |
| 4.3 | The relative weight of the criteria in light of my distinction ... | 81 |
| 4.3.1 | The importance of exactness | 86 |
| 4.4 | The internal structure of the criteria of adequacy | 88 |
| 4.4.1 | The internal structure of Carnap's account | 88 |
| 4.4.2 | Brun's interpretation of the internal structure | 89 |
| 4.4.3 | The internal structure of Brun's account | 94 |
| 4.5 | Concluding remark | 95 |
| 5 | An overview of different versions of explication | 96 |
| 5.1 | Overview | 96 |
| 5.2 | Oppenheimian explication: explication for similarity and exactness | 97 |
| 5.3 | Explication for exactness | 99 |
| 5.4 | Explication for exactness and fruitfulness | 100 |
| 5.5 | Operational explication: explication for empirical testability | 100 |
| 5.6 | Explication for Gerring's eight criteria | 102 |
| 5.7 | Explication for the Brülde–Tengland criteria of adequacy ... | 104 |
| 5.8 | Concluding remarks | 107 |

| | | |
|----------|---|------------|
| 6 | An overview of conceptions of explication | 108 |
| 6.1 | Overview | 108 |
| 6.2 | The main divide: is there a privileged language? | 108 |
| 6.3 | Explication in a privileged language..... | 109 |
| 6.3.1 | Does explication require Carnapian tolerance?..... | 109 |
| 6.3.2 | Carnapian language engineering: proposing linguistic conventions for the utility of empirical science | 111 |
| 6.4 | Quine's conception of explication: simplification of total theory 111 | |
| 6.4.1 | Three kinds of defective nouns | 113 |
| 6.4.2 | Quinean explication..... | 116 |
| 6.4.3 | Differences between Carnapian and Quinean explication..... | 121 |
| 6.5 | Concluding remarks | 121 |
| 7 | The place of explication in the field of conceptual engineering ... | 122 |
| 7.1 | Overview | 122 |
| 7.2 | The history of the term 'conceptual engineering' | 124 |
| 7.3 | Conceptual engineering, conceptual ethics and ameliorative analysis..... | 127 |
| 7.4 | Two paradigmatic examples of conceptual engineering | 130 |
| 7.5 | Motivations for conceptual engineering..... | 131 |
| 7.6 | The implementation challenge and conceptual activism..... | 132 |
| 7.7 | The continuity challenge | 135 |
| 7.8 | Concluding remarks | 137 |
| 8 | Broadening the fruitfulness criterion | 139 |
| 8.1 | Overview | 139 |
| 8.2 | Problems with Carnap's notion of fruitfulness | 140 |
| 8.2.1 | True universal statements | 140 |
| 8.2.2 | Explanation | 141 |
| 8.2.3 | Carnap's view of cognitive significance..... | 142 |
| 8.2.4 | Not only exceptionless generalizations are useful for prediction and explanation..... | 144 |
| 8.2.5 | To detect laws is not the only aim of science | 145 |
| 8.3 | Alternative criteria of fruitfulness..... | 146 |
| 8.3.1 | Overview of the accounts | 146 |

| | | |
|-----------|--|------------|
| 8.3.2 | New knowledge..... | 146 |
| 8.3.3 | Answers to significant questions..... | 148 |
| 8.3.4 | Approaching relevant theoretical goals | 149 |
| 8.4 | Contextualism versus generalism | 150 |
| 8.5 | Concluding remarks..... | 152 |
| 9 | Towards a modified version of explication..... | 153 |
| 9.1 | Overview..... | 153 |
| 9.2 | Locating my account in the literature | 153 |
| 9.2.1 | Explication of social concepts..... | 153 |
| 9.2.2 | How my account deviates from Brun..... | 155 |
| 9.3 | The proposed modifications of the criteria of adequacy | 158 |
| 9.4 | Defining criteria..... | 162 |
| 9.4.1 | Similarity | 162 |
| 9.4.2 | Explicitness..... | 163 |
| 9.5 | Good-making criteria..... | 164 |
| 9.5.1 | Sharpness..... | 164 |
| 9.5.2 | Connectedness..... | 165 |
| 9.5.3 | Fruitfulness | 168 |
| 9.5.4 | Intersubjectivity | 169 |
| 9.5.5 | Simplicity..... | 171 |
| 9.6 | How to evaluate the standards of evaluation..... | 172 |
| 9.7 | Concluding remarks..... | 173 |
| 10 | Evaluating my proposal with an example | 175 |
| 10.1 | Overview..... | 175 |
| 10.2 | Explicating the concept <i>democratic</i> | 175 |
| 10.3 | Is explication the best strategy to deal with the concept <i>democratic</i> ? | 177 |
| 10.4 | First step: clarifying the explicandum..... | 181 |
| 10.4.1 | From noun to gradable adjective..... | 181 |
| 10.4.2 | Generality requirement..... | 184 |
| 10.4.3 | Normative neutrality..... | 184 |
| 10.4.4 | Logical form and domain..... | 185 |
| 10.4.5 | Why does it matter how the D-words are used in ordinary language? | 186 |
| 10.5 | Second step: specifying the explicatum | 186 |
| 10.6 | Evaluation of the explicatum..... | 187 |

| | | |
|-----------|--|------------|
| 10.6.1 | Defining criteria: similarity and explicit rules of use | 187 |
| 10.6.2 | First good-making criterion: sharpness | 188 |
| 10.6.3 | Second good-making criterion: connectedness..... | 188 |
| 10.6.4 | Third good-making criterion: fruitfulness | 189 |
| 10.6.5 | Fourth good-making criterion: intersubjectivity | 189 |
| 10.6.6 | Fifth good-making criterion: simplicity..... | 190 |
| 10.6.7 | Summary of the evaluation..... | 191 |
| 10.7 | Comparisons..... | 191 |
| 10.8 | Concluding remarks..... | 192 |
| 11 | Explication as a philosophical method | 193 |
| 11.1 | Overview | 193 |
| 11.2 | The standard model of philosophical methodology | 193 |
| 11.3 | The standard model of philosophical explication | 198 |
| 11.4 | What is the motivation for philosophical explication?..... | 199 |
| 11.5 | Should explication exhaust the philosophical toolbox? | 202 |
| 11.6 | How can explication tackle philosophical problems?..... | 203 |
| 11.7 | The broadened account of explication as a philosophical method..... | 206 |
| 11.8 | Open questions and future research | 209 |
| | References | 211 |

Acknowledgements

I would like to begin by expressing my deep gratitude to my main supervisor, Tobias Hansson Wahlberg. It has been a luxury to have such a knowledgeable, engaged, and caring supervisor. I am grateful for all the time and effort you have so generously invested in reading and commenting on my work, and for taking the initiative to regularly meet and discuss it. Your comments and advice have been invaluable. Thank you, Tobias!

Thank you also to my assistant supervisor Erik J. Olsson. I have learned so much from your work and from your insightful feedback and from our discussions on explication. Thank you for your warm generosity as a supervisor, personally and as the chair of the Higher Seminar.

I am thankful to all participants of the Higher Seminar in Theoretical Philosophy for constructive and insightful comments on my presentations and writings. A special thanks, for closely reading and writing comments on my work, to Hubert Hågemark, Jiwon Kim, Andreas Stephens, Asger Kirkeby-Hinrup, Balder Ask Zaar, Anton Emilsson, and Frits Gävertsson.

I want to thank Martin Jönsson for all the insightful comments on my work and the words of encouragement along the way. Thank you also for organizing lovely dinners and game nights, and lunches at the department, and for creating a sense of community in the subject and at the department.

I am grateful to our fantastic administrative staff for all the support you have provided me over the years. Thanks to Anna Östberg, Annah Smedberg-Eivers and Anna Cagnan Enhörning.

The PhD seminar on Friday afternoons has been a forum for valuable feedback and lively conversations, as well as great source of fun and camaraderie during the years. The conversations have often continued late into the night at Inferno or at one of the many dinner parties that Niklas Dahl has generously hosted. Thank you to all participants!

I would like to thank Max Minden Ribeiro, Hubert Hågemark, Melina Tsapos, Jakob Stenseke, Niklas Dahl, and Andreas Stephens for making B478 into a lovely office (and at times seminar room or living room). I also want to thank my “classmate” Jiwon Kim for being such an impressive and supportive and fun colleague and friend, as well as a literal rock star.

Thank you to everyone in the broader community of PhD students (and former and future ones as well) at the department. You are all incredibly talented, funny, and kind, and I will cherish the years I got to spend in this community for the rest of my life. Thank you: Max Minden Ribeiro, Hubert Hågemark, Melina Tsapos, Jakob Stenseke, Niklas Dahl, Andreas Stephens,

Jiwon Kim, Balder Ask Zaar, Martin Sjöberg, Anton Emilsson, Marianna Leventi, Shervin Mirzaeighazi, Jenny Magnusson, Alexander Velichkov, Vidar Bratt, Agnès Baehni, Carl-Johan Palmqvist, August Olsen, Robert Pål-Wallin, Elsa Magnell, Signe Savén, Samantha Stedtler, Ellen Davidsson, Frits Gåvertsson, Sarah Koeglsperger, Mark Bowker, Marta Johansson Werkmäster, Gloria Mähringer, Anton Wrisberg, Mattias Gunnemyr, Trond Arild Tjøstheim, Maybí Morell Ruiz, Thibault Boehly, Pierre Klintefors, Alexander Tagesson, Matthew Tompkins, Lona Lalic, Simon Grendeus, Amandus Krantz, and many more.

Moving on to more personal debts of gratitude, I want to thank Benjamin Aspelin for all the quests for knowledge of the most varying kinds that we've embarked on together since our youth. It has been one of the great joys of my life to analyse and figure out the world and its inhabitants together with you, and I have benefited greatly from spending all that time with your superior mind. I want to thank my friends in Gothenburg and Helsinki for long and rewarding friendships and for inspiring me by being so annoyingly productive and brilliant writers and researchers. Thank you Per Oleskog Tryggvason, Johan Alfonsson, Axel Kronholm, Jakob Sandberg, Evelina Johansson Wilén, Carl Wilén, Vanja Carlsson, Mattias Lehtinen, Valter Sandell-Maury, Olivia Maury and Ina Kauranen. Thank you to Emil Wikström for the many years of rewarding friendship and conversations. My life has been immensely enriched by your unfailing wit.

It is with never-ending gratitude that I dedicate this book to my mother Maria Malin and my father Henrik Österblom. I do not know where to begin thanking you for all your love and support. I have truly got to lead the life of my dreams, and in so many distinct ways it is all thanks to you.

Thank you to my sister Joanna Österblom, my favourite sister, for your friendship and all the love and support you have given me. Thank you for all the long nights of conversation when we have analysed the world together and explored ideas, with the invigorating sense of urgency you get when things are really at stake.

Finally, and most of all, I want to thank Vilma Nyberg. Since the day I met you, I've been happy. Thank you for sharing your life with me and thank you for your patience with me in these hectic and challenging times. I look forward to all the times to come. You are the light of my life, and I love you.

Abstract

In this thesis I critically evaluate and develop the method of explication. To explicate a concept is, roughly, to replace it with a similar but more theoretically useful concept. The method was first articulated in the 1940s by Rudolf Carnap (1891–1970) but had been pursued in science and philosophy long before that. I begin the thesis with a critical evaluation of Carnap’s own writings on explication, and comment on contemporary interpretations of Carnapian explication. I discuss the relative importance of the criteria of adequacy for explication, and the internal structure among them. I also give an overview of various versions and conceptions of explication beside Carnap’s. As part of this task I will include accounts of concept formation in the social sciences which previously have not been treated as explications, and I argue that they may plausibly be understood as versions of explication. Most importantly, I will propose new ways to develop and modify Carnap’s method. The modifications are partly based on insights from the literature on concept formation in the social sciences. In this sense I am broadening explication. Although I adapt Carnap’s method for purposes beyond those intended by him, my project is in certain aspects in the spirit of Carnap. While I modify and supplement his criteria of adequacy, I retain the ideal of a common standard of evaluation for concepts in distinct fields of inquiry. The overall aim is to shed light on the question of how we should deal with concepts that are defective (at least in some contexts) and yet indispensable to our cognitive lives.

Abbreviations of Carnap's works

- Aufbau* Carnap, Rudolf. 1967 [1928]. *Der Logische Aufbau der Welt*. Berlin-Schlachtensee: Weltkreis-Verlag, 1928. Translated as *The Logical Structure of the World* by Rolf A. George. Original edition, Los Angeles: University of California Press. Reprint, Chicago and La Salle, IL: Open Court, 2003.
- LFP Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: Chicago University of Chicago Press.
- M&N Carnap, Rudolf. 1956 [1947]. *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd ed. Chicago: University of Chicago Press.
- RSE Carnap, Rudolf. 1963b. "Replies and Systematic Expositions." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 859-1013. LaSalle, IL: Open Court.

1 Introduction

1.1 Opening remarks

To explicate a concept is, roughly, to replace it with a similar but more theoretically useful concept. In this thesis I critically evaluate and develop the method of explication, which was first articulated in the late 1940s and early 1950s by Rudolf Carnap (1891–1970).¹ A Carnapian explication is the replacement of a concept that is already in use, either in ordinary life or in science, with a substitute that has some similarity to the old concept but is more exact and fruitful and as simple as the first requirements permit. Although explications are omnipresent in the history of science and philosophy, Carnap was the first to articulate explicit principles and criteria for how to perform explications well.² As Erich Reck points out, “until recently Carnap was one of very few thinkers who elaborated on methodological aspects [concerning explication]; hence the value of his corresponding remarks” (Reck 2024, 128). This is the main motivation for my focus on Carnap in the thesis. Although Carnap arguably practiced explication before he coined the term, it was in an 1945 paper on probability that he first began to describe his own philosophical

¹ Carnap borrows the term ‘explication’ from Kant and Husserl, but he clarifies that he uses the term in a different sense: “The procedure of explication is here understood in a wider sense than the procedures of analysis and clarification which Kant [and] Husserl [...] have in mind” (LFP, 3). Beaney (2007) notes that “[f]or Kant, ‘explication’ simply meant the unpacking of a complex concept into its constituent concepts” (Beaney 2007, 213). He also notes that “Carnap’s apparent acknowledgement of Husserl’s influence is misleading” (ibid.), since “[b]oth Carnap’s and Husserl’s conceptions involve the idea of precisification, but while for Carnap this involves ‘replacing’ our vague ordinary concepts with scientifically defined ones, for Husserl we work within our ordinary understanding to elucidate its essential structures” (ibid.). For a thorough discussion of the issue, see Beaney 2004.

² It should be mentioned that Cordes and Siegwart (2018) trace what they take to be explicit remarks on the method of explication back to Johann Heinrich Lambert (1728–1777), although Lambert did not use the term ‘explication’. They claim that “Lambert merits special mention as his remarks predate what is usually considered the starting point of the systematic treatment of explication [Carnap’s LFP] by over 170 years” (Cordes and Siegwart 2018).

method as ‘explication’ (1945b, 513).³ Shortly thereafter he wrote elaborately about the method in relation to his work on semantics and inductive logic.

During the last decade there has been a surge of interest in Carnapian explication, in the wake of a more general reappraisal of Carnap’s views and as a part of the surge of interest in conceptual engineering more generally (see chapter 7).⁴ I will critically evaluate not only Carnap’s own writings on explication but also other philosophers’ interpretations of Carnapian explication, as well as other philosophers’ attempts to improve and develop the method. I will also give an overview of the various versions and conceptions of explication beside Carnap’s. As part of this task I will include notable contributions to the methodological literature on concept formation in the social sciences and reformulate methods of concept formation in terms of explication. It is natural to look at some of the tasks of concept formation in the social sciences as forms of explication, and it could bring clarity to think of them explicitly in such terms. Most importantly, I will propose new ways to develop and modify Carnap’s method, partly based on criteria and discussions of concept formation in the social sciences. I will incorporate criteria from these methods to complement Carnap’s method. This is one of the ways in which I am broadening explication.

There is also a related secondary aim to the thesis, which is to unify conceptual standards across different areas of inquiry, and to make connections between hitherto unconnected methodological work on concept formation and concept revision. In the thesis, I make connections between two bodies of literature which, despite having important things in common, have rarely been treated together.⁵ On the one hand there are Carnap’s (and like-minded philosophers’) writings on explication; on the other hand there are writings by social scientists and political theorists on criteria for good concepts in the

³ Carus judges that “[t]hough [Carnap] does not mention explication by name until 1945, and does not describe it in any detail until 1950, all the tools required for it are essentially available by the late 1930s” (Carus 2007, 37–38).

⁴ A general re-evaluation of logical empiricism began in the 1980s and took off in the 1990s, with a heavy focus on Carnap’s works. See Limbeck-Lilienau and Uebel 2022 for an overview. The rise in historical interest in Carnap and logical empiricism can also be seen as a part of a more general ‘historical turn’ in analytic philosophy (see for example Reck 2013a and Lapointe and Pincock 2017).

⁵ A recent exception is Reuter, Herfeld, and Brun (2023), and the contributions to their topical collection in *Synthese* on concept formation in the natural and social sciences. Another recent exception is Cappelen (2023), whom I draw heavily on as well as criticize in chapter 10.

social sciences.⁶ The methodology of concept formation is an area of inquiry characterized by an ongoing exchange between philosophy and the sciences. Instructions for concept formation in the special sciences, particularly the social sciences, are often informed by philosophical work on definitions and conceptual clarification and analysis. In the other direction, philosophers inspired by the successes of natural science, such as Carnap, have looked to the practices of concept formation in natural and mathematical science in their quest to improve the methods of their own trade. In the present thesis, too, I use work on concept formation to modify Carnap's criteria. In the other direction, I use insights from philosophical work on explication and conceptual engineering to propose a new way to understand some of the tasks of concept formation in the social sciences.

When a concept causes trouble in contexts of inquiry, there are different possible strategies to deal with it (see 10.3 for a discussion of such strategies). It is possible that the best way to deal with some concepts is to abandon or eliminate them altogether (for example, Cappelen advocates abandonment of the concept *democracy*, as we will see in chapter 10). For other concepts, the best option may be to improve them (as I attempt to do with the concept *democratic*, also in chapter 10). It is presumably easier to improve concepts for the purpose of inquiry if we have explicit criteria for what makes a concept good for inquiry. This is where my suggested modifications of the criteria of adequacy enters the picture. I hope that my thesis will shed new light on how we should deal with concepts that are troublesome and yet seem indispensable to our cognitive lives.

To wrap up my opening remarks, I quote Sven Ove Hansson's description of the central role of analyses and definitions in the philosophical skill set and of how that skill set can be useful in other fields:

Careful analysis and development of our own terminology is an essential part of modern philosophy. Definitions and conceptual analysis provide us with philosophical tools in the form of precise concepts that can be used in philosophical arguments. In addition, definitions are an important part of our contributions to other disciplines. In interdisciplinary co-operations, it is often the role of philosophers to work out precise definitions and distinctions. (Hansson 2006, 5)

⁶ I have in mind work by authors such as Felix Oppenheim (1961, 1981), Giovanni Sartori (1970, 1984), John Gerring (1999, 2001, 2011), Gary Goertz (2006, 2020) and David Collier (Collier and Gerring 2009).

The method of explication supplies the link between analyses of concepts that are already in use and definitions of new, technical concepts. To understand that link is important for the improvement of our own terminology in philosophy, and such an understanding could also be an important part of the contributions that philosophers bring to other disciplines. Before I present my project in more detail, I give a brief presentation of Carnap's account of explication.

1.2 Carnap's views on explication in brief summary

Carnap took methods of concept revision in science and mathematics as a model for his preferred way of improving philosophical concepts. He emphasized that the method he described and prescribed is used frequently by philosophers, scientists, and mathematicians, but that they “do not often discuss explicitly the general rules which they follow implicitly” (LFP, 7). In *Meaning and Necessity* (hereafter M&N), which was first published in 1947, Carnap describes explication as “the task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept” (M&N, 7-8). Note that in his brief characterizations of the method, such as the one just quoted, Carnap only mentions that the new concept should be more exact and does not mention other criteria, such as fruitfulness. Carnap also introduces the terminology, which I will use continually in this thesis, in which the old concept is called the *explicandum* (pl. explicanda) and the new concept is called the *explicatum* (pl. explicata):

We call this the task of explicating, or of giving an **explication** for, the earlier concept; this earlier concept, or sometimes the term used for it, is called the **explicandum**; and the new concept, or its term, is called an **explicatum** of the old one. (M&N, 8, original emphasis)

Carnap's most in-depth exposition of explication is to be found in the first chapter of his 1950 book *Logical Foundations of Probability* (hereafter LFP). In this work he specifies his four criteria for an adequate explicatum, viz. similarity, exactness, fruitfulness, and simplicity:

- (1) The explicatum is to be **similar** to the explicandum in such a way that, in most cases in which the explicandum has so far been used, the

explicatum can be used; however, close similarity is not required, and considerable differences are permitted.

- (2) The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an **exact** form, so as to introduce the explicatum into a well-connected system of scientific concepts.
- (3) The explicatum is to be a **fruitful** concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept).
- (4) The explicatum should be as **simple** as possible; this means as simple as the more important requirements (1), (2), and (3) permit. (LFP, 7, my emphasis)

In an explication of a concept, the old concept is replaced—at least in the relevant theoretical contexts—by a new concept which is provided with explicit rules for its use. The new concept should according to Carnap be similar to the old one in the sense that it can be used—again, in relevant contexts—in most cases where the old concept was used. According to Carnap, the deviations from the old concept are motivated by gains in exactness and fruitfulness. If two or more candidate explicata are equally exact and fruitful Carnap recommends choosing the simplest concept.

In section 1.4, I give several examples of explication in various fields. The examples that Carnap himself gives of paradigmatic explications performed by philosophers are:

- Gottlob Frege’s and Bertrand Russell’s analyses of number terms (e.g., “two” defined as the class of pair-classes). (M&N, 8; RSE, 935; LFP, 17)
- Frege’s and Russell’s analyses of phrases involving the definite article (phrases of the form “the so-and-so”). (M&N, 8)
- Alfred Tarski’s semantical definition of the concept of truth. (M&N, 8; LFP, 5)

However, it is an open question whether all of these examples actually are to be considered as examples of the method of explication rather than other forms

of analyses or definitions.⁷ Carnap simply calls these examples ‘explications’, but he does not address the fact that until that point they had been described as ‘analyses’ or ‘definitions’, including by Frege, Russell, and Tarski themselves.⁸ Definitions can be and often are involved in explications, but it should be emphasized that an explication is not merely a definition. When Carnap first introduced the term ‘explication’, he described it as a “redefinition” (Carnap 1945b, 513), although later he clarifies that in an explication it is the explicatum that is being defined, not the explicandum. He writes in LFP that “if the explication consists in giving an explicit definition, then both the definiens and the definiendum in this definition express the explicatum, while the explicandum does not occur” (LFP, 3). The structure of explications will be clarified further in chapter 2.

1.3 How I use the term ‘explication’

I will use the term ‘explication’, when used without the prefix ‘Carnapian’, in a broad sense that includes but is not restricted to Carnapian explication. In that broad sense, I use ‘explication’ for *any method whereby a concept is deliberately replaced, at least in certain contexts of inquiry, with a similar concept which is judged to be better for the purposes of inquiry.*

This usage excludes for example conceptual analysis and cases of concept replacement for legal or political purposes, but it includes methods of concept replacement for theoretical purposes that differ from Carnapian explication. In chapter 5 I give an overview of different versions of explications.

In addition to the mentioned terms ‘analysis’ and ‘definition’, there is a cluster of terms in the vicinity of explication. Here is a list of such terms:

- analysis
- logical analysis
- logical construction
- logical reconstruction

⁷ See Reck’s 2007 for a discussion of whether the Frege–Russell conceptions of numbers should be regarded as analyses (in the strong sense of being the right construction of numbers) or as explications.

⁸ In chapter 6 I address these issues further.

- rational reconstruction
- definition
- redefinition
- precisising definition
- stipulative definition
- conceptual clarification
- conceptual determination
- conceptual revision
- conceptual engineering
- conceptual re-engineering
- language engineering
- linguistic engineering
- conceptual reconstruction
- ameliorative analysis

Some of these terms have been used as synonyms for ‘explication’ or have been used to describe explication. Some of them are meant to pick out different methods in the vicinity of explication or even competing methods. Unfortunately, many of them have been used in both of these ways (e.g., ‘rational reconstructions’ and ‘conceptual clarification’). During the course of the thesis, I will clarify the senses in which most of these terms are related to the term ‘explication.’

1.4 Examples of explication

1.4.1 Examples of explication in empirical science

Carnap discusses examples of explication in a broad range of fields. His examples from the empirical sciences are the concepts *fish*⁹ (1950, 5-6), *temperature* (1950, 9-15), and *IQ* (Carnap 1945b, 516). The first example consists in the replacement of ‘fish’ in the sense of sea creature (which includes sea mammals such as whales and dolphins) with ‘fish’ in the sense of cold-blooded gill-bearing aquatic vertebrate¹⁰ (which excludes, e.g., sea mammals such as whales and dolphins). The second example consists in the replacement of the classificatory concept *warm*, first by the comparative concept *warmer* and finally by the quantitative concept *temperature*. The third example consists in the replacement of the comparative concept *higher intelligence* by the quantitative concept *IQ*.

A fairly recent and very public example of explication in science is the explication of ‘planet’; besides Carnap’s own examples, this is a particularly oft-mentioned example of explication or conceptual engineering. In 2006 the International Astronomical Union (IAU) decided on a definition of the term ‘planet’, which Mauro Murzi (2007) was quick to discuss as an example of explication. Murzi comments that the “situation is very interesting for the philosopher of science who can be eyewitness of a real process of explication of a scientific concept” (Murzi 2007, 1). In his 2014, Peter Ludlow discussed the same example, but without reference to Murzi or to the method of explication. Since then it has been widely used as an example of explication, e.g., by Cordes and Siegwart (2018) and—with reference to Cordes and Siegwart—by Olsson (2021). In recent literature it is most extensively treated by Mark Pinder (2022b) in a paper on fruitfulness which I will engage with in chapter 8.

At the 26th General Assembly of the IAU, the assembly agreed upon a definition of planet as a “a celestial body that

⁹ I am following Brun (2016, 1214, n. 6) in using italics to indicate that I do not refer to, in this case, the animals labelled by the term ‘fish’ but to the concept of fish. Carnap uses upper-case initial letters (Fish) for the same purpose.

¹⁰ The definition given by Carnap is now obsolete. Justus comments on the example, with reference to Ereshefsky (2000), that: “[f]rom a contemporary perspective, this example is rather dated. Phylogeny, rather than phenotypic similarity, is the primary reason mammals are excluded from *Piscis* today” (Justus 2012, 164).

- (a) is in orbit around the Sun,
- (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and
- (c) has cleared the neighbourhood around its orbit” (IAU 2006).

To the outrage of parts of the public sphere this decision demoted Pluto from planethood since Pluto does not fulfil (c).¹¹ (For opposition to the definition by planetary scientists, see Metzger et al. 2020.)

What motivated the explication? This is how the IAU motivates the need for a new definition:

Contemporary observations are changing our understanding of planetary systems, and it is important that our nomenclature for objects reflect our current understanding. This applies, in particular, to the designation “planets”. The word “planet” originally described “wanderers” that were known only as moving lights in the sky. Recent discoveries lead us to create a new definition, which we can make using currently available scientific information. (IAU 2006)

1.4.2 Examples of explications in philosophy and mathematics

I now present a few examples of explications in philosophy. Although philosophers may have tended to be more explicit about their explications than scientists, Carnap remarks that they tend not to spell out the general rules they follow:

Philosophers, scientists, and mathematicians make explications very frequently. But they do not often discuss explicitly the general rules which they follow implicitly. (Carnap 1950, 7)

As late as 2007, Theo Kuipers made the same observation, writing that “[a]lthough the term ‘explication’ is not often used by philosophers, it is clear that [...] they are practicing concept explication in a more or less explicit and rigorous way” (Kuipers 2007a, vii). He goes on to reflect on reasons for “the reluctance to use the word ‘explication’”, and identifies at least three reasons:

First, the word itself may be found a bit too affected. Second, making the application of the method explicit may not only lead to rather cumbersome texts, but also appear to be a difficult task. Finally, many philosophers do not

¹¹ For a popular account of these events written by one of the main protagonists, see Tyson 2009.

like to be associated with logical empiricists that introduced ‘(concept) explication’ around 1950 as a technical term for this philosophical method, viz. Rudolf Carnap and Carl Hempel. (Kuipers 2007a)

Despite the prevalence of explications in philosophy, I will stick to the examples mentioned by Carnap, partly to avoid controversial questions of interpretation at this stage and partly because Carnap’s own examples are frequently discussed in the literature on explication and will appear in discussions throughout the thesis.

In M&N, Carnap gives non-empirical examples of what he considers to be explications, such as the previously mentioned Frege’s and Russell’s conceptions of the natural numbers in terms of classes, Frege’s and Russell’s analyses of definite descriptions, and Tarski’s semantical definition of truth (M&N, 8).¹² Carnap also mentions the Vienna Circle mathematician Karl Menger’s (1943) definition of ‘dimension’ as an example of explication, and he cites Menger as someone who has explicitly discussed general rules to follow for redefinitions of scientific concepts (LFP, 7).

Notable among Carnap’s own explications are, e.g., the explications of ‘degree of confirmation’, ‘analytic truth’, and ‘empirical meaningfulness of theoretical terms’ (Carnap 1956, 49). As pointed out by Jonah Schupbach (2017), Carnap suggests that symbolic logic may be understood as a series of explications. While Carnap only mentions metalogical concepts as examples of logical explications, Schupbach notes that the logical constants may also be considered as explications of the corresponding ordinary language expressions:

Carnap [1958, 2] suggests that the whole project of symbolic logic essentially involves a series of explications. This is true when it comes to metalogical concepts like Entailment—thus, Carnap defines a formally precise concept of L-Implication (Logical Implication) and then he writes: “L-Implication is our explication for the traditional concept which is usually called ‘implication’ or ‘logical implication’ or ‘entailment’, and whose inverse is ordinarily referred to by such terms as ‘logical consequence’, ‘deducibility’ and the like” (p. 20). But the formal concepts defined as part of the logic itself may also be seen as explications. Accordingly, one might interpret the formal concepts designated by the connectives, Λ , \vee , and \rightarrow as explications of the concepts of Negation, Conjunction, Disjunction, and Conditionality as used in ordinary language. (Schupbach 2017, 676)

¹² As previously mentioned, it is a matter of debate whether these projects should be understood as analyses or explications (and what the difference is).

1.5 Why spend time on interpretations of Carnap?

For the aims of this thesis, there are some philosophical benefits to be gained from engaging with the history of philosophy, specifically work on Carnap and logical empiricism. In the paragraph quoted below, Erich Reck suggests some of the philosophical benefits of historical work:

Why might an analytic philosopher want to ‘do analytic philosophy historically’? Why all the extra effort, in other words? One quick answer is: to counteract the distortions and other limiting effects of stereotypes, quasi-historical legends, and rational reconstructions. Another answer has already come up as well: to de-familiarize us from our current concepts and assumptions so as to be able to question them fruitfully. Further benefits include: to recognize fine, otherwise missed nuances in the views of past philosophers, especially ones that can still play a role today; to understand better, and partly to recover, what the agenda of analytic philosophers was, is, and should be; to motivate corresponding projects further, by becoming aware of their original contexts, their developments, and possible extensions; and to evaluate more accurately the importance, as well as the limits, of the results that have already been achieved. (Reck 2013, 13)

These are all good reasons to engage more closely with historical texts. The aim is to recover and better understand the agenda in developing explication as a philosophical method, e.g., the broader philosophical agenda of logical empiricism. As summarized by Alan Richardson (2008, 90), “[l]ogical empiricism was, in the first instance, a project meant to secure the scientific status of philosophy, to find a place for philosophy within a scientific culture.” Despite major changes in Carnap’s philosophical views the ideal of scientific philosophy is a constant in his body of works. It is present already in Carnap’s first major book *Der Logische Aufbau der Welt* (The Logical Structure of the World), hereafter *Aufbau*, which was published in 1928. As noted by Richardson: “In both its motivational remarks and its actual procedures, Carnap’s [*Aufbau*] exhibits a concern with a metaphilosophical project, a project of bringing conceptual standards already extant in the sciences into philosophy” (Richardson 2008, 90).¹³ The same concern takes a new shape in Carnap’s later writings on explication.

The historical interest I take in Carnap and his contemporaries comes both from a wish to counter misapprehensions about Carnap’s views and from a wish to appropriate what is valuable and viable in their writings. The task of

¹³ See also Richardson’s 2023.

interpretation is a means for the task of appropriation. The terminology of *interpretation* and *appropriation*, in this context, was introduced by James Andrew Smith (2021), who relates these terms specifically to contemporary work on Carnap. As Smith observes, these two approaches are commonly combined by philosophers who engage with Carnap's work:

In considering the work of a philosopher on a given topic, we can ask two different questions. We can ask what I will call an interpretation question: what are the philosopher's views on this topic? We can also ask what I will call an appropriation question: are there basic ideas in the philosopher's work on this topic—ideas at least similar to the philosopher's actual views—which are defensible or at least worthy of consideration? Many philosophers today wish to answer both types of questions about the work of Rudolf Carnap. His work forms a starting point for contemporary metaontology and the currently burgeoning literature on conceptual engineering, to give two examples. Contemporary philosophers engaged in such projects want to understand Carnap's work as well as consider or defend basic ideas he has to offer. (Smith 2021, 1)

I too embrace the dual aims described by Smith. I want both to understand Carnap's actual views and motivations and to critically evaluate them and appropriate what is viable and promising. As noted by Reck in the paragraph quoted earlier, a reason to go back to the work of historical philosophers is to counteract caricatured and skewed portrayals of their views. Since this thesis is concerned with a method which originated in the logical empiricist movement, and which was the main methodology of its surviving members from the 1940s onwards, a natural first step is to make sure not to engage with a caricature but with a careful reconstruction of the views that Carnap and other logical empiricists actually held.

Important reconsiderations of logical empiricism have been undertaken since at least the 1980s, and this endeavour continues to shed light not only on the actual views of these past philosophers but also on what are the available options for viable views. Michael Friedman writes in one of his early re-evaluations of logical empiricism that since the demise of the movement, "it has naturally been customary to view logical positivism as a kind of philosophical bogeyman whose faults and failings need to be enumerated (or, less commonly, investigated) before one's favored 'new' approach to philosophy can properly begin" (Friedman 1991, 505). He continues that "we have seen in recent years a veritable flowering of historically oriented reconsiderations of logical positivism" (*ibid.*), and adds that thanks to these reconsiderations "it has become clear—not at all surprisingly, of course—that

the above-mentioned post-positivist reaction gave birth to a large number of seriously misleading ideas about the origins, motivations, and true philosophical aims of the positivist movement” (Friedman 1991, 506). Since the fall of logical empiricism, deeply misrepresentative accounts of Carnap’s views have been part of the received view in philosophy, and many of them are still being repeated.¹⁴

1.6 Explication beyond Carnap

Carnap’s last extensive discussion of explication was in his reply to Strawson in the volume dedicated to Carnap in *The Library of the Living Philosophers* series edited by Arthur Schilpp. Although it was published in 1963 he had finished writing his replies in 1958 (Brun 2016, 1214, n. 5; Creath 1990a, 448-449). It is arguable, and it is argued by Brun (2016), that Carnap’s view of explication at that time had changed from the earlier view expressed in LFP in 1950, where the adequacy of an explication is measured by the four requirements (see 1.2 above). After Carnap’s last writings on explication, other philosophers have both used and—to a lesser degree—discussed the method. In some cases, however, they have conceived of it in a way that differs substantially from Carnap’s conception, even when they have referred to it as Carnap’s method. The most prominent example of such a philosopher is W. V. O. Quine, who discusses and advocates the method of explication in his *Word and Object* (1960). According to Quine, the task of explication is to find acceptable objects for referentially defective but useful nouns: “We fix on the particular functions of the unclear expression that make it worth troubling about, and then devise a substitute, clear and couched in terms to our liking, that fills those functions” (Quine 1960, 258-259). Put in such general terms, the method seems familiarly Carnapian. However, that Quine and Carnap have different conceptions of explication has been clearly shown by Martin Gustafsson (2006, 2014), and more by recently Jonas Raab (2024). In their comparisons between Carnap and Quine it becomes clear that agreement that explication is an appropriate philosophical method can coincide with disagreement regarding the epistemological and metaphysical significance of the results. However, it is an open question to what degree the method and the views and attitudes concerning it are independent, and how changes in one

¹⁴ An example is given by Carus (2019, 339–340), who corrects Robert Brandom’s portrayal of Carnap as a “representationalist”.

affect the other. Where motivations differ, two philosophers engaged in the same method of explication may nevertheless take themselves to be engaged in different projects or programmes. My task in chapters 5 and 6 is to give an overview of different versions and conceptions of explication. With such an overview I hope to contribute to a clearer view of Carnap's motivations and to clarify what options there are on the methodological menu in the vicinity of Carnapian explication. To be able to give an overview of different versions of explication, I need to use the term 'explication' in the broad sense.

1.7 The objectives of the thesis

As already described in earlier sections, in contemporary philosophy there has been a rise in the interest in and popularity of the method of explication (see 3.4 for a selection from the literature). In the vast and growing literature on explication, I aspire to contribute with:

- (1) An interpretation of how to understand the criteria of adequacy proposed by Carnap.
- (2) A systematic overview of extant versions of explication, including methods of concept formation which have not previously been regarded as forms of explication.
- (3) A proposal for how to modify and develop Carnap's criteria of adequacy.

The aims of the thesis can be divided—although not sharply—into historical, classificatory, and methodological aims. The historical aim is to give a plausible interpretation of Carnap's motivations for developing and pursuing the method in the way he did, and to argue against some of the extant interpretations of Carnap's writings on explication.

The classificatory aim is to clarify the borders and relations between explication and other methods in the vicinity of explication, as well as different versions of explication. Most importantly, I aim to systematize various extant versions and conceptions of explication. I will compare these versions and conceptions with Carnap's.

The methodological aim, finally, is to develop and broaden the method of explication by proposing modifications of the criteria of adequacy.

1.8 My position in relation to extant literature on explication

I will briefly explain how my project fits into the extant literature on explication, and in particular how it differs from extant accounts. The most immediate source of inspiration for my project is Brun's development of Carnapian explication in his influential article "Explication as a Method of Conceptual Re-engineering" (2016). However, I take a different path from Brun. While I do agree with Brun that there is a need for an account of explication that is more relaxed than Carnap's 1950 account—and I try to develop such an account in chapter 9—I disagree with Brun regarding in what way it should be relaxed.

Further, I argue against the contextualist view advocated by Mark Pinder. He argues, following Philip Kitcher, that the criteria of adequacy should be specified on a case-by-case basis. I think that a list of general criteria should be retained, but that Carnap's list of criteria should be modified.

Regarding the role of explication as a philosophical method, I have drawn a lot from works by Pinder (e.g., his 2020) and Erik J. Olsson (2015, 2017, 2021). Olsson (2021) introduces the term "explicationist philosophy" for philosophical definitions that give weight to all of Carnap's four criteria:

Explicationist philosophy refers, in my terminology, to this general view on the nature of philosophical definitions, which, in an updated version, may recognize other kinds of concept than just logical or empirical, e.g. legal or ethical concepts. Explicationist philosophy should be understood to imply that all four requirements on an explicatum be given substantial positive weight. (Olsson 2021, 44)

While updated, Olsson's term 'explicationist philosophy' follows Carnap's original method closely, and more closely than other prominent accounts of explication in the current literature. In this thesis I will propose a modification of Carnap's four requirements. Hence, I do not want to tie the term 'explication' or the term 'explicationist philosophy' to Carnap's four requirements. In short summary, my account is more in line with Carnap's view in LFP than Pinder's and Brun's accounts, and less in line with it than Olsson's account.

1.9 The plan of the thesis

In chapter 2, I present and discuss in more detail Carnap's writings on explication, the structure of Carnapian explication and Carnap's motivations for introducing explication.

In chapter 3, I discuss other philosophers' commentary and criticism of the method. P. F. Strawson is the first author to critically engage with explication, claiming that explication does not solve philosophical problems but merely changes the subject. I will devote a section to the exchange between Strawson and Carnap that took place in the Schilpp volume (1963) dedicated to Carnap.

In chapter 4, I introduce a distinction, inspired by a distinction made by James Justus (2012), between the immediate and ultimate purpose of explication. I use the distinction to interpret and clarify Carnap's views on the purpose of explication. Before I introduce the distinction as I will use it, I present the source of inspiration, which is Justus's discussion of the putatively different aims of conceptual analysis and explication. I also discuss the relative weight or importance of the criteria of adequacy and the internal structure among the criteria of adequacy.

In chapter 5, I give an overview of extant versions of explication, individuated on the basis of criteria of adequacy. I include methods of concept formation that have not been regarded as explications and argue that they may plausibly be understood as forms of explication.

In chapter 6, I give an overview of extant conceptions of explication. Different conceptions are here individuated by differences in views about the epistemological and metaphysical significance of explication. I focus on Quine's discussion of the explication in *Word and Object*.

In chapter 7, I discuss the place of explication among other methods of conceptual engineering. I trace the history of the term 'conceptual engineering' to Carnap's work. I argue that at least two of the central debates in the literature on conceptual engineering are not directly relevant for explication.

In chapter 8, I discuss the role of fruitfulness in explication, and what kind of fruitfulness we should strive for in explications. While I recognize the shortcomings of Carnap's original notion of fruitfulness, I reject the extant alternative accounts of what it is that makes an explication fruitful.

In chapter 9, I propose a modified version of Carnap's method explication, with modifications of Carnap's criteria and new additional criteria.

In chapter 10, I evaluate my criteria with an example. The explicandum in my example is the concept *democratic*.

In chapter 11, I conclude the dissertation with a discussion of the role of explication as a method in philosophy, and a discussion of remaining questions and problems for future research.

2 Carnapian explication

2.1 Overview

In this chapter I give an overview of Carnap's writings on explication, describing his views and his motivation for developing the method. I will use and refer to historical work on Carnap's philosophy, but at this stage of the thesis I try to avoid taking a stance on contested questions regarding how to interpret Carnap. The aim of the chapter is to give the reader the relevant background for contemporary discussions about and developments of Carnapian explication. I end the chapter with a broader discussion of the role of explication within Carnap's conception of scientific philosophy.

2.2 Before explication: rational reconstruction

Before discussing Carnap's writings on explication, I briefly address Carnap's methodology before 1945 (when he introduced the term 'explication'), namely the method of rational reconstruction. In the appendix to *Logical Foundations of Probability* (hereafter LFP),¹⁵ published in 1950, Carnap distinguishes between rational reconstruction and explication. There he seems to regard explication as targeting concepts and rational reconstruction as targeting whole systems of belief (LFP, 576). Brun comments that in LFP

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

¹⁵ I reintroduce the abbreviations in every chapter the first time I mention an abbreviated title.

there are, in addition to concepts, other sources of inconsistency, vagueness and theoretical ineptness that need to be dealt with. (Brun 2016, 1236)

On that view, explication could be seen as a sub-version of the more general method of rational reconstruction. Specifically, explication would be the version of rational reconstruction that is concerned with concepts. However, as Brun also points out, Carnap sometimes distinguishes ‘explication’ from ‘rational reconstruction’ and sometimes uses the terms interchangeably. Brun emphasizes that Carnap was well aware of the difference between dealing with a single concept and with a system or theory. But Brun notes that over time Carnap’s terminology fluctuated with his focus:

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]). (Brun 2016, 1236)

When Carnap looked back at his own work in 1961, in the preface to the second edition of the *Aufbau*, he seemed to regard rational reconstruction as a *precursor* to the method of explication, or even as the same method:

By rational reconstruction is here meant the searching out of new definitions for old concepts. The old concepts did not ordinarily originate by way of deliberate formulation, but in more or less unreflected and spontaneous development. The new definitions should be superior to the old in clarity and exactness, and, above all, should fit into a systematic structure of concepts. Such a clarification of concepts, nowadays frequently called “explication”, still seems to me one of the most important tasks of philosophy, especially if it is concerned with the main categories of human thought. (*Aufbau*, v)

What should we make of this? Since it is useful to distinguish between explication and rational reconstruction in the way Carnap does in LFP, I think that we should disregard Carnap’s comments quoted above as myopic with regard to changes in his own philosophical and methodological development, perhaps caused by his enthusiasm regarding the new method and terminology of explication.

There is a stark contrast between Carnap’s view in the preface to the second edition of the *Aufbau* and his view in LFP. In the above-mentioned appendix to LFP, Carnap regards rational reconstructions to be applied to some “body of generally accepted but more-or-less vague beliefs” and explication to be applied to “the concepts involved in those beliefs” (LFP, 576). From this

remark I gather that Carnap at least at this point in time considered explication to be a sub-version of rational reconstruction, rather than a replacement of it or an alternative term for the same method. On such a view, to improve our concepts and systems of concepts through explication is a step in the ulterior task of improving our body of beliefs through rational reconstruction.

Carus, however, draws a sharp line between rational reconstruction and explication in his interpretation of Carnap's philosophical development.¹⁶ Carus also recognizes important similarities between rational reconstruction and explication. Let us begin with the similarities before attending to the differences. In Carus's characterization of rational reconstruction, it is a

process of replacing intuitive, initially rather vague concepts by more precise ones. We still use ordinary language for practical communication, and for practical decision making. But the effect of rational reconstruction is to improve or upgrade the ordinary language [...] by replacing its loose, soft concepts with harder and more explicit ones [...]. (Carus 2007, 15)

With these striking similarities, what, then, are the differences? In Carus's interpretation, the rational reconstruction programme rested on the "hope that there could be a single, permanent logical framework for the whole of knowledge" (Carus 2007, 20). Carnap abandoned that hope when he adopted the Principle of Tolerance in late 1932¹⁷. He formulated the new attitude as a principle in *The Logical Syntax of Language* (1937 [1934]). In §17 he gives the "Principle of Tolerance", stating that "[i]t is not our business to set up prohibitions, but to arrive at conventions" (1937 [1934], 51), and further that:

In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (Carnap 1937 [1934], 52)

However, to draw the line between explication and rational reconstruction in the way Carus does raises several problems of interpretation. For example, it is arguably that some of the examples that Carnap gives of paradigmatic cases of explication would not qualify as explication. I return to these questions in chapter 6.

¹⁶ As Brun (2016, 1225, n. 24) points out.

¹⁷ Arguably, his first expression of the attitude occurred in his 1987[1932].

Michael Beaney (2013, 237) ascribes to Carnap the first use of the term ‘rational reconstruction’, as a term for a philosophical method. Carnap introduced the German term ‘rationale Nachkonstruktion’ in his *Aufbau* (1967 [1928]). Since then the term has been used by a host of philosophers in more or less the same sense.¹⁸ Beaney summarizes different uses of the term and proposes the following, as “the definition that best captures all of its uses”:

A rational reconstruction of a (purported) body of knowledge or conceptual scheme or set of events is a redescription and reorganization of that body or scheme or set that exhibits the logical (or rational) relations between its elements. (Beaney 2013, 253)

While Beaney’s definition reflects Carnap’s use in LFP, which is how I will use the terms, to avoid confusion it is important to be aware that Carnap later seems to regard “rational reconstruction” as another previously used term for the method of explication. I end this section by quoting, with endorsement, the following conclusion by Beaney:

There is room for dispute about the extent to which ‘explication’ means the same as ‘rational reconstruction’. [...] But there is a common methodological core, which Carnap highlights [...]: redefining our old concepts and systematizing them to make clear their logical relations. (Beaney 2013, 240)

2.3 Carnap’s writings on explication

As previously mentioned, the term ‘explication’ first appears in Carnap’s writings in relation to his work on probability. Beaney (2004) tracks the very first appearance of the term in Carnap’s work to the paper “The Two Concepts of Probability” (1945b), published in June 1945. Beaney also makes the interesting observation that in Carnap’s article “On Inductive Logic” (1945a), published in April 1945, Carnap is still using the term ‘rational reconstruction’ when he refers to his own philosophical method (Beaney 2004, 134).

In “Two Concepts of Probability” Carnap introduces the notion of explication, in relation to the problem of probability, in the following way:

¹⁸ Richard Rorty (1984), e.g., uses the term ‘rational reconstruction’ for an approach in the history of philosophy, to be contrasted with historical reconstruction.

The problem of probability may be regarded as the task of finding an adequate definition of the concept of probability that can provide a basis for a theory of probability. This task is not one of defining a new concept but rather of redefining an old one. Thus we have here an instance of that kind of problem—often important in the development of science and mathematics—where a concept already in use is to be made more exact or, rather, is to be replaced by a more exact new concept. Let us call these problems (in an adaptation of the terminology of Kant and Husserl) problems of *explication* [...] (Carnap 1945b, 513)

In light of his later descriptions of the method, it is a misleading formulation that the “task is not one of defining a new concept”. As his later formulation in the same quotation makes clear, the task *is* to define a new concept suitable to replace the old concept. As an example of the method Carnap mentions the Frege–Russell definition of natural numbers.

Thus, for instance, the definition of the cardinal number three by Frege and Russell as the class of all triples was meant as an explication; the explicandum was the ordinary meaning of the word ‘three’ as it appears in every-day life and in science; the concept of the class of all triples [...] was proposed as an explicatum for the explicandum mentioned. (Carnap 1945b, 513)

Later, he also mentions two cases of “scientific explication”, namely ‘temperature’ replacing ‘warmer’ (Carnap treats this example at length in LFP) and ‘IQ’ replacing ‘higher intelligence’.

In *Meaning and Necessity* (hereafter M&N) Carnap discusses explication in more detail and puts it to use in his attempt to explicate semantical and modal concepts, for example the notion of *logical truth*,¹⁹ which he explicates as *L-truth*. Carnap gives the following informal condition for an explicatum to be an adequate replacement of *logical truth*: it needs to be defined so that a sentence in a semantical system is L-true if and only if the semantical rules of the system are sufficient to establish its truth²⁰ (1956 [1947], 7). Although Carnap first uses the term ‘explication’ for his work on probability, Steve Awodey suggests that his concept *L-true* may be seen as a paradigmatic case of explication since “it was perhaps the difficulty in nailing down just this notion that first led to the conception of explication” (Awodey 2012, 131).

¹⁹ Carnap uses *logical truth* as interchangeable with *necessary truth* and *analytic truth*.

²⁰ He proposed the following definition for L-truth in a semantical system S1: “A sentence $\mathcal{E}I$ is L-true (in S1) = Df $\mathcal{E}i$ holds in every state-description (in S1)” (M&N, 10).

In LFP, Carnap returns to his work on probability, and now devotes the whole first chapter to an exposition of the method of explication, a chapter that is Carnap's longest and most detailed discussion of the method of explication. The second most important source for Carnap's views on explication is his reply to P. F. Strawson's criticism of explication in *The Philosophy of Rudolf Carnap*, which was the volume dedicated to Carnap in the *Library of the Living Philosophers* series, edited by Paul Arthur Schilpp (1963). Strawson's contribution is titled "Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy" (Strawson 1963, 503–518) and Carnap's reply in the "Replies and Systematic Expositions" (RSE) is titled "P. F. Strawson on Linguistic Naturalism" (RSE, 933–940). By 'linguistic naturalism' Carnap refers to "the method of describing and analyzing the actual usage of words in everyday language" (RSE, 933). Here Carnap further develops his view of the role of explication in philosophy and its relation to ordinary language philosophy. It turns out to be his final in-depth commentary on the method.

While the first chapter of LFP is where Carnap outlines the method in greatest detail, his reply to Strawson is the best source for his metaphilosophical view of the role of explication. This is where he most explicitly addresses how explication relates to other ways of doing philosophy. His attitude of tolerance²¹ does not apply only to one's choice of linguistic framework but extends also to methodology. At least, this is the case regarding the apparent conflict with ordinary language philosophy. Instead of conflict he ordains coexistence, if not cooperation:

It is certainly more fruitful, instead of wasting time deprecating the method of the other side, to work out some mode of peaceful coexistence of the two movements, and if possible, to cooperate. We all agree that it is important that good analytic work on philosophical problems be performed. Everyone may do this according to the method which seems to be the most promising to him. The future will show which of the two methods, or which of the many varieties of each, or which combinations of both, furnishes the best results. (Carnap 1963b, 940)

In the following subsection my aim is to characterize the method of Carnapian explication and spell out Carnap's own views about its role.

²¹ Carnap's Principle of Tolerance regarding linguistic frameworks guides his work from 1932 onwards, although he first formulated the Principle of Tolerance in *The Logical Syntax of Language* (1937 [1934]).

2.4 The structure and motivation of explication

Carnap develops the methodology of explication in his attempt to clarify the concept of probability, or rather the two concepts of probability which he gives the names “probability₁” (degree of confirmation) and “probability₂” (relative frequency). In LFP, Carnap expresses the general kind of problem that motivates explications in the following way:

There is a certain term (‘confirming evidence, ‘degree of confirmation’, ‘probability’) which is used in everyday language and by scientists without being exactly defined, and we try to make the use of these terms more precise or, as we shall say, to give an explication for them. (LFP, 2)

In this motivation for explication, Carnap’s description of explication could easily be mistaken for a description of an analysis or a precisifying definition. However, he soon makes clear that explication is not a kind of definition, but something else. Here is Carnap’s perhaps most succinct summary of the task of explication.

The task of **explication** consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. (LFP, 3, original emphasis)

Note that Carnap in the summary above shifts from describing the task as a transformation of a (single) concept, into describing the task as replacing one concept with another. The slightly ambiguous way in which Carnap phrases this forebodes one of the main controversies about the method, namely the nature of the connection between the given concept and the new concept (as we will see in 3.2 where I discuss Strawson’s critique of explication). Regardless of exactly how they are connected, Carnap immediately makes it clear that he has two separate concepts in mind, using the above-mentioned terminology of explicandum and explicatum:

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, original emphasis)

Carnap switches between talking about terms and talking about concepts without any clear systematicity. In both early and recent literature on explication, however, it has been debated whether the explicandum and the

explicatum should be regarded as terms or concepts, i.e., as linguistic entities or as non-linguistic entities (see 2.8 for further discussion).

The explicandum is regarded by Carnap as a “more or less inexact” concept or term that is used in everyday language or in science. In M&N, Carnap describes the explicandum as a “vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development” (M&N, 7). When Carnap talks of an “earlier stage” he seems not to mean that it has to be an actually surpassed stage, but that the one who pursues the explication aims at a more developed stage. And when he talks about the explicandum being “prescientific” (LFP, 5), this should not be taken as a belief in a sharp distinction between prescientific and scientific language and contexts, as he clarifies later (RSE, 936). Whether or not an explication is needed for a concept seems to be relative to the present purposes of inquiry. This means that Carnapian explication is iterative. A concept which today is an adequate explicatum may tomorrow be an explicandum in need of a more exact and fruitful replacement.²²

It should also be noted that explications of terms or concepts used in everyday life and language are only called for when the terms or concepts have a desired role to play in some scientific or theoretical context. To improve everyday communication is never the primary purpose of the method although it might be an indirect benefit of it if the new scientific concepts are adopted in everyday life. In RSE, Carnap emphasizes that there is no sharp boundary between scientific purposes and everyday purposes, exemplifying this with reference to the scientific explicatum “temperature” for the everyday expression “very hot”:

Suppose the statement “it will probably be very hot tomorrow at noon” is made for the purpose of communicating a future state to be expected, perhaps with regard to practical consequences. The use of the explicatum “temperature” instead of “very hot” in the above statement makes it possible to fulfil the same purpose in a more efficient way: “the temperature tomorrow at noon will probably be about so and so much”. (RSE, 936)

This illustrates how scientific explications of a piece of everyday language may seep back into everyday language to serve the same purpose as the former chunk of language, in a situation where precision or objectivity is especially important. Other examples that Carnap gives of scientific terms that have been accepted into everyday language are “at 4:30 P.M.” and “speed” used as a

²² The point that Carnapian explication is iterative has been made, specifically in relation to exactness, by Dutilh Novaes and Reck (2017, 201).

quantitative term (RSE, 936). I will now go on to discuss the process of performing an explication, which consists of two main steps.

2.5 The two steps

2.5.1 The first step: clarification of the explicandum

The process of explication consists of two steps: the first step is the clarification of the explicandum and the second step is the specification of the explicatum. When we have a “problem of explication” (LFP, 3) or, in other words, when we have encountered a scientific or theoretical need for an explication, the first step is to clarify the explicandum. Carnap stresses the importance of this step, writing that “we must, in order to prevent the discussion of the problem from becoming entirely futile, do all we can to make at least practically clear what is meant as the explicandum” (LFP, 4). Carnap’s emphasis on the importance of this step is partly motivated by his belief that philosophers “very frequently violate this requirement”:

They ask questions like: ‘What is causality?’, ‘What is life?’, ‘What is mind?’, ‘What is justice?’, etc. Then they often immediately start to look for an answer without first examining the tacit assumption that the terms in question are at least practically clear enough to serve as a basis for an investigation, for an analysis or explication. (LFP, 4)

A way of making it practically clear what is meant is to give examples of intended and unintended uses of the term. Carnap gives two examples of how to clarify explicanda, using the terms “salt” and “true”. Here is the first one:

I might say, for example: “I mean by the explicandum ‘salt’, not its wide sense which it has in chemistry but its narrow sense in which it is used in the household language”. (LFP, 4–5)

With such a clarification achieved, the task of finding an explicatum—“sodium chloride” or “NaCl”—can begin. Here is the other example:

I might say, for example: [...]“I am looking for an explication of the term ‘true’, not as used in phrases like ‘a true democracy’, ‘a true friend’, etc., but as used in everyday life, in legal proceedings, in logic, and in science, in about the sense

of ‘correct’, ‘accurate’, ‘veridical’, ‘not false’, ‘neither error nor lie’, as applied to statements, assertions, reports, stories, etc.” (LFP, 5)

Once again, it is only after this clarification that the proposed replacement enters the picture. In this case, the example of an explicatum that Carnap gives is Tarski’s semantical definition of “true”. It should be noted that although Carnap uses Tarski’s definition as an example of explication it is not obvious that Tarski is aiming for the same goal as Carnap. While Carnap holds that “it is not required that an explicatum have, as nearly as possible, the same meaning as the explicandum” (M&N, 8), Tarski stresses that his definition of truth “aims to catch hold of the actual meaning of an old notion” (Tarski 1944, 341).²³ It might be that Tarski’s remark amounts merely to a desire for some degree of similarity. Nevertheless, it is not obvious whether their approaches agree or conflict with each other on this point.

2.5.2 The second step: the specification of the explicatum

Carnap contrasts the task of explication to the tasks of solving ordinary scientific and mathematical problems, where under favourable conditions both “the datum and the solution” are formulated in exact terms. His examples of ordinary problems where “the datum” is exact are “What is the product of 3 and 5?” and “What happens when an electric current goes through water?” (LFP, 3). In a problem of explication, on the other hand, “the datum, viz., the explicandum, is not given in exact terms; if it were, no explication would be needed” (LFP, 3–4). When a need for explication appears, there is no exactly formulated problem and yet there is an expectation of an exactly formulated solution. Carnap readily admits that this is a puzzling situation:

Since the datum is inexact, the problem itself is not stated in exact terms; and yet we are asked to give an exact solution. This is one of the puzzling peculiarities of explication. (LFP, 4)

From this it follows, according to Carnap, that there are no correct or incorrect solutions to problems of explication and that the idea of a correct explication is nonsensical.

It follows that, if a solution for a problem of explication is proposed, we cannot decide in an exact way whether it is right or wrong. Strictly speaking, the

²³ The contrast between Tarski and Carnap is pointed out by Hanna (1968, 28).

question whether the solution is right or wrong makes no good sense because there is no clear-cut answer. (LFP, 4)

Instead, explications are to be judged based on satisfaction, i.e., based on their usefulness. A satisfactory explicatum allows us to do what we want to do.

The question should rather be whether the proposed solution is satisfactory, whether it is more satisfactory than another one, and the like. (LFP, 4)

How, then, to judge whether an explicatum is satisfactory? Carnap proposes four requirements or criteria of adequacy²⁴, which a concept must fulfil to a “sufficient degree” to be an adequate explicatum, viz. similarity to the explicandum, exactness, fruitfulness, and simplicity.

- (1) The explicatum is to be **similar** to the explicandum in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted.
- (2) The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an **exact** form, so as to introduce the explicatum into a well-connected system of scientific concepts.
- (3) The explicatum is to be a **fruitful** concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept).
- (4) The explicatum should be as **simple** as possible; this means as simple as the more important requirements (1), (2), and (3) permit. (LFP, 7, my emphasis)

The first criterion concerns the connection between the explicandum and the explicatum. The other three criteria concern the explicatum and its relation to the target system of concepts or target theory. In the process of choosing or constructing an explicatum, trade-offs are to be made between the first three criteria. Simplicity cannot be gained at the expense of similarity, exactness, or fruitfulness. Hence, the simplicity criterion is only relevant when two or more candidate explicata fare equally well on the other criteria. As we can gather

²⁴ I am using “criteria of adequacy” or just “criteria” when referring to what Carnap calls “requirements”. There are many interchangeable terms for these criteria used in the literature on explication. Some authors call them “conditions” or “desiderata”.

from (2) and from other remarks by Carnap, to give a definition is a possible but not necessary part of the characterization of the explicatum, but in such a definition both the definiens and the definiendum “express the explicatum” (LFP, 3).

We can now see that there are two different descriptive moments in an explication, one in the first step of explication and another in the second step of explication. The descriptive part in the first step, the clarification of the explicandum, is typically a form of conceptual analysis. It was acknowledged by both Carnap (RSE) and Strawson (1963), and argued by Frank Tillman (1965, 383), that in this task there is a bridge between the conceptual analysis of ordinary language philosophy and explicationist philosophy. For example, Carnap’s explication of logical probability is not merely a proposal for a new convention but is intended to reflect intuitive judgements of inductive rationality. Regarding what reasons there are for accepting any axioms of inductive logic, Carnap writes that the “reasons are based upon our intuitive judgements concerning inductive validity, i.e., concerning inductive rationality of practical decisions (e.g., about bets)” (RSE, 978). In recent literature, it has been proposed by proponents of experimental philosophy that the first step of explication should be performed as an empirical study of intuitions since the methodology of conceptual analysis has been discredited.

The descriptive part in the second step is not addressed by Carnap, his criteria of adequacy for explications are based on descriptions of features of previously successful scientific concepts. He extracts his criteria from actual examples from the history of science. Some of his justifications for the conditions of adequacy for explication are purely descriptive, for example when he remarks that “scientists appreciate simplicity in their concepts” (LFP, 7). In this sense, the criteria of adequacy for explications which were proposed by Carnap could be considered to be both descriptive and prescriptive. They are meant to describe actual concepts in science, and because of the success of science they are prescribed for concept formation in philosophy.

2.6 Three kinds of concepts

An explicatum may according to Carnap belong to one of these three kinds of concept: classificatory concepts, comparative concepts, and quantitative concepts. A classificatory concept divides things or cases into two or more mutually exclusive kinds (or classifications). This is “the simplest and least effective kind of concept” (LFP, 12). Comparative concepts allow us to say of

things or cases that they have more or less of a quality or variable, i.e., they allow us to make comparisons without the use of numerical values. These are “more powerful” (ibid.) than classificatory concepts but not as powerful as quantitative concepts. Quantitative concepts allow for the characterization of something by the ascription of numerical values. They are powerful because “they enable us to give a more precise description of a concrete situation and, more important, to formulate more comprehensive general laws” (ibid.).

The kind most commonly used in everyday thinking is the classificatory kind, and Carnap claims that in scientific progress these are often (but not always) replaced by comparative and quantitative concepts.

In prescientific thinking classificatory concepts are used most frequently. In the course of the development of science they are replaced in scientific formulations more and more by concepts of the other two kinds, although they remain always useful for the formulation of observational results. (LFP, 9)

In ordinary life, examples of the use of classificatory concepts are “when the things surrounding us are described as warm or cold, big or small, hard or soft, etc., or when they are classified as houses, stones, tables, men, etc.” (LFP, 9). Concerning the use of classificatory concepts in scientific contexts, Carnap exemplifies this with concepts such as those used “when substances are divided into metals and nonmetals, and again the metals into iron, copper, silver; likewise, when animals and plants are divided into classes and further divided into orders, families, genera, and, finally, species” (ibid.).

Comparative concepts are of course common in everyday language as well: Carnap’s first line of examples of everyday classificatory concepts can be turned into comparative concepts, as when things are described as warmer or colder or as bigger or smaller than other things. In Carnap’s semantics, the classificatory concepts are properties and the comparative concepts are relations:

A comparative concept is always a relation. If the underlying classificatory concept is a property (e.g., Warm), the comparative concept is a dyadic relation, that is, one with two arguments (e.g. Warmer). (LFP, 10)

The quantitative concepts are ranked the highest in Carnap’s hierarchy, as they are “no doubt the most effective instruments in the scientific arsenal” (LFP, 9). Carnap’s examples are “length, length of time, velocity, volume, mass, force,

temperature, electric charge, price, I.Q., infantile mortality, etc.” (ibid.).²⁵ As his examples clearly show, some quantitative concepts are also used in ordinary life and some have been so used throughout the history of civilization (length, length of time, price, etc.). Therefore, these three kinds of concepts cannot easily be classified as either prescientific or scientific kinds of concepts.

The explicatum may take any of these forms, and one way of performing an explication is to switch from a kind of concept that is lower in the rank to a kind of concept that is higher in the rank. So, one may replace a classificatory concept with a comparative or quantitative concept or replace a comparative concept with a quantitative concept. Carnap’s example of this process is the earlier mentioned replacement of sensory concepts (*warm, cold*, etc.) with the concept of *temperature*. So, we have first the replacement of *warm* with *warmer*, and second the replacement of *warmer* with *having a higher temperature*. Here is Carnap’s fictitious description of the development of a more precise language:

The state of bodies with respect to heat can be described in the simplest and crudest way with the help of classificatory concepts like Hot, Warm, and Cold (and perhaps a few more). We may imagine an early, not recorded stage of the development of our language where only these classificatory terms were available. Later, an essential refinement of language took place by the introduction of a comparative term like ‘warmer’. [...] Finally, the corresponding quantitative concept, that of temperature, was introduced in the construction of scientific language. (LFP, 12)

The descriptive precision gained by adopting quantitative concepts is a form of exactness that is not covered by the official exactness criterion discussed above. With a quantitative concept (and a precise thermometer) one may decide that one room is 0.1 °C warmer than another, while based on heat sensations on the skin the two rooms would be judged to be equally warm. This discriminatory power seems to be a different form of exactness than in the exactness criterion.²⁶ However, this could not be a general criterion for explications, since only some explications involve a change from non-quantitative to quantitative concepts. In other examples of explication, the

²⁵ There is a further distinction related to quantitative concepts which was introduced by Stevens (1946), namely between interval variables (or scales) and ratio variables (or scales). Carnap’s list of examples includes both kinds. I return to these distinctions at the end of this section.

²⁶ As we will see below, Brun uses the term ‘precision’ for this form of discriminatory power to distinguish it from the exactness criterion, which he interprets as “requiring that the explicatum is not more vague than the explicandum” (Brun 2016, 1223).

explicatum belongs to the same kind of concept as the explicandum, such as the explication of the classificatory concept *fish* with the classificatory concept *piscis*, or the explication of the classificatory concept *table salt* with the classificatory concept *sodium chloride*.

Since I aim to develop Carnap's views of explication in light of theorizing about concept formation in the social sciences, a historical sidenote is in order here. Carnap's three kinds of concepts are analogous to the scales of measurement by psychologist Stanley Smith Stevens. In a seminal 1946 paper Stevens introduces four scales of measurement, namely the nominal, the ordinal, the interval, and the ratio scale (Stevens 1946, 678–680). These scales of measurement or levels of measurement are today part of the basic training for social scientists, and they are included in virtually every textbook on social research.²⁷ Nominal scales represent objects as belonging to mutually exclusive and exhaustive categories. For two objects *a* and *b*, the only information we get from a nominal scale is whether *a* and *b* belong to the same category or not. Ordinal scales represent objects as having more or less of a quality or variable. A frequently cited example of an ordinal scale in the physical sciences is the Mohs scale for the hardness of minerals, which is a scratch test based on which materials can scratch others (Babbie 2016; Tal 2015, §3.2). These two are the categorical scales and the remaining two are the numerical scales. Among the numerical scales, the difference between the interval and ratio scale is that interval scales have arbitrary zero points, while the zero point on ratio scales indicates the absence of the measured quantity. For example, the Celsius temperature scale is an interval scale while the Kelvin temperature scale, with an absolute zero point (i.e., the absence of thermal energy) is a ratio scale.

There are striking similarities between Carnap's three kinds of concept and Stevens's scales of measurement. And it is hardly a coincidence that analogous ideas appear simultaneously in both Carnap's and Stevens works, since they collaborated when Carnap were at Harvard, before he wrote LFP. As George Millers notes in his biographical writings on Stevens, “[i]n 1940 [Stevens] and Rudolf Carnap organized a monthly discussion group at Harvard on the Unity of Science” (Miller 1975, 442).

I will now discuss the four criteria of adequacy one by one.

²⁷ For a few examples, see Earl Babbie (2016, 139-142), Alan Bryman (2008, 321), John Gerring (2011, 167-172), and Tim May (2011, 112). In social research the scales are often introduced as different types of variables.

2.7 The four criteria of adequacy

2.7.1 The similarity criterion

In his list of criteria in LFP, previously quoted in 1.2., Carnap gives what may be considered his official formulation of the similarity criterion (and the other criteria). He states that:

The explicatum is to be similar to the explicandum in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used. (LFP, 7)

In RSE, Carnap explicitly mentions that one should be able to use the explicatum for the same *purpose* as the explicandum. (RSE, 936). Carnap points out in LFP that an obvious reason for only requiring that the explicatum can be used in *most* of the same cases as the explicandum, is that the explicatum is more exact than the explicandum: “Since the explicandum is more or less vague and certainly more so than the explicatum, it is obvious that we cannot require the correspondence between the two concepts to be complete coincidence” (LFP, 5). The exactness of the explicatum is bought at the expense of its similarity with the explicandum. Discrepancy in the degree of exactness between explicandum and explicatum is not only unavoidable but also desirable.²⁸

Similarity is not only traded for increased exactness but also for increased fruitfulness. Carnap makes clear that “most cases” does not mean as many cases as possible given the difference in exactness, since “close similarity is not required, and considerable differences are permitted” (LFP, 7). Carnap maintains that given the difference in exactness it would be too strong a demand to require that the explicatum should be as similar as possible.²⁹

one might perhaps think that the explicatum should be as close to or as similar with the explicandum as the latter’s vagueness permits. However, it is easily

²⁸ As Gustafsson points out, “that there are certain differences between the original concept and the substitute might be said to be the very point of the procedure” (Gustafsson 2006, 61). The remark concerns Quine’s conception of explication but in his 2007 Gustafsson makes the same point even more strongly concerning Carnap’s conception of explication (Gustafsson 2007, 43).

²⁹ Already in M&N, Carnap remarks that “it is not required that an explicatum have, as nearly as possible, the same meaning” (M&N, 8).

seen that this requirement would be too strong, that the actual procedure of scientists is often not in agreement with it, and for good reasons. (LFP, 5)

The reason Carnap gives against such a strong requirement is that it is not the actual practice of scientists, “and for good reasons” (LFP, 5). The reason is that fruitfulness too may be gained at the expense of similarity to the explicandum. Hence, whether the first criterion is sufficiently satisfied cannot be judged independently of the third criterion. It depends on the degree of fruitfulness whether or not an explicatum is sufficiently similar. To illustrate this point Carnap refers to the scientific explication of *fish*, which is one of his most detailed examples of an explication of an empirical concept:

Let us consider as an example the prescientific term ‘fish’. In the construction of a systematic language of zoölogy, the concept Fish designated by this term has been replaced by a scientific concept designated by the same term ‘fish’; let us use for the latter concept the term ‘piscis’ in order to avoid confusion. [...] The zoölogists found that the animals to which the concept Fish applies, that is, those living in water, have by far not as many other properties in common as the animals which live in water, are cold-blooded vertebrates, and have gills throughout life. Hence the concept Piscis defined by these latter properties allows more general statements than any concept defined so as to be more similar to Fish; and this is what makes the concept Piscis more fruitful. (LFP, 6)

The point in bringing up this example is, once again, for Carnap to show that deviation from the meaning of the explicandum is motivated if there is fruitfulness to be gained. And below he makes the point that it was the fruitfulness of *piscis* that motivated zoologists to choose this concept rather than some other concept that is closer to the prescientific concept *fish*.

Instead of the Piscis they could have chosen another concept—let us use for it the term ‘piscis*’—which would likewise be exactly defined but which would be much more similar to the prescientific concept Fish by not excluding whales, seals, etc. What was their motive for not even considering a wider concept like Piscis* and instead artificially constructing the new concept Piscis far remote from any concept in the prescientific language? The reason was that they realized the fact that the concept Piscis promised to be much more fruitful than any concept more similar to Fish. (LFP, 6)

It is not clear to me, however, how the scientist could have chosen a concept much more similar to a prescientific concept that would be equally or almost equally exactly defined. There would be plenty of vague cases if ‘fish’ were

defined, e.g., as applying to x if and only if x is an animal and x lives in the sea. What about animals that live sometimes or most of the time in the sea?

In relation to Carnap's previously mentioned example of 'temperature', he also gives a clear everyday illustration of why the similarity criterion is formulated as it is. Recall the three kinds of concepts, described above, which Carnap distinguishes between, namely classificatory, comparative, and quantitative concepts. In the following discussion, the quantitative concept *temperature* is the explicatum for the comparative concept *warmer*. It is then required, according to Carnap, that "[t]he concept Temperature is to be such that, in most cases, if x is warmer than y (in the prescientific sense, based on the heat sensations of the skin), then the temperature of x is higher than that of y " (LFP, 12–13). In what kind of situation is the explicatum allowed to differ? Here is an everyday scenario from Carnap in which the explicatum is allowed to differ from the explicandum.

Suppose I enter a moderately heated room twice, first coming from an overheated room and at a later time coming from the cold outside. Then it may happen that I declare the room, on the basis of my sensations, to be warmer the second time than the first, while the thermometer shows at the second time the same temperature as at the first (or even a slightly higher one). Experiences of this kind do not at all lead us to the conclusion that the concept Temperature defined with reference to the thermometer is inadequate as an explicatum for the concept Warmer. On the contrary, we have become accustomed to let the scientific concept overrule the prescientific one in all cases of disagreement. (LFP, 12–13)

Since judgements about room temperature now are ultimately decided by thermometers and not our senses, Carnap goes on to claim that the term has undergone a change in meaning. Using the prescientific meaning of the term, it would presumably have been contradictory to say that "the living room felt warmer than before but it wasn't warmer than before", while this is a sensible thing to say using the scientific meaning of 'warmer' in the second occurrence of the word.

In other words, the term 'warmer' has undergone a change of meaning. Its meaning was originally based directly on a comparison of heat sensations, but, after, the acceptance of the scientific concept Temperature into our everyday language, the word 'warmer' is used in the sense of 'having a higher temperature'. Thus the experience described above is now formulated as follows: "I believed that the room was at the second time warmer than at the first, but this was an error; the room was actually not warmer; I found this out with the help of the thermometer". (LFP, 13)

Carnap also spells out how a requirement that assigns all the weight to similarity and none to fruitfulness would be formulated. For the same case of explication, it would be required that “the concept Temperature is to be such that, if x is not warmer than y (in the prescientific sense), then the temperature of x is not higher than y ” (LFP, 13). If we were committed to a methodological approach closer to traditional conceptual analysis, we would perhaps require this from an analysis of ‘warmer’ in terms of temperatures. However, here too Carnap observes that we are used to letting the scientific explicatum overrule the prescientific explicandum: “When the difference between the temperatures of x and y is small, then, as a rule, we notice no difference in our heat sensations. This again is not taken as a reason for rejecting the concept Temperature” (LFP, 13). It is clear that such a difference is not only an acceptable price to pay, but that on the contrary it is part of the point of conceptual revision. To gain a scientifically more fruitful explicatum is reason enough to deviate from some of the uses of the everyday explicandum that otherwise may be preserved.

Carnap does not see it as a problem that the degree of similarity varies from case to case. Concerning “the relation between explicatum and explicandum which must be required for an adequate explication”, he writes in his reply to Nelson Goodman (1963) in the Schilpp volume:

It seems justifiable to me to make different requirements for different situations. Although sometimes synonymy in the strong sense might be required, this does not seem necessary to me in general; in most cases logical equivalence is sufficient, and perhaps even this is not necessary. For the system in the Aufbau I required identity of extensions. Goodman regards this requirement as too strong and suggests replacing it by a certain kind of isomorphism. This may in many cases be a good requirement. (RSE, 945–6)³⁰

While this flexible spirit is characteristic of Carnap’s philosophy, Brun (2016, 1222) has pointed out that it is a surprising comment from Carnap that “sometimes synonymy in the strong sense might be required”. It is surprising because Carnap in his reply to Strawson in the same volume, as well as in earlier work, states that “[t]he only essential requirement is that the explicatum be more precise than the explicandum” (RSE, 936). However, an explicatum cannot both be synonymous with the explicandum and more precise than it. It

³⁰ Brun (2016, 1221) notes that in set-theoretical explications of the natural numbers, there are several equally good explicata. Since the extensions of these explicata do not overlap with each other but are all acceptable replacements of the explicandum, it can evidently be permitted for an explicatum to have no extensional overlap at all with the explicandum.

is also hard to see how a synonymous explicatum could be more fruitful than the explicatum. In these cases, with all the weight on the similarity criterion, there seems to be no room for increased exactness and fruitfulness, and explication would collapse into conceptual analysis and face the paradox of analysis (see 11.2).

Brun concludes that the similarity in general cannot be specified with reference to the extensions of the explicanda and the explicata but must be specified with reference to the intended purposes of the explicatum (Brun 2016, 1221–1222). I agree and I doubt that there is a point in trying to formulate the relation between explicandum and explicatum in exact terms.

2.7.2 The exactness criterion

The exactness criterion, as Carnap states it in his official formulation of the criteria of adequacy, is that:

The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an exact form, so as to introduce the explicatum into a well-connected system of scientific concepts. (LFP, 7)

Related to this criterion Carnap remarks in RSE that “exactness and clarity are best achieved by a certain degree of systematization” (RSE, 936), and that, therefore “the explicatum usually belongs to a systematic conceptual framework” (ibid.). Brun (2016, 1217) points out that while the target system of concepts according to Carnap does not have to be framed in a formal or ideal language, this has been falsely assumed to be the case by many commentators, both approving and critical ones. We may add another critical commentator to Brun’s list, namely Giovanni Boniolo (2003). Boniolo claims that “Carnap’s explication needs formalization” (2003, 296). Against that view, it becomes clear in the following quotation from Carnap that the kind of conceptual framework he has in mind can be found in natural language as well:

the system may be of a rather elementary kind as, for instance, the system of numerical words in everyday language. The use of symbolic logic and of a constructed language system with explicit syntactical and semantical rules is the most elaborate and most efficient method. *For philosophical explications the use of this method is advisable only in special cases, but not generally.* (RSE, 936, my emphasis)

On the previous page, Carnap writes that “[t]he explicatum may belong to the ordinary language, although perhaps to a more exact part of it” (RSE, 935).

Brun (2016, 1222) identifies five distinct aspects of exactness in Carnap's account. First there is exactness in the sense of giving explicit rules for how to use the explicatum in the target system of concepts. Arguably, 'explicitness' is a better term than 'exactness' for this aspect.³¹ In Brun's view, to formulate "the rules for using the explicatum-term explicitly in terms of the target system is just a necessary condition for giving an explication" (1222). He distinguishes what he calls "a necessary condition for giving an explication" from what he calls "necessary conditions of adequacy" (more on this in 4.4.2). As further conditions on how the rules of use must be formulated, Brun introduces two "necessary conditions of adequacy",³² namely (1) that the rules are unambiguous and (2) that the rules do not lead to paradoxes or contradictions. In Brun's own formulation: "That the rules are unambiguous and do not lead to paradoxes or contradictions are then necessary conditions of adequacy" (1222). Regarding (2), Carnap comments in RSE that the explicatum must not yield logical paradoxes:

in spite of practical skill in usage, people in general, and even mathematicians before Frege, were not completely clear about the meaning of numerical words. Clarity is here understood in a stricter sense than in ordinary language; this sense does not require the ability to give a definition, but it requires that the usage does not lead to logical paradoxes. (RSE, 935)

Regarding the condition of unambiguity, Brun refers to page 4 in LFP. However, there Carnap merely discusses the importance of disambiguating the *explicandum* term, but regarding the explicatum it can be assumed that by requiring exact rules of use Carnap requires that they are unambiguous.

So far we have three aspects of exactness, all of which are related to the step of providing explicit rules of use for the explicatum. The first aspect is explicitness, the second is unambiguity and the third is consistency. The first aspect is just that explicit rules of use should be provided, while the second and third aspects specify the ways in which that must be done. As Brun explains, these criteria differ from Carnap's criteria in the sense that they do not admit of trade-offs and are not a matter of degree (Brun 2016, 1223).

³¹ Ludlow distinguishes between narrowing a word and what he calls "explicitifying" a word. If we stipulate that horses are not athletes, we have narrowed the word "horse". If we stipulate that cars are not athletes we have not narrowed the word since no one thought cars were athletes. We have merely introduced an explicit definitional component (Ludlow 2014, 88).

³² In Brun 2016, the terms "conditions of adequacy" and "criteria of adequacy" seem to be used interchangeably.

Brun's fourth aspect of exactness is what Carnap lists as his official exactness criterion, interpreted by Brun as the requirement that the explicatum should not be more vague than the explicandum. Brun writes:

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. For this, I use "exact", whereas "precise" is reserved for the precision and discriminating power of comparative and quantitative concepts. (Brun 2016, 1223)

There are two questions to be raised in relation to this aspect of exactness. The first question concerns if the explicatum has to be *more* exact than the explicandum, as most interpreters of Carnap think, or if it is sufficient that the explicatum is *not less* exact than the explicandum, as Brun interprets the criterion (see 4.4.2 for further discussion).

The second question concerns whether the term 'exact' in Carnap's official second criterion should be understood in the sense of sharpness or narrowness. These are the two main forms of exactness. To make a term more exact can be to sharpen it in the sense of reducing its vagueness or to narrow it in the sense of restricting the range of things it applies to. I take the terminology of sharpness and narrowness from Ludlow (2014).³³ As he points, we can narrow a word meaning without sharpening it and sharpen a word meaning without narrowing it. The consensus among interpreters of Carnap is that he intends the first sense, i.e., to make the explicatum sharper than the explicandum in the sense of reducing vagueness. Some interpreters have, however, falsely thought that reducing vagueness is the whole point of Carnap's method.³⁴

The fifth aspect of exactness, for which Brun reserves the term "precision", is the discriminating power of comparative and—especially—quantitative concepts. In LFP, Carnap discusses comparative and quantitative concepts in isolation from his discussion of exactness and the other criteria of adequacy. Brun takes Carnap's discussion of the benefits of quantitative concepts to suggest the need for further criteria, such as theoretical scope.

Some commentators have questioned the centrality of exactness for explication. Dutilh Novaes and Reck argue (2017) that "exactness is less fundamental than might appear at first" (2017, 197). I agree that Carnap's general characterizations of explication seem misleading, since in these

³³ Hansson (2018, 13–14) presents the same distinction in terms of definiteness (what Ludlow calls sharpness) and restrictedness (what Ludlow calls narrowness). I will use Ludlow's terminology.

³⁴ E.g., Martin (1973), as pointed out by Brun (2020, 934, n. 16).

characterizations he exclusively mentions exactness or precision. He writes that explication is a method whereby “a concept already in use is to be made more exact or, rather, is to be replaced by a more exact new concept” (Carnap 1945b, 513), that explication is “the task of making more exact a vague or not quite exact concept” (M&N, 7), that the “task of **explication** consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second” (LFP, 3, original emphasis), and that “An explication replaces the imprecise explicandum by a more precise explicatum” (RSE, 935).

These characterizations do not reflect his discussions and examples of explication, in which fruitfulness is emphasized. The concept *temperature* is not only superior due to its discriminating power but also due to its fruitfulness:

The quantitative concept Temperature has proved its great fruitfulness by the fact that it occurs in many important laws. This is not always the case with quantitative concepts in science, even if they are well defined by exact rules of measurement. For instance, it has sometimes occurred in psychology that a quantitative concept was defined by an exact description of tests but that the expectation of finding laws connecting the values thus measured with values of other concepts was not fulfilled; then the concept was finally discarded as not fruitful. (LFP, 14)

These remarks may suggest that exactness is only desirable as a means to fruitfulness, which is the ultimate goal of explication. However, the view that fruitfulness has priority over exactness appears to contradict not only the previously quoted formulations by Carnap but also the stronger claim, made in (RSE), that precision is the only essential requirement: “The only essential requirement is that the explicatum be more precise than the explicandum” (RSE, 936). If these remarks are to be taken at face value, then fruitfulness without more exactness does not seem to be enough for explication. Despite Carnap’s remark above, some interpreters of Carnap claim that fruitfulness is the purpose or the agenda of explication. While recognizing Carnap’s emphasis on precision, Justus (2012) claims that exactness is important because it tends to enhance fruitfulness:

Explication requires increasing precision—at one point Carnap even says that it is the “only essential requirement” [RSE, 936]—but its value does not derive from precision alone. Rather, precision is paramount because it usually enhances fruitfulness. Vague concepts are rarely, if ever, components of the well-confirmed generalizations the fruitfulness criterion targets. (Justus 2012, 168–9)

Justus rightly emphasizes the close connection between the exactness criterion and the fruitfulness criterion. Joshua Shepherd and Justus (2015) further emphasize the connection between precision and fruitfulness in scientific explications, since precision is required for testable predictions:

Without sufficiently precise concepts, it is difficult if not impossible to derive predictions from statements containing them. Without such predictions, in turn, statements cannot be confirmed or disconfirmed. (Shepherd and Justus 2015, 388)

Hence, both exact and fruitful concepts are needed to achieve epistemic success. However, to understand the role of precision in relation to fruitfulness, it is important to keep apart two of the senses in which a concept can be precise, namely by being exactly defined and by being quantitatively defined. Shepherd and Justus (2015, 308, n. 8) comment in a footnote that Carnap recognized that precisification is not always conducive to fruitfulness in science. The paragraph they refer to is the one quoted previously in this section, where Carnap mentions that there are examples of quantitative concepts that have failed to be fruitful, e.g., in psychology. The lesson to take away from this is that according to Carnap we should not be discouraged from trying to explicate a concept even if we cannot find or construct a quantitative explicatum (LFP, 14). *Quantitatively defined* concepts are not always conducive to fruitfulness, but that is unrelated to the question of whether more *exactly defined* concepts are always conducive to fruitfulness. Carnap merely recommends the use of quantitative concepts when possible. With quantitative concepts it is often easier to give exact definitions and to connect them to other concepts with exact laws.

2.7.3 The fruitfulness criterion

In the discussion of exactness in the previous section I already raised questions about the fruitfulness criterion. Here is again Carnap's official formulation of it:

The explicatum is to be a fruitful concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept). (LFP, 7)

I mentioned above that fruitfulness is what motivates differences between the explicandum and the explicatum that are larger than the unavoidable difference

stemming from the fact that the latter is more exact than the former. Similarity to the explicandum is traded for the ability to formulate laws and theorems. Several authors claim that fruitfulness is the main criterion or the main purpose of Carnapian explication. For example, Shepherd and Justus claim that fruitfulness “is the agenda” (2015, 388) and Dutilh Novaes and Reck (2017) call fruitfulness “arguably the most important requirement for a successful Carnapian explication” (2017, 201) and “ultimately the most significant requirement for an explication” (ibid., 207). These are bold interpretational claims since Carnap repeatedly characterizes explication as the task of gaining more exact concepts, and claims that precision is the only essential requirement. I treat these questions further in chapter 4.

Pinder (2022b) distinguishes between two different ways of understanding the primacy of fruitfulness:

First, fruitfulness might be weighted more heavily than the other desiderata, so that a small increase in fruitfulness is preferable to a larger increase in (say) exactness. Or, second, fruitfulness might subsume other desiderata so that, in particular, exactness and simplicity are only desirable as a means to fruitfulness. (2022, 914)

On the first understanding, let us call it the ‘different weights’ account, fruitfulness is given more weight than the other criteria. On the second understanding, let us call it the ‘subsumption account’, the other criteria are subsumed under fruitfulness, as means to the end of acquiring more fruitful concepts. While the ‘different weights’ account would merely affect how we construct and choose explicata, the subsumption account would change more fundamentally how we understand the project of explication. On that account we would have to depart from Carnap’s characterization of explication as the task of replacing a less exact concept with a more exact concept, and instead define it as the task of replacing a less fruitful concept with a more fruitful concept. As an interpretation of Carnap’s views, neither account is plausible. However, the accounts could be taken as proposals for better ways to understand explication. While the ‘different weights’ account might be a plausible modification of Carnapian explication, I reject the ‘subsumption’ account both as an interpretation of Carnap and as a methodological proposal.

In chapter 8, I discuss at length problems with Carnap’s notion of fruitfulness and alternative proposals.

2.7.4 The simplicity criterion

The fourth and last requirement is the simplicity criterion. Here, again, is Carnap's official formulation of the criterion: "The explicatum should be as simple as possible; this means as simple as the more important requirements (1), (2), and (3) permit" (LFP, 7). Concerns about simplicity enter the picture only in cases where two or more candidate explicata are equally exact and fruitful and equally similar to the explicandum:

In general, simplicity comes into consideration only in a case where there is a question of choice among several concepts which achieve about the same and seem to be equally fruitful; if these concepts show a marked difference in the degree of simplicity, the scientist will, as a rule, prefer the simplest of them. (LFP, 7)³⁵

Carnap does not discuss the benefits of simplicity but merely gives the description above regarding the preferences of scientists. Further, although he remarks that "scientists appreciate simplicity" (ibid.), Carnap thinks that simplicity "is only of secondary importance" because "[m]any complicated concepts are introduced by scientists and turn out to be very useful" (ibid.).

Concerning the nature of conceptual simplicity, Carnap suggests the two following ways of measuring the simplicity of a concept, namely

- (a) "by the simplicity of the form of its definition", and
- (b) "by the simplicity of the forms of the laws connecting it with other concepts" (ibid.).

One may safely assume that Carnap thinks that the benefits are purely practical, in the sense that more simply defined concepts and simpler laws are easier to use.

In the literature on simplicity as a theoretical virtue, simplicity as a feature of theories is typically divided into

- (i) syntactic simplicity, sometimes called elegance, measured by the number and complexity of hypotheses or principles in a theory (Baker

³⁵ The fact that Carnap here only mentions fruitfulness might be taken as an indication that he ascribes more importance to this criterion than he seems to do in the general characterizations of the method, where he only mentions exactness.

2022) or by the number of “basic theoretical principles” (Keas 2018, 2775);

- (ii) ontological simplicity, sometimes called parsimony, measured by the number of entities or number of kinds of entities assumed by a theory (Baker 2022; Schindler 2018, 14); and
- (iii) ideological simplicity, understood as the number of primitive concepts in a theory (Cowling 2013; Quine 1951a, 14).

Brun (2016, 1224) comments that Carnap refers to syntactic simplicity. Indeed, the form of a definition and the form of laws seem to be considerations that pertain to the syntactical level. Nevertheless, to describe these features syntactic similarity might be confusing since Baker (and others) uses the term ‘syntactic simplicity’ in the sense of a measure of “the number and complexity of hypotheses” (Baker 2022). This sense of simplicity is not directly applicable to concepts. I therefore hesitate to use the notion of syntactic simplicity to characterize Carnap’s simplicity criterion.

2.8 Concepts versus terms

Carnap is explicit that he uses “concept” to refer not to words or phrases but to their meanings (LFP, 7–8). However, he does not discuss the problems, which are later raised by other authors, regarding the explicanda/explicata (there are problems both for viewing them as terms and as concepts), except when he stresses the difference between mere formalization and interpretation, concluding that for a “genuine explication” of probability “an interpretation is essential” (LFP, 16).³⁶ Although I try to avoid contested questions of interpretations in this chapter, I will make a short exception for the issue of concepts versus terms.

An early commentator who raises this issue is Joseph Hanna (1968). He is explicit that he wishes to abandon Carnap’s treatment of the explicanda and explicata as concepts and instead treat explicanda and explicata as linguistic items. As he acknowledges, this is in line with Quine (1960, 258), who talks about “defective nouns” as the objects of explications. Hanna seems to want

³⁶ As pointed out by (Brun 2016, 1216).

explication to be applicable not only to nouns but to any grammatical category, as he proposes to

view the method as applying fundamentally to linguistic terms (i.e. predicates, operators, names, etc.) and as applying only derivatively to concepts. Concepts are mysterious entities and it is best to avoid them when the same point can be made by referring to predicates. Thus, our task is to explicate ‘explication,’ rather than to explicate the concept of explication. (Hanna 1968, 30)³⁷

The issue of concepts versus terms seems not to have worried Carnap much. To repeat, he writes in M&N (and similarly in LFP) that the “earlier concept, or sometimes the term used for it, is called the **explicandum**; and the new concept, or its term, is called an **explicatum** of the old one” (M&N, 8, original emphasis). Regarding his understanding of concepts, Carnap explains that he is not talking about linguistic items:

The word ‘*concept*’ is used in this book as a convenient common designation for properties, relations, and functions. [Note that (a) it does not refer to terms, i.e., words or phrases, but to their meanings, and (b) it does not refer to mental occurrences of conceiving but to something objective.] (Carnap 1950, 7-8)

Since Carnap rejects the view of concepts as mental representations, it would seem that he is left with abstract objects. Discussing the same issue in M&N, Carnap writes that he uses ‘concept’ as referring to “something objective that is found in nature and that is expressed in language by a designator of nonsentential form” (M&N, 21). He is quick to comment that it should “by no means be regarded as an attempt toward a solution of the old controversial problem of the universals” (M&N, 22). However, Maher (2010) and Brun (2016) has brought to light a tension in the method that causes problems both for the concept-view and for the term-view. This problem is neglected both by Carnap in his defence of the concept-view and by Hanna in his defence of the term-view. Maher has pointed out that “[o]ne of the main tasks in clarifying the explicandum is to distinguish [...] different meanings and identify the one that is the explicandum. Therefore, what is being explicated is really a concept, not a term” (Maher 2010, 17). Brun notes that the clarification of the explicandum must be understood as a procedure of going from an

³⁷ In the contemporary debate, a similar sentiment is expressed by Cappelen: “I don’t think a theory of conceptual engineering is hostage to a theory of what concepts are. And that’s good because the nature of concepts is one of the most disputed topics in philosophy and psychology: there’s a plethora of theories, no agreement on theoretical role, and significant pressure towards eliminativism [Machery 2009]” (Cappelen 2018, 105).

explicandum-as-a-term to an explicandum-as-a-concept. At least this is the case when disambiguation is needed since terms but not concepts can be ambiguous, but it is a concept that is replaced by the explicatum.

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 [...] 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. (Brun 2016, 1216)

In summary, the process of explication may begin when we take interest in a term. During the first step we may select one out of several ways of using that term, i.e., we identify a concept. That is the explicandum-concept which, in the second step, is replaced by the explicatum-concept.

2.9 The ideal of scientific philosophy in Carnap's conception of explication

2.9.1 Explication as the method of scientific philosophy

I end this chapter with a reflection on Carnap's broader motivations for introducing the method of explication. Specifically, I address the role of explication in Carnap's conception of scientific philosophy. Carnap pursued several explications of traditional philosophical concepts, such as *necessary truth* (see 2.3). He writes in his reply to Charles Morris in the Schilpp volume that "many problems concerning conceptual frameworks seem to me to belong to the most important problems of philosophy [...] especially [...] the most general frameworks containing categorial concepts which are fundamental for the representation of knowledge" (RSE, 862). Schupbach summarizes Carnap's motivation for introducing explication:

Carnap's motivation for introducing explication was his longstanding dissatisfaction with prevailing concepts; he believed that such concepts tend to be vague, confused, and inexact. This is problematic because these concepts, in virtue of such inherent imprecision, ultimately hinder epistemic progress in philosophy and science. Whatever the means by which our prevailing concepts

were instilled in us, it was evidently not via an active, intentional attempt to optimize their effectiveness toward the pursuit of knowledge. Our working concepts thus are generally not tailored to the search for scientific laws and logical theorems. We would do much better at representing the world, Carnap believed, by taking an active role in developing and optimizing the conceptual schemes that we apply in thinking about the world. Explication is the philosophical method that Carnap develops to guide us to improved concepts; it allows us to construct more exact concepts for the purpose of improving their epistemic fruitfulness. (Schupbach 2017, 677-678)

Regarding expressions of ordinary language Carnap comments, that they “may be terms which in most cases are used without any difficulty” but that “it may occur that in certain critical contexts the ordinary usage leads to difficulties, unanswerable questions, even contradictions, demonstrating the surprising fact that people are not completely clear about the correct usage” (RSE, 935). Carnap’s examples of such terms are numerical, spatial and temporal terms, as well as ‘true’ and ‘entailment’. By replacing such concepts with better substitutes, futile disputes about how they are to be used may be avoided. As Shepherd and Justus (2015, 387) have pointed out, “explication was developed to place the use of concepts in philosophy on as rigorous and secure methodological footing as science”. Such radical and therapeutic ambitions for new philosophical programmes have a long history. In Rorty’s introduction to the classic 1967 anthology *The Linguistic Turn* (1967b), he summarizes attempts that have been made to once and for all overcome philosophical quarrels:

The history of philosophy is punctuated by revolts against the practices of previous philosophers and by attempts to transform philosophy into a science—a discipline in which universally recognized decision-procedures are available for testing philosophical theses. In Descartes, in Kant, in Hegel, in Husserl, in Wittgenstein’s *Tractatus*, and again in Wittgenstein’s *Philosophical Investigations*, one finds the same disgust at the spectacle of philosophers quarreling endlessly over the same issues. The proposed remedy for this situation typically consists in adopting a new method: for example, the method of “clear and distinct ideas” outlined in Descartes’ *Regulae*, Kant’s “transcendental method,” Husserl’s “bracketing,” the early Wittgenstein’s attempt to exhibit the meaninglessness of traditional philosophical theses by due attention to logical form, and the later Wittgenstein’s attempt to exhibit the pointlessness of these theses by diagnosing the causes of their having been propounded. (Rorty 1967a, 1)

Carnap is of course one of these philosophers who are objecting to “the spectacle of philosophers quarrelling endlessly over the same issues”, and his diagnosis of what has gone wrong comes first at an early stage, inspired by Wittgenstein’s *Tractatus*, and then at a later stage in which he identifies the problem as a confusion between theoretical questions internal to a linguistic framework and practical questions external to a linguistic framework.

Carnap’s method of explication is motivated by a radical programme to replace traditional philosophy with another activity. The proposal is motivated by pessimism about traditional philosophy and the desire to avoid futile controversy and debate. But there is also the positive side of the programme, namely the view that traditional philosophical activity can be replaced by concept formation conducted in the image of scientific concept construction, i.e., optimism that methods of scientific concept formation can be successfully imitated in the formation of philosophical concepts. Gustafsson (2014) summarizes this dual motivation for Carnap, which he describes as a “both radical and grandiose” vision:

In one fell swoop, finding artificially constructed replacements for inherited concepts both dissolves age-old metaphysical disputes and thereby opens up an entirely new and much more productive, positive, and distinct role for philosophy to play—that of inventing various logico-linguistic frameworks and exploring their formal properties [Friedman 2007, 16]. (Gustafsson 2014, 514)

The details and formulations of the new activity change during his career, but there is a continuity in the proposal that philosophical debates should be dismissed and replaced with constructions of language forms which are to be evaluated not for their correctness but for their utility and expedience for the purposes of science. In the 1930s Carnap proposed to replace philosophy with *wissenschaftslogik*, logic of science, in the sense of syntactic analysis of scientific language. Michael Friedman (2008) summarizes the anti-philosophical programme in Carnap’s *Logical Syntax of Language*:³⁸

Carnap aims to use the new tools of metamathematics definitively to dissolve all such metaphysical disputes and to replace them, instead, with the much more rigorous and fruitful project of language planning, language engineering—a project which, as Carnap understands it, has no involvement whatsoever with any traditional epistemological program. Indeed, as Carnap clearly and emphatically states in *Logical Syntax*, the new discipline he here

³⁸ First published in German in 1934 as *Logische Syntax der Sprache* and published in English translation in 1937 as *Logical Syntax of Language*.

calls *Wissenschaftslogik* (the logic of science) “takes the place of the inextricable tangle of problems one calls philosophy” [1937 (1934), Sect. 72]. (Friedman 2008, 393)

Given that one does make the choice to embrace the programme of explication, there is still a need to justify the further choice to pursue it for the purposes of science rather than for some other purpose, as well as the choice to apply Carnap’s four criteria of adequacy rather than some other criteria.

In a paper in *Metaphilosophy*, Hannes Leitgeb (2013), at the time editor-in-chief of *Erkenntnis*, explained the decision to change the subtitle of *Erkenntnis* to “An International Journal of Scientific Philosophy”:

We recently changed the subtitle of *Erkenntnis* from “An International Journal of Analytic Philosophy” to “An International Journal of Scientific Philosophy.” In a sense, this was old news: “scientific philosophy” had been among the terms of art that characterized the Logical Empiricist movement from the start—exemplified by the work of Rudolf Carnap and Hans Reichenbach, the founders of the journal. But it was also good news: “scientific philosophy” expresses the attitude of philosophizing that is associated with the journal more adequately than the less specific “analytic philosophy.” And I am convinced that one important way of doing philosophy in the future will be scientific. (Leitgeb 2013, 267)

In the paper, Leitgeb notes that there are at least three different conceptions of scientific philosophy, namely: (1) philosophy for the sake of science, (2) philosophy as part of science, and (3) philosophy done (partially) by scientific methods. As Leitgeb further notes, conception (1) is Carnap’s conception of scientific philosophy: it “is very close to what the Vienna and Berlin circles had in mind, and especially Carnap: philosophy as an auxiliary discipline that reflects, on the metalevel, science proper, with the aim of reforming and improving the language and logic of science wherever appropriate.” (2013, 267-268).

A fourth sense of scientific philosophy is the one used by Timothy Williamson. It has elements of both (2) and (3), but since Williamson has a broader understanding of ‘scientific’ than Leitgeb, the methodological commitments involved are different. He writes:

I do not use the term ‘science’ in the narrow sense of natural science, which involves the use of experimental and observational methods. For instance, mathematics is a science, but not a natural science. I use ‘science’ in the much broader sense of systematic, critical, evidence-constrained inquiry into how things are on some topic. In that sense, the rigorous study of history based

mainly on written documents also counts as a science, though not as a natural science. *Thus scientific philosophy is philosophy done as a science, but that does not require the assimilation of philosophy to a natural science. Where appropriate, scientific philosophy will use the methods of modern logic and mathematics, or those of history.* (Williamson 2021, 77, my emphasis)³⁹

It becomes motivated to modify Carnap's account, if one disagrees with Carnap on the answer to one or both of the following questions: (i) In which fields of inquiry should we look for paradigmatic examples of concept formation to imitate? (ii) What are the characteristic virtues of the concepts in these fields? A proposal to broaden the method of explication may be prompted by one's taking a different view than Carnap regarding questions (i) or (ii). As we will see, there are good reasons to deviate from Carnap in the answers to both (i) and (ii).

2.10 Concluding remarks

I have discussed Carnap's writings on explication, including the structure of his proposal, his motivations for favouring explication, and its role in his ideal of scientific philosophy. In the next chapter I discuss Carnap's exchange with Strawson and give an overview of other philosophers' commentary and work on Carnap's method.

³⁹ While Leitgeb has a narrower understanding of 'scientific' than Williamson, he does not claim that scientific philosophy is the only appropriate way of doing philosophy:

I do not want to say that scientific philosophy should be the only way of doing philosophy. [...] As long as some minimal criteria are satisfied that ought to be characteristic of all kinds of academic research—to speak as clearly as possible, to defend some theses, to put forward arguments, to give concrete examples, to systematize one's thoughts, to compare one's view with rival ones, to stay critical of one's theory, and so on—let many flowers bloom, also and especially, in philosophy. It is just that scientific philosophy is a particularly beautiful flower; and it is blossoming. (Leitgeb 2013, 275).

3 Early commentary and debate on Carnap's method

3.1 Overview

In this chapter I discuss Carnap's exchange with Strawson and give a survey of other philosophers' comments and work on Carnap's method of explication, up until the interest in it exploded in the last ten to fifteen years. I begin with and focus on the exchange with Strawson because it is the source of much contemporary debate both in the literature on explication and in the literature on conceptual engineering in general.

3.2 The exchange between Strawson and Carnap

In this section I discuss the exchange between Strawson and Carnap in the 1963 Schilpp volume, where Strawson rejects (1963, 503–518) and Carnap defends (RSE, 933–940), explication as a plausible philosophical method. It is currently debated how Strawson's objection to explication should be understood. Hence, the exchange is not only of interest as historical background to the current debate but is also a part of what is being debated (see e.g., Pinder 2020b and Koch 2023).

Strawson's challenge is arguably still one of the main challenges for those advocating the use of explication as a philosophical method. In recent literature on conceptual engineering the objection is often referred to as the problem of topic preservation or topic continuity (Cappelen 2018, 107–121) or as 'the topic shift worry' (Kitsik 2020, 1043). In *Fixing Language*, Cappelen (2018) devotes three whole chapters to the challenge, claiming that a response to it is "a central task for any theory of conceptual engineering" (2018, 99, n. 3). Koch (2023), however, claims that Strawson's challenge does not pose a problem for conceptual engineers and that they should stop worrying about how to preserve topics or subject matters. I return to this issue in chapter 7.

Strawson expresses his critique in a striking analogy. According to Strawson, the problem with explication as a method for solving philosophical problems is that the solution is irrelevant because it changes the subject. The following two quotations are the central expressions of Strawson's objection. First as an analogy:

it seems prima facie evident that to offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant—is a sheer misunderstanding, like offering a text-book on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (Strawson 1963, 505)

In Strawson's view, 'typical' philosophical problems are rooted in the unclear mode of functioning of ordinary concepts, and therefore they cannot be solved by scientific concepts:

typical philosophical problems about the concepts used in non-scientific discourse cannot be solved by laying down the rules of use of exact and fruitful concepts in science. *To do this last is not to solve the typical philosophical problem, but to change the subject.* (Ibid., 506, my emphasis)

Two comments are in order here. First, it is Strawson who brings the notion of subject matter into the picture, while Carnap, in his similarity criterion, only talks about the explicatum having to be useful in most cases where the explicandum is useful, or about its being useful for the same purposes. Second, Strawson's issue with explication seems to be directed at its use as a philosophical method, as a way of clarifying philosophical problems. In Strawson's view, philosophical problems arise from our usage of concepts in ordinary language. For example, he takes issue with the replacement of *warm* with *temperature*:

I may mention again Carnap's own example of the clarification of the prescientific concept of warmth by the introduction of the exact and scientifically fruitful concept of temperature. Sensory concepts in general have been a rich source of philosophical perplexity. How are the look, the sound, the feel of a material object related to each other and to the object itself? Does it follow from the fact that the same object can feel warm to one man and cold to another that the object really is neither cold or warm nor cool nor has any such property? (Ibid., 506)

These problems cannot be solved, according to Strawson, by the introduction of scientific concepts such as temperature, wavelength, and frequency (*ibid.*). Carnap counters with an example in which age-old philosophical problems could be answered only when formal precision and sophistication were introduced. The problems are Zeno's paradoxes, resulting from a "misinterpretation of the expressions describing motion" (RSE, 935). For their solution,

certain parts of mathematics are needed which go far beyond elementary arithmetic, such as the theory of real numbers, the concept of the limit of a series, and finally the proof that certain infinite series are convergent, i.e., that every member of the series is greater than zero and nevertheless the sum of the whole series is finite. In this case, the perplexities were formulated in the natural language. But the diagnosis consists in the demonstration that certain apparently valid forms of inference involving the infinite are fallacious and lead to contradictions. (RSE, 939–940)

Carnap's point is that the perplexities cannot be solved using the concepts that gave rise to them, and that new concepts with new rules of use are required to prevent the contradictions. He illustrates this point more generally with an analogy of his own. While Strawson likened explication to offering a textbook in physiology to some who talks metaphorically about the human heart, Carnap likens explication to offering or using a more suitable tool for a specific task:

A natural language is like a crude, primitive pocketknife, very useful for a hundred different purposes. But for certain specific purposes, special tools are more efficient, e.g. chisels, cutting-machines, and finally the microtome. If we find that the pocket knife is too crude for a given purpose and creates defective products, we shall try to discover the cause of the failure, and then either use the knife more skillfully, or replace it for this special purpose by a more suitable tool, or even invent a new one. The naturalist's thesis is like saying that by using a special tool we evade the problem of the correct use of the cruder tool. But would anyone criticize the bacteriologist for using a microtome, and assert that he is evading the problem of correctly using a pocketknife? (RSE, 938–939)

Is this an appropriate response to Strawson's objection? To repeat, Strawson's objection is that a formal explication is irrelevant for "one who seeks philosophical illumination of essential concepts of non-scientific discourse" (Strawson 1963, 5). Patrick Maher (2007, 332–333) defends explication against Strawson's critique, but objects to Carnap's knife analogy, arguing that nobody would criticize the bacteriologist who uses a microtome only "because the bacteriologist's problem was not about the pocketknife". He goes on:

However, the relevant analogy for “one who seeks philosophical illumination of essential concepts of non-scientific discourse” is someone who seeks knowledge of proper use of the pocketknife; Carnap has offered nothing to satisfy such a person. (Maher 2007, 333)

With the example of Zeno’s paradoxes in mind, one could perhaps reply to Maher that there is nothing that could satisfy such a person, since no illumination can come out of an everyday language conception of infinity treated in natural language. And the problem arises precisely because the person is not content with using the pocketknife in the crude way that is appropriate for a pocketknife.⁴⁰

Furthermore, Strawson explicitly states that philosophical problems are not about how to use everyday expressions but about the puzzles that arise when we use them:

it is characteristic of philosophers’ perplexities and questions that they are felt and raised by people who know very well how to use the expressions concerned, who have no practical difficulties at all in operating with the concepts in question. (1963, 508–509)

Strawson stresses this point further, claiming that he finds in Carnap’s writings

evidence of a lack of sympathy with, and even of understanding of, that need for the elucidation of concepts which can coexist with perfect mastery of their practical employment. Now this is precisely the need for their philosophical elucidation. (1963, 509)

This seems to be a genuine point of disagreement, from which much of their methodological differences follow. Strawson’s point is that since philosophical perplexities arise from concepts that people know how to use, the suitable method for solving the perplexities is not to work on and try to improve the concepts themselves. Carnap explicitly objects to this claim in his reply:

Strawson believes that philosophical problems are raised by people “who know very well how to use the expressions concerned”. I should rather say that these people usually believe they know this very well, but often deceive themselves. (Carnap 1963b, 935)

So Carnap thinks that people often *falsely believe* that they know how to use certain expressions, and that this leads them into avoidable perplexities.

⁴⁰ See 4.2 for further discussion.

According to Carnap's diagnosis the problem is twofold. The first problem is that people use their expressions in a confused way and the second problem is that they think that they know how to use these expressions. To fix the first problem by offering explications, one must begin with the second problem:

The first step in helping these people consists in leading them to the insight that something is wrong with their use of certain expressions, that it involves confusions or even inconsistencies. Frequently the puzzle concerns expressions of ordinary language [...]. These may be terms which in most cases are used without any difficulty. But then it may occur that in certain critical contexts the ordinary usage leads to difficulties, unanswerable questions, even contradictions, demonstrating the surprising fact that people are not completely clear about the correct usage. (RSE, 935)

Once again, Zeno's paradoxes would be an example of this "surprising fact". Going back to Strawson's critique, it is clear that Strawson is mainly concerned with the idea that scientific explicata would replace ordinary language explicanda in everyday situations. Solutions to Zeno's paradoxes are seldom required in daily life, so Strawson would perhaps not be so concerned with that example. However, he thinks that

it seems in general evident that the concepts used in non-scientific kinds of discourse could not literally be *replaced* by scientific concepts serving just the same purposes; that the language of science could not in this way *supplant* the language of the drawing-room, the kitchen, the law courts and the novel. [...] [I]t seems to require no argument to show that, in most cases, either the operation would not be practically feasible or the result of attempting it would be something so radically different from the original that it could no longer be said to be fulfilling the same purpose, doing the same thing. (Strawson 1963, 505)

Whether or not there is an actual conflict here seems to hinge on the phrase "in most cases". Carnap would surely agree that scientific language could not and should not replace all of everyday language. The point is to replace everyday language in scientific and theoretical contexts to serve scientific and theoretical needs. As mentioned in section 2.2, Carnap notes that the explicatum can sometimes also fulfil the same purpose as the explicandum in more efficient ways also in daily life. His examples were the use of *temperature* and *speed* (as a quantitative concept), and expressions such as '4:30 PM'. But his recognition of such cases does not make it into a general rule or aim for explication to replace everyday language. Just as explication is motivated by usefulness, the integration of explicata in everyday language is for Carnap also

a matter of usefulness. This brings us to the last issue of this section, which is the boundary between scientific and non-scientific language.

As we have seen, Strawson emphasizes the inability of scientific concepts to replace non-scientific language, claiming that it either would not be feasible or would not achieve the same purpose. The way Carnap understands this disagreement with Strawson is as a disagreement about the boundaries between everyday language and scientific language:

I have the impression that Strawson's view is based on the conception of a sharp separation, perhaps even a gap, between everyday concepts and scientific concepts. I see here no sharp boundary line but a continuous transition. (RSE, 934)

Specific purposes call for specific tools, and the harder the problem the more sophisticated a tool is needed. With this background established I will return throughout the thesis to the debate and the challenge posed by Strawson.

3.3 Early writing on explication

Carnap's notion of explication was quickly picked up by philosophers in his sphere, such as W. V. O. Quine (1951b, 25), Carl Hempel (1952, 11–12), and Paul Oppenheim and John Kemeny (1952; Kemeny 1959, 105–109). In Hempel's *Fundamentals of Concept Formation in Empirical Science* (1952), he discusses explication with reference to Carnap's LFP. Comparing explications to real definitions, Hempel concludes that explication "combines essential aspects of meaning analysis and of empirical analysis" (Hempel 1952, 11). He formulates the aim of the task in the following way:

Taking its departure from the customary meanings of the terms, explication aims at reducing the limitations, ambiguities, and inconsistencies of their ordinary usage by propounding a reinterpretation intended to enhance the clarity and precision of their meanings as well as their ability to function in hypotheses and theories with explanatory and predictive force. Thus understood, an explication cannot be qualified simply as true or false; but it may be adjudged more or less adequate according to the extent to which it attains its objectives. (Hempel 1952, 12)

Although Hempel does not announce any wish to deviate from Carnap, he does not follow Carnap to the letter here. For one thing, he mentions disambiguation as one of the aims of explication, whereas according to Carnap's methodology

it is a precondition for an explication that the explicandum already is disambiguated, or is being disambiguated in the first step, i.e., in the clarification of the explicandum, rather than as an end result.

Hempel not only discusses explication but also performs explications. In LFP, Carnap discusses Hempel's 1945 paper "Studies in the Logic of Confirmation" as an example of explication, although in that paper Hempel does not yet use the term 'explication'.⁴¹ Later, Hempel himself regards his influential Deductive-Nomological account of scientific explanation as "in the nature of an *explication*" (Hempel 1965, 489, original emphasis), although not as a completed explication: "Actually, our explicatory analysis has not even led to a full definition of a precise 'explicatum'-concept of scientific explanation; it purports only to make explicit some especially important aspects of such a concept" (ibid.).

Another important early commentator on explication is Quine, who briefly discusses "what Carnap calls *explication*" (1951b, 25) in "Two Dogmas of Empiricism", but fully embraces explication in *Word and Object* (1960). Quine appears to consider his conception of explication to be more or less the same as Carnap's (1960, 259, n. 4), although, as we will see in chapter 6, there are significant differences between them: for instance, as Gustafsson (2006, 2014) emphasizes, Quine but not Carnap ascribes an ontological importance to explication (2014, 509), since in Quine's view explication is a tool for reforming our total theory to be as ontologically sparse as possible (2006, 58).

Apart from Hempel and Quine, various other philosophers also picked up on Carnap's ideas at the time. In a paper commenting on the Strawson–Carnap exchange, Frank A. Tillman (1965) finds a one-sided basis for cooperation between the methods, claiming that the ordinary language analyses, or "linguistic portrayals" as he calls them, could be of help in the first step of explication, that is with the clarification of the explicandum (Tillman 1965, 382–3). In a review of the Schilpp volume dedicated to Carnap, Peter Achinstein (1966) focuses on explication. However, Tillman does not comment on Strawson's objection that explication changes the subject. Shortly before Carnap's death in 1970, Hanna published the paper "An Explication of 'Explication'" (1968), in which he claims that Carnap holds inconsistent views about when an explication fulfils the similarity criterion of adequacy, and proposes both a disambiguation of 'explication' and an explication of the terms. As we saw already in 2.8, Hanna wishes to abandon Carnap's treatment of the explicanda and explicata as concepts and instead consider them to be

⁴¹ As previously mentioned, 1945 is the year in which the term 'explication' in Carnap's sense first appeared in print.

linguistic items. And shortly after Carnap's death, Michael Martin (1973) proposed to extend the method of explication from concepts to theories. However, his account has restricted applicability since, as Brun has pointed out, Martin assumes that the only goal of explication is to reduce vagueness (Brun 2020, 934, n. 16).

For a couple of decades after this, not much attention was paid to the method, apart from sporadic mentions. For example, in Wesley Salmon's *Scientific Explanation and the Causal Structure of the World* (1984), Salmon credits Carnap's LFP for convincing him of the importance of an informal clarification of the explicandum, in his case *scientific explanation*, before one proceeds to give a more precise account of it.

I have been convinced for some time that many recent philosophical discussions of scientific explanation suffer from a lack of what Rudolf Carnap called "clarification of the explicandum." As Carnap has vividly shown, precise philosophical explications of important concepts can egregiously miss the mark if we do not have a sound prior informal grasp of the concepts we are endeavoring to explicate. (Salmon 1984, x)

Although some philosophers and scientists were performing explications or something close to it in these interval years, there is not much explicit reflection on the method or discussions of its problems and its role in theoretical work. As Carnap was well aware, this is just a return to the normal state: "Philosophers, scientists, and mathematicians make explications very frequently. But they do not often discuss explicitly the general rules which they follow implicitly" (1950, 7).

3.4 Renewed interest in explication

Among the writings that have played an important role in sparking the renewed interest in explication that we are witnessing today, there is a clear Chicago-based lineage going back to Carnap himself. In 1992 Carnap's former student in Chicago Howard Stein published the paper "Was Carnap Entirely Wrong, After All?", in which he set out to reassess the criticism brought by Quine against Carnap and the program of logical empiricism. The modest title of his paper is probably telling for how outdated Carnap's philosophy was regarded to be at the time. After an assessment of Quine's criticism and of Carnap's positions, he goes on to treat "the character of Carnap's philosophy, as it is manifested in his later writings" (*ibid.*, 275). Stein's motivation for this

addendum is that he “believe[s] that Carnap is a far subtler and a far more interesting philosopher than he is usually taken to be” (ibid.). An interesting aspect of Stein’s paper is that he is likely the first or among the first to discuss explication in relation to Carnap’s distinction between internal and external questions, introduced in the paper “Empiricism, Semantics, and Ontology”, published in 1950. In 2007, Stein’s student A. W. Carus discussed and historically contextualized explication at book-length in *Carnap and Twentieth-Century Thought: Explication as Enlightenment* (2007), and Carus’ interpretation of Carnapian explication is extensively discussed in the anthology *Carnap’s Ideal of Explication and Naturalism* (2012), edited by Pierre Wagner. A notable contribution is Reck’s “Carnapian Explication: A Case Study and Critique.”

Before Carus’s book, Beaney (2004) had given an overview of Carnapian explication and the influences of his predecessors in the anthology *Carnap Brought Home: The View from Jena* (Klein and Awodey 2004). Eric Loomis and Cory Juhl (2006) wrote a detailed encyclopaedia entry on explication, where they characterized Carnapian explication both as a form of “conceptual clarification” and as a form of “linguistic engineering”. In contrast, Schupbach (2017) reserves the term ‘concept clarification’ for Oppenheimian explication (see 5.3) and uses the term ‘concept engineering’ to distinguish Carnapian explication from mere ‘concept clarification’. In relation to probability theory, Patrick Maher discussed Carnap’s explication of probability as well as the method of explication in a series of papers (2006; 2007; 2010). In his 2007 “Explication Defended”, he defends Carnapian explication from Strawson’s challenge, among other things, and posits it as a viable method specifically for formal philosophy. Martin Gustafsson discusses Quine’s conception of explication in *Word and Object* in a 2006 article (see 2.4), and in a subsequent text (2014) compares Quine’s conception of explication with Carnap’s conception. As mentioned in 1.5.1, Murzi (2007) considers the redefinition of ‘planet’ in 2006 to be a case of Carnapian explication; and in recent years this case has become a stock example in the literature on explication and conceptual engineering. Theo Kuipers (2007a) characterizes explication as “an important, though largely implicit, method in philosophy of science”. Reck (2007) discusses the Frege–Russell numbers in terms of Carnapian explication. Jörgen Sjögren (2011) discussed Carnapian explication at length in his doctoral thesis “Concept Formation in Mathematics”.

In the last fifteen years or so, the literature on explication has grown quickly.⁴² The most detailed and systematic analysis and critical discussion of Carnap's own views of explication is probably George Brun's 2016 paper "Explication as a Method of Conceptual Re-engineering". Brun notes that Carnap in LFP lays out his views of explication in a way that "raises a number of questions of interpretation and systematic problems" (2016, 1213). Nevertheless, Brun's aim with the paper is to develop an account of explication which is "Carnapian in spirit but not tied to his specific philosophical projects of semantics and inductive logic" (1212). To solve systematic problems in a way that remains within Carnap's spirit, Brun uses Carnap's clarificatory commentary in the Schilpp volume (RSE).

3.5 Concluding remarks

This concludes the summary of notable literature on explication. Immediately after Carnap introduced the method of explication and its terminology, other philosophers adapted the terminology, adapted the method (Hempel 1952; Kemeny 1959), contested the method (Strawson 1963; Kemeny and Oppenheim 1952), and, with more or less explicit awareness of doing so, modified the method (Kemeny and Oppenheim 1952; Quine 1960). After a period of decline of influence and interest in Carnap and logical empiricism in general, a first small wave of renewed interest in Carnapian explication occurred in the 1990s and 2000s, while an explosion of new interest occurred in the 2010s. Drawing on discussions in the recent literature on explication, I devote the next chapter to a discussion of the purpose of explication, and the internal structure of the criteria of adequacy.

⁴² I have not sought to provide not an exhaustive list but a selection of notable contributions to the literature on explication.

4 The purpose of explication

4.1 Overview

In this chapter I discuss questions regarding the purpose of explication. Explications are judged not for their correctness but for their utility for some purpose. To have a clear view of the purpose is therefore important when we construct and evaluate explications of particular concepts. It is also important when we evaluate and develop the standards of evaluation appealed to when we judge explications of particular concepts. This latter task is the main focus of most of the remaining chapters. Regarding this task, it is important to distinguish between differing views about how to successfully perform an explication and differing views about the intended use of an explicated concept. To clarify this issue I introduce a distinction, based on a distinction made by Justus (2012), between the *immediate purpose* of explication and the *ultimate purpose* of explication. I argue in 4.3 that some recent interpretations of Carnapian explication are flawed because these two purposes are not held apart.

Further, I will distinguish and discuss two questions related to purpose which are often not held apart in the literature. The first question regards the relative weight given to the criteria of adequacy in explications of particular concepts. Presumably the purpose is what determines which criteria of adequacy are most important, so the balance varies between different projects of explication. The second question regards the internal structure between the criteria of explication, which is constant. How much relative weight one should give to the criteria may depend on one's purposes (for example, exactness may be the most important criterion for one purpose, while fruitfulness may be the most important criterion for another purpose). In contrast, the structure does not vary between explications of particular concepts. On Carnap's account, for example, simplicity is always of secondary importance and there is no purpose such that most weight should be given to simplicity. However, as we will see, extant accounts of explication differ with regard to the internal structure.

In the following section, I introduce the already mentioned distinction between immediate and ultimate purpose. I use it later in the chapter to clarify

debates, and I comment on interpretations by other authors regarding the purpose of Carnapian explication. In 4.3, I discuss the relative weight of the criteria, and in 4.4 I discuss the internal structure among the criteria of adequacy. I spell out the internal structure of Carnap's criteria, as well as the internal structure of Brun's influential interpretations of Carnap's account, and the internal structure of Brun's own account of explication, which is the most comprehensive and ambitious development made of Carnap's account. I end with a preliminary exposition of the internal structure of my own proposal to modify Carnap's account, which I develop further in chapter 9.

4.2 Immediate and ultimate purpose

As a tool both for interpreting Carnap's views and developing the method, I introduce in this section a distinction between the immediate purpose of explication and the ultimate purpose of explication. The distinction is inspired by Justus (2012), who discusses the putatively different aims of concept formation in science and conceptual analysis in philosophy. Before I introduce my distinction between immediate and ultimate purpose I present Justus's discussion.

The context in which Justus distinguishes between immediate and ultimate aim is his discussion of Strawson's critique of the use of explication in philosophy. Justus argues, against the interpretation by Loomis and Juhl (2006), that Carnap's response to Strawson is successful. Loomis and Juhl think that "it is not obvious that Carnap's analogy [of pocketknives and microtomes (see 3.2)] adequately answers Strawson's charge of the irrelevance of explication for unravelling perplexity involving ordinary notions" (Loomis and Juhl 2006, 291). They give two reasons, or, as they call it, "obvious rejoinders":

- (1) pocketknives are not replaceable by microtomes for most ordinary uses and
- (2) someone who was having trouble using a pocketknife in an ordinary circumstance would not be helped in the least by being shown the workings of a microtome. (Ibid.)

Justus is not impressed by the objections. He correctly points out that the reason someone is having trouble using a pocketknife (i.e., a natural language

term) might be that they expect it to do more than it was designed (or evolved) to do. Justus replies:

Although this may be true, its critical force against explication and Carnap's instrumental view of language is unclear. Continuing the analogy, if the pocketknife corresponds to a term of natural language denoting a concept and difficulty arises in an ordinary circumstance—i.e. one in which the term is normally used—then it seems the troubled concept user does not understand the term's actual meaning and should consult a lexicon. If the problem persists, it is likely the individual is expecting more of the concept than it delivers in ordinary circumstances where it functions unproblematically. (Justus 2012, 176)

In Justus's words the critique presupposes a "putative sharp distinction between philosophical and scientific concepts" (176) and also the belief that the analysis of philosophical concepts and scientific concepts "requires disparate evaluative standards" (ibid.). Hence, the problem as Strawson sees it lies not with explication in science but only with the use of explication in philosophy. Therefore, Justus continues, "Strawson could agree with Carnap about the advantages of precision for properly scientific concepts such as 'ecological stability,' but insist that exclusively philosophical concepts such as 'knowledge' or 'personal identity' require a different approach" (ibid.). Is there such a sharp distinction between scientific and philosophical concepts as the Strawsonian critique seems to presuppose? Justus opposes the idea of a sharp line, arguing that philosophy of science is a field where the distinction is particularly blurred:

Many concepts central to philosophy of science seem to defy label as strictly philosophical or scientific and it is generally unclear on what basis, besides mere disciplinary division, such a distinction between putative types of concepts could be drawn. There are numerous examples: 'confirmation,' 'disposition,' 'inference,' 'law,' 'model,' 'natural kind,' 'observation,' 'probability,' 'representation,' 'space,' and 'time' among others. (Justus 2012, 177)

The list could be continued with further concepts, e.g., 'causation'. With regard to these concepts in the philosophy of science, Justus argues that explications of them and conceptual analyses of them share the same "ultimate goal". He distinguishes between what he calls the ultimate goal and what he calls the "*immediate aim*" of analyses of concepts in philosophy of science. The immediate aim according to Justus is to improve "the understanding of their role (or potential role) in scientific practice" (Justus 2012, 177), thus improving

the understanding of “the bearing empirical and theoretical results have on characterizations of these concepts and statements containing them” (ibid.). In contrast to that immediate aim “the ultimate payoff in understanding [that] philosophical work affords is measured in the same enduring currency scientists use to measure success: well-confirmed generalizations” (ibid.). Justus admits that the goal of finding well-confirmed generalizations “may seem orthogonal to the purpose of analysis in philosophy of science,” but he maintains that “this conflates distance with irrelevance” (ibid.). While Justus recognizes that “[a]nalyzes by philosophers of science and scientists often differ in degree of abstraction and generality”, he concludes that they “do not differ in kind or in their ultimate goal” (ibid.). Here Justus seems to advocate for the naturalist view that philosophy and science are methodologically continuous, which is closer to Quine than Carnap (see 2.9). However, I will not address that issue further, since the point of presenting Justus’s distinction was to borrow it for my own purposes. It would be useful to distinguish between immediate and ultimate purpose not only when discussing the purpose of conceptual analysis in philosophy of science but also when discussing the purpose of explication in general.

Hereafter I apply Justus’s distinction internally to explications. The *immediate purpose* of explication is to acquire a sufficiently similar and useful concept to take the place of the explicandum. The ultimate purpose is whatever we want to achieve *by the use* of our explicated concepts. However, contra Justus, I think that the ultimate purpose should be considered broader than well-confirmed generalizations. In the ultimate sense the purpose of explication is to gain new knowledge, or advance inquiry, or improve our epistemic situation. To fulfil the ultimate purpose is not solely a task for the explicator but a task for the research community as a whole. To make this difference between the two kinds of purpose clearer we may also distinguish between the narrow process of explication and the wide process of explication. The narrow task of explication is completed when the immediate purpose is fulfilled. The wide task is to continually enhance our epistemic situation with the use of the explicatum. This task is not exclusive to the explicator and it may continue indefinitely.⁴³

⁴³ In an ideal situation this process would stop only if either the concept was to be replaced in a second explication and then abandoned (which would presumably happen if it were considered obsolete in all contexts of inquiry), or, if a final stage of science was to be reached.

In recent literature on explication, it is debated what the purpose of explication is⁴⁴ but it is not always clear which of the two purposes it regards. In my interpretation of Carnap's view, put in my terminology of immediate and ultimate purpose, the immediate purpose of explication is to achieve concepts that are both exact and potentially fruitful, as well as sufficiently similar to the explicandum and, secondarily, as simple as the fulfilment of these former criteria permits. The ultimate purpose of explication is according to Carnap the same as the purpose of science (detecting well-confirmed generalizations and proving theorems).

4.3 The relative weight of the criteria in light of my distinction

Now that I have introduced the distinction between immediate and ultimate purpose, I will address questions regarding the relation between (1) the purpose of explication according to Carnap and (2) the relative weight given by Carnap to each criterion. Plausibly, (2) is derivative of (1). The weight given to a criterion should be conducive to the purpose of the task. Consider a comparison with conceptual analysis. If the purpose of the task is to uncover the hidden meaning of a term, then similarity between analysandum and analysans should be prioritized over e.g., scientific fruitfulness. Not much is said by Carnap regarding the relative weight of the three primary criteria. He does not explicitly address the relative importance of them. (Carnap considers simplicity to be of secondary importance, but the status of the simplicity criterion is treated in the following section which is devoted to the internal structure among the criteria.) How Carnap should be interpreted on this issue is an open question with various answers in the literature.

but he does not address any priorities among the other three criteria. As mentioned in 2.5.2, when Carnap characterises explication colloquially he only mentions exactness, which would suggest that he takes it to be the most important criterion. In *Meaning and Necessity* (hereafter M&N), Carnap characterizes explication as “the task of making more exact a vague or not quite exact concept” (M&N, 7–8) and in *Logical Foundations of Probability* (hereafter LFP) as the task of “transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second” (LFP, 3). In “Replies and Systematic Expositions” (hereafter RSE) in the Schilpp

⁴⁴ Sometimes it is called “the agenda” of explication (Shepherd and Justus 2015).

volume, he even remarks that “[t]he only essential requirement is that the explicatum be more precise than the explicandum” (RSE, 936). These formulations indicate that Carnap takes exactness to be the most important criterion, even though he makes no explicit mention of different priorities among the three primary criteria. Such a natural interpretation seems to be made by Wesley Salmon, when he remarks that “[u]nless the explicatum is precise it does not fulfil the purpose of explication, namely, the replacement of an imprecise concept by a precise one” (Salmon 1989, 5). However, there is a debate regarding the priority between the criteria. Despite the apparent plausibility of Salmon’s interpretation given the quoted definitions by Carnap, it is clear from textual evidence elsewhere in Carnap’s writings that fruitfulness should be included as well.

As noted by Schupbach (2017, 676), it would be tempting to make the interpretation that Carnap places the three primary criteria on equal footing. However, Schupbach emphatically rejects that interpretation:

Carnap gives us some, but not much, explicit guidance on how properly to balance the desiderata. He notes that simplicity is of secondary importance to similarity, exactness, and fruitfulness. [...] He does not, however, make any parallel statements about the relative importance of the first three desiderata. It is tempting to interpret Carnap’s silence on this matter as an indication that he places similarity, precision, and fruitfulness all on equal footing, in the sense that no one of these desiderata is always to be preferred over and above the others. However, Carnap’s uses and examples of explication, taken together with his general philosophical inclinations, rule out this interpretation. (Schupbach 2017, 676–7)

Ruling out this interpretation, Schupbach concludes that “Carnap [...] subordinates similarity to fruitfulness” (2017, 677), and he ascribes to Carnap the view that “one’s explicatum is satisfactory to the extent that it is as similar to the explicandum as fruitfulness allows” (*ibid.*) On Schupbach’s interpretation of the relative weight given to the criteria by Carnap, one may conclude that fruitfulness is the purpose of explication; and in some of Carnap’s discussions of specific cases of scientific explication, it does indeed seem as if fruitfulness is the most important criterion. This interpretation of Carnap is also suggested by Reck (2012) and argued for by Dutilh Novaes and Reck (2017). However, such a view is far from uncontroversial, and diverges from interpretations of Carnap made by Olsson and arguably by Brun and Pinder.

We begin with Olsson who in his 2017 observes that:

The methodology of explication should be understood to imply that all four requirements on an explicatum be given positive weight. After all, Carnap himself refers to the conditions on an explication as “requirements”. (Olsson 2017, 30)

Further, he concludes that “The crucial observation here is that the goodness of an explication is a matter of satisfying all four desiderata, as a package, to as high a degree as possible” (ibid.). However, as Olsson acknowledges, that leaves open the question of how to weigh the requirements against each other. He adds that:

Carnap does not give any rule for how to weigh the different considerations against each other in cases where there are different plausible ways of explicating the same explicandum. His only advice, as we saw, is that simplicity should generally be the least important concern. (Olsson 2017, 36)

The suggestion given by Olsson, both as an interpretation of Carnap and as a recommendation, is that the purpose of one’s investigation determines how the weight should be distributed between the requirements.

[I]t is plausible to think that the relative weight of the desiderata will be, to some extent, guided by the context. Thus, one context may require a very exact explication of knowledge, e.g. for the purpose of AI programming. In another context, a relatively unpolished explicatum may be sufficient if it proves to be empirically or otherwise fruitful. [...] How we choose to proceed depends considerably on the purpose of the investigation. (Olsson 2017, 36)

In summary, on Olsson’s interpretation, a good explication requires that all criteria are satisfied as a package to as high a degree as possible, and it is the purpose of the task at hand that determines the relative importance of the three primary criteria, and therefore what counts as possible. Presumably, though, our purposes cannot determine that the criterion of simplicity should be prioritized above the first three criteria. Therefore, I discuss the status of simplicity in the section devoted to the internal structure among the criteria.

Now, how can this plausible interpretation be squared with the view that fruitfulness is the purpose of explication? Let us move on to Pinder, who distinguishes between the question of what it is to be an explication and the question of what it is to be a good explication. Regarding the view that fruitfulness is the most important criterion, advocated by Schupbach (2017), Pinder has the following to say:

Carnap is typically held to have prioritised the fruitfulness requirement over similarity, precision and simplicity. Thus, for example, Schupbach writes that “Carnap plays favorites with regards to his desiderata, prioritizing fruitfulness over similarity” [2017, 678] and Dutilh Novaes and Reck write that “fruitfulness is ultimately the most significant requirement for an explication overall” [2017, 202]. The thought, as I understand it, is twofold. *First, a concept requires a sufficient degree of fruitfulness, similarity, precision and simplicity to count as an explicatum. Second, given multiple candidate explicata—that is, concepts with a sufficient degree of fruitfulness, similarity, precision and simplicity—fruitfulness is prioritised as the most important factor in determining which of those candidates to choose.* (Pinder 2020b, 958, my emphasis)

While Olsson and Pinder agree on what it takes for a concept to count as an explicatum (that all four criteria are satisfied), we can now see that there is at least one significant difference in Olsson’s and Pinder’s respective views. Olsson’s view is that the *goodness* of an explication requires satisfaction of all four criteria as a package to *as high a degree as possible*. Presumably, on his account, given multiple candidate explicata we should choose the one that satisfies all four criteria to the highest degree possible given the purpose at hand. If the purpose requires that most weight is given to exactness, then exactness is the most important factor in determining which candidate to choose. On Pinder’s view, for a concept to count as a candidate explicatum it has to satisfy the four criteria. However, when we have several candidate explicata and the question arises of which to choose, i.e., which is the best one, Pinder thinks that we should prioritize fruitfulness.⁴⁵ He does not, however, think that we should build the priority of fruitfulness into the method:

Importantly, though, we ought not to build the prioritisation of fruitfulness into the method of explication per se. An explication is performed so long as the explicandum is replaced with a candidate explicatum, whichever factors one subsequently prioritises in determining which candidate explicatum to choose. (Pinder 2020b, 958, my emphasis)

Dutilh Novaes and Reck (2017) maintain both (i) that “it is evident that the criterion of exactness is at the core of Carnapian explication” and (ii) that

⁴⁵ We may further compare Olsson’s interpretation with Pinder’s view (which will be discussed in chapter 7). In a sense they share the pluralistic view that the context and purpose of inquiry determines the explication, but with an important difference. Olsson thinks that the purpose determines the relative weight given to the criteria, e.g., to prioritize the fruitfulness criterion, but Pinder thinks that the purpose determines what the specific measure of fruitfulness should be.

“fruitfulness is ultimately the most significant requirement for an explication overall” (ibid., 202). While they acknowledge that exactness is at the core, they claim that “it is not clear how one could make a compelling case for the desirability of exactness as a goal in and of itself” (ibid., 211). They suggest that the best Carnapian reply to this challenge is to “*subordinate the desideratum of exactness to that of fruitfulness*, i.e., to argue that exactness can be useful in various circumstances” (2011, original emphasis). Similarly, when commenting on Carnap’s view, Shepherd and Justus (2015, 388) claim that:

Precision for precision’s sake is not the agenda. Rather, enhancing precision usually enhances fruitfulness, which is the agenda.

Indeed, I agree that the goal is not precision for precision’s sake and exactness is not desirable as a goal in and of itself. However, the same may be said of fruitfulness. The goal is not fruitful concepts for the sake of fruitful concepts, at least given Carnap’s account of what a fruitful concept is, namely one that facilitates the formulation of universal statements. It would be hard to make a compelling case for the desirability of many universal statements in and of itself, but in Carnap’s view, universal statements are necessary both for prediction and explanation and hence for the advancement of scientific knowledge.

Dutilh Novaes and Reck (2017) do not only advocate the view that exactness should be subordinated to fruitfulness, but they also argue for a broader notion of fruitfulness (more on this in chapter 7). According to them,

Carnap’s view seems to be that an explication is useful or fruitful when it delivers ‘results’ that could not be delivered otherwise (or with much more difficulty), i.e. with the explicandum alone. What this suggests is a conception of explication as a method for *discovery* [...] The goal is to produce *new knowledge*. (Dutilh Novaes and Reck 2017, 206)

However, Dutilh Novaes and Reck are no longer talking about the immediate purpose here. Rather, when taking fruitfulness and usefulness (for discovery) to be the same thing, they seem to conflate the immediate and the ultimate purpose. Carnap’s view, as I understand it, is that an explicatum is useful for gaining new knowledge through prediction and explanation, in virtue of being both fruitful and exact. He formulates the criteria of exactness and fruitfulness the way he does in order for explicata to serve the ultimate purpose. Since universal statements according to Carnap are required for new knowledge, the fruitfulness criterion serves the ultimate purpose. If fruitfulness is equated with

the prospect of achieving new knowledge, in whatever way, then the immediate and the ultimate purpose have been conflated.

One could perhaps argue that the immediate purpose of explication is to construct a concept suitable for the formulation of many universal statements and that the exactness of the concept is of subordinate importance. However, as an interpretation of Carnap's view of explication this seems too drastic. After all, there is still the fact that Carnap emphasizes exactness but not fruitfulness in his general characterizations.

Both exactness and fruitfulness are needed for the ultimate purpose. While only the immediate purpose needs to be fulfilled for an explication to be successful, the hope is that the explicatum will be picked up and put to use in explanatory or predictive or theorem-proving work, or that the explicator is able to do so. But, as I understand Carnap's views, an explicator sharing Carnap's views, who constructs an exact and fruitful concept has fulfilled her duty *qua* explicator, even before the concept is put to further use in empirical or logical inquiry.

I take it that fruitfulness on Carnap's account is not a goal in and of itself, just as exactness is not a goal in and of itself. Both are means toward the ultimate goal of scientific knowledge. On my interpretation, then, we need to hold apart the purpose of the *method of explication* in the bigger picture and the purpose of an *explicator* in a specific project. In the latter case, the purpose is to achieve more exact and fruitful concepts which also fulfil the similarity criterion to a sufficient degree and are as simple as the other criteria permit. Whether these methodological steps will eventually contribute to the ultimate purpose, e.g., scientific knowledge, is an issue which may fall outside of the immediate control of the explicator. Hence, it should not be built into the criteria of adequacy.

4.3.1 The importance of exactness

In the discussed arguments for the primacy of fruitfulness the focus has been on explications in the developed sciences, where the standards of exactness are already high. In line with Carus (2007, 286), I argue that in philosophy and the less developed sciences the need to emphasize exactness is higher than in the developed sciences, because there is less exactness and less strong norms of exactness. Carus describes the contrast between the role of explication in the more developed and the less developed sciences as follows:

In the case of the more developed sciences [...] explications will often, perhaps nearly always, be *motivated* by theoretical interests. They will be driven less by

puzzlement about a concept in an evolved language than by the desire to extend a constructed language. (Carus 2007, 286)⁴⁶

And in contrast to the developed sciences, Carus describes the role of explication in the less developed sciences in the following way:

in less developed sciences, whose concepts are still largely embedded in informal usage and defined by it, explication will be driven more by the effort to make those evolved concepts more precise. There, no single system is accepted yet as the standard explicatum language [...] and explications are often highly controversial because any given decision concerns not just a particular explicatum [...] but fixes the entire language in which this and (potentially) *all future* explicata are to be embedded as a part. (Carus 2007, 286)

Regarding the role of exactness in explication, I defend the following three claims:

- (a) The method of explication should be broadened in order to be suitable for less developed sciences.
- (b) It is more important to emphasize exactness in less developed sciences.
- (c) The exactness criterion should be relaxed compared to Carnap's criterion.

At first sight (a)–(c) may appear to be in tension. For example, given (a) and (b) it may seem surprising or unintuitive to argue for (c). I will now clarify (b) and (c), hoping to show not only that they are not in tension with each other but that they are mutually supportive.

Let us begin with (b), the claim that it is more important to emphasize exactness in the less developed sciences. Since exactness is taken for granted in the more developed sciences, the need to emphasize exactness is larger in the less developed sciences. Perhaps the need is even larger because in some of the less developed sciences there are research paradigms that are suspicious of or even hostile to exactness and clarity. My reason for pressing this point is that exactness has been demoted in recent literature on explication, in which fruitfulness is emphasized at the expense of exactness (see chapter 4).

Now I will explain (c), the claim that Carnap's criterion of exactness should be relaxed. The reason to relax Carnap's criterion of exactness is that it is too demanding for some legitimate contexts of inquiry. In the less developed and

⁴⁶ In some cases, explications in developed sciences are motivated by new empirical discoveries that make a previously well-functioning concept into a troublesome concept.

exact sciences and fields, e.g., fields that study humans as cultural beings, the demand for a “well-connected system of concepts” may not be feasible or even desirable. Hence, the need for a relaxed criterion of exactness, where it is enough to specify the intended rules of use to some degree of exactness. Note that when I talk about a more relaxed criterion, I do not mean to minimize the importance of the criterion. By a relaxed criterion of exactness I mean a criterion with a wider range of ways to be fulfilled.

Therefore, to hold both (b) and (c) is in line with my aim to broaden explication so that it can be used to improve the concepts in less developed fields of inquiry than the ones Carnap had in mind. The exactness criterion should be relaxed in order to be more suitable for all contexts in which explications are, or should be, driven by the effort to make our evolved concepts more exact.

4.4 The internal structure of the criteria of adequacy

4.4.1 The internal structure of Carnap’s account

Once again, the only explicit remark from Carnap about internal structure among his four criteria is that simplicity is of secondary importance. As mentioned, however, in RSE he singles out exactness when he states that “[t]he only essential requirement is that the explicatum be more precise than the explicandum” (RSE, 936). Earlier, in LFP, where Carnap lists his official criteria, he merely says that the explicatum should be given rules of use in an exact form in order to “introduce the explicatum into a well-connected system of scientific concepts” (LFP, 7). He does not single out exactness from the other criteria, and he does not include the view, expressed in RSE, that the explicatum necessarily must be *more* exact than the explicandum (although he emphasizes that in his general characterizations of the method). As Brun points out, Carnap’s own example of the concept *piscis* may be considered a counterexample to that view (as I discussed in 2.6.1). However, when it comes to exactness in the sense of explicitness, in cases of pre-theoretical concepts the explicatum will naturally be more precise merely by the stipulation of rules of use.

If we take at face value Carnap’s comment that more exactness is essential (Brun advises against it, as we will see below) we get the view that exactness is a defining feature of explication while fruitfulness is what makes for a good explication. This is the structure that we can gather from Carnap’s explicit

remarks. Without the similarity criterion, however, we would not be dealing with an explication but with concept formation from scratch (if it were absent) or conceptual analysis (if it were stronger). Hence, I think that both similarity and exactness should be considered as defining features of Carnapian explication. The structure should then be:

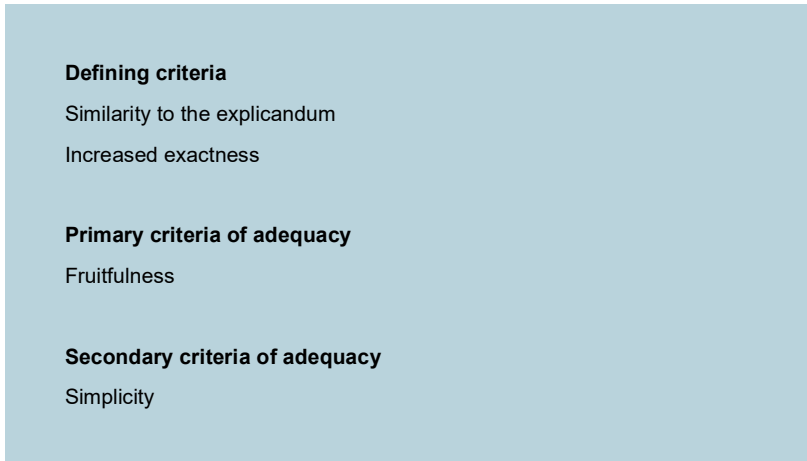


Fig. 1. The structure of Carnap's account on my interpretation.

Before I leave the interpretational part of this section, and proceed to a comparison of the way Brun proposes to modify the method with the way I propose to modify it, I will first discuss Brun's interpretation of Carnap, which differs from both Olsson's and Pinder's. As far as I know, Brun is the first to argue that the criteria should be divided into necessary and optional criteria, applying this division both to his interpretation of Carnap and in his own development of the method. Therefore, before I go on to present my own modification of the criteria and the internal structure of my account, I address both Brun's interpretation of the structure among Carnap's criteria, and his own modification of Carnap's account.

4.4.2 Brun's interpretation of the internal structure

Brun raises the issue of necessary versus non-necessary criteria in his discussion of Carnap's exactness criterion (Brun 2016, 1222–1224). Since I presented Brun's discussion of exactness in 2.7.2, I will merely summarize it quickly here. He identifies five distinct aspects of exactness in Carnap's account. First there is exactness in the sense of giving explicit rules for how to

use the explicatum in the target system of concepts, and in Brun's view, this is just a "necessary condition for giving an explication" (1222). As conditions on how the rules of use must be formulated, Brun introduces two "necessary conditions of adequacy": (1) that the rules are unambiguous and (2) that the rules do not lead to paradoxes or contradictions. Regarding the condition of unambiguity, Brun adds in a footnote, with reference to Hempel (1952), that unambiguity includes that the logical form of the explicatum needs to be determined. Hempel calls this the requirement of syntactical determinacy, considering it "an important but frequently neglected requirement, which applies to analytic definitions and explications as well as to nominal definitions". Brun includes this aspect in recipe for explication, and I consider it an important addition to Carnap's recommendations. Therefore, I will quote at length Hempel's discussion of the importance of specifying the logical form. The requirement of syntactical determinacy is, according to Hempel, that:

A definition has to indicate the syntactical status, or, briefly, the syntax, of the expression it explicates or defines; i.e., it has to make clear the logical form of the contexts in which the term is to be used. Thus, e.g., the word 'husband' can occur in contexts of two different forms, namely, 'x is a husband of y' and 'x is a husband'. In the first type of context [...] the word 'husband' is used as a relation term; it has to be supplemented by two expressions referring to individuals if it is to form a sentence. In the contexts of the second kind [...] the word is used a property term, requiring the supplementation by only one individual name to form a sentence. (Hempel 1952, 12–13)

As an example of a syntactically *indeterminate* term, Hempel discusses the term 'vital force':

The term 'vital force' is used so loosely that not even its syntax is shown; no clear indication is given of whether it is to represent a property or a scalar or a vectorial magnitude, etc.; nor whether it is to be assigned to organisms, to biological processes, or to yet something else. The term is therefore unsuited for the formulation of even a moderately precise hypothesis or theory; consequently, it cannot possess the explanatory power ascribed to it. (Hempel 1952, 14)

As an example of a case in which we can improve a concept by determining its syntax, Hempel points to the concept of probability:

The definitions given in older textbooks, which speak of "the probability of an event" and thus present probabilities as numerical characteristics of individual events, overlook or conceal the fact that probabilities are relative to, and change

with, some reference class (in the case of the statistical concept of probability) or some specific information (in the case of the logical concept of probability) and thus are numerical functions not of one but of two arguments. Disregard of this point is the source of various “paradoxes” of probability, in which “the same event” is shown to possess different probabilities, which actually result from a tacit shift in the reference system. (Hempel 1952, 14)

I agree with Brun that these considerations raised by Hempel are necessary at the stage of specifying the explicatum. One should at least specify what kind of property or relation the term expresses, and how many arguments it takes, and one should specify what kinds of entity the term is supposed to be assigned to. I try to adhere to Hempel’s requirements in my explication of *democratic* in chapter 10.

After these defining and necessary conditions, Brun comes to the fourth and the fifth aspect of exactness. The fourth is what Carnap lists as his official exactness criterion, interpreted by Brun as the requirement that the explicatum should not be more vague than the explicandum. The fifth aspect of exactness, for which Brun reserves the term “precision”, is the discriminating power of comparative and—especially—quantitative concepts.

Before I spell out the structure of Brun’s interpretation and modification of Carnap’s criteria, I need to address his rejection of Carnap’s claim that it is an essential requirement that an explicatum should be more precise than the explicandum. Despite Carnap’s explicit remark that increased exactness is essential, Brun rejects it, holding it to be incompatible with what Carnap writes elsewhere. Specifically, it seems to be in tension with what Carnap writes in his reply to Goodman in RSE and with what Carnap writes in LFP about the example of *piscis* (I have briefly addressed both issues in 2.6.1). Carnap seems to think that in some situations the similarity criterion requires synonymy, but that it is justified to make different requirements in different situations. Regarding the relation between explicatum and explicandum, Carnap remarks in RSE that “[a]lthough sometimes synonymy in the strong sense might be required, this does not seem necessary to me in general; in most cases logical equivalence is sufficient, and perhaps even this is not necessary” (RSE, 945). However, it is incompatible to hold that there may be situations where synonymy is required and also to hold that it is a necessary requirement to increase exactness. In defence of his case, Brun puts more emphasis on the *piscis* example:

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*. (Brun 2016, 1223)

Therefore Brun thinks that the only plausible version of Carnap's exactness criterion is the minimal requirement that "the explicatum must not be more vague than the explicandum" (Brun 2016, 1223). Even with such a humble version of the criterion, Brun stresses its importance:

This requirement of exactness is a non-trivial feature of explication. It distinguishes explication from other forms of concept change and conceptual re-engineering [...]. 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. (Brun 2016, 1223).

Brun does not spell out the structure (in my terminology) that he has in mind, but, as I see it, on his account we end up with the following. The defining condition is that the explicatum is given explicit rules of use. The necessary criteria of adequacy are unambiguity and consistency, and that the explicatum is at least as exact as the explicandum.

Further, the optional criteria of adequacy related to exactness are, according to Brun, exactness in the sense of reducing vagueness (which Carnap intended as a necessary criterion of exactness), and exactness (called by Brun "precision") in the sense of discriminating power (as exemplified in the way quantitative concepts allow us to make fine-grained distinctions between degrees).

Beyond exactness, Brun lists Carnap's criteria of fruitfulness and simplicity, but also expands the lists with further theoretical virtues that he takes to be implied by Carnap's comments:

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. (Brun 2016, 1224)

Regarding simplicity, Brun does not propose any changes to Carnap's view. He merely mentions that e.g., Goodman argues that "simplicity plays not a secondary but a pivotal role in theory assessment" (ibid.), without further comment about the relation between theory assessment and concept assessment.

The structure of Carnap's account according to Brun's interpretation is then the following:

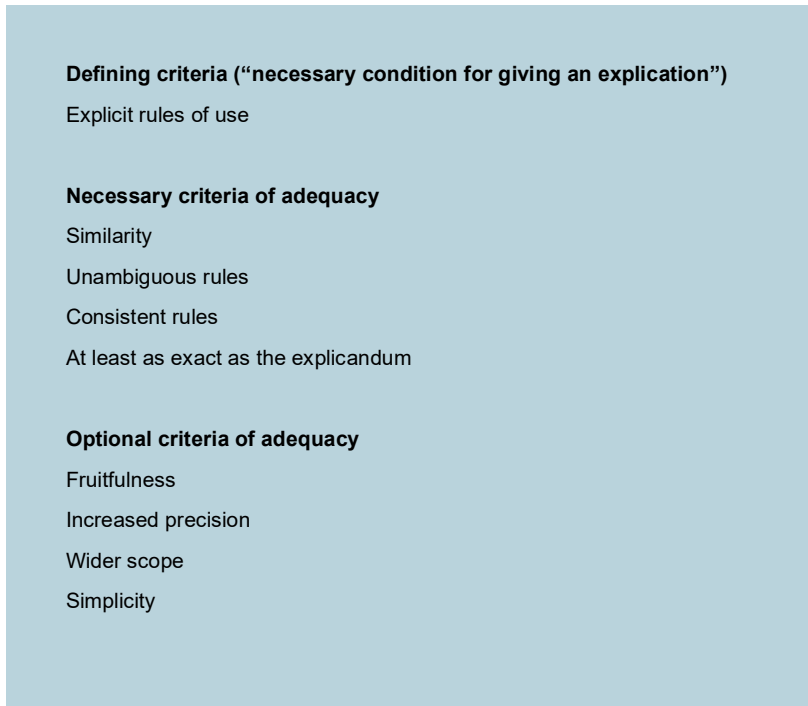


Fig. 2. The structure of Brun's interpretation of Carnap's account.

Commenting on Carnap's later view in RSE, Brun writes that “[Carnap] 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 explicatum must meet” (Brun 2016, 1225). Plausibly, similarity should be considered either a defining criterion or a necessary criterion of adequacy. The role of similarity in the structure is exceptional since it is a demand on the relation between the old and the new concept, while all the other criteria are demands on the features of the new concept (although that includes its relations to other concepts in the target system of concepts). I agree with Brun (and Carnap) that regarding similarity “[t]here is no general rule and we need to decide on a case-by-case basis what similarity requires” (Brun 2016, 1221). The difference in this regard between the similarity criterion and the other criteria is highlighted by Cordes and Siegwart (2018), who write that “[The other three requirements] are not specific to a given explication. In an

intuitive sense, exactness, fruitfulness, and simplicity are signs of quality for any introduction of a term” (Cordes and Siegwart 2018). Now, while these features are signs of quality for any concept, depending on the purpose at hand they may be interpreted and weighted in very different ways. That is how I view the criteria I propose in chapter 9.

4.4.3 The internal structure of Brun’s account

Beyond the criteria mentioned or implied by Carnap, Brun suggests that further criteria should be added in light of the discussions about theoretical virtues sparked by Kuhn (1977). Brun writes:

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. (Brun 2016, 1224)

By suggesting virtues such as ability to predict novel phenomena and visualizability, it is clear that Brun does not think of these as good features for any explicatum to have to some degree and in some balance with the others. Rather, he seems to suggest an account that is more in line with Kitcher’s and Pinder’s contextualism, where what the relevant criteria of adequacy are depends on the needs in the particular context in which an explication is pursued. Here, then, is Brun’s development of Carnap’s account:

Defining criteria

Explicit rules of use

Necessary criteria of adequacy

Similarity

Unambiguous rules

Consistent rules

At least as exact as the explicandum

Optional criteria of adequacy

Fruitfulness

Increased precision

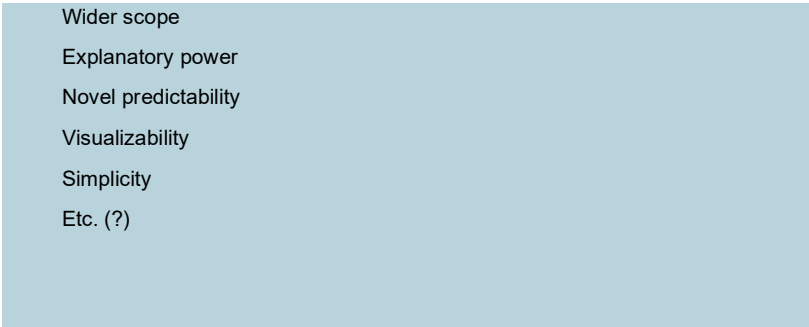


Fig. 3. The structure of Brun's development of Carnap's account.

I chapter 9 will propose my own modification of the criteria and my view of their internal structure. Afterward I will use the set-up from this section to compare the internal structure of my account with the respective internal structures of Carnap's and Brun's accounts. In short, I propose the following structure: similarity and explicitness are necessary for an explication, and for there to be any point in pursuing that explication, there should be an improvement with regard to a package of good-making features.

4.5 Concluding remark

In this section I have discussed different interpretations of what the purpose of Carnapian explication is. I introduced the distinction between immediate purpose and ultimate purpose to distinguish between the question of what a good explicatum should be like and the question of what the intended use of a good explicatum is. With that distinction to hand, I criticized extant interpretations of Carnap's views of the purpose of explication. I have related distinguished two different questions of purpose, namely the question of the internal structure of the criteria of adequacy and the question of how to prioritize between the criteria of adequacy. I have given an interpretation of the internal structure of Carnap's account and compared it to Brun's interpretation of the internal structure of Carnap's criteria as well as to the internal structure of Brun's own development of the method of explication. In light of these reflections, it is time to give an overview of different extant versions of explication.

5 An overview of different versions of explication

5.1 Overview

In this section I give a systematic overview of extant versions of explication, individuating the versions on the basis of their criteria of adequacy. Besides the aim of systematization, the intended contribution of this chapter is to show that some extant methods of concept formation that have not been conceived in terms of explication can plausibly and beneficially be framed as versions of explication.

I also discuss methods of explication in political theory, something which has not been discussed much in relation to Carnapian explication and conceptual engineering.⁴⁷ I attempt to bring together the literature on explication and conceptual engineering with the literature on concept formation and definitions in social science. There are insights worth carrying over in both directions. My hope is that the chapter should contribute to a better overview of the options on the menu of philosophical methods and methods of concept formation in general, across disciplines, and that it could contribute to better and more self-aware conceptual practices.

What all the versions of explication treated in this chapter have in common is that they demand some degree of similarity between explicandum and explicatum, and where they differ is with regard to the other criteria—those criteria that supposedly make the explicatum better for inquiry than the explicandum and hence justify the deviation from it. The broad collection of methods which I am here calling ‘explication’ could perhaps also be called method of ‘conceptual re-engineering for the purposes of systematic inquiry’.⁴⁸

⁴⁷ There are exceptions, such as Tengland 2008 and Cappelen 2023.

⁴⁸ I understand the distinction between conceptual engineering and conceptual re-engineering (the latter is Brun’s term, although he refer to Carus for the term ‘re-engineering’ (Brun 2016, 1233, n. 41), in the following way: re-engineering is the more narrow task of intentionally revising or replacing (with similarity) an old concept with a new concept that

For the sake of brevity, I simply use ‘explication’. The downside of my choice is a risk of terminological confusion since ‘explication’ is often used in the narrower sense of Carnapian explication (i.e., with weight given to all and only Carnap’s criteria of adequacy). Given my use of ‘explication’ in this chapter, Carnapian explication is one among several versions of explication, individuated from others by the priority given to exactness and fruitfulness.

5.2 Oppenheimian explication: explication for similarity and exactness

I begin the overview with the form of explication which is the most similar to conceptual analysis, but yet distinct from it. It is a form of explication that prioritizes two criteria: similarity to the explicandum and exactness. Schupbach (2017) calls such a method “Oppenheimian explication”, after the paper by Kemeny and Oppenheim (1952) in which they aim to give an explication of the intuitive concept *degree of factual support*. In the paper, Kemeny and Oppenheim adopt Carnap’s recently coined terminology of ‘explication’, ‘explicandum’, and ‘explicatum’, but they explicitly object to Carnap’s method, and when they specify their own methods they write that they “hope to set a standard for explication” (Kemeny and Oppenheim 1952, 308). Schupbach describes Oppenheimian explication as concept *clarification*, in contrast to Carnapian explication which he describes as concept *engineering*. He summarizes the differences in the following way:

John Kemeny and Paul Oppenheim [1952] proposed an alternative use for explication. In their particular project, Kemeny and Oppenheim focus on the prevailing “intuitive concept” of Factual Support, a concept which scientists commonly apply—as Kemeny and Oppenheim argue—when weighing the evidence for and against a theory. Their explicative goal is to “clarify” this concept by “making it precise.” As Kemeny and Oppenheim themselves emphasize, this constitutes an important departure from what we might call “Carnapian explication”—the use of explication for concept engineering. In “Oppenheimian explication,” we are not primarily interested in engineering fruitful new concepts to take the place of our explicanda. Rather, we place our

is useful for some purpose, while conceptual engineering includes at least also the activities of conceptual elimination and conceptual innovation (concept formation from scratch or de novo engineering), and possibly also the activity of evaluating and implementing concepts.

focus on a prevailing concept (the explicandum) and the overarching goal is to illuminate rather than to replace this concept. (Schupbach 2017, 678)

While Carnap prioritizes fruitfulness over similarity, Kemeny and Oppenheim prioritize similarity over fruitfulness. Schupbach continues:

In other words, Kemeny and Oppenheim recommend that we prize similarity over fruitfulness. This makes sense in Oppenheimian explication, where the goal is concept clarification rather than concept engineering. Any explication only has a chance of illuminating the nature, implications, etc. of some target concept (i.e., the explicandum) to the extent that the end result of the explication (i.e., the explicatum) retains a significant enough similarity to that concept. An explicatum may be as precise and fruitful as you like, but if this comes at the cost of disconnecting it from the target concept, then it will not likely teach us anything new about that concept. One might have successfully engineered an eminently fruitful new concept in this case, but that concept will have little to do with the explicandum. (Schupbach 2017, 679)

What, then, is the difference between Oppenheimian explication and conceptual analysis? Schupbach emphasizes that in Oppenheimian explication there are two concepts involved and the goal is to clarify the old concept *by means of* the new concept.

This is not to say that Oppenheimian explication just amounts to conceptual analysis. The distinction Carnap makes between explication and analysis [...] holds whether explication is used for concept engineering or concept clarification. While Oppenheimian explication resembles analysis insofar as it prizes similarity over and above any requirement of fruitfulness, it is unlike analysis in that it gives logicomathematical or scientific precision the same priority. *The overarching goal here is not to redescribe the explicandum, but to clarify it by means of a new concept that is both more precise than, and substantially similar to, the explicandum.* (Schupbach 2017, 679, n. 7, my emphasis)

Although similarity to the old concept is prioritized, together with exactness, a new concept is specified. Hence, we are not revealing the meaning of the old concept but clarifying it by the means of a new concept. As we are dealing with two concepts, the terminology of “explicandum” and “explicatum” is appropriate for this form of concept clarification or Oppenheimian explication, and it should be counted among the various methods of explication.

5.3 Explication for exactness

Brun argues that precisising definitions should be regarded as a form of explication instead of a form of definition. He writes:

The very idea of ‘precisising’ 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 ‘precisising’ 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. (Brun 2016, 1232)

Hence, we have another version of explication that balances similarity with exactness, just as Oppenheimian explication does, but in this case exactness is prioritized over similarity. Since Carnap only mentions exactness in his general characterizations of explication, many commentators have taken Carnapian explication to be merely the process of making concepts more exact. A previously mentioned example is Salmon (1989, 5). Another example is Frederick Suppe’s (2017) characterization of explication, where he distinguishes explications from definitions but describes them as precisising definitions:

Definitions of concepts are closely related to explication, where imprecise concepts (*explicanda*) are replaced by more precise ones (*explicata*). The *explicandum* and *explicatum* are never equivalent. In an adequate explication the *explicatum* will accommodate all clear-cut instances of the *explicandum* and exclude all clear-cut instances of noninstances. [...] In many scientific cases, definitions function more as explications than as meaning specifications or real definitions. (Suppe 2017, 77–78)

For many purposes the criterion of exactness suffices, e.g., when for judicial purposes the term ‘adult person’ is defined as a person who has attained the age of 18 years. For systematic inquiry, however, exactness is often not enough. The point is illustrated by Carnap’s example of quantitative concepts in psychology (see 2.5.1).

5.4 Explication for exactness and fruitfulness

Carnapian explication seeks to increase exactness and fruitfulness, and, secondarily, simplicity. Increases in exactness and fruitfulness motivate differences between the explicatum and the explicandum, but since some similarity is required, the method does not amount to mere concept innovation or concept formation from scratch. Since Carnapian explication has been thoroughly discussed in the previous chapters, I now merely direct the reader to chapters 1–4.

5.5 Operational explication: explication for empirical testability

Felix Oppenheim⁴⁹ proposes what he calls “reconstructions”, or “explications”, or “explicative definitions” of several political concepts.⁵⁰ For the method, he refers back to Hempel’s *Fundamentals of Concept Formation in Empirical Science*—in which Hempel, in turn, refers back to Carnap’s LFP. Oppenheim’s method is very much in the spirit of Carnapian explication, but he does not refer to or adopt Carnap’s list of criteria in LFP. He does, however, give his own list of four criteria of adequacy (or, as he also calls them, criteria of “validity”). In accordance with Hempel’s writings (except that Oppenheim occasionally uses the word “valid” as a synonym of “adequate”), Oppenheim writes that

Explicative definitions deal with terms already in use and try to make their meanings more precise and explicit. While not strictly true or false either, explicative definitions can be, in a broader sense, valid or invalid. (Oppenheim 1961, 6)

However, there is a difference between Hempel’s and Oppenheim’s accounts of what it is for an explication to be adequate or “valid”. Hempel sticks to Carnap’s requirement of similarity, exactness, and fruitfulness in the form of “the ability to function in hypotheses and theories with explanatory and

⁴⁹ Felix Oppenheim is the son of Paul Oppenheim, whose “Oppenheimian explication” was discussed in 5.3.

⁵⁰ For example, the concept of freedom (1961), and the concepts of power, equality, and interest (1981).

predictive force” (Hempel 1952, 12), whereas Oppenheim deviates from these criteria. In his view, “To be valid, such a definition [an explicative definition] must explicate the concept it defines, must be operational, must be fruitful, and must be valuationally neutral” (Oppenheim 1961, 6). The first requirement, “to explicate the concept it defines”, is another way to phrase Carnap’s similarity criterion. The rest of his criteria, however, deviate from Carnap’s formulations, given their explicit and heavy emphasis on operationalist strictures. The criteria are specifically formulated for the purpose of explicating political concepts in operational terms. Here are Oppenheim’s criteria stated in list form:

1. The proposed definitions will constitute adequate explications of the concepts to which they refer.
2. The proposed definitions will fulfil the requirements of operationalism, that is, they will be stated exclusively in empirical terms, so that statements in which the concepts thus defined occur will become empirically testable.
3. The proposed definitions will transform so vague a notion as freedom or even power into a fruitful empirical concept. The fruitfulness (and hence the validity) of scientific concepts depends not only on their serviceability for the formulation of new empirical laws, but also on the possibility of connecting them with each other in a significant way.
4. The last requirement of a scientifically valid definition—the absence of valuational connotation—is part of the canon of operationalism. To be operational, a definition must be formulated in terms used uniformly, with a high degree of agreement, by different scientists. (Oppenheim 1961, 7–8)

If it were not for the fact that the term ‘Oppenheimian explication’ is already used to refer to the account of explication proposed by his father Paul Oppenheim, together with Kemeny (see 5.3), I would propose that label for the version of explication proposed by Felix Oppenheim; but since that label is already occupied, I will call the latter Operational explication. I wish hereby to highlight the requirement that according to Felix Oppenheim a definition (of a political concept) should be operational. And as Oppenheim points out, it follows from the requirement of operationalism that definitions of political concepts are also required to be non-evaluative, which is the other criterion emphasized by Oppenheim which is not discussed explicitly by Carnap (although it is in line with Carnap’s views).

5.6 Explication for Gerring's eight criteria

I now treat John Gerring's account of concept formation (Gerring 1999, 2001). I think that his account is an important complement to Carnap's view, and I take inspiration from it when I propose modifications of the method of explication in chapter 9. Gerring does not refer to Carnap's method of explication or use Carnap's terminology. Nevertheless, I believe that his account can plausibly be regarded as a method of explication. His eight criteria are familiarity, resonance, parsimony, coherence, differentiation, depth, theoretical utility, and field utility. Gerring revises the list in his (2001) book on social science methodology. In the second edition published in 2011, which is "dramatically revised and expanded" (Gerring 2011, xx), Gerring revises the list again, to the following seven criteria: resonance, domain, consistency, fecundity, differentiation, causal utility, and operationalization (Gerring 2011, 117). I will stick to his 1999 account, since I find it most suitable for philosophy and for generalization beyond social science. For example, there he mentions operationalizability as one way in which a concept may be differentiated and bounded, but in the books he elevates operationalizability as criterion in itself. Further, the criterion of differentiation in the 1999 account is more analogous to the exactness criterion, and can naturally be applied in non-empirical inquiry. While Gerring thinks of these as criteria for concept formation rather than concept revision, he emphasizes in the first criterion, familiarity, that "There should be, in any case, a demonstrable fit between new and old meanings of a given term" (Gerring 1999, 368). Gerring also thinks of his criteria as gradable, claiming that "the criterion of *familiarity* must be understood, like other criteria, as a matter of degree" (ibid.), and in the spirit of explication he emphasizes that deviation from ordinary usage is beneficial but comes with a cost. Rejecting what he calls the ordinary-language approach to concept definition, he notes that "although ordinary usage may be an appropriate place to begin, it is not usually an appropriate place to end the task of concept formation."⁵¹ Yet he recognizes that "all departures from natural

⁵¹ It should be noted that not even an ordinary language philosopher such as Austin believes that ordinary use should be the last word. As Austin points out in "A Plea for Excuses": "Certainly ordinary language has no claim to be the last word, if there is such a thing. It embodies, indeed, something better than the metaphysics of the Stone Age, namely, as was said, the inherited experience and acumen of many generations of men. But then, that acumen has been concentrated primarily upon the practical business of life. If a distinction works well for practical purposes in ordinary life (no mean feat, for even ordinary life is full of hard cases), then there is sure to be something in it, it will not mark nothing: yet this is likely enough to be not the best way of arranging things if our interests are more extensive or intellectual than the ordinary. And again, that experience has been derived only from the

language impose costs, and should not be adopted lightly”.⁵² Gerring’s familiarity criterion is clearly analogous to Carnap’s similarity criterion. I think that it would be an improvement to explicitly introduce the “clarification of the explicandum” step into the process here, since it is often not the case that there is a single, uniform everyday use of a term.

Gerring also thinks of concept formation as “a set of *tradeoffs*—a tug of war among these criteria” (Gerring 1999, 367). Here are all eight, formulated by him as questions:

1. **familiarity**: How familiar is the concept (to a lay or academic audience)?
2. **resonance**: Does the chosen term ring (resonate)?
3. **parsimony**: How short is a) the term and b) its list of defining attributes (the intension)?
4. **coherence**: How internally consistent (logically related) are the instances and attributes?
5. **differentiation**: How differentiated are the instances and the attributes (from other most-similar concepts)? How bounded, how operationalizable, is the concept?
6. **depth**: How many accompanying properties are shared by the instances under definition?
7. **theoretical utility**: How useful is the concept within a wider field of inferences?
8. **field utility**: How useful is the concept within a field of related instances and attributes?

The criterion of parsimony is analogous to Carnap’s simplicity criterion, although Gerring does not mention as part of parsimony the simplicity of the laws connecting a concept to other concepts. The criteria of differentiation and theoretical utility together amount to something analogous to Carnap’s

sources available to ordinary men throughout most of civilised history: it has not been fed from the resources of the microscope and its successors. And it must be added too, that superstition and error and fantasy of all kinds do become incorporated in ordinary language and even sometimes stand up to the survival test (only, when they do, why should we not detect it?). *Certainly, then, ordinary language is not the last word: in principle it can everywhere be supplemented and improved upon and superseded. Only remember, it is the first word*” (Austin 1956, 11, my emphasis).

⁵² By “departure from natural language” he seems to mean departure from ordinary usage. He does not mean departure from natural language to formal languages.

exactness criterion (exact rules of use that introduce the concept in a well-connected system of concepts). The criteria of coherence, depth, and theoretical utility amount to something analogous to Carnap's fruitfulness criterion.

Gerring recognizes that concepts may be formed for a wide variety of purposes in the social sciences and that this is a "highly contextual process" (1999, 366). But, he argues, it does not follow "that we should toss up our hands and conclude that concept formation is a matter of 'context'" (ibid.). On the contrary, Gerring believes "that it *is* possible to arrive at a single account of concept formation within the social sciences that is at once comprehensive and reasonably concise" (1999, 367). While different criteria are emphasized and deemphasized in different tasks of concept formation, "the suppression of one or more demands does not go unnoticed by other social scientists" (ibid.). This point is important for the case I make later in the thesis, that it is valuable to retain general criteria of explication. In an account of explication with general criteria of adequacy, the explicator may emphasize or suppress different criteria depending on their goals, but the criteria themselves are not relative to particular goals (although they may be regarded as relative to the ultimate goals of systematic inquiry).

5.7 Explication for the Brülde–Tengland criteria of adequacy

Bengt Brülde and Per-Anders Tengland, both in collaboration and in separation, have used roughly the same set of criteria⁵³ (with some variations) for definitions of concepts such as *health* (Brülde 2000; Brülde and Tengland 2003; Tengland 2007), *illness* (Brülde and Tengland 2003), *mental disorder* (Brülde 2010), *empowerment* (Tengland 2008), *trust* (de Fine Licht and Brülde 2021).

In Tengland's definition of empowerment (2008), and in other later works, he refers to Carnap's LFP and describes his method as an explication, saying that he will be "giving an 'explication' of the concept, which means that one starts with the common use of a concept (using some of the criteria mentioned), but 'sharpens' it through stipulation (using some of the other criteria mentioned), in order for it to be useful for practical and for scientific work" (Tengland 2008, 79). Hence, Tengland explicates the concept of

⁵³ In some works they call them "conditions" or "desiderata".

empowerment, and his criteria may be regarded as criteria of adequacy for explication. The criteria given by TEngland for an adequate explication of the concept of empowerment are:

1. **The language criterion** “says that the definition produced should not differ too much from ordinary language” or from “how the concept is used in various professional contexts” (TEngland 2008, 78).
2. **The uniqueness criterion** is to maintain “the uniqueness of the concept in relation to other important concepts” (TEngland 2008, 79). This criterion is partly analogous to Gerring’s criterion of differentiation.
3. **The value criterion** “states that whatever value, positive or negative, that is attached to the concept should be reflected in the definition” (ibid.). This a possible way of formulating the similarity criterion (which TEngland and Brülde calls ‘the language criterion’), but I think that it depends on the purpose of the explication whether or not the value of an evaluative explicandum should be preserved (see 9.4.1).
4. **The homogeneity criterion** is that the “definition should preferably pick out some homogenous kind of characteristic” (ibid.). Sometimes this is called “the simplicity condition” (e.g., in Brülde 2010), and I will adopt that terminology later, but it should not be confused with criterion (7) below.
5. **The theory criterion** is that the “definition should also preferably be formulated as a theory or principle which helps us pick out that which belongs to the concept” (ibid.).
6. **The precision criterion** is that the definition “should be produced in well-known and well-defined terms” (ibid.).

7. **The simplicity criterion** is here understood as the criterion that “we allow as few exceptions, or additions, to the definition as possible (ibid.)”

8. **The goal criterion** is, for the concept of empowerment, that it “should be useful in describing how professionals in relevant areas (i.e., in those areas where the concept is used) work, and it should be in harmony, or at least compatible, with the normative goals for existing professional practices.” (ibid.).

There are some variations depending on the concept, but roughly the same eight criteria are used in the above-mentioned definitions by Brülde and Tengland. Although Brülde does not refer to his own definitions as explications, he puts forth roughly the same set of criteria (with small variations) for his definitions as the one Tengland used in the explication of ‘empowerment’ mentioned above. Hence, I will call these eight criteria the Brülde–Tengland criteria.

Besides the fact that Tengland himself considers these definitions in terms of explication, there are two further reasons why it seems reasonable to understand the Brülde–Tengland methodology in terms of explication. First, similarly to Gerring, their first criterion, the “language criterion” or “ordinary language criterion”, is analogous to Carnap’s similarity criterion. Although the list of criteria used by Brülde and Tengland varies slightly between works, they always include the language criterion. Second, for the task of defining health, Brülde introduces an alternative version of philosophical analysis, which is suited for social concepts, and is in the spirit of Carnap’s recommendation to move from classificatory concepts to comparative concepts.

In general their methodology is highly relevant for my project in this thesis since (a) their criteria are partly based on criteria for adequate operationalizations of social science concepts, and (b) they put forth a cluster of criteria which they take to be generally applicable for philosophical definitions. I return to the value criterion and the simplicity criterion in chapter 9.

5.8 Concluding remarks

I have given an overview of versions of explications. Later, in chapter 9, I develop my own account, or version, which is a relaxation of Carnap's account with additional criteria, partly inspired by accounts presented in this chapter. In contrast to extant proposals to relax Carnap's account, such as Pinder's (8.3.4), I think that we should strive for a unified method of explication for all contexts of inquiry. In the next chapter, I will discuss different conceptions of explication.

6 An overview of conceptions of explication

6.1 Overview

In a specific case of explication, two philosophers may agree on every step of the process but nevertheless have different views about what is achieved by the process. The aim of this chapter is to give an overview of extant *conceptions* of explication. Different conceptions are here individuated by (sufficiently important) differences in views about the epistemological and metaphysical *significance* of explication. In my terminology, two philosophers who have different views about what we achieve by giving explications have different conceptions of explication. Such views are not always explicitly stated. Hence, I use the phrase ‘conception of explication’ for an implicit or explicit view or attitude regarding the *significance of explication*.

For example, Carnap gives the theories of descriptions by Russell, Frege, and others as examples of explications. One might think that an explication of definite descriptions in everyday language in terms of existential quantification relieves us from ontological commitment to non-existent objects. Hence, on this conception, the explication is taken to involve the important achievement of slimming down one’s ontology. As we will see below, and as Gustafsson (2006) has argued, Quine conceives of explication as a method for eliminating ontological commitments (Gustafsson 2006, 58; Quine 1960). I begin with a section on what I take to be the main dividing view: the question whether there is a privileged language of inquiry.

6.2 The main divide: is there a privileged language?

While an explication is a matter of adequacy and not a matter of correctness, an important difference in view is whether or not one thinks that there is a correct language or privileged language for inquiry (in a stronger sense than a

most practical language). For Carnap, both the choice of language and the choice of explicata within a language are pragmatically decided. It is possible, however, to regard the choice of explicata to be pragmatic but the choice of language (in which to couch an explicatum) to be privileged. Hence there are two different questions about correctness. We have, first, the question of whether there is a correct explicatum. The negative answer to this question is what distinguishes explications from e.g., conceptual analyses and lexical definitions. But there is a second question, which arguably would have been answered positively by Frege, Russell and Quine and negatively by Carnap after 1932, namely, whether there is a uniquely correct or privileged language in which an explicatum should be constructed. It is notoriously difficult to interpret Frege on this issue⁵⁴ and I will leave him out of the following discussions, focusing on Carnap's and Quine's conceptions of explication.

6.3 Explication in a privileged language

6.3.1 Does explication require Carnapian tolerance?

I now consider the view that there is a privileged language for explications to be couched within. On such a view there may be many possible explicata for an explicandum, none more correct than the other, but only a single privileged language or logic to formulate these explicata in. On a plausible interpretation, this is the view held by Frege and Russell and by Carnap himself until 1932. Arguably, it is in another sense also the view of explication held by Quine.

Carus argues that Quine “conflated explication with a particular linguistic proposal”, with the consequence that Quine “returned, in effect, to where Carnap had been before 1931, the rational reconstruction phase of the Vienna Circle” (Carus 2007, 265). Contrary to Carus, I think it more terminologically plausible to consider Quine to be engaged in a form of explication, as he takes himself to be, but with a conception of it which differs from Carnap's (this is the terminology in Gustafsson 2006; 2014). Such a use of the term would be in line with Carnap's use. Carnap does not seem to think that explication requires the pragmatic view that he himself endorses after 1932, and which he later calls ‘language engineering’. For one thing, he mentions Frege's and Russell's analyses of natural numbers and Russell's analysis of definite descriptions as examples of explication. Carus, however, makes much of the

⁵⁴ As Reck has pointed out (2007, 38-39).

contrast between rational reconstruction, the method used before the Principle of Tolerance, and explication. Brun summarizes Carus' view:

In [Carus's] interpretation, rational reconstruction rests on the “hope that there could be a single, permanent logical framework for the whole of knowledge” [Carus 2007, 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 [Carus 2007, 263–265]. (Brun 2016, 1237, n. 47)

It was in order to mark this distinction that Creath coined the term ‘conceptual engineering’ (1990a) as a label for Carnap’s view of philosophy after 1932. Creath also makes much of the shift in Carnap’s philosophy represented by the introduction of the Principle of Tolerance, seemingly more so than Carnap himself did (as seen in, e.g., the preface to the second edition of the *Aufbau*). Creath claims that “[t]he development of Carnap's radical conventionalism and pragmatism really is a watershed and really does begin a new period in his philosophy” (Creath 1990b, 409).

As Brun has made clear, sometimes Carnap distinguishes between ‘rational reconstruction’ and sometimes he equates them:

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]). (Brun 2016, 1236)

Brun adds in a note to the quoted paragraph that:

In the foreword to the second edition of *The Logical Structure of the World* [*Aufbau*, 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 [*Aufbau*, § 100]. (Brun 2016, 1236, n. 47)

In the same note, Brun (2016, 1236, n. 47) also points out that Hempel claims that logical empiricists used ‘explication’, ‘logical analysis’, and ‘rational reconstruction’ in the same sense. This stands in contrast to Creath (2012), who distinguishes between ‘logical analysis’ and ‘explication’. According to Creath, logical analysis “is simply an attempt to uncover the logical structure of various claims and concepts under discussion. [...] Its aim is to reveal structure that is already there” (Creath 2012, 162).

Due to the important differences between Quine and the rest with regard to epistemological views, I will treat the Quinean conception in a separate section. This is in line with Peter Hylton's claim that "Quine's canonical notation is a rebirth of something like Russell's idea of a logically perfect language, although in the context of quite different epistemological views" (Hylton 2004, 105).

6.3.2 Carnapian language engineering: proposing linguistic conventions for the utility of empirical science

Carnap's principle of tolerance was first articulated in the protocol sentence debate in 1932 and given its full expression in 1934 in *Der Logische Syntax der Sprache* (The logical syntax of language). Richard Creath has convincingly argued that with the idea of tolerance a genuinely new period begins in his philosophy, a period that spans both his syntactical period and his semantical period and that lasted for the rest of his life. I direct the reader to chapter 2 for discussions of Carnap's views.

6.4 Quine's conception of explication: simplification of total theory

Each elimination of obscure constructions or notions that we manage to achieve, by paraphrase into more lucid elements, is a clarification of the conceptual scheme of science. The same motives that impel scientists to seek ever simpler and clearer theories adequate to the subject matter of their special sciences are motives for simplification and clarification of the broader framework shared by all the sciences. *Here the objective is called philosophical, because of the breadth of the framework concerned; but the motivation is the same. The quest of a simplest, clearest overall pattern of canonical notation is not to be distinguished from a quest of ultimate categories, a limning of the most general traits of reality.* (Quine 1960, 161, my emphasis)

From the quotation above we can see that Quine's view of explication differs substantially from Carnap's views, since Quine regards the clarification of scientific language as a part of the quest for the traits of reality. Quine's main discussion of explication in *Word and Object* (1960) is in section §53, with the title "The ordered pair as a philosophical paradigm" (257–262). In that section

Quine explicitly discusses the task of offering an “analysis” or “explication” of a “hitherto inadequately formulated ‘idea’ or expression” (1960, 258). This is how he characterizes explication, beginning with what is *not* involved in the task.

We do not claim synonymy. We do not claim to make clear and explicit what the users of the unclear expressions had unconsciously in mind all along. We do not expose hidden meanings, as the words ‘analysis’ and ‘explication’ would suggest; we supply lacks. We fix on the particular functions of the unclear expression that make it worth troubling about, and then devise a substitute, clear and couched in terms to our liking, that fills those functions. (Quine 1960, 258–9)

Quine thinks that his clarificatory comments about what explication is not, i.e., a method for exposing hidden meanings, are motivated since his methodological view has not been widely held before him: “[p]hilosophical analysis, explication, has not always been seen this way.” However, he adds in a footnote (Quine 1960, 259, n. 4):

By Carnap, yes; see *Meaning and Necessity*, pp. 7 f.

This is Quine’s only explicit reference to Carnap’s conception of explication in §53. Surprisingly, he makes no reference to Carnap’s main treatment of explication in *Logical Foundations of Probability* (LFP) despite his evident familiarity with it. In his correspondence with Carnap he had followed the development of the book, and in “Two Dogmas of Empiricism” (1951, 23, n. 4) Quine refers to Carnap’s treatment of analyticity in terms of state-descriptions in LFP (70ff).⁵⁵ More importantly, there is evidence that Quine was familiar also with Carnap’s treatment of explication in the first chapter of LFP. Quine did a referee report for the University of Chicago Press on Carnap’s manuscript, at the time titled “Probability and Induction”. In it, Quine mentions as an “incidental achievement” the discussion of explication: “Incidental achievements: Clarifies nature of philosophical explication in general” (Creath 1990, 400). Despite this, in *Word and Object* Quine does not discuss Carnap’s views of explication. Based on Quine’s remark in the footnote quoted above, that Carnap is a predecessor who sees explication in the same way as Quine does, we can assume that with regard to the method of explication Quine takes himself to be engaged in roughly the same method as Carnap, even though their conceptions of what is achieved by it differ.

⁵⁵ Jonas Raab (2024, 2047) directed my attention to this note.

6.4.1 Three kinds of defective nouns

In the sections that precede §53, Quine discusses defective nouns which do not call for explication, and in §53 and the sections that succeed it he discusses defective nouns that call for explication. Among the defective nouns that are not worthy of explication there are both those that don't serve a theoretical purpose and those that, more surprisingly, do serve a theoretical purpose. All in all, there are therefore three kinds of defective nouns: those with no purpose worth preserving, those with a purpose that can be preserved without explication, and those with a purpose which for their preservation requires explication. In Quine's view, explication is not called for when a purpose can be fulfilled without any object to be admitted in our universe of discourse. The need for explication arises only when expressions have purposes that do require a kind of object to be admitted in our universe of discourse.

Before discussing more theoretically interesting nouns, Quine begins, for instructive purposes, with 'sake' and 'behalf'. These illustrate well what is defective about defective nouns. What, according to Quine, is the reason for not bothering to explicate 'sakes' and 'behaves'?

It is that 'sake' and 'behalf' have their uses only in the clichés 'for the sake of' and 'on behalf of' and their variants; hence these clichés can be left unanalyzed as simple prepositions. (Quine 1960, 244)

These are the most flagrant examples of nouns that are defective in the sense that they sometimes occur as terms (i.e., as referring to objects) but do not do so in a systematic way. To understand the role of explication in Quine's philosophical project, it is worth lingering on these mundane examples of defective expressions. Earlier in chapter seven, in section §48, Quine makes the point in more detail:

Even a superficially termlike occurrence is no proof of termhood, failing a systematic interplay with the key idioms generally. Thus we habitually say 'for the sake of', with 'sake' seemingly in term position, and never thereby convict ourselves of positing any such objects as sakes, for we do not bring the rest of the apparatus to bear: we never use 'sake' as antecedent of 'it', nor do we predicate 'sake' of anything. 'Sake' figures in effect as an invariable fragment of a preposition 'for the sake of', or 'for 's sake'. (Quine 1960, 236).

These words do not cause any theoretical problems. Nouns that are defective in a similar manner but slightly trickier to deal with, are units of measure:

Units of measure turn out somewhat like sakes and behalves. ‘Mile’, ‘minute’, ‘degree Fahrenheit’, and the like resemble ‘sake’ and ‘behalf’ in being *defective* nouns: they are normally used only in a limited selection of the usual term positions. Their defectiveness, though less extreme than that of ‘sake’ and ‘behalf’, is easily exposed in absurd interrogation. Are miles alike? If so, how can they count as many? And if they cannot, what of the two hundred between Boston and New York? (Quine 1960, 244)

Finding no purpose served by making units of measure accessible to variables of quantification, Quine proposes to accommodate them as parts of relative terms, such as ‘length in miles’, temperature in degree Fahrenheit’, and in addition to his own examples, presumably, ‘duration in minutes’.

Thus instead of ‘length of Manhattan = 11 miles’ we would now say length-in-miles of Manhattan = 11’ (form ‘*F* of *b* = *a*’) or ‘11 is length-in-miles of Manhattan’ (form ‘*Fab*’). (Quine 1960, 245)

For two more philosophically charged and controversial examples of this phenomenon, Quine turns to unactualized possibles and facts. Just as with miles, there is perplexity of identity concerning unactualized possibles. Quine’s reluctance to accept objects that lack a clear identity criterion is summed up in “Speaking of Objects” in the slogan “No entity without identity” (Quine 1957, 20). This discussion also echoes Quine’s discussion of unactualized possibles in ‘On What There Is’ (1948), where he interrogates the identity criterion for “the possible fat man in that doorway” (1948, 23), leading to the conclusion that the concept of identity is inapplicable to unactualized possibles and that there is no sense in “talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another” (ibid., 23–4). In *Word and Object*, Quine’s example is “the possible new church on that corner”, and his solution is to let the modal operator govern whole sentences about possibilia such as possible churches, instead of governing just the objects.

A sentence about possible churches can usually be paraphrased satisfactorily enough into a sentence that treats of churches and is governed, as a whole, by a modal operator of possibility. (1960, 245)

Thereby, we absorb “the ‘possible’ of ‘possible object’ appropriately into the context and so no treating ‘possible object’ as a term” (ibid.). In order to get on to more challenging cases, I will skip Quine’s discussion of ‘fact’ and just let him conclude that

We were able to abjure sakes, measures, unactualized possibles, and facts without a pang, having satisfied ourselves that to admit them would serve no good purpose. (Quine 1960, 248)

We now have a clearer view of what Quine means by ‘defective noun’. Those hitherto discussed have according to Quine no theoretical purpose or function that needs preserving. But when there is a purpose to preserve, things get trickier.

Examples are not far to seek, on the other hand, of supposed objects that are absurd or troublesome and yet such that their banishment from the domain of values of our variables threatens to impair our apparatus. (Quine 1960, 248)

While our theoretical apparatus would be the worse for the rejection of the nouns in this second category, their purpose can be maintained without admission of any substitute object available to our variables of quantification. Quine’s examples of such objects are infinitesimals, ideal objects (mass points, frictionless surfaces, isolated systems), and geometrical objects (points, curves, surfaces, and solids). Beginning with infinitesimals, which were posited in differential calculus to deal with rates such as instantaneous velocities, Quine explains the problem thus:

What does it mean to say of a particle that at a momentary time t its velocity is ten feet a second? Not precisely that during some actual period of s seconds (say a hundredth of a second), spanning t , the particle traverses the appropriate distance of $10s$ feet (a tenth of a foot); for the velocity may change during that and every period. (Ibid.)

Quine goes on to explain how the problem is solved by calculus, invented by both Newton and Leibniz:

Newton and Leibniz answered, in their differential calculus, by positing infinitesimals: quantities infinitely close to zero and yet, absurdly enough, distinct from one another. [...] Though the idea of infinitesimals was absurd, the differential calculus, in which infinitesimals were reckoned as values of the variables, gave true and valuable results. (Quine 1960, 248)

To benefit from the success of differential calculus, it seemed, one had to pay the price of accepting absurd objects in one’s universe of discourse. However, that further problem was solved by the mathematician Karl Weierstrass (1815–1897):

The conflict was resolved by Weierstrass, who showed by his theory of limits how the sentences of the differential calculus could be systematically reconstrued so as to draw only on proper numbers as values of the variables, without impairing the utility of the calculus. (Quine 1960, 248)

Another example is that of ideal objects such as mass point, frictionless surface and isolated systems. They are useful devices in mechanics, but they are also contradicted by physics:

Just as infinitesimal numbers were contrary to arithmetic, so a point with mass, a surface without friction, or a system immune to outside forces would be contrary to physical theory. (1960, 249)

To avoid commitment to these entities, Quine suggests that one may paraphrase one's talk of them in the following way:

When one asserts that mass points behave thus and so, he can be understood as saying roughly this: that particles of given mass behave the more nearly thus and so the smaller their volumes. When one speaks of an isolated system of particles as behaving thus and so, he can be understood as saying that a system of particles behaves the more nearly thus and so the smaller the proportion of energy transferred from or to the outside world. (1960, 249)

In these cases we have, according to Quine, elimination but not explication: "Explication is elimination but not all elimination is explication" (Quine 1960, 261). According to Quine, these examples lack the "parallelism of function" required between the troublesome expressions and the replacements required to count as explications.

Showing how the useful purposes of some perplexing expression can be accomplished through new channels would seem to count as explication just in case the new channels parallel the old ones sufficiently for there to be a striking if partial parallelism of function between the old troublesome form of expression and some form of expression figuring in the new method. (Quine 1960, 261)

6.4.2 Quinean explication

Finally, then, we have the third kind of defective noun: those that call for explication. That is, defective nouns which for their purpose to be maintained require that we accept some substitute objects with a parallel function in our universe of discourse. Absurd or mysterious objects, referred to in some

occurrences by the troublesome nouns, are to be replaced by non-mysterious objects. Beside ordered pairs, treated in §53 (1960, 257–262), the examples of explication that Quine discusses are natural numbers and mental states, both treated in 54§ (1960, 262–266).

Furthermore, Quine briefly mentions, without discussion, three more examples of explication: singular descriptions, indicative conditionals, and quantifiers (1960, 260–261). As pointed out by Raab (2024, 2049), these are surprising examples since they are not nouns. Quine leaves this discrepancy with his previous emphasis on defective *nouns* uncommented. However, in the examples that he discusses in detail he follows the pattern of treating defective but useful nouns. Here is, in Quine’s words, the situation that calls for explication:

A pattern repeatedly illustrated in recent sections is that of the defective noun that proves undeserving of objects and is dismissed as an irreferential fragment of a few containing phrases. But sometimes a defective noun fares oppositely: its utility is found to turn on the admission of denoted objects as values of the variables of quantification. In such a case our job is to devise interpretations for it in the term positions where, in its defectiveness, it had not used to occur. (Quine 1960, 257)

Quine’s paradigm example of explication is the set-theoretical explication of ordered pairs, “a device for treating objects two at a time as if we were treating objects of some sort one at a time” (Quine 1960, 257). An ordered pair is, as Randall Dipert (1982) informally puts it, an entity which (a) contains two other entities, and (b) is such that one of these two entities is identifiable as the ‘first’, the other as the ‘second’ (Dipert 1982, 354). Ordered pairs are typically used for “assimilating relations to classes, by taking them as classes of ordered pairs” (Quine 1960, 257). As an example, Quine gives the father relation:

The father relation becomes the class of just those ordered pairs which, like $\langle \text{Abraham}, \text{Isaac} \rangle$, have a male and one of his offspring as their respective components. (Ibid.)

Quine emphasizes that ‘ordered pair’ is a good illustrative example since, as he puts it, “mathematicians pretty deliberately introduced it, subject in effect to the single postulate:

(1) If $\langle x, y \rangle = \langle z, w \rangle$, then $x = z$ and $y = w$. (1960, 258)

Dipert formulates the postulate for ordered pairs as a biconditional instead of a conditional,

$\langle A, B \rangle = \langle C, D \rangle$ if and only if

$$A = C \text{ and } B = D$$

He comments that “No other property of ordered pairs has ever been shown to be necessary or desirable” (Dipert 1982, 354). With such a precise expression of its only relevant feature, which seems to give as clear a criterion of identity as one could wish for, what is the problem with the noun ‘ordered pair’? There is a philosophical problem, namely that we do not know what they are. This according to Quine is not a problem. He quotes dismissively C. S. Peirce’s answer that an ordered pair is a “mental Diagram”. If we, like Peirce, try to give a philosophical account of ordered pairs we are missing Quine’s point, since he thinks that ‘ordered pair’ is a defective noun and needs to be dealt with as such. Quine writes that we should “face the fact that ‘ordered pair’ is (pending added conventions) a defective noun, not at home in all the questions and answers in which we are accustomed to imbed terms at their full-fledged best” (1960, 258). Postulate (1) given above offers no remedy to this:

Pending added conventions, the expressions of the form ‘ $\langle x, y \rangle$ ’ are, like ‘ordered pair’ itself, defective nouns, their normal occurrences being limited to special sorts of context where (1) can be exploited. (1960, 258)

Despite the precise criterion of identity given by (1), we are still in need of an explication of ‘ordered pair’, which means finding a kind of object that we are willing to admit in the place of ordered pairs. This is how the ordered pair differs from infinitesimals and ideal objects, in that “it is central to the purposes of the notion of ordered pair to admit ordered pairs as objects” (1960, 258). Objects are required to preserve the usefulness and purpose of ordered pairs. For example, “[i]f relations are to be assimilated to classes as classes of ordered pairs, ordered pairs must be available on a par with other objects as members of classes” (ibid.). In other mathematical uses of ordered pairs there are similar demands: “the very point of the ordered pair is its role of object—of a single object doing the work of two” (ibid.). This is why ordered pairs pose a different

challenge than infinitesimals or ideal objects. The role cannot be played without commitment to objects because it is central to their usefulness that they are objects available as members of sets. Or even stronger, this is their sole purpose: “A notion of ordered pair would fail of all purpose without ordered pairs as values of the variables of quantification” (ibid.)

For a solution to this problem we need an acceptable object for the noun to refer to in all occurrences.

The problem of suitably eking out the use of these defective nouns can be solved once and for all by systematically fixing upon some suitable already-recognized object, for each x and y , with which to identify $\langle x, y \rangle$. (1960, 258)

Such objects are given by the set-theoretical constructions of ordered pairs. In 1914, Norbert Wiener offered a definition formulated in the type-theoretic terminology of A. N. Whitehead’s and Bertrand Russell’s *Principia Mathematica*. Quine puts it in terms of classes in the following way, commenting that it is “nearly enough” how Wiener’s explication was originally put: “ $\langle x, y \rangle$ is identified with the class $\{\{x\}, \{y, \emptyset\}\}$ ” (Quine 1960, 258).⁵⁶ In 1921, Kazimierz Kuratowski offered the now established set-theoretical definition, sometimes called the ‘Wiener–Kuratowski definition’ (Dipert, 256), in which $\langle x, y \rangle$ is identified with $\{\{x\}, \{x, y\}\}$. Quine gives an example of yet another construal of ordered pair, given in terms of number theory, and he remarks that there is an infinite variety of ways to do so, where each version fulfils postulate (1). The multitude of versions is not a problem since there does not have to be a “unique right analysis” (1960, 260). All versions are right, as long as they fulfil (1), and all other features are unimportant. Quine sums up his discussion of the ordered pair in the form of a story:

On this and other points, the nature of explication as illustrated by the ordered pair may be made wholly evident by retelling the story of Wiener, Kuratowski, and the ordered pair in a modified terminology. In the beginning there was the notion of the ordered pair, defective and perplexing but serviceable. Then men found that whatever good had been accomplished by talking of an ordered pair

⁵⁶ Quine does not use ‘ \emptyset ’ as symbol for the empty set, but a symbol that has fallen out of use. Otherwise this is how Quine writes Wiener’s explication.

$\langle x, y \rangle$ could be accomplished by talking instead of the class $\{\{x\}, \{y, \emptyset\}\}$ —or, for that matter, of $\{\{x\}, \{x, y\}\}$. (Quine 1960, 260)⁵⁷

After this, Quine turns to the natural numbers. The set-theoretical construal of natural numbers is analogous to the construal of ordered pairs (or rather the other way around, for historical accuracy). For Quine, in this case also it is a question of finding acceptable objects to play the desired role.

Frege dealt with the one question, as Wiener did with the other, by showing how the work for which the objects in question might be wanted could be done by objects whose nature was presumed to be less in question. (1960, 262).

Finding adequate explications of natural numbers “means that natural numbers, in any distinctive sense, do not need to be reckoned into our universe in addition. [...] It is thus borne in on us, as in the case of ordered pairs, that explication is elimination”. This observation about elimination, that Quine repeats many times in §53 and §54, is also his segue to the mind–body problem, and specifically a physicalist explication of mental states, which is the last example of explication that he discusses.

That explication is elimination, and hence conversely that elimination can often be allowed the gentler air of explication, is an observation about a philosophical activity that far transcends the philosophy of mathematics, even if the best examples are there. Before we drop the topic, we may do well to note the bearing of that observation on the philosophical issue over mind and body. Let me lead up to the matter with a defense of physicalism. (Quine 1960, 264).

The difference between this example and the set-theoretical examples is the lack of available paraphrases. As Gustafsson (2006, 65) puts it, there is no Wiener and Kuratowski of neurology. But Quine does not regard this as a problem.

When Frege explains numbers as classes of classes, or eliminates them in favor of classes of classes, he paraphrases the standard contexts of numerical expressions into antecedently significant contexts of the corresponding expressions for classes [...] But when we explain mental states as bodily states, or eliminate them in favor of bodily states, in the easy fashion here envisaged,

⁵⁷ Once again, in the quote, for the empty set I use a different symbol than Quine uses in the original.

we do not paraphrase the standard contexts of the mental terms into independently explained contexts of physical terms. (Quine 1960, 266)

The language remains unchanged, but the relevant expressions “merely come to be thought of as taking physicalistic rather than mentalistic complements” (ibid.). Physical body states are less mysterious objects than mental states, and in this aspect there is a similarity with the previous cases. Without an available paraphrase, however, it is not as easily seen as in previous cases why this should earn the label of explication.

6.4.3 Differences between Carnapian and Quinean explication

An arguably superficial difference is that Quine talks about troublesome expressions, while Carnap regards the explicanda and explicata as concepts. Some philosophers have not regarded the issue as superficial (Hanna 1968; Maher 2010; Brun 2016). As observed by Jonas Raab (2024, 2049, n. 6), Carnapian explication is iterative while Quinean explication is non-iterative. For Carnap, explicata may be formulated in natural language or as comparative concepts. Such explicata may be further explicated by quantitative concepts or concepts couched in formal language. In Quinean explications, on the other hand, the explicandum is always defective and the explicatum (which he calls the explicans) is always couched in canonical notation. The only defectiveness that Quine seems concerned with in explicanda is inconsequential behaviour in relation to idioms of quantification. So, once a chunk of language is regimented in canonical notation⁵⁸ it cannot be defective, and so it cannot be an explicandum. Quinean explication is a non-iterative, binary affair while Carnapian explication is an iterative, non-binary affair where success is measured by satisfaction of the criteria to a sufficient degree.

6.5 Concluding remarks

I have discussed different conceptions of explication, with focus on Quine’s conception. I now turn, in the following chapter, to the broader methodological programme and practice of conceptual engineering.

⁵⁸ First-order predicate logic with identity and without individual constants.

7 The place of explication in the field of conceptual engineering

7.1 Overview

In the following chapter I locate the method of explication within the broader activity of conceptual engineering, in the sense in which the term ‘conceptual engineering’ is commonly used today which includes activities such as conceptual activism (to be explained below). Further, I compare the challenges of explication with the challenges of conceptual engineering in that broad sense. I argue that two of the most hotly debated issues in the literature on conceptual engineering are not particularly relevant for explication, in line with recent claims made by Eklund (2024). These issues are the *implementation challenge*, regarding how to secure uptake of a proposed concept, and the *continuity challenge*, or Strawsonian challenge, regarding how to secure topic continuity (see 3.2 for background). With some qualifications, to be explained below, neither implementation in the linguistic community nor topic continuity are required for an explication to be successful.

In the literature on conceptual engineering, Carnap’s method of explication is considered to be either a paradigm example of conceptual engineering (Cappelen and Plunkett 2020, 4–5; Chalmers 2020, 6), or a precursor to it (Scharp 2013, 4). Explication is commonly regarded as a form of conceptual engineering for *theoretical* purposes. Carnap’s writings on explication around 1950 are considered the first articulations and systematic expositions of a method of conceptual engineering, and therefore a pioneering contribution to the project of developing a theory of conceptual engineering, which is the task Cappelen undertakes in his influential *Fixing Language* (2018). Although conceptual engineering has been a prevalent practice both within and outside of philosophy (more or less so depending on how we construe it), it is only recently that a metaphilosophical field of research devoted to this practice has developed: the theory of conceptual engineering. In this field, challenges and methodological options regarding various projects of conceptual engineering

are articulated and debated (Cappelen 2018, 4). It would be expected that this field of research sheds light on Carnap's method of explication since it is considered a paradigm example of conceptual engineering. For example, Justus (2012, 175) comments that "[a]s a method of conceptual engineering, explication accords no privilege to the conceptual apparatus we inherit and which provides the fodder for traditional conceptual analysis." Furthermore, the term 'conceptual engineering' originated in Richard Creath's writings on Carnap's view of philosophy (1990a),⁵⁹ based on Carnap's own use of the engineering metaphor when describing his own work (see 7.2). Nevertheless, I will argue in this chapter that at least two of the central debates regarding conceptual engineering are mostly irrelevant for explication.

As the term 'conceptual engineering' is used in contemporary literature, it encompasses projects done for any purpose, not merely scientific or theoretical purposes. Conceptual engineering projects may be pursued for, e.g., ethical, social, or political purposes. Hence, 'conceptual engineering' has become an umbrella term for a wide collection of projects and activities.

Critique of the prospects for a general study of conceptual engineering has been raised before. Can such a broad and heterogeneous collection of methods and projects be an interesting subject of research? Eklund (2021) raises, without endorsement, the possibility to doubt that what is called conceptual engineering is unified enough to be an interesting subject of study (his examples of Sally Haslanger's and Kevin Scharp's respective projects will be discussed below):

Even if individual projects of conceptual engineering, such as Haslanger's and Scharp's, are of interest, one can think that these projects are dissimilar, in such a way that nothing is gained by treating these projects as parts of a bigger, unified project, conceptual engineering. The projects may be no more unified than the areas of philosophy of gender and race on the one hand and philosophy of logic, on the other, are unified. To be sure, in both cases, we talk about revising and replacing concepts, but the means of assessing the proposals are so different that there is no unified philosophical project here. (Eklund 2021, 20)

I do, however, think that something is gained by treating such projects as parts of a broader, unified project. Owing to the focus on foundational,

⁵⁹ Carnap himself used the engineering metaphor for his view of philosophy, e.g., in *Meaning and Necessity* (M&N, 43). He also described the task of constructing linguistic frameworks as "language engineering" (in his intellectual autobiography in the Schilpp volume published in 1963 (Carnap 1963a, 66)). See 7.6 for the history of the term 'conceptual engineering'.

metaphilosophical issues regarding these projects it has become clearer what differences there are between them and what the methodological options are. But by now the umbrella of conceptual engineering is entangled with many issues that are not very relevant for explication.

7.2 The history of the term ‘conceptual engineering’

In a survey by Isaac, Koch, and Nefdt (2022) of the conceptual engineering literature they note that the term ‘conceptual engineering’ was independently coined in Carnap scholarship and in metaphilosophy, referring to Simon Blackburn’s *Think* (1999), which is a popular introduction to philosophy. While the term ‘conceptual engineering’ might have been independently coined by Blackburn, before that the term ‘conceptual engineering’ was used by Creath to describe Carnap’s view of philosophy. The term ‘conceptual engineering’ was coined by Richard Creath (1990a) in his commentary on Carnap’s view of philosophy. In his introduction to *The Quine–Carnap Correspondence*, Creath claims that “[p]hilosophy, on [Carnap’s] model, becomes a kind of conceptual engineering [...]” (1990a, 7), and later in the same introduction he comments that after Carnap’s move to semantics Carnap “still held that philosophy is a form of conceptual engineering [...]” (1990a, 31). Although Carnap himself did not use the exact term ‘conceptual engineering’, he used the term ‘language engineering’ to describe the practical (as opposed to theoretical) problems involved in constructing and choosing a linguistic framework. He writes that “[w]hether or not [introducing a linguistic framework] is advisable for certain purposes is a practical question of language engineering, to be decided on the basis of convenience, fruitfulness, simplicity, and the like” (Carnap 1963a, 66). Before that, Carnap used the engineering metaphor in *Meaning and Necessity* (M&N):

I should prefer not to use the word ‘ontology’ for the recognition of entities by the admission of variables. This use seems to me to be at least misleading; it might be understood as implying that the decision to use certain kinds of variables must be based on ontological, metaphysical convictions. In my view, however, the choice of a certain language structure and, in particular, the decision to use certain types of variables is a practical decision like the choice of an instrument; it depends chiefly upon the purposes for which the instrument—here the language—is intended to be used and upon the properties of the instrument. *I admit that the choice of a language suitable for the purposes of physics and mathematics involves problems quite different from those*

involved in the choice of a suitable motor for a freight airplane; but in a sense, both are engineering problems, and I fail to see why metaphysics should enter into the first any more than into the second. (M&N, 43, my emphasis)

The connection between engineering and explication is heavily emphasized by Carus in his influential *Carnap and Twentieth-Century Thought: Explication as Enlightenment* (2007, 38). Carus also emphasizes the continuity between Carnap's later projects of explication and the earlier projects of 'rational reconstruction', and he makes the case that the programmes of explication and of rational reconstruction are continuations of the Enlightenment spirit and the 'engineering and revolutionary spirit' of the *École Polytechnique* in the first years of the nineteenth century (Carus 2007, 14).

[t]he central thrust of rational reconstruction was not to clarify the conceptual scheme embedded in, and reinforced by, natural language, but to decide what to replace that inherited scheme with. [...] Even at this early stage it was seen—appropriately to its origins in the *École Polytechnique*—as more of an *engineering* task. (Carus 2007, 16–17)

On Carus's rather speculative historical account, there is an engineering idea, or spirit, which originates in the Enlightenment and which blossoms at the *École Polytechnique* after the French Revolution, and which then is reawakened in a radical form in the programme of rational reconstruction in early logical empiricism, and turned into the programme of explication in late logical empiricism. Regardless of the putative connection to the Enlightenment, it is clear that the engineering attitude is central to Carnap's philosophical outlook. Richardson, for example, is critical of Carus's account of Carnap as an intellectual descendant of the engineers at *École Polytechnique*,⁶⁰ but he nevertheless endorses a (different) view of Carnap as a "philosophical engineer" (Richardson 2013, 71–72). Reck observed in 2013 that "[v]ery recently, the notion of 'conceptual engineering' has become central for interpreting Carnap" (2013b, 15–16). This, then, is the origin of the

⁶⁰ Richardson writes that "Carus [...] provides little evidence that Carnap was aware of, or inspired by, the group of French engineers that Carus himself evokes in discussing Carnap's views" (Richardson 2013, 60) and "I do not think that Carus's Enlightenment engineering project is Carnap's project" (Richardson 2013, 71). To be fair, though., Carus recognizes that what he calls the Carnapian ideal of explication "is not straightforwardly available from the published (or indeed, unpublished) writing" and that while Carus claims that "[i]t was something toward which Carnap approached, in his later year" he acknowledges that it "never quite crystallised, probably not even in his own mind" (Carus 2007, 38).

engineering terminology in Carnap's work and in interpretations of Carnap; there is, however, another history of the term which runs somewhat in parallel.

Scharp and Eklund appear to have played central roles in the popularization of the term outside of Carnap scholarship. In his book *Replacing Truth* (2013), Scharp describes his method as "conceptual engineering". While Scharp does mention Carnap's method of explication as a precursor to conceptual engineering, he does not, however, trace the term 'conceptual engineering' back to Creath (1990a) or to Carnap's engineering terminology; instead he traces the term back to Blackburn (1999) and Robert Brandom (2001), both of whom appear to have coined the term independently of Creath and of each other. When Cappelen in his 2018 mentions sources for the term, he only mentions Blackburn's 1999 and Eklund's 2014 and 2015.⁶¹ Cappelen and Plunkett (2020, 2, n. 3) refers to Blackburn and Brandom. To credit Blackburn (or Brandom) for coining the term is a common mistake in the literature, but it has been pointed out and corrected by David Chalmers (2020, 6).⁶²

The credit given to Blackburn and Brandom is unfortunate not only because Creath was the first to coin the term but also because neither Blackburn nor Brandom use the term for the kinds of projects that are now most associated with it. Blackburn uses it in the introduction to his popular book *Think* as a more hands-on sounding term for philosophy, writing that the philosopher "studies the structure of thought" just as the engineer "studies the structure of material things" and in both cases this yields knowledge of "what would happen for better or worse if changes were made" (1999, 2). This does not include the construction of new concepts or languages in the way advocated and pursued by, e.g., Carnap and Scharp. Brandom's use of the term is also not obviously precursory to the prevalent contemporary use. Instead, Brandom refers to what Isaac, Koch, and Nefdt (2022, 10, n. 2) describe as a "naturalized

⁶¹ Eklund (2014) himself refers to Blackburn 1999 as the source for the term, although he refers to Chalmers 2011 as the inspiration for doing philosophy as conceptual engineering: "David Chalmers [2011] forcefully stresses what seems to me to be a good and important point concerning philosophical methodology: while philosophers often have been concerned with our actual concepts or the properties or relations they stand for, philosophers should also be asking themselves whether these really are the best tools for understanding the relevant aspects of reality, and in many cases consider what preferable replacements might be. Philosophers should be engaged in conceptual engineering" (Eklund 2014, 293). In a footnote Eklund adds that "Chalmers does not himself use the locution 'conceptual engineering.' I take that locution from Blackburn [1999]" (Eklund 2014, 312, n. 1).

⁶² Chalmers writes that "Incidentally, in the recent literature on conceptual engineering, the credit for the term 'conceptual engineering' is often given to Simon Blackburn (1999). The credit should really go to Richard Creath (1990)" (Chalmers 2020, 6).

version of conceptual analysis”, with paradigm examples in the works of Fodor, Dretske and Millikan. In Brandom’s own words:

The enterprise in which they are jointly engaged is not so much one of conceptual analysis as it has been traditionally understood as one of conceptual engineering. That is, instead of thinking about what ordinary people or even sophisticated philosophers already mean by terms such as ‘representation’, they appeal to the tools of the special sciences (for instance, information theory and evolutionary biology) to describe abstractly, but in criticizable detail, how one might craft a situation in which some state arguably deserves to be characterized as ‘representationally contentful’ in various important senses. (Brandom 2001, 587)

However, in the sentences that follow he describes an aspect of the method which arguably makes it into more than just naturalized conceptual analysis—and rather into a method in the same spirit as the projects discussed under the contemporary umbrella of conceptual engineering:

Insofar as the theories are good ones, they may shed light on how human knowers actually work. But their immediate aim is a broader one: to say what would count as doing the trick, rather than how we manage to do it. (Brandom 2001, 587)

While the project described by Brandom shares the spirit of conceptual engineering in the contemporary sense, his use is still not the obvious precursor to the use of the term in the recent philosophical theories of conceptual engineering. Hence, Creath was the first to use the term ‘conceptual engineering’ in the sense that is relevant for the field of conceptual engineering and Creath’s use describes Carnap’s methodological outlook and attitude. Despite the Carnapian origin of the term, I think that the field of conceptual engineering is preoccupied with many debates that are mostly irrelevant for explication. In sections 7.6 and 7.7 I make that case. Now I move on to introduce the field of conceptual engineering.

7.3 Conceptual engineering, conceptual ethics and ameliorative analysis

In Cappelen’s characterization of conceptual engineering, it is concerned with the *assessment and improvement of our concepts (or representational devices)*

(Cappelen 2018, 1). By now this is probably the standard characterization, used frequently in the introductions of papers on conceptual engineering.

A compatible characterization is given by Scharp, who takes conceptual engineering to be the task of “actively changing some aspect of our concepts—eliminating bad ones, deciding which ones we should use, and which words should express them” (Scharp 2020, 396–397). A third way of using the term, proposed by Chalmers, is this:

I’ll suggest that conceptual engineering should be understood as the projects of designing, evaluating, and implementing concepts. [...] I’m inclined to think that each of the three taken alone is at least a distinctive and important mode of conceptual engineering, and that conceptual engineering is most powerful when all three modes are combined. (Chalmers 2020, 4)

In contrast to Cappelen and Scharp, Chalmers does not include the negative part of identifying defects and eliminating bad concepts and he includes the further activity of implementation. In the literature on conceptual engineering, this activity is called conceptual activism.

Further, there are two activities which go by other names but nevertheless can plausibly be included under the label ‘conceptual engineering’: conceptual ethics and ameliorative analysis. Conceptual ethics is a philosophical programme or method which is close to conceptual engineering or a part of conceptual engineering. In a pair of papers published in 2013, Burgess and Plunkett introduce the term:

Which concepts should we use to think and talk about the world and to do all of the other things that mental and linguistic representation facilitates? This is the guiding question of the field that we call ‘conceptual ethics’. (Burgess and Plunkett 2013, 1091)

While the term is new, they claim that the case can be made that “the field is already quite active, with contributions coming in from areas as diverse as fundamental metaphysics and social/political philosophy” (ibid.)

In a paper published in 2000, Haslanger proposed ameliorative analyses of gender concepts (*man*, *woman*) and race concepts. She describes the method in the following way:

we begin by considering more fully the pragmatics of our talk employing the terms in question. What is the point of having these concepts? What cognitive or practical task do they (or should they) enable us to accomplish? Are they effective tools to accomplish our (legitimate) purposes; if not, what concepts would serve these purposes better? (Haslanger 2000, 33)

She exemplifies this with the concept of knowledge, noting that the question “What is knowledge?” may be construed in various ways:

One might be asking: What is *our* concept of knowledge? (looking to a priori methods for an answer). On a more naturalistic reading, one might be asking: What (natural) kind (if any) does our epistemic vocabulary track? Or one might be undertaking a more revisionary project: What is the point of having a concept of knowledge? What concept (if any) would do that work best? (Haslanger 2000, 32)

While the activity of conceptual engineering is old and prevalent (again, depending on how we construe it) both within and outside of philosophy, the theory of conceptual engineering is a new field. As noted by Scharp, “[a]lthough there are plenty of instances of conceptual engineering in the history of philosophy, it hasn’t really been a focus of attention” (Scharp 2020, 397).

In the next section I go on to discuss and compare two differently motivated projects of conceptual engineering. First, it should be noted that there is a broader common motivation for the programme of conceptual engineering as a whole, described here by Eklund:

The concepts we have are the ones we have ended up with because of various biological and cultural factors. By some measure they have proven themselves, since we keep using them. But still, why should the concepts we actually have be the best conceptual tools for describing and theorizing about the relevant aspects of reality? Maybe philosophy should rather be concerned with *conceptual engineering*: it should study what concept best plays the theoretical role of our concept of truth and what features this concept has, what concept best plays the theoretical role of our concept of knowledge and what features this concept has, etc. (Eklund 2017, 192)

The general motivation for conceptual engineering, then, is to make our concepts better for our purposes. I will now compare two projects of conceptual engineering, one which is motivated by theoretical purposes and one which is motivated by political purposes.

7.4 Two paradigmatic examples of conceptual engineering

The field of conceptual engineering is often introduced with one epistemically motivated example and one politically motivated example, and often with the following two examples: Scharp's (2013) proposal to replace the concept of *truth*, for theoretical purposes, and Haslanger's (2000) proposal to replace the concept of *woman* (and *man*), for political purposes.⁶³ In his introduction to conceptual engineering in philosophy, Eklund chooses these two examples partly because of the stark difference between them; juxtaposed they give a sense of the width of conceptual engineering (Eklund 2021, 16).

Scharp argues that our ordinary concept of *truth* is inconsistent. He thinks that the alethic paradoxes are not the results of faulty reasoning but the result of the inconsistent constitutive principles of the concept. To avoid the paradoxes, and to facilitate natural language semantics (these are the aims), Scharp proposes to replace truth by two other concepts: *ascending truth* and *descending truth*. These new concepts have the following constitutive principles (here in the simplified form given in Isaac, Koch, and Nefdt 2022, 7):

1. From $\langle p \rangle$ infer $\langle \langle p \rangle$ is ascending true \rangle .
2. From $\langle \langle p \rangle$ is ascending true \rangle and $\langle \langle p \rangle$ is safe \rangle infer $\langle p \rangle$.
3. From $\langle \langle p \rangle$ is descending true \rangle infer $\langle p \rangle$.
4. From $\langle p \rangle$ and $\langle \langle p \rangle$ is safe \rangle infer $\langle \langle p \rangle$ is descending true \rangle .

The use of *ascending truth* and *descending truth* instead of *truth* prevents the liar paradox and other alethic paradoxes from arising. As Isaac, Koch, and Nefdt explains:

For our purposes, it matters only that using ASCENDING TRUTH and DESCENDING TRUTH instead of TRUTH makes it impossible to derive the liar paradox and other logical paradoxes related to truth. In other words, instead of solving the paradoxes directly, for instance, by denying premises from which they follow, Scharp provides an axiomatization of truth-like replacement

⁶³ Or possibly a project motivated by the theoretical purposes of so-called critical theory and social constructionism – but such purposes are entangled with political purposes.

concepts that prevent the derivation of the liar paradox. (Isaac, Koch, and Nefdt 2022, 8)

Over to Haslanger's project. Haslanger proposes to redefine the concept *woman* in the following way:

S is a *woman* iff_{df} S is systematically subordinated along some dimension (economic, political, legal, social, etc.), and S is "marked" as a target for this treatment by observed or imagined bodily features presumed to be evidence of a female's biological role in reproduction. (Haslanger 2000, 230).

The motivation of the definition is to raise to salience the view that gender is socially constructed and linked to oppression. By making it salient this would presumably help in overcoming the oppression.⁶⁴ The examples of Scharp's project and Haslanger's project illustrate the difference between conceptual engineering for epistemic aims and for non-epistemic aims.⁶⁵ For the latter but not for the former some change in the linguistic behaviour in the general population is required to meet the aims.⁶⁶

7.5 Motivations for conceptual engineering

Eklund (2021, 17–18) distinguishes between two ways in which a concept may be better than or preferable to another, by being free from a defect possessed by the other and by serving some purpose better than another. First, there may be defects in the first concept. In that case, a non-defective concept is preferable (assuming sufficient similarity otherwise). Examples given by Eklund are inconsistent concepts, indeterminate concepts, and concepts for which the use of them presupposes a false view Eklund gives the following three examples of defects in concepts: inconsistency, indeterminacy and the presupposition (by the use of the concept) of a false view (ibid.). We also have vagueness, which is the defect emphasized by Carnap. I suggest that

⁶⁴ However, in her 2006, Haslanger instead views the definition as a revelation of the meaning of 'woman', in line with semantic externalism.

⁶⁵ Scharp himself writes: "Haslanger's amelioration project is obviously a conceptual engineering project, and there are clear similarities between her project and mine" (Scharp 2020, 400). I will, however, focus on the important differences.

⁶⁶ Once again, as Haslanger understands the project in her later writings, this may not be required.

presupposition of contested views and presupposition of untestable views could be added to the list of epistemic defects. Empty terms are defective in a way. However, in many cases, there is no need to replace empty terms with anything, as in the case of ‘phlogiston’ mentioned above. However, there are also theoretically useful empty terms, such as ‘frictionless plane’ and ‘point particle’. Conceptual engineering is needed only when there is a need for something to take the place of the old concepts. Hence, when it comes to reasons for adopting and forming new concepts, there are two important judgements to be made: first a judgement that there is a useful role to be played by a concept similar to the (at least partially) abandoned one; second, a judgement about what kind of concept would be suited for that role.

Second, a concept may serve a specific purpose better than another, even if there are no outright defects in the first concepts. Then, the second concept is preferable given some specific purpose. Eklund does not mention the distinction between epistemic and non-epistemic purposes, but he gives the following examples of both. A concept may be non-epistemically better than another because it is “more likely to generate positive social change” (Eklund 2021, 18). A concept may be epistemically better than another concept because “it better tracks what is explanatorily powerful or inductively useful (ibid.)”

7.6 The implementation challenge and conceptual activism

Many of the debates and issues given most attention in the field of conceptual engineering are mainly relevant (if relevant at all) as regards certain projects that are pursued for social and political aims. To make my point I will focus on two of the most debated problems, namely the so-called implementation challenge and the continuity challenge. I begin with the former.

The implementation challenge has given rise to various debates that stand at the forefront of the literature on conceptual engineering. In a recent paper on the implementation challenge by Matthieu Queloz and Friedemann Bieber (2021), the authors introduce the subject of conceptual engineering in a way that I may use as an illustration of the problem that I am trying to highlight in this chapter, i.e., that the field in which conceptual engineering is studied and debated is unsuitable for methodological discussions regarding explication. Queloz and Bieber write that “[c]onceptual engineering, with its ambition not merely to analyze, but to assess and alter people’s concepts, is intimately tied

up with issues in political philosophy” (Queloz and Bieber 2021, 670). The problem is that they identify the target as “people’s concepts”.

While some projects that are now called conceptual engineering have the ambition of altering people’s concepts, or have the alteration of people’s concepts as a success condition, this is certainly not the ambition of paradigmatic projects of explication or of Scharp’s replacement of the concept of truth.⁶⁷ Ascribing such an ambition to conceptual engineering projects in general is either descriptively false or a prescriptive proposal to use the term ‘conceptual engineering’ in a way that excludes explication and other forms of epistemically motivated conceptual engineering. With that said, I now present the implementation challenge, i.e., the challenge of how to secure uptake of a proposed concept. To do so, I turn to Cappelen and Plunkett’s overview of the steps in the process of conceptual engineering.

In Cappelen and Plunkett’s (2020) overview, the first step is to detect a conceptual deficiency. Once a defective concept is detected, there are four options (or ‘ameliorative strategies’): (i) do nothing about it, (ii) abandon it, (iii) improve it, and (iv) replace it. The options (ii)–(iv) raise a further option, namely to engage in what they call ‘conceptual activism’ (2020, 4):

Once you have settled an ameliorative strategy, you might want to do some work to implement it, that is, you might want to engage in a bit of activism on behalf of your ameliorative strategy. If that’s something you want to do, it raises an ‘implementation challenge’: how are ameliorative strategies best implemented? (Cappelen and Plunkett 2020, 3)

It is one thing to figure out what the meaning of a term should be, and another thing to implement a change in meaning. According to semantic externalism, which Cappelen endorses, “[t]he process of conceptual engineering is governed by factors that are not within our control: no individual or group has a significant degree of control over how meaning change happens” (Cappelen 2018, 72). In some cases of conceptual engineering, the aims require that the old ordinary language term is preserved with a different meaning. Arguably, this is the case for the conceptual engineering of some socially and legally important concepts such as *marriage* and *family*. Since Cappelen thinks that changes in meaning happen through “inscrutable external factors that we lack control over”, he does not believe that conceptual engineering is a feasible project. However, Cappelen thinks that this should not stop us from pursuing it, just as we generally should not stop making normative judgements merely

⁶⁷ Hence, such projects are *not* intimately tied up with issues in political philosophy.

because we don't know how to change the world. Similarly, we should not stop making proposals about what words should mean just because we do not know how to change meanings. Cappelen illustrates the point with Haslanger's proposal:

suppose I'm right and Haslanger has little or no control over the meaning of the English word 'woman'. That doesn't mean that her normative proposal for what 'woman' ought to mean is wrong or in some way misguided. She could be right about what it ought to mean, despite the fact that there is no algorithm for how to implement that change. (Cappelen 2018, 75)

On Cappelen's view, conceptual engineering is not more futile than any other normative enterprise, but the stage of implementation or conceptual activism is futile. Many debates about implementation have unfolded. Koch has argued that conceptual engineering would be futile without control over meaning change, but he argues that some control is possible:

For example, in order to make 'woman' apply to the social kind suggested by Haslanger, we have to make it the case that people associate descriptions with the term 'woman' that mark it as a social kind of the envisaged sort. Doing this may not be an easy task, of course, but it is something about which we, as a linguistic community, possess collective long-range control. (Koch 2021, 345)

In the paper mentioned in the beginning of this section, Queloz and Bieber (2021) question the very framing of the implementation challenge. They point out that it would be a democratic problem if the implementation of concepts were within the control of conceptual engineers:

Whether or not we have control over concepts themselves, however, an underlying assumption shared by many different views—which is arguably implicit already in the framing of conceptual engineering as facing an implementation 'challenge'—is that lack of control over conceptual uptake, that is, over which concepts people in fact come to use, is a regrettable shortcoming. (Queloz and Bieber 2021, 671)

This is just a small sample of the responses and debates regarding the implementation challenge.⁶⁸ However, the implementation of a new concept is not important if the success only depends on epistemic gains, such as the ability to make new or more exact predictions, or the ability to connect the concept to already well-confirmed generalizations. Scharp is also explicit that his

⁶⁸ Some further contributions: Jorem 2021; Pinder 2021; Deutsch 2020.

replacement of truth is not intended as a full replacement. It is only intended to be replaced in some theoretical contexts:

I want it to be clear, right from the start, that I do not advocate eliminating truth from our conceptual repertoire. I am not trying to persuade people to stop using the word ‘true’. There is no need for flyers or public service announcements. For most purposes, the risk posed by our concept of truth is negligible; so it is reasonable to use truth, despite its defect, in most situations. (Scharp 2013, 2)

Something analogous could be said of most paradigmatic examples of explication. One might accept that there is no need to alter people’s concepts for epistemic projects to succeed but nevertheless maintain that there is a need to alter the concepts of one’s peers. Hence, it may be argued, the difference is merely in scope: some projects require broad implementation and some projects only require narrow implementation. Nevertheless, the challenges involved are different.

Eklund has recently made a similar point, by dividing conceptual engineers into the *activists*, concerned with changing which concepts we employ, and the *theoreticians*, concerned with changing which concepts we focus on as objects of study (Eklund 2024, 4). He observes that “[t]he implementation challenge may [...] be more of an issue for the activists” (ibid.).

7.7 The continuity challenge

The challenge initially posed by Strawson against Carnap’s method of explication (3.2) has got a life of its own in the field of conceptual engineering. Cappelen thinks of it as a serious challenge to conceptual engineering as he understands the method. In *Fixing Language*, he devotes three whole chapters to the challenge, claiming that a response to it is “a central task for any theory of conceptual engineering” (2018, 99, n. 3).

The challenge discussed by Cappelen is, however, misleadingly called the Strawsonian worry because it is assumed to be based on Strawson’s objection to Carnap’s method of explication. In the version discussed by Cappelen and after him in the theory of conceptual engineering, it was first raised by Haslanger (2000):

In asking what *race* is, or what *gender* is, our initial questions are expressed in everyday vocabularies of race and gender, so how can we meaningfully answer these questions without owing obedience to the everyday concepts? [...] Given

the difficulty of determining what “our” concept is, it isn’t entirely clear when a project crosses over from being explicative to revisionary, *or when it is no longer even revisionary but simply changes the subject*. (Haslanger 2000, 34)

Cappelen lumps Strawson’s objection together with Haslanger’s worry, and calls it the Strawsonian worry. It has also misleadingly become known as Strawson’s challenge. We saw above that he thinks that it is a serious challenge to conceptual engineering. As Koch (2021) has shown, the challenges articulated by Strawson, on the one hand, and Haslanger and Cappelen, on the other, are different challenges. The Carnap–Strawson debate is about what philosophy is for. According to Carnap, philosophers should construct conceptual tools for the purpose of scientific progress. According to Strawson (1992), “just as the grammarian [...] labours to produce a systematic account of the structure of rules which we effortlessly observe in speaking grammatically, so the philosopher labours to produce a systematic account of the general conceptual structure of which our daily practice shows us to have a tacit and unconscious mastery” (Strawson 1992, 7). Strawson’s objection is not that there is a problem regarding the limits of conceptual revision, but that revision as such is always philosophically irrelevant because the goal of philosophy is to illuminate the concepts we have (Koch 2023, 2128).

In contrast, Haslanger is asking about when a project is no longer revisionary but changes the subject. Cappelen asks: how much revision is too much? Strawson’s view is that any revision is too much. The continuity of subject matter may be a challenge for projects where the success conditions of the new concepts require successful communication in a broader community—perhaps *family* and *marriage* are such concepts. But for the method of explication, it is not a problem if the subject changes. The similarity criterion is intended to ensure that a pocketknife is not replaced by a hammer but by a surgical knife or microtome. It is not intended to preserve our attention on the pocketknife.

Another possible criticism against the similarity criterion in Carnapian explication would be the one taken by Cappelen (2018, 180–1), viz. to question the very idea of concepts having such things as purposes or aims or functions or jobs that can be preserved in conceptual revisions and replacements. In direct response to Cappelen, Amie Thomasson (2020, 442) argues that functions of concepts have a central role to play in the kind of pragmatic approach to conceptual engineering that she defends. In a way that seems faithful to Carnap’s views, she also maintains that functions save us from Strawson’s challenge, since “thinking of concepts or words in functional terms provides ways of legitimating the feeling that we haven’t simply ‘changed the subject’ when we engage in conceptual engineering” (443). Eklund goes

further, and similarly to the point Eklund made regarding the implementation challenge, he argues in response to the continuity challenge that it “may be a serious challenge for anyone who wants to defend the idea that one addresses the same topic before and after conceptual revision” but he rejects the idea, and concludes that “it would be in the spirit of conceptual engineering to say that changing the topic is sometimes exactly the point” (Eklund 2024, 8). For my purposes, it is sufficient so say that for explication the point is to focus and work with new concepts, and if our original questions or subject or topics are changed to more worthwhile questions or subjects or topics, it is not a problem—on the contrary.

7.8 Concluding remarks

In summary, some of the problems that are at the centre of attention in the conversations on conceptual engineering do not arise for explication. The challenges for theoretically motivated concept revisions are different than the challenges for projects of concept revision where the aims are social and political and where the success therefore depends on uptake in the broader linguistic community. Uptake is a success condition for the paradigmatic cases of socially or politically motivated conceptual engineering. But problems regarding how to implement a concept in the linguistic community do not concern explication.

The important questions regarding explication are questions about our epistemic and theoretical purposes and about what kinds of concepts are conducive to our epistemic and theoretical purposes. These questions are separate (but presumably with no clear-cut line) from questions, for example, about what kinds of concepts are conducive to our political, legal, social, and etiquette-related goals (and what those goals should be).

For all projects of conceptual engineering, including explication, the success is measured practically, by what is achieved by the new concept and whether it allows us to do what we want to do—hence the engineering metaphor. Still, there is an important albeit not clear-cut difference between epistemic and theoretical purposes and other purposes such as social and political purposes. For social and political projects, the relevant consequences come about through people’s use of the new concepts. In those cases, the desired changes in mental states and behaviour and social structures can only come about through people’s use of the new language. For epistemic projects, the relevant consequences are what we can do with the new linguistic device. The success

condition requires only that one can do something with the new linguistic device that one could not do with the old one, such as make new or more precise predictions, establish deductive and inductive links to laws and theories, pursue natural language semantics, and so on.⁶⁹ The task is to make epistemic tools, and the success condition is a fulfilled by a tool fit for its purpose. For political projects, the task is to change things in the social world by means of a new linguistic tool, and the success conditions require change in the social world.

Implementation requires that we can control the meaning of our expressions. Hence, implementation becomes a problem to the extent that we cannot control the meaning of expressions (however, if we could control it that would raise other problems with regard to democracy and autonomy, as discussed by Queloz). However, in an explication the meaning of the new term is stipulated by a definition or by the explicitly given rules of use (in line, I think, with Pinder's (2020a, 2021) speaker-meaning account of conceptual engineering). Whether or not people start to use the new expression according to those rules is not of immediate concern for aims of explication, and has no part in the success condition of projects of explication. Hence, the literature on conceptual engineering, if applied to the method of explication, is largely engaged in problems that distract from the problems that explication face. Nevertheless, there is a need to explore the method beyond the prescription given by Carnap. I embark on this task in the following chapters. I begin the next chapter with a discussion of the problems with the fruitfulness criterion and then present and criticize revisions that have been proposed for it.

⁶⁹ In Carus and Leitgeb's (2021) interpretation, a fruitful concept according to Carnap is a concept that can be: "usefully applicable in scientific or philosophical theorizing and discourse, e.g., in the formulation of lawlike statements, or by creating deductive or inductive links to established theories that are themselves sufficiently clear, exact and successful."

8 Broadening the fruitfulness criterion

8.1 Overview

Among Carnap's four criteria,⁷⁰ the fruitfulness criterion is the one that is most deeply connected to views held by Carnap which today are widely rejected. In Carnap's view, a fruitful concept should be useful for the formulation of universal statements, which he understands as "empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept" (LFP, 7). It has been pointed out that this notion of fruitfulness is unsuitable for many scientific fields (Kitcher 2012, 197), and that it excludes normative philosophical purposes (Shepherd and Justus 2015, 393–4; Olsson 2017, 4–5). In light of these problems, philosophers who are mostly sympathetic to Carnap's conception of explication have proposed revisions that go beyond or even against Carnap's views. Some authors have merely criticized aspects of Carnap's notion of fruitfulness (Brun 2016; Olsson 2017), while others entirely reject it and propose alternative notions of fruitfulness (Kitcher 2008; Pinder 2022b). One of the revisions of Carnap's notion of fruitfulness which I will discuss, the one by Dutilh Novaes and Reck (2017), is an *interpretational* revision supposed by the authors to be in accordance with Carnap's views, while the other two revisions, by Kitcher (2008) and by Pinder (2022b), clearly deviate from Carnap's views.

Before I address these alternative accounts, I will discuss issues with Carnap's notion of fruitfulness. The issues which I discuss below in relation to the fruitfulness criterion are (i) the epistemic role of universal statements in Carnap's criterion, (ii) the role of explanations in relation to Carnap's criterion, (iii) the role of normative and metaphysical statements, (iv) the role of historical kinds in predictions and explanations, and (v) the role of laws in the aims of science.

⁷⁰ They are discussed at length in 2.7.

After my discussion of these problems, I present extant alternative accounts of fruitfulness, and I end the chapter with a discussion of the shortcomings of these alternative accounts.

8.2 Problems with Carnap's notion of fruitfulness

8.2.1 True universal statements

A minor issue with Carnap's formulation, raised by Olsson (2017, 4, n. 3), is that Carnap does not mention truth in his formulation of the criterion. Carnap's requirement is merely that the explicatum should be useful for the formulation of universal statements, but he does not mention that those statements need to be true. As Olsson remarks, "One would have expected 'many *true* universal statements'" (Olsson 2021, 3, n. 2). Further, Olsson comments that, taken literally, Carnap's formulation implies that "any old universal statement will do", rather than true universal statements. As Olsson also points out "[Carnap's] clarification in terms of 'empirical laws' and 'logical theorems' suggests that the statements in question must be plausibly true and also of a certain theoretical standing" (2017, 4, n. 3). I agree that it is reasonable to take it as implied that also the non-logical universal statements are expected to be of a certain theoretical standing. However, a possible reason for not including the question of truth in the criterion would be that the truth, or rather the degree of confirmation (in Carnap's terminology), of an empirical law can only be found out at a later stage when observable predictions have been made and tested. As Carnap remarks in his *Philosophical Foundations of Physics* (1966), "an empirical law, if it is a tentative hypothesis, confirmed only to a low degree, would still be an empirical law although it might be said that it was rather hypothetical." (1966, 227). At the stage of explication it might be premature to make judgements about the epistemic status of the law statements, and therefore such a requirement might thwart scientific progress.

Regarding the expected role of universal statements in Carnap's criterion, it may be illuminating to take a further look at Carnap's distinction between empirical laws and theoretical laws. He characterizes empirical laws as

laws containing terms either directly observable by the senses or measurable by relatively simple techniques. Sometimes such laws are called empirical generalizations, as a reminder that they have been obtained by generalizing results found by observations and measurements. They include not only simple

qualitative laws (such as, “All ravens are black”) but also quantitative laws that arise from simple measurements. (Carnap 1966, 227)

In contrast, he characterizes the theoretical laws by the unobservable nature of what their terms (putatively) refer to. Theoretical laws are about entities which “cannot be measured in simple, direct ways”, such as “molecules, atoms, electrons, protons, electromagnetic fields” (Carnap 1966, 277). The supreme value of theoretical laws, according to Carnap, is their power to predict new empirical laws. When empirical laws which are derived from a theoretical law are confirmed, the theoretical law is thereby indirectly confirmed. On Carnap’s view, then, the goal of theoretical inquiry is to be able to explain known empirical laws and predict new empirical laws. New empirical laws are confirmed or disconfirmed through experiment and observation. Hence the explicator would not have to worry about the epistemic status of the empirical laws *qua explicator*. Nevertheless, there would be no point to the enterprise if the newly formulated universal statements were not at least considered to be promising hypotheses.

8.2.2 Explanation

In his writings about natural laws Carnap lays equal emphasis on both of the uses we have for empirical laws, namely, to explain known facts and to predict unknown facts; theoretical laws, meanwhile, are used for explaining and predicting empirical laws. Carnap endorses a covering-law, or deductive-nomological, account of explanation (1966, 7). On Carl Hempel’s classic deductive-nomological account of scientific explanation, we explain a phenomenon by deducing the sentence describing it from a set of premises that includes at least one law which is necessary for the deduction (Hempel 1965, 335–376; Lipton 2004, 25–27). Given such a view, the requirement that our explicata are useful for the formulation of laws seems sufficient to serve the aims of science since it facilitates both the task of explaining and the task of predicting.

However, there are a host of well-known problems with covering-law theories of explanation.⁷¹ And if we reject the covering-law account of explanation it is no longer guaranteed that Carnap’s fruitfulness criterion facilitates the task of providing new explanations. There is less appeal in a fruitfulness criterion that would not have facilitated some of the paradigmatic

⁷¹ As Peter Lipton puts it, more strongly, the covering law account “faces debilitating objections” (2004, 27).

examples of good scientific explanations. Depending on which account of explanation we adopt, we need to change accordingly our notion at least of fruitfulness, so that a fruitful concept facilitates both new prediction and new explanations. The role of explication in facilitating explanations becomes particularly important if we allow non-logical and non-empirical explicata, such as concepts in metaphysics. Since we cannot use metaphysical concepts to make predictions, but only to (putatively) explain phenomena, the expected benefits of an explicatum would be entirely tied to explanation.

8.2.3 Carnap's view of cognitive significance

Carnap's notion of fruitfulness is too restrictive because he equates non-logical concepts with empirical concepts, thereby leaving out, for example, normative concepts, an issue that has been raised by several authors (Olsson 2017, 4; Brun 2016, 1224; Shepherd and Justus 2015, 394). I quote from Olsson's summary of the problem:

Details of Carnap's account can be questioned from a contemporary perspective. One example is the division between "logical" and "nonlogical" concepts, whereby, crucially, "nonlogical" is treated as synonymous with "empirical". This picture leaves out, among other things, the important categories of ethical and legal concepts, not to mention presumably evaluative epistemological expressions such as "good reason" or "justification". Behind Carnap's treatment lies presumably the thought that explication only concerns scientific concepts, and the further view, widely shared at the time, that ethical or legal concepts, to the extent that they cannot be reduced to logical or empirical concepts, do not belong to science, properly so called. (Olsson 2017, 4)

While Olsson in general does not deviate much from Carnap's account of explication, he concludes with regard to this problem that:

A modern reader can happily dismiss the identification of the nonlogical with the empirical alluded to above as a relic of 20th century positivism and proceed on the assumption that irreducibly epistemological, ethical and legal concepts, if such there be and presumably there are, are just as amenable to explication as logical and empirical concepts are once the criteria of fruitfulness are correspondingly broadened. (Olsson 2017, 4–5)

However, such a move to broaden the fruitfulness criterion comes with potential problems.

For example, if one dismisses Carnap's assumptions about meaningful theoretical questions and statements, one needs to reconsider what the motivation is to do philosophy in the form of explication, in contrast to other ways of doing philosophy. It follows, of course, from Carnap's view of cognitive significance that the universal statements by which fruitfulness is measured are restricted to logical and empirical statements (including theoretical statements in the empirical sciences, which according to Carnap are connected to observable statements through rules of correspondence). Carnap's non-cognitivism about other propositions in other domains such as normative ethics and metaphysics dictates that normative concepts are excluded. In Carnap's view, universal statements formulated with ethical concepts would not be cognitively significant. While the following quotation is from an earlier period of Carnap's thinking, it would presumably be representative of the later Carnap too:

The propositions of normative ethics, whether they have the form of rules or the form of value statements, have no theoretical sense, are not scientific propositions (taking the word scientific to mean any assertive proposition). [...] Therefore we assign them to the realm of metaphysics. (Carnap 1935, 25–26)

Carnap regarded metaphysical statements, which he takes to include moral statements, as pseudo-statements devoid of cognitive content. What about theoretical laws? For the nonlogical concepts, Carnap only mentions empirical laws as suitable universal statements. Presumably, theoretical laws should also be included as cognitively significant universal statements, since they obtain cognitive significance through the empirically testable laws derived from them. In his "The Methodological Character of Theoretical Concepts" (1956), Carnap sets out to clarify the nature of theoretical language and the relation between theoretical language and observation language. He takes the problem of providing a criterion of significance for theoretical language to be "very serious", and describes such a criterion as giving "exact conditions which terms and sentences of the theoretical language must fulfill in order to have a positive function for explanation and prediction of observable events and thus to be acceptable as empirically meaningful" Carnap (1956, 38). While Carnap here admits that "the connection between the observation terms and the terms of theoretical science is much more indirect and weak than it was conceived either in my earlier formulations or in those of operationism" (1956, 53), he explicitly rejects the view, famously advocated by Quine (1951b) and commonplace today, that science is continuous with metaphysics. Regarding the weak connection between observation terms and theoretical terms Carnap writes:

From this fact, some philosophers draw the conclusion that, once the earlier criteria are liberalized, we shall find a continuous line from terms which are closely connected with observations, e.g., ‘mass’ and ‘temperature’, through more remote terms like ‘electromagnetic field’ and ‘psi-function’ in physics, to those terms which have no specifiable connection with observable events, e.g., terms in speculative metaphysics; therefore, meaningfulness seems to them merely a matter of degree. (1956, 39)

Carnap, however, maintains that “also in the theoretical language, it is possible to draw an adequate boundary line which separates the scientifically meaningful from the meaningless” (1956, 40). To eliminate or dissolve what he regarded as futile disputes about metaphysical statements was a necessary part of the reformation of philosophy advocated by Carnap and the other logical empiricists, and linguistic engineering in the form of explication was intended to replace philosophical debates.

Hence, to broaden the fruitfulness criterion so that it applies beyond theoretical laws in empirical science to universal statements in metaphysics would require the rejection of Carnap’s empiricist criterion of meaning. However, if this cornerstone of Carnap’s outlook and motivation is abandoned we also lose what Carnap regarded as one of the central benefits of pursuing philosophy as explication. For Carnap, the benefit was to dissolve what he regarded as futile disputes over traditional philosophical concepts by replacing them with new concepts with firm empirical or logical underpinnings. If we allow as explicatum, for example, normative or metaphysical concepts, we need to reconsider why we should bother with explication. What motive are we left with to pursue philosophy as explication and not in some other ways? If the fruitfulness criterion is broadened in this way, a new account is needed of what the benefits are of explication over other methodological paradigms of philosophy. I return to these questions in the concluding chapter of the thesis.

8.2.4 Not only exceptionless generalizations are useful for prediction and explanation

There are many interesting kinds which do not facilitate the formulation of universal statements but nevertheless facilitate predictions and explanations. To illustrate, I turn to Ruth Garrett Millikan’s study of substance concepts (2001). Millikan compares eternal kinds and historical kinds, and notes that “[m]any kinds of interest to social scientists, such as ethnic, social, economic, and vocational groups, are historical kinds” (2001, 22). Further, she points out that historical kinds “are not likely to ground many, if any, exceptionless

generalizations” (ibid.). Despite not facilitating universal generalizations, they are clearly useful:

Historical substances are not likely to ground exceptionless generalizations. But many substances interest us not because they afford such reliable inductions, but because they afford so many inductions. They bring a great wealth of probable knowledge with them. (Millikan 2001, 25)

While we cannot formulate universal generalizations with historical kind concepts, we may nevertheless improve such concepts with explicata that allow for more inductions than the explicandum. Therefore, we should not restrict our measure of fruitfulness to a concept’s ability to formulate universal statements.

8.2.5 To detect laws is not the only aim of science

The fifth and last problem regarding the fruitfulness criterion which I will discuss is a problem that has led proponents of explication to entirely reject Carnap’s account of fruitfulness and propose alternatives to it. From the perspective of philosophy of biology, Kitcher claims that “Carnap’s own elaboration of how to think about the scientific purposes of a concept needs refinement” because Carnap “ties the fruitfulness of a concept to its utility in formulating laws and universal generalizations” (2008, 115). The problem according to Kitcher is that since the aims of the life sciences cannot be understood in those terms, Carnap’s notion of fruitfulness is not applicable for explications in the life sciences. Kitcher diagnoses the problem as a narrow focus on physics by Carnap and his fellow logical empiricists:

Inspired by examples from the physical sciences, Carnap and his contemporaries developed a picture of the sciences according to which the principal aims are to identify laws of nature. That picture doesn’t suit parts of the physical sciences, and it’s deeply problematic for the biological, earth, and human sciences. (Kitcher 2008, 115)

Against the view that the aim of science is not to identify laws of nature, Kitcher argues that our own changing interests are the only standard for the purposes of inquiry:

There’s no higher standard to which our concepts are to answer than the efficient satisfaction of the purposes of inquiry; those purposes are set, not by nature, but by us, and they evolve with changes in our knowledge and our

society. To think that there's a structure set in nature to which the concepts must conform is, I believe, an article of metaphysics—and a form of metaphysics we are much better off without. (Kitcher 2008, 119)

If the aims of some of the sciences are not facilitated by the formulation of new laws then Carnap's fruitfulness criterion arguably needs to be supplemented or broadened. Also, if Carnap has the wrong view about the aims of all or some sciences, and his criterion is intended to serve those misconceived aims, there is a need to modify his criterion.

I will now discuss extant alternatives to Carnap's notion of fruitfulness, including Kitcher's own positive proposal to understand fruitfulness as the provision of answers to significant questions. Later, in chapter 9, I will propose that we should adopt a broader formulation of Carnap's fruitfulness criterion.

8.3 Alternative criteria of fruitfulness

8.3.1 Overview of the accounts

The critical revisions of the fruitfulness criterion begin with Kitcher's above-mentioned account of fruitfulness as answers to significant questions, which he developed in the paper "Carnap and the Caterpillar" (2008). As a development of Kitcher's account, Pinder (2022b) proposes an account of fruitfulness as facilitation of progress towards relevant theoretical goals. A less revisionary account of fruitfulness is provided by Dutilh Novaes and Reck (2017). They propose an interpretation of Carnap in which fruitfulness is understood as the facilitation of the production of new knowledge. While they do not themselves propose it as an alternative criterion but merely as an interpretation of Carnap, Pinder treats their proposal as an alternative criterion of fruitfulness. Regardless of the intentions of Dutilh Novaes and Reck, their proposal can plausibly be treated as an alternative account of fruitfulness. I begin with a section where I present the interpretation by Dutilh Novaes and Reck, followed by a section each for the revisions proposed by Kitcher and by Pinder.

8.3.2 New knowledge

Dutilh Novaes and Reck think that Carnap's fruitfulness criterion "is somewhat under-developed", including "its relationship to the other criteria"

(2017, 197). While they do not claim to “present a new textual exegesis of Carnap’s work” (ibid.), they nevertheless claim that their interpretation is to be found in Carnap’s own writings:

Carnap seems to expect that the explicatum will be able to reveal certain things about the phenomenon in question that the explicandum could not reveal, by means of the newly established connections to other concepts. In other words, Carnap’s view seems to be that an explication is useful or fruitful when it delivers ‘results’ that could not be delivered otherwise (or with much more difficulty), i.e. with the explicandum alone. (Dutilh Novaes and Reck 2017, 205–6)

Hence, they conclude, explication is a method for discovery, for producing new knowledge.

What this suggests is a conception of explication as a method for discovery, as opposed to a method for testing or justification alone. The goal is to produce new knowledge about the phenomena to which the explicandum pertains. (Dutilh Novaes and Reck 2017, 206)

While Dutilh Novaes and Reck do not themselves propose this as an alternative fruitfulness criterion, Pinder does present it as such. In his formulation, the criterion is that: “A concept is fruitful insofar as it facilitates the production of new knowledge about the phenomena to which it pertains” (2022b, 918).

A crucial difference between the ‘new knowledge’ criterion and Carnap’s original criterion is that only the latter provides instructions for how an explicator may facilitate discoveries, results, and new knowledge: namely, by constructing explicata that are helpful for the formulation of law-statements. The ultimate aim of producing new knowledge is not the immediate concern of the explicator at the stage of constructing an explicatum. When the fruitfulness of an explicatum is measured in terms of new knowledge produced with its help, its fruitfulness can only be evaluated in the long run. Therefore, the criterion is no longer instructive. In contrast to Carnap’s criterion, it does not give instructions for how conceptual work might contribute to the ultimate purpose of obtaining new knowledge. I will raise the same critique for the other accounts that I discuss below.

Another point, which I clarified in section 8.2.1 where I discussed the question of truth in relation to the fruitfulness criterion, is that on Carnap’s view the question of truth does not seem to enter the picture at the stage of explication, but only at the later stage when testable predictions confirm or

disconfirm the universal statements formulated with the explicatum. It would be premature to require knowledge, and hence truth, in the criteria of adequacy.

I end this section with an exegetical and rather pedantic remark about Dutilh Novaes and Reck's interpretation of Carnap. They seem to consider the sole purpose of Carnap's fruitfulness criterion to be the facilitation of testable predictions, since they write that "[t]he idea of an explicatum leading to many universal statements may be understood in terms of predictive power and testability: if it allows for the formulation of such statements, it allows for many predictions that can be tested" (Dutilh Novaes and Reck 2017, 205). This interpretation is part of the motivation for their revision of the account. However, since Carnap endorses a nomological model of explanation where at least one law-statement is required in every explanation (1966, *passim*), if an explicatum allows us to formulate new law-statements it does not only help us make new testable predictions but also helps us provide explanations for phenomena that we could not explain before the explication. On Carnap's view, the goal to be achieved with the help of fruitful explicata is both new explanations and new predictions. Therefore, on Carnap's view, the "results" or the "new knowledge" are in the form of predictions or explanations, and both require laws.

8.3.3 Answers to significant questions

As already seen in section 5.2.3, Kitcher criticizes Carnap's fruitfulness criterion for being unapplicable in the life sciences as well as in many other fields. The reason, in Kitcher's words, is that the criterion "ties the fruitfulness of a concept to its utility in formulating laws" (2008, 115) and Kitcher believes that the aims of these scientific fields cannot be understood in those terms. Instead, Kitcher proposes that "we conceive of the aims of the sciences in terms of provision of answers to significant questions, where the sources of significance are various, sometimes practical, sometimes in terms of the satisfaction of disinterested curiosity." He adds that "with this revision, however, I endorse Carnap's requirement of fruitfulness" (Kitcher 2008, 115). By generalizing the criterion in this way, Kitcher ensures that it is applicable to any scientific field or line of inquiry. Substantive accounts of fruitfulness are only derived from a significant question formulated in a specific inquiry. Pinder (2022b, 916) regards it as a benefit of Kitcher's account that the notion of fruitfulness is general and is only made substantive relative to specific cases when a significant question has been formulated. In his own account, to be discussed further in the next section, Pinder attempts to "combine generality with substantive detail" in a similar fashion (*ibid.*). However, I find problems

with the aspect of Kitcher's account that Pinder considers to be a benefit, and I believe that those problems are inherited in Pinder's proposal. In the last section of this chapter, 8.4, I address these problems.

8.3.4 Approaching relevant theoretical goals

Pinder states his Relevant-Goals criterion of fruitfulness as follows (2022b, 918):

An explicatum is fruitful insofar as its replacement of the corresponding explicandum would facilitate, through the ordinary course of inquiry, progress towards achieving relevant theoretical goals.

The criterion is both general and case-specific. The requirement to facilitate progress towards relevant goals applies to all cases, but substantive accounts which can be used to judge the fruitfulness of a concept are of course relative to the specific relevant theoretical goals. Hence, this is a general criterion that "can be used to derive more substantive accounts for specific explications by specifying what the relevant theoretical goals are in each case" (Pinder 2022b, 914). Since what counts as a relevant theoretical goal varies from case to case, Pinder takes his account to be compatible with the notions of fruitfulness proposed by Carnap, and by Kitcher and by Dutilh Novaes and Reck. He thinks that there could be specific cases where the relevant theoretical goals are, respectively, "to formulate universal statements; to answer significant questions; and to produce new knowledge about the phenomena to which the explicatum pertains" (Pinder 2022b, 919, n. 13). However, for these other notions of fruitfulness to be compatible with Pinder's account they must be considered case-specific and not considered as exhaustive criteria of a fruitful concept.

Since fruitfulness on this account is relative to particular relevant theoretical goals, it raises the question of what we should count as a relevant theoretical goal. Pinder's answer is to contextualize the notion of relevant theoretical goal: "Contextualism is the view that, for any given explication, the relevant theoretical goals are the *explicator's theoretical goals*" (922). On Pinder's view, "explicators have a significant degree of control [...] over the measure of fruitfulness that applies in a given case" (924). However, he stresses that "[t]his does not imply that anything goes" (ibid.). Since Pinder understands

theoretical goals in terms of theoretical values,⁷² which are “general characteristics of good scientific theories” (ibid.), his account is indirectly connected to previously successful inquiry but is not directly derived from previously successful features of concepts.

In the next section I criticize the contextualism that is employed in these accounts, with a focus on Pinder’s account.

8.4 Contextualism versus generalism

In this section I compare contextualism with the opposing view, which we may call ‘generalism’. Generalism, then, is the view that there are generic features to be fulfilled to some degree by all explications for them to count as satisfactory. Before defending generalism, I will address problems contextualism. On Pinder’s contextual notion of fruitfulness, the criterion of fruitfulness is not attempts to capture previously successful features of scientific concepts. In this regard, his account deviates from Carnap’s account. Carnap attempts to specify features that previously have led to scientific success, namely exactness, fruitfulness, and less importantly simplicity. The previous success of these features in scientific concepts is the rationale for adopting the same method of concept formation in philosophy. Without such a tie to scientifically fruitful concepts, explication seems to lose some of its legitimacy as a philosophical method. It is presumably the previous successes of exact and fruitful concepts that legitimizes explication as a method for philosophy. Shepherd and Justus (2015) emphasize this aspect of explication, claiming that “[p]rioritizing precision and fruitfulness over strict preservation of conceptual content reflects methodology in science and as the unparalleled exemplar of epistemic success in human inquiry Carnap thought philosophy should follow suit” (2015, 388). This is what motivates Carnap’s programme of explication in philosophy. Without this guiding and legitimizing connection to features that have previously led to epistemic success, the motivation for this way of doing philosophy is weakened. Besides the loss of legitimacy the

⁷² A worry about circularity may arise here since one of the canonical theoretical values, or theoretical virtues, is fruitfulness. However, Pinder does not include fruitfulness in his non-exhaustive list of examples of theoretical values. He mentions internal consistency, coherence with other accepted theories, evidential accuracy, scope, explanatory power and simplicity. The kind of fruitfulness discussed by Kuhn (1974), and others, as a theoretical virtue that may guide theory choice, is however different from Carnap’s notion of fruitfulness.

criterion also loses some of its instructive role. Without instructions that are both substantive and generally applicable for how to proceed in a task of explication (at least with regard to fruitfulness), it is hard to see how the criteria of adequacy help the explicator. While Carnap's criterion needs to be broadened and revised, a tie should be kept to specific features of concepts that previously have led to epistemic success, not just in specific inquiries but in general.

The proposed revisions discussed in this chapter do not provide guidance for the kind of explications for which I think Carnap articulated the method, namely explications in philosophy and the less developed parts of inquiry. As Gerring points out, "methodology is central to the disciplines of the social sciences in a way that it is not to the natural sciences" (2011, xxiii). A significant part of the appeal of explication in philosophy is that it supposedly provides a way of doing philosophy that is more likely to yield progress than other ways of doing philosophy. To play that role, we need to articulate the features that make our concepts good for inquiry, also in the less developed knowledge disciplines. Felix Oppenheim makes a similar point in relation to his defense of reconstructionism, i.e., the method of giving explicative definitions of political concepts. He remarks that "if I propose some explicative definition, I have the burden not of proving that it is true, but of justifying it in terms of standards that themselves are not under discussion" (Oppenheim 1981, 181).

For philosophers sympathetic to explication, the point of investigating the method of explication rather than just pursuing explication is presumably to develop a methodology that is better than Carnap's original method for fulfilling the same ultimate purpose. These methodological investigations should therefore ideally result in a method that can be taught to students and included in textbooks on methodology in philosophy and perhaps in textbooks on methodology in the human and social sciences. If we think that it would be ill-advised to include Carnap's four criteria, as they are stated by him, in for example a textbook on philosophical methodology or in a textbook on concept formation in other fields, we should develop a better method which could be included such a textbook.

In the balancing act between applicability and instructiveness, I tilt further towards the latter than the accounts treated in this chapter. For a criterion of adequacy for explication to be useful it should be applicable in many fields and contexts but it should also be instructive.

8.5 Concluding remarks

I have addressed a conflict that arises between, on the one hand, views in which explication is considered a method for meeting the needs that arise in specific inquiries and, on the other hand, views that criteria for adequate explications can and should be articulated for inquiry in general. The question is whether criteria of adequacy vary from case to case, or not. While some revision of the fruitfulness criterion is needed, I claim that too much of the instructive purpose of methodological inquiry is lost when general and substantive criteria are abandoned. The revisions that I have discussed in this chapter are attempts to make the criterion of fruitfulness more applicable than Carnap's criterion, which is considered too narrowly applicable. However, by drawing from historical examples of successful conceptual revisions in a variety of fields, I believe that one can formulate general principles for how explication should be pursued. This is what Carnap tried to do and what I set out to do in the following chapter.

9 Towards a modified version of explication

9.1 Overview

It is now time to present my own account of explication, in which I modify and expand Carnap's criteria of adequacy. In 9.2, I discuss how the account I propose differs from Carnap and from the account proposed by Brun (2016), which is the most comprehensive and ambitious contemporary development of Carnap's account. In 9.3, I give an overview of the criteria and present the internal structure of my version of explication, building on the discussion of internal structure in chapter 4. I divide the criteria of adequacy into defining and good-making criteria. In 9.4, I present the defining criteria and in 9.5 I present the good-making criteria. I end the chapter with a discussion of how to evaluate my version of explication against other versions of explication, which leads into the next chapter, where I put my criteria to use in an explication.

9.2 Locating my account in the literature

9.2.1 Explication of social concepts

In this section I discuss how my proposed modification of Carnap's account is related to Carnap's original version and to the account proposed by Brun (2016). Relatedly, my interpretation of Carnap differs from Brun's interpretations and I will clarify these differences as well. First, I address how my proposed modifications deviate from Carnap's account.

I propose to retain relaxed versions of Carnap's criteria and to add new criteria, to accommodate a greater variety of ways in which a concept may be improved for inquiry. In particular, I expand Carnap's list of criteria, which was mainly based on concept formation in the natural and mathematical

sciences, with criteria which are appropriate to a broader range of social and human sciences. It is natural to discuss concept formation in the social sciences in relation to philosophical explication, since the problem of validity for social science concepts bear similarities to concerns regarding the similarity requirement for explications.

In the social sciences much weight and focus is given to what is called the validity⁷³ of a measure, in the sense that a measure is valid to the extent that it adequately reflects the concept it is intended measure (Babbie 2016, 148-5). In Hempel's formulation, a test is valid if it provides "a correct characterization of the feature to which [a] term refers in ordinary usage" (Hempel 1952, 48). He notes the contrast between the social sciences and physics, where validity is not a desideratum:

Physics, for example, would not have attained its present theoretical strength if it had insisted on using such terms as 'force', 'energy', 'field', 'heat', etc., in a manner that was "valid" in the sense discussed above. [...] Indeed, it is largely a matter of historical accident, and partly one of convenience, that the terms of conversational English are used in the formulation of abstract theories [...]" (Hempel 1952, 49)

While concept formation in the natural sciences could be lexically and conceptually divorced from everyday language, in the social sciences the link to everyday language is often intimate and important, just as it is in (much of) philosophy. Hence, it is plausible in many of these fields to understand the task of concept formation as a form of explication (as I argued in chapter 5). There is, of course, plenty of conceptual invention of entirely new technical terms in philosophy and the social sciences as well, so I do not suggest that concept formation should be pursued exclusively in the form of explication. However, there are many concepts in need of improvement whose relation to the terms of conversational natural language are not merely a matter of historical accident or convenience. In many cases, the study of these concepts falls in the intersection between philosophy and the social sciences, e.g., the study of the concepts of *health*, *well-being*, *trust*, and *democratic*.

In the fifth chapter I began a task of unification when I reframed Gerring's work on concept formation in terms of explication and conceptual engineering, and when I compared Felix Oppenheim's reconstructions of political concepts with Carnap's method. In this chapter I strive to continue that task when I propose modifications of Carnap's criteria of adequacy.

⁷³ Not to be confused with logical validity.

9.2.2 How my account deviates from Brun

While I agree with Brun that there is a need for an account of explication that is modified and in some ways more relaxed than Carnap's account in LFP (and I try to develop such an account in the next chapter), I disagree with Brun regarding *how* the criteria should be relaxed. I also agree with Brun's assessment that "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" (Brun 2016, 1224). However, I disagree regarding how that should be supplemented. Brun recommends putting more focus on the target theory and then evaluating the explication by the theoretical virtues of the resulting theory. Hence, Brun looks for additional criteria in the canon of theoretical virtues. In contrast, I recommend maintaining the distinction between on the one hand good features in a concept and on the other hand good features in a theory. For supplementation of the criteria, I do not mainly look at the theoretical virtues—features that makes a scientific theory good—but I look at work on social science concepts, i.e., work on what makes a concept good for social research.

Further, Brun argues that the method of explication should be relaxed in line with Carnap's later view, which Brun interprets as more pragmatic. I also object to parts of Brun's interpretation of Carnap. To give important background for the ensuing discussion, I here repeat from chapter 2 a few remarks about Carnap's most important writings on explication. Carnap's most in-depth exposition of explication is found in *Logical Foundations of Probability* (hereafter LFP), published 1950, and the second most important source for Carnap's views on explication is his reply to Strawson in "Replies and Systematic Expositions" (hereafter RSE), published in 1963 in the Schilpp volume dedicated to his work. Carnap's replies were written already in the mid-1950s and completed in 1958 (Creath 1990a, 448–449). That RSE was written not much more than five years after LFP was brought to my attention by Brun (2016, 1214, n. 5). Nevertheless, Brun argues that Carnap's views changed substantially between these works. Brun makes the following three interpretational claims:

- (a) there are internal tensions in Carnap's views in LFP,
- (b) Carnap's views changed substantially towards a more pragmatic perspective between LFP and RSE (Brun 2016, 1225), and
- (c) Carnap's discussion in RSE can be read as resolving the tensions in LFP.

In Brun's interpretation, the official account, i.e., the account given in LFP, has been relaxed in RSE. Brun summarizes the changes in Carnap's account as expressed in RSE in the following way:

[Carnap] 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. (Brun 2016, 1226)

However, there are reasons to be cautious about whether or not Carnap meant to relax his account by emphasizing philosophical uses of explication. First, as Reck points out, among the goals that motivated Carnap's conception of explication were the goals to "erase the lines between logic, philosophy, and the sciences" and "thereby to transform philosophy into 'mathematical' or 'scientific philosophy'" (Reck 2024, 143). Hence, the mere fact that Carnap stresses the importance of explication in philosophy is not obviously evidence of a relaxation of his account. There is a further reason to be cautious. Although Carnap gives examples of explications in the empirical sciences where formal methods are not required (e.g., the fish example), as Gustafsson (2014) observes it is important for Carnap's specifically *philosophical* uses of explication to incorporate explicata into formal semantical systems. Gustafsson addresses Carnap's apparent relaxation of his standards of exactness in his reply to Strawson in RSE, but concludes that

even if Carnapian explication does not always or even in most cases require that the explicatum is given by rules that assign to it a determinate function in a formally exact calculus or semantic system, such formally precise explicata *are* needed if our aim is to make adequately sharp distinctions between framework and assertion, analytic and synthetic sentences, and so on. And such distinctions are precisely what Carnap thinks we need to make in order to get rid of the most stubborn and seemingly irreconcilable philosophical quarrels [...] (Gustafsson 2014, 514)

Hence, when Carnap emphasizes the philosophical usefulness of explication it is not by itself a relaxation of the requirement of exactness; it could also be an emphasis on the usefulness of exactness for philosophical purposes. In summary, there is reason both to interpret Carnap's endorsement of philosophical uses of explication in a restricted sense and to take his "conciliatory rhetoric" (Gustafsson 2014, 514) towards Strawson with a grain of salt.

Regarding Carnap's pragmatic development, Brun claims more specifically that Carnap in his new and more pragmatic view in RSE endorses the following two points:

- (1) "the requirements an adequate explication has to meet cannot be specified in general but only with respect to a specific task of explication". (Brun 2016, 1125)
- (2) "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." (Ibid.)

Since I object to Brun's interpretation of Carnap regarding both (1) and (2), I do not think that there is sufficient evidence to support the claim (b), i.e., that Carnap's views changed substantially between LFP and RSE. Regarding (1), I do not think that there is sufficient textual evidence in Carnap's writings that he adopted such a view. In a sense, the four criteria (and especially the similarity criterion) have to be specified with respect to specific tasks of explication, but that was clearly assumed by Carnap already in LFP when he provided general requirements, as the variations in his examples illustrate.

Regarding (2), I believe that Carnap already held this view in LFP. The view that the choice of the explicatum is a practical decision to be taken based on its intended purposes is compatible with LFP. That the choice of an explicatum is a practical decision is not only in line with what Carnap writes in LFP, but it is a view clearly expressed in his "Empiricism, Semantics, and Ontology", also published 1950. It is a view that Carnap endorses throughout his works at least from 1932 onwards. The view that such a practical decision needs to be "taken in view of the specific problems the explicatum is expected to solve" seems to me to be an uncontroversial claim, and it is compatible with the discussions and examples in LFP. Further, it is compatible with the general requirements for explications in Carnap's LFP. While I think that Carnap endorses (2) in both LFP and RSE, it is not clear to me that (1) follows from (2). The stronger claim of (1) is a defensible position (although I advise against it) but I do not think that there is sufficient evidence that it is the view of Carnap in RSE.

How, then, are these interpretational questions related to Brun's own account? Brun develops his own account of explication in line with his interpretation of Carnap, but he notes that the interpretational claim is not needed to justify his own project. Brun thinks that there are independent

reasons to develop the method in line with (his interpretation of) RSE, and that it is therefore charitable to interpret Carnap as advocating the more pragmatic position:

[S]ince 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.* (Brun 2016, 1227, my emphasis)

Hence, there are two interpretational options regarding RSE: either to interpret Carnap as advocating a new position in it or to interpret Carnap as clarifying the earlier position in LFP. In Brun's view, we get a better interpretation by choosing the second option and thereby resolve the tensions in LFP in line with RSE. Hence, he claims, it would be a charitable interpretation of Carnap to do so. I also think that the second option is the right choice, but in contrast to Brun I think that RSE should be interpreted as being mostly in line with LFP (and Carnap's later writings in general). I propose to develop explication while retaining a list of general criteria of adequacy, in line with LFP. I do, however, agree with Brun that the criteria should be supplemented and relaxed.

In summary, I disagree with Brun's interpretation that Carnap substantially changes his views but I agree with Brun's claim that there are independent reasons to develop his account of explication. For example, I think that Carnap's criteria should be modified and extended since they are not suitable for important and unavoidable contexts of inquiry, viz. contexts that fall outside formal language construction, logico-mathematical inquiry, and parts of natural and empirical science. However, I partly disagree with the idea that we should tie the criteria of adequacy to the target theory. Because of this point of disagreement, I propose to develop the method differently than Brun, more in line—in some respects—with Carnap's account in LFP.

9.3 The proposed modifications of the criteria of adequacy

I propose to modify Carnap's account in two ways. First, I divide the criteria into defining and good-making criteria. Second, I modify and supplement Carnap's criteria. The similarity criterion is the only one which is unchanged,

although in relation to it I discuss aspects of similarity that are not included in Carnap's discussion. I divide the exactness criterion into three separate criteria: explicitness (which I count as a defining criterion), sharpness and connectedness. I relax the fruitfulness criterion and I propose to add a criterion of intersubjectivity, which retains the spirit of Carnap's ideal of unified science, while relaxing it. Both connectedness and intersubjectivity are new in the sense that I take Carnap's presuppositions and make them into explicit (but weaker) independent criteria. I add a substantial criterion of simplicity that differs from Carnap's simplicity criterion. I have added further features to the simplicity criterion, and with these features it is motivated to raise that criterion to the same level of priority as the other good-making criteria.

The defining criteria I propose are *similarity* and *explicitness*. The good-making criteria are *sharpness*, *connectedness*, *fruitfulness*, *intersubjectivity*, and *simplicity*. I will devote one section to each criterion, discussing the criterion in question in detail and providing motivations to adopt it. First, though, I give the following brief "bullet point" statement of these criteria in the same manner in which Carnap gives his criteria in LFP:

Defining criteria

- (1) The explicatum should be *similar* to the explicandum in such a way that it is useful for similar purposes as the explicandum. Differences are required and considerable differences are permitted.
- (2) The explicatum should be provided with *explicit* rules of use.

Good-making criteria

- (3) The explicatum should be at least as *sharp* as the explicandum, and preferably sharper.
- (4) The explicatum should be *connected* to relevant systems of concepts.
- (5) The explicatum should be *fruitful* in the sense that it is conducive for the formulation of generalizations and lawlike statements, or for the creation of links to established theories.
- (6) The explicatum should be *intersubjective* in the sense that it is easy to judge if the concept is applicable or not, so that different users will use the concept in the same way and so that a single user would use it in the same way over time.
- (7) The explicatum should be *simple*. Explicata with fewer dimensions are simpler than explicata with more dimensions.

These criteria are intended as methodological guidelines for explications, and there is no sharp line between the bullet point statements of the criteria (1–7

above) and the additional remarks and considerations that will follow.⁷⁴ I simply try to put the most important and generally applicable parts in the bullet point formulations. Importance and applicability do not always line up, of course. For example, there are very important considerations regarding evaluative and normative aspects of concepts in relation to the similarity criterion (9.3.2.1), but since there is no generally applicable instruction to gain from it, I have left it out of my bullet point formulation. Nevertheless, I think that a responsible explicator should take the discussed questions of value and normativity into consideration.

We are now in a position to compare the internal structure of my proposed account with the structure of Carnap’s and Brun’s accounts, discussed in 4.4. The internal structure of the criteria on my account has a two-part structure consisting of defining criteria and good-making criteria. In my broad use of ‘explication’,

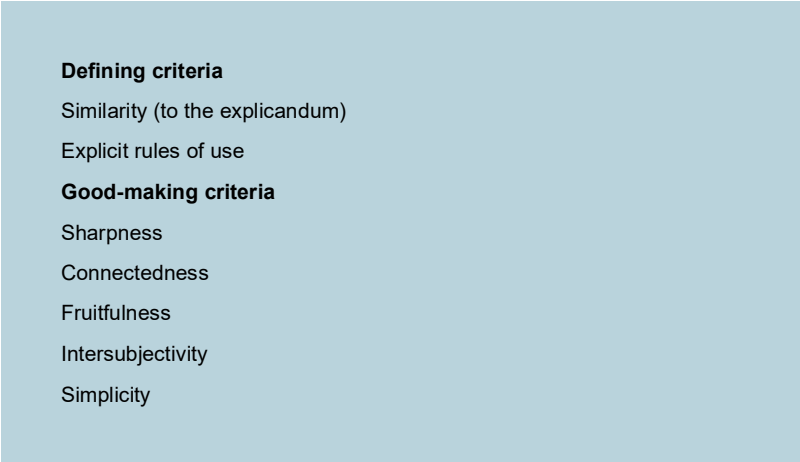


Fig. 4. The structure of my modification of Carnap’s account

To accommodate my broader characterization of explication (1.3), I have split the criteria into two kinds. To give an explication is to satisfy the first two, and two formulate the rules to in order to improve the concept for the purposes of inquiry. Hence, the different versions in chapter 5 count as explication. My proposal is that good concepts for inquiry are those that satisfy the five good-

⁷⁴ I understand Carnap’s criteria and discussions of them in the same way. There is no sharp line between what is in the “official” numbered formulation of them (1–4) and the discussions on the side.

making criteria above. Depending on the purposes of specific tasks of explication, the criteria may be specified in different ways and assigned relative weights in accordance with the purposes. In contrast to the structure of Brun's account, I think that all criteria are general in the sense that they should be assigned some positive weight regardless of the particularities of a project of explication.

Although the criteria I propose are more relaxed than Carnap's criteria they are similar to Carnap's criteria in that they are also intended to be generally applicable across disciplines and contexts of inquiry. The reason to relax his Carnap's criteria should be relaxed because there are important contexts of inquiry for which they are not suitable, namely contexts of inquiry outside formal language construction, logico-mathematical inquiry, and parts of natural and empirical science (for a discussion, see chapter 8).

As I have clarified above, my account is heavily based on both Carnap's criteria and Brun's proposal for a relaxed version of explication. My criteria are also based on and inspired by some of the criteria for concept formation in the social sciences and philosophical definitions of social concepts, proposed and advocated by e.g., Gerring (5.7) and Tengland (5.8). I am also using Ludlow's distinction between explicifying and sharpening word meanings (see 2.7.2).

As mentioned, I have divided the criteria into two kinds. They can also be divided based on the way in which they are modifications of Carnap's account. Thus, the criteria I am proposing can be divided into the following three kinds:

- Criteria that are relaxed versions of Carnap's criteria.
- Criteria that are new in the sense that I take some of Carnap's presuppositions, and make them into explicit, independent criteria, to be traded against the other criteria (rather than taking them to be assumed requirements on meaningful or scientific discourse).
- Criteria that are new in the sense that they have not been proposed by either Carnap or previously in the literature on Carnapian explication but have been discussed elsewhere in relation to conceptual analysis and concept formation.

9.4 Defining criteria

9.4.1 Similarity

I endorse Carnap's similarity criterion almost as it is, with one small modification and one substantial addition. The small modification is that in my formulation of the criterion, I mainly try to make it clearer what the point of the criterion is, and how it makes explication different from other conceptual tasks. In Carnap's discussion of similarity, he remarks that "close similarity is not required, and considerable differences are permitted." However, Carnap's formulation is not strong enough to reflect his own views as he expresses them elsewhere, viz. the entirely reasonable view that close similarity is not desired and differences are not only permitted but *necessary* for the task to succeed. Gustafsson (2006, 61; 2007, 43) emphasizes this point. In line with Carnap's views, I propose to change the formulation of the criterion so that it is made clear that due to the nature of explication, some differences are necessary and considerable differences are permitted when justified by the improvements.

Hence, my formulation of the criterion is:

- The explicatum should be similar to the explicandum in such a way that it is useful for similar purposes as the explicandum. Differences are required and considerable differences are permitted.

A few further comments about evaluative concepts are in order. In a Carnapian explication of an evaluative concept, the explicandum would be replaced by a descriptive explicatum. Felix Oppenheim has pursued this task of "reconstructing" or giving "explicative definitions" of political concepts in value-neutral terms (5.5). Since the goal is not conceptual analysis, we are not required to preserve the evaluative status (if there is one) of the explicandum. However, we could, in principle, also replace a value-neutral concept with an evaluative concept. And we could replace a value-neutral explicandum with a value-neutral explicatum or replace an evaluative explicandum with an evaluative explicatum. Depending on our needs, the evaluative status of a concept can be removed, preserved or even introduced.

In some cases it is not clear whether the explicandum is evaluative or merely appears to be an evaluative concept. If the clarification of the explicandum reveals that it merely appears to be evaluative or normative or that is misleadingly used in such a way, then we presumably want to construct a value-neutral or descriptive explicatum. (Although one would not be required

to do so.) Some evaluative concepts seem to be irreducibly evaluative,⁷⁵ and in some cases the purposes of inquiry require that an adequate explicatum is given evaluative features. For example, Tengland (2007) appeals to a ‘value criterion’ in his definition of the concept of health. He claims that a “definition should capture the positive value that the concept of health has” (Tengland 2007, 258). Arguably, a systematic inquiry involving the concept of health needs to keep the evaluative aspect of it, i.e., its positive prudential value, not only in order to satisfy the similarity criterion but also in order to clarify its relation to rights and responsibilities. If the explicandum is evaluative and the evaluative aspect of it is required for it to play the desired role, then the explicatum should be evaluative too. The task of accommodating irreducibly evaluative or normative concepts is a part of the method of explication that needs to be developed.

Presumably, if the normative role is central for why we care about a concept, then an explication of it requires that the normative value is preserved. Nothing in Carnap’s explicit formulation of his similarity criterion prevents the preservation of evaluative features of the explicandum in the explicatum. Naturally, however, this was not one of his concerns, given his non-cognitivist view regarding evaluative language.⁷⁶ Since it is compatible with the letter of Carnap’s criterion, I find no need to propose any changes to the similarity criterion to accommodate evaluative and normative features. As the criterion is formulated by Carnap, one is free to include such features in considerations of similarity.

9.4.2 Explicitness

My explicitness criterion is the result of splitting Carnap’s exactness criterion into several criteria. For comparison, we have here, once again, Carnap’s exactness criterion:

The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given *in an exact form, so as to introduce the explicatum into a well-connected system of scientific concepts*. (LFP, 7, my emphasis)

⁷⁵ I mean “irreducible” in the sense that in order to serve our needs they need to have evaluative or normative features.

⁷⁶ See e.g., Carnap’s “Intellectual Autobiography” in the Schilpp volume (1963a, 81 ff.) for his later views on the topic.

In my explicitness criterion (see below) I have removed the parts that are in italics in the quotation above, i.e., the requirements that the rules are exact and that the explicatum should be introduced into a well-connected system of scientific concepts. My criterion is simply what is left:

- The explicatum should be provided with explicit rules of use.

9.5 Good-making criteria

Below I present the criteria that make an explicatum good. The higher the extent to which they are fulfilled as a package, the better.

9.5.1 Sharpness

I have relatively little to say about sharpness compared to the other good-making criteria. What counts as a sufficiently sharp concept depends on our purposes. Ludlow (2014, 44) exemplifies with the example of the definition of ‘planet’ (see 1.4.1). The third condition for a celestial body to count as a planet was that it “has cleared the neighbourhood around its orbit”. However, without an (arbitrary) quantitative definition of what counts as having cleared the orbit from debris, the definition is not sharp. In our solar system, however, there is a large gap between objects that “completely dominate their orbital zone” (Soter 2006, 2514) and those that “live amid a swarm of comparable bodies” (ibid.). Therefore, as Ludlow points out, the “definition of ‘planet’ narrowed the meaning by excluding cases in our solar system, but it did not sharpen it because there was no real attempt to sharpen the notion of a clear orbit” (Ludlow 2014, 44). Hence, the definition is sharp enough for our solar system but not for other solar systems where orbits are in the process of being cleared (ibid.). The general lesson from Ludlow is that:

To sharpen a meaning is to modulate it in a way that avoids borderline cases in that context. Thus a meaning that is sharp in one context with a few borderline cases could fail to be sharp in a context with a lot of borderline cases. (Ludlow 2014, 89)

In short, it seems generally desirable to give sharp rules of use for the explicatum, and how sharp they should be, i.e., how much weight this criterion should be given relative to the others, depends on the context and purpose of

the explication. I do not think that sharpness is more important than the other good-making criteria. I will follow Brun's recommendation that the explicatum should not be less sharp than the explicandum. My bullet point formulation of the sharpness criterion is thus:

- The explicatum should be at least as *sharp* as the explicandum, and preferably sharper.

9.5.2 Connectedness

That the explicatum should be incorporated into “a well-connected system of scientific concepts” is a reasonable requirement when it comes to many explications in the natural sciences and when it comes to the kind of explications that Carnap typically performed, i.e., metalogical and methodological concepts introduced into semantical systems. In other contexts, the requirement is too strong. As previously mentioned, Carnap acknowledges in his RSE that:

The use of symbolic logic and of a constructed language system with explicit syntactical and semantical rules is the most elaborate and most efficient method. *For philosophical explications the use of this method is advisable only in special cases, but not generally.* (RSE, 936, my emphasis)

I think that connectedness should be treated as a matter of degree and that it should be an independent criterion. Further, I think that “system of scientific concept” should be understood in a broad sense of ‘scientific’, as systematic inquiry.

A concept may have more or less sharp rules of use and more or less connections to other more or less relevant and important concepts. Hence, connectedness should be distinguished from both explicitness and sharpness. It should also be distinguished from fruitfulness. The fruitfulness criterion also involves the ability to make connections to established theories, but fruitfulness is a measure of what can be achieved with the concept. Fruitfulness is a measure of what you can do with the concept (e. g., make useful generalizations), while connectedness is a state of the system of concepts to which the explicatum belongs, or—when it is not justified to talk of a system—it is a measure of its amount and quality of relations to other important concepts. If we can define an explicatum as a hypernym of an empirically useful concept, we may not be able to formulate empirical laws with the explicatum but the explicatum is nevertheless connected to hyponyms with

which we can formulate laws. By being more general than its hyponyms the explicatum may bring conceptual and theoretical order and unity and ensure that investigations into its sub-concepts are investigations of the same phenomenon.

Say that we want to give an operationalist explication of a social concept *C* in purely behavioural terms as *C**. The conditions of application for the term may then be sharp, and inductive links between the frequency of *C** and the frequency of other behaviours can be established. But *C** likely lacks connections to important concepts typically associated with *C*. It is not only a cost in similarity to the explicandum, but also a cost in connectedness.

There are other ways of making connections between concepts, for example, by explicitly formulating important consequences of application (in line with Brandom's work on inference tickets). Such consequences may include not only inferences between descriptive statements, e.g., from “*x* is red” to “*x* is coloured”, but also inferences to normative and evaluative statements, e.g., from “*x* got a disease” or “*x* is ill” to “*x* is worse off, *ceteris paribus*”.⁷⁷ Inferences that are relevant include but are not restricted to analytical inferences. There are also stereotypical or probable inferences, such as the inference from “*x* is a dog” to “*x* barks when excited”. The latter example is taken from Foster and Ichikawa (2023), who consider it a “reliable stereotype”, and comment that “[i]n suitable circumstances, this inference ticket will give knowledge” but that “[o]ther inference tickets are less benign—they correspond to stereotypes we have good reason to resist” (Foster and Ichikawa 2023, 7–8). There are also more controversial inferences that are often (at least in certain contexts) implicitly taken to follow. An example may be the inference from “*x* is not ill” to “*x* does not have the right to receive health care”. In an explication of “illness” intended to be useful for health research one may stipulate such an inference as a rule of use for “illness”, if the optimal balance between the similarity criterion and the other criteria allows it. The point is to make explicit inferences which are typically licensed by the use of the explicandum and which may be relevant for one's purposes, and then decide whether or not to include them in the specification of the explicatum.

As mentioned, in many cases sharpness go hand in hand with connectedness (e.g., in Carnap's own explications where he constructs an artificial target language), but they do not always go hand in hand and sometimes they are

⁷⁷ Dominic Murphy writes on the concepts of disease and health that: “There may [...] exist [...] importantly different stresses on the kinds of value judgements that different theorists think are part of disease categories and their application. Typically, the relevant normative claim is taken to apply to the life of the person whose health is under discussion—it is bad for you to be that way” (Murphy 2015).

even in conflict. Operational definitions, for example, are sharp but they are often not sufficiently connected and fruitful. For many explicanda, the most exact rules of use one could provide for the explicatum would be in the form of an operational definition, i.e., to define it in terms of a test operation. But as Hempel (1966) and others after him have pointed out, a requirement to define scientific concepts operationally leads to a proliferation of concepts given the operationalist maxim that “different operational criteria should be regarded as characterizing different concepts” (Hempel 1966, 92). Such a requirement conflicts with the quest for fruitfulness and connectedness. As Hempel puts it, the maxim of operationalism “would defeat one of the principal purposes of science; namely the attainment of a simple, systematically unified account of empirical phenomena.” (Hempel 1966, 92).

On the other hand, the avoidance of operational definitions comes with its downsides, as argued by Elina Vessonen (2021), in her interesting discussion of the relation between Carnapian explication and operationalism as methods in psychology. She points out that:

With all the bad press operationalism has received, a scientist who wants their work respected is better off being vague about what their measure measures instead of declaring operationalist commitments. [...] But vagueness leads to confusion. And [...] there is debate about whether or not psychologists’ bread-and-butter validation methods reliably ensure that measures indeed track non-operational concepts.

To counteract these negative consequences, Vessonen proposes the following form of operationalism:

My concrete proposal, then, is that researchers should openly admit to operationalism – or otherwise come up with an argument to support a richer, test-independent reading of the claims they make. The rhetorical move where operational conceptions magically give rise to non-operational claims does not serve psychology well.” (Vessonen 2021, 10634).

Perhaps such illegitimate claims are worse for the state of a field of inquiry than the proliferation of exact concepts. Vessonen suggests the following remedy:

The least scientists can do, I think, is that when they make scientific claims on the basis of a given operational concept, they clearly index the claim to the relevant measure. For example, a claim about the relation between income and well-being would take the form: “Well-being is associated with income in such

and such a manner, when well-being is constructed in terms of satisfaction of preferences as expressed in questionnaire Q.” (Vessonen 2021, 10635).

For support for such a view of good conceptual conduct, Vessonen refers to Leland Wilkinson (1999), who expresses similar views (while distancing himself from operationalism). Regarding how to name variables, Wilkinson writes that:

the name “IQ test score” is preferable to “intelligence” and “retrospective self-report of childhood sexual abuse” is preferable to “childhood sexual abuse.” Without such precision, ambiguity in defining variables can give a theory an unfortunate resistance to empirical falsification. Being precise does not make us operationalists. It simply means that we try to avoid excessive generalization. (Wilkinson 1999, 596).

The precision prescribed by Vessonen and Wilkinson is intended to be a safeguard against unfounded claims and excessive generalizations. On the other hand, excessive precision in the definitions may lead to concepts that are not useful for us and do not allow us to make connections to other relevant concepts. There seems to be hard cases where trade-offs have to be made between these criteria, which is why I have separated sharpness criterion from the connectedness criterion.

My criterion of connectedness is the following:

- The explicatum should be connected to relevant systems of concepts.

9.5.3 Fruitfulness

I propose to split the fruitfulness criterion into two criteria. There are two reasons for this. First, Carnap’s fruitfulness criterion is too narrowly applicable. Second, it would be an improvement to make Carnap’s empiricist strictures into a criterion for intersubjectivity instead of observability (for non-logical concepts) and an independent and explicitly stated criteria which can be satisfied to different degrees. Generally, it is better to make one’s theoretical presuppositions explicit.

To repeat, Carnap’s criterion is that: “The explicatum is to be a fruitful concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept)” (LFP, 7). The need for a broader notion of fruitfulness has been discussed thoroughly in chapter 8. My criterion is partly

based on Carus and Leitgeb's (2021) interpretation of Carnap's view of fruitfulness, set out in their *Stanford Encyclopedia of Philosophy* entry on Carnap. In their interpretation, a fruitful concept is a concept that can be: "usefully applicable in scientific or philosophical theorizing and discourse, e.g., in the formulation of lawlike statements, or by creating deductive or inductive links to established theories that are themselves sufficiently clear, exact and successful" (Leitgeb and Carus 2021). The quoted description of fruitfulness is presented by Carus and Leitgeb as an interpretation, but I consider it to be a different and broader notion of fruitfulness than Carnap's notion. In their formulation the requirement that the explicatum should be useful for the formulation of universal statements removed and replaced with "lawlike statements", as an example of usefulness. Therefore, it is a substantial relaxation of the criterion. In their formulation, the requirement to facilitate the formulation of either empirical laws or logical theorems has been given up.

I endorse Carus and Leitgeb's characterization of fruitfulness but I consider it to be a new, relaxed fruitfulness criterion. Hence, my formulation of the fruitfulness criterion is:

- The explicatum should be fruitful in the sense that it is conducive for the formulation of generalizations and lawlike statements, or for the creation of links to established theories.

9.5.4 Intersubjectivity

Carnap embeds his empiricist views in the fruitfulness criterion by assuming that any legitimate explicatum is either a logical concept or an empirical concept, since they are required to facilitate the formulation of "universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept)" (LFP, 7).

I propose adding an explicit criterion of intersubjectivity to be balanced against the others. My intersubjectivity criterion is the result of removing Carnap's empiricist restriction on the fruitfulness criterion and making it into an independent criterion to be traded against the other criteria. With such a criterion, the ideal of intersubjectivity is preserved while not being a restriction on explication. With such a criterion, an aspect of the empiricist ideal of intersubjectivity is preserved, without the absolute requirement of observability of explications of non-logical concepts. A low degree of intersubjectivity may be traded for high scores on the other criteria.

The criterion is inspired by the reliability criterion proposed by Tengel (2007), who appeals to a reliability criterion in his definition of health.⁷⁸ According to my criterion, the explicatum is to be intersubjective in the sense that it should be easy to judge if the concept is applicable or not, so that different users will use the concept in the same way and so that a single user will use it in the same way over time.

With this criterion I try to make explicit the empiricism already assumed in Carnap's criterion of fruitfulness, and relax it. I consider observability to be one among several ways in which a concept might be intersubjective, and I will consider intersubjectivity to come in degrees. Hence, the empiricist strictures presupposed by Carnap's account are removed. Observable concepts as well as logical and mathematical concepts are highly intersubjective; however, there are plenty of concepts which are neither. A concept such as *culture* cannot plausibly be replaced either by observable or by logical or mathematical concepts, but it is still a candidate for explication. Kroeber and Kluckhohn (1952) identified 164 definitions of culture, and Gerring and Barresi (2009) describe "culture" as "a term which has plagued the social sciences for over a century" (Gerring and Barresi 2009, 242). Nevertheless the role played by the concept seems to be indispensable. As summarized by Mark Risjord: "Humans recognize and name group-level differences, and such differences can be very important aspects of the way people interact. Understanding human behavior thus requires something like the culture concept" (Risjord 2012, 398).

On this account, observability or measurability are not requirements on an explicatum but the lack of these features arguably renders an explicatum less intersubjective and hence one needs to make up for it by higher satisfaction of other criteria.

Why is intersubjectivity not already covered by the exactness criterion? Because my exactness criterion only requires explicit rules of use for the explicatum, and explicit rules of use do not always ensure that the concept is applied in the same circumstances by different users, or that the same entailments of application are recognized by different users. To illustrate, let us consider Sally Haslanger's famous ameliorative definition of *woman* (and *man* and *gender*), which I presented in 7.4. Haslanger proposes to redefine the concept *woman* in the following way:

⁷⁸ Reliability is here to be understood in the sense that it is used in statistics and social science, where the measurement of a phenomenon is judged for its validity and reliability.

S is a *woman* iff_{df} S is systematically subordinated along some dimension (economic, political, legal, social, etc.), and S is “marked” as a target for this treatment by observed or imagined bodily features presumed to be evidence of a female’s biological role in reproduction. (Haslanger 2000, 230)

Haslanger’s definition may be considered a successful project of conceptual engineering, as argued, e.g., by Pinder (2022a), for the purpose of “understanding ourselves and those around us as deeply molded by injustice and to draw the appropriate prescriptive inference” (Haslanger 2000, 13; Pinder 2022a, 334). However, as an explication, with my criteria, it would score low on intersubjectivity despite being given as an explicit definition. Presumably, the exactness criterion is fulfilled by the definition. But there are concepts in the definiens which are highly contested, such as *social subordination*. It goes against the intersubjectivity criterion to define an explicatum in terms with regard to which agreement in usage is less likely than in the case of the explicandum.

Here is my formulation of the intersubjectivity criterion:

- The explicatum is to be intersubjective, in the sense that it should be easy to judge if the concept is applicable or not, so that different users will use the concept in the same way and so that a single user would use it in the same way over time.

9.5.5 Simplicity

I propose two modifications to Carnap’s criterion in which he states that “The explicatum should be as simple as possible; this means as simple as the more important requirements (1), (2), and (3) permit” (LFP, 7).

I propose to include more than the simplicity of the rules of use for the explicatum or of the laws connecting the explicatum to other concepts. I propose that included in the criterion should be the substantial feature of dimensionality, or homogeneity. As defined by Hedden and Nebel (2024), “A concept F is multidimensional just in case whether and to what extent something is F depends on how it stands along multiple underlying dimensions, or respects, of F-ness” (Hedden and Nebel 2024, 265). Here are two of the examples they give, namely of the concepts of biodiversity and inequality:

Biodiversity is widely held to be multidimensional, depending not only on the number of species present in the ecosystem, but also their phylogenetic and morphological diversity [...]. Inequality is multidimensional, depending on inequality of income, inequality of resources, and inequality of opportunity, among other things, each of which may have multiple aspects [...]. (Hedden and Nebel 2024, 267)

Applied to my discussion of the simplicity criterion, we can say that a concept is simpler the fewer dimensions it has. For example, Tengland proposes a definition of health in terms of both ability and well-being. In his evaluation of his own definition (or theory, as he phrases it), Tengland concludes that his definition fares well on other criteria but that its multidimensionality is a weakness.

The major problem for the theory is homogeneity. Not only are there two dimensions, ability and well-being, but ability is not a homogenous category, since it consists of different kinds of abilities (intentional and non-intentional), dispositions, motivation, and some acquired mental states (mainly correct perception and self-confidence). How bad this is for the theory others will have to judge. (Tengland 2007, 278)

Since dimensionality does not seem to be merely a matter of convenience, my simplicity criterion is raised to the same level of importance as the other criteria. While simplicity in the form of homogeneity as well as in other forms may be traded for gains in the other criteria, it should not merely be a secondary feature to help us choose between otherwise equally good explicata.

Here is my formulation of the simplicity criterion:

- The explicatum should be simple. Explicata with fewer dimensions are simpler than explicata with more dimensions.

9.6 How to evaluate the standards of evaluation

Before I end this chapter and move on to a concrete example in chapter 10, where I explicate the concept *democratic*, I am first going to discuss what it would take for me to make the case that my set of criteria is superior to Carnap's and others' sets of criteria. By giving an explication of *democratic* using the criteria that I proposed in this chapter, I seek to achieve not only a good explicatum, but also, and mainly, seek to achieve something else: to make

the case that the criteria I have proposed are up to the task, and that they are better than alternative sets of criteria. How does one make the case for a set of criteria of adequacy? In that task, we are two steps removed from actually using concepts for inquiry. At the ground level, we are using concepts for the purposes of inquiry, e.g., for making a prediction. One level up, we are evaluating our concepts and trying to improve them through explications. Two levels up, we are debating which criteria we should adhere to when performing explications. In other words, we are evaluating different standards of concept evaluation. For a conclusive argument for the superiority of a set of criteria, one would have to show that for any explicandum, one would get a better result using that set of criteria than any other set of criteria. Such a demonstration is not feasible. Furthermore, the criteria are by design vague enough to leave plenty of room for judgement. As Sjögren (2011) points out regarding Carnap's criteria:

The criteria that an explicatum ought to satisfy are (deliberately?) vague. This has the advantage that there are no (or few) formal obstacles to the process of explication. The scientist or philosopher can concentrate on the content and not on whether he is formally doing the right thing. (Sjögren 2011, 19)

The criteria are merely guiding the creative, inventive task of coming up with good concepts. Hence, I cannot conclusively demonstrate the superiority of my criteria over Carnap's by showing that for a given explicandum, my criteria would yield a better explicatum than Carnap's criteria would yield. Nevertheless, even with deliberately vague criteria it surely matters *which* deliberately vague criteria we are using. We need to evaluate them somehow. What we can do, and what I will do in the next chapter, is to show with an example the features of concepts that are emphasized by my criteria, and to argue that they bring something to the table that is lacking in Carnap's criteria.

9.7 Concluding remarks

I have now presented and motivated my proposed modified criteria. In summary, the criteria I propose differs from Carnap's in that I have:

- Split the exactness criterion into three separate criteria: explicitness, sharpness and connectedness.
- Broadened the fruitfulness criterion, in line with many recent accounts of explication

- Made intersubjectivity into an independent and explicit criterion regarded as a matter of degree (in contrast to the empiricist strictures assumed in Carnap's fruitfulness criterion).
- Expanded the simplicity criterion to include dimensionality as a measure of simplicity, and raised simplicity to the same level of priority as the other criteria.
- Added considerations about the normative and evaluative features of concept (although these considerations are not a part of the bullet point formulations of my criteria).

I think that there are reasons to retain a set of general criteria of adequacy. However, our conceptual needs can vary a lot from one context of inquiry to another. To accommodate that we need flexible criteria, where the degree and balance may shift greatly from case to case. This raises the worry that my generalist account collapses into contextualism. What is the difference between contextualism and my version, which we may call generalism, where there is a list of rather vague criteria, to be interpreted and balanced differently from case to case? I believe that even with very flexible criteria, there is a benefit in having common and explicitly stated criteria to hold ourselves and each other accountable to. There is a point in having a shared framework for conceptual improvement, and a list of criteria to go through both when we are constructing concepts and when we are evaluating our own and others' explications.

In chapter 10 I give a concrete example of how to explicate a concept using my criteria. The aim is to show how the criteria may be used and to shed further light on the motivations for my modifications of Carnap's criteria.

10 Evaluating my proposal with an example

10.1 Overview

The main task of this chapter is to evaluate my criteria with a concrete example, in the form of an explication of the concept *democratic*, when it is used as a property of decision procedures. While the first aim of the chapter is to provide a good explication of *democratic*, the second aim of the chapter is to use that explication to illustrate and argue for the benefits of my criteria. I will first argue for the benefits of the explication itself, and then I evaluate how my explication illustrates the benefits of my proposed criteria. The main way to do this is by pointing to shortcomings of alternative accounts.

10.2 Explicating the concept *democratic*

I am explicating the concept *democratic*, expressed by the adjective ‘democratic’. I will explicate the qualitative concept *democratic*, as it is used in sentences such as “*x* is democratic” (where *x* ranges over decision methods). The explicatum I suggest should take the place of the explicandum is the comparative concept *more democratic than*, as it is used in sentences such as “*x* is more democratic than *y*” (where *x* and *y* range over decision methods). I will mostly talk about explicating the concept, but at times when it is more natural I will talk about explicating the word or term used to express the concept. In particular, I will focus on words rather than concepts when I clarify the explicandum and when I discuss the views of other philosophers, mainly Capelen, who refer to the words ‘democratic’ and ‘democracy’.

The explication is based on Thomas Christiano's definition of the term 'democracy' (Christiano and Bajaj 2022),⁷⁹ but with significant modifications to be explained below. I have adopted from Felix Oppenheim (1971) and from Herman Cappelen (2023) the idea that the property concept *democratic* is more fundamental than the concept *democracy*. In my explication, therefore, *democratic* is the primary concept and *democracy* is derived from it. However, my proposal to explicate *democratic* is made in response to Cappelen's recent *The Concept of Democracy: An Essay on Conceptual Amelioration and Abandonment* (2023). In the book, Cappelen argues that we should abandon the words 'democracy' and 'democratic' and the concepts they express: *democracy* and *democratic* (Cappelen 2023, 3). Cappelen refers to these words as the "D-words". He raises problems with the use of the D-words in both public and academic discourse and he argues that those problems are so severe that the best remedy is to abandon them altogether. In this chapter I am drawing heavily from Cappelen's work on the D-words.

Cappelen explicitly rejects the strategy of amelioration, e.g., in the form of explication. He discusses Christiano's definition as the most promising candidate for amelioration but ultimately rejects it. I accept some of the problems raised by Cappelen as serious challenges. However, I argue that my explication based on Christiano's definitions meets the worries that Cappelen gives as reason for abandonment of the D-words. I argue that my explicatum has the potential to provide a conceptual basis for democratic theory and potentially facilitate further conceptual unification in political philosophy and political science. I defend my explication by arguing that it satisfies the criteria which I proposed in chapter 9. Before I give my explication I address different attitudes or strategies to take toward contested or defective concepts, and I will argue for amelioration as an appropriate strategy regarding the D-words.

⁷⁹ The definition was given in 2008 in a previous version of the *SEP* entry on "Democracy" with Christiano as the sole author. I will refer to the revised entry written by Christiano and Sameer Bajaj (2022), but in order to facilitate engagement with Cappelen's discussion of the 2008 version, in the main text I refer to the definition as "Christiano's definition. There would be a risk of confusion if the same definition was sometimes referred to as Christiano's definition and sometimes as Christiano and Bajaj's definition.

10.3 Is explication the best strategy to deal with the concept *democratic*?

While I accept and build on many of the insights in Cappelen’s book, I reject the main case that the D-words should be abandoned. In the introduction, Cappelen outlines four ways in which a theorist may respond to a contested concept, such as democracy (Cappelen 2023, 11). There are:

- Libertarians
- Ameliorators/explicators
- Cynics
- Abolitionists

These are strategies that political theorists may adopt with regard to contested concepts. **Libertarians** refuse to engage in meaning-policing and say “let a thousand definitional flowers bloom” (ibid.). **Ameliorators** try to improve language, typically with an admiring eye toward conceptual conduct in natural science. **Cynics** think that political language is so corrupt and empty that there is no point in trying to fix it. **Abolitionists**, finally, think that we should abandon certain concepts (typically high-level abstract and vague concepts) and instead try to anchor political discourse in “careful descriptions, analyses, and assessments of particular policies, structures, and actions” (ibid., 14).⁸⁰

⁸⁰ As Cappelen points out, not all cases of conceptual abandonment are philosophically interesting. For example, we abandon expressions when we for some reason lose interest in and forget the thing the expression is about:

“Archibald Stansted Hall” is the name of a man who died 248 years ago. He lived an unremarkable life. He was a butcher who lived in London and died from a heart attack on June 5, 1773. He left behind two children and his wife. It’s probably been more than two hundred years ago since anyone has used “Archibald Stansted Hall”. The reason for that is entirely mundane: interest in the man was gradually reduced, eventually to a point at which no one remembered him. No one wanted to talk about Archibald. Since he was of no concern, his name was abandoned. (Cappelen 2023, 18)

Cappelen mentions this as the most obvious and trivial way in which expressions are abandoned, to be contrasted with philosophically interesting ways in which expressions are abandoned in both ordinary speech and scientific language. However, while the example of Archibald used by Cappelen is amusing, I think it is not a good example of a trivial instance of a phenomenon with more interesting instances. I consider it an example of a categorically different phenomenon than normative abandonment. Abandonment, in the philosophically interesting sense, requires some kind of normative reason for abandonment, not reasons having to do with lack of interest or oblivion.

Among these strategies, Cappelen advocates abolitionism regarding ‘democracy’. I will advocate amelioration in the form of explication. To defend my choice I will give reasons to reject the other strategies, beginning with cynicism. I agree with Cappelen’s verdict that it is defeatist and hyperbolic:

Surely, not all political speech is empty of clear thought, even though some salient examples are. If you’re a theorist, a public intellectual, a serious journalist, or an honest politician, you don’t want to self-identify as a bullshitter whose aim is simply to spew empty rhetoric. You’ll want to do better. (Cappelen 2023, 13)

Next, we have libertarianism, which is a position widely held among political theorists and academics, prominently defended by W. B. Gallie (1956). I also agree with Cappelen that libertarianism should be rejected and I think that Cappelen successfully dismantles Gallie’s case (Cappelen 2023, 201–9). Furthermore, Cappelen observes that “Even though Gallie’s view has been very influential and is widely accepted by academics, the implications for real-world politics are overlooked” (Cappelen 2023, 11). In his vivid formulation, if everyone, is free to use ‘democracy’ with different meanings, “It makes public discourse involving ‘democracy’ a depressing and never-ending orgy of verbal dispute” (Cappelen 2023, 12).

Then we come to my preferred option. Why not try to ameliorate the concept of democracy? Cappelen entertains the idea:

the ameliorator tries to improve “democracy” by giving it a clear, appropriate, and fixed meaning. It’s hard to see how to achieve this, but one strategy can be modeled on what often happens in the sciences. When scientists (e.g., in physics, medicine, or economics) use an expression with a determined meaning, it has that meaning independently of whether it is used in a messy way by the rest of the linguistic community. [...] For example, what “weight” means in physics is fixed in a clear and precise way, independently of how that term is used by ordinary speakers of English (where “mass” and “weight” are often used interchangeably). What “arthritis” means in English depends on how medical experts use that term, independently of whatever confused views ordinary speakers have about arthritis. (Cappelen 2023, 12)

What, then, is the problem with this strategy? In Cappelen’s summary, his case against amelioration of ‘democracy’ is the following:

In theory, amelioration is a promising strategy, but the problem is that the way “democracy” is tied to norms, values, and practical policies makes it very different from “weight” and “arthritis”. *Since there’s no convergence on the*

relevant norms, values, and practical policies, we're simply not going to get a convergence on the meaning of "democracy" (in the way we have for "weight" and "arthritis"). Even among political scientists, there are deep and irresolvable disagreements about how to define "democracy". We find everything from Schumpeterian minimalists (where the only thing that matters is an occasional election) to quantitative maximalists (where a broad range of cultural, normative and institutional factors are quantified as part of the various democracy indices.) (Cappelen 2023, 12-13, my emphasis)

In summary, the two main arguments of Cappelen's case against ameliorating "democracy" are:

- (1) In contrast to the situation in developed sciences, there is no consensus among the putative experts about how to define 'democracy'.
- (2) Since people (including political theorists and other candidate experts on democracy) do not converge on norms, values, and practical policy, people (including political theorists and other candidate experts on democracy) will not converge on terms such as 'democracy' that are tied to norms, values, and practical policy.

With regard to (1), I agree but I think that in principle we could get to a situation where there is consensus among experts. With regard to (2), I think we can explicate "democratic" in such a way so that theorists who do not converge on norms, values and practical policy nevertheless can converge on how to use the term 'democratic'.

Among political theorists, Felix Oppenheim is a pioneer in the explicationist or reconstructionist strand of the ameliorative strategy.⁸¹ In his explication of the concept of freedom, Oppenheim (1961) considers the option of abandoning it. In contrast to Cappelen's proposal to abandon 'democracy', Oppenheim advises against abandonment of 'freedom' and other terms for political concepts:

Naturally, the more inflated the meaning of a word in everyday usage, the more difficult it is to adapt it to the requirements of a scientifically useful conceptual scheme. It might, therefore, be asked why we do not simply give up the word 'freedom'. After all, natural scientists did not hesitate to abandon words like 'phlogiston' and 'epicycle', when the theories in which these terms functioned were abandoned. 'Freedom' seems just as impractical as a scientific term, not

⁸¹ Although I have mentioned Felix Oppenheim previously in the chapter and elsewhere in the thesis, I will frequently use his full name to distinguish him from his father Paul Oppenheim.

because freedom does not exist, but because there are too many phenomena labeled with this word. However, the concept of freedom remains with us a result of a long tradition of political thought and action of which we are inescapably the heirs. (Oppenheim 1961, 7)

Oppenheim does not consider abandonment to be an option. Instead he proposes amelioration in the form of explication:

We want to translate what others have had to say about freedom into a more systematic and more precise language. And so we have no choice but to keep the word and to reinterpret it. For better or for worse, the vocabulary of political science is bound to remain to a large extent prescientific or nonscientific. To make such words do the work of valid concepts in an empirical science of politics, it becomes all the more necessary to subject them to the purifying process of an adequate explication. (Oppenheim 1961, 7)

I follow Oppenheim in his optimism about the ameliorative strategy, but I reject his operationalism.⁸²

Finally, I want to highlight an observation by Cappelen that he takes to be in support of the strategy of abandonment. I think that the same observation should be considered as an argument for the importance of general criteria in ameliorative explications. Researchers should not only try to define their concepts in useful ways, but they should also try to figure out together what criteria an adequate definition should satisfy. If there was an expectation that conceptual investigations were preceded by a discussion of criteria of adequacy and followed by an evaluation, perhaps the situation described below by Cappelen would occur less often:

it's striking how two of the world's leading theorists of democracy—Christiano and Estlund—take such radically different approaches to the definitional challenge. Christiano wants to respect both generality and neutrality [see below]. Estlund does not. That raises the concern that they are talking past one

⁸² Also, I don't think it should be a criterion of success for an explication that one can translate what others have said about the concept into the new language. Explication does not require translation. As Carnap remarks: "The former concept has been succeeded by the latter in this sense: the former is no longer necessary in scientific talk; most of what previously was said with the former can now be said with the help of the latter (though often in a different form, not by simple replacement)" (LFP, 6). I accept Carnap's view with one modification: I do not think that one should have to be able to say most of what previously was said using the explicandum, and not even that one should be able to say the most important things previously said. I find it sufficient if *many* of the most important things previously said can be said with the help of the explicatum.

another—they are not even talking about the same thing. They are using the word “democracy,” but even these two speakers—at the highest intellectual level—fail to communicate. (Cappelen 2023, 166)

I think that there is a remedy to the situation described by Cappelen. The remedy is to carefully clarify the intended uses of the explicandum, and to explicitly state which criteria an explicatum is expected to satisfy (beyond sufficiently performing the role of the explicandum). One may discover that there is not a single concept which can perform all roles we expect our explicanda to play, but that is not itself a reason to abandon the concept.

10.4 First step: clarifying the explicandum

10.4.1 From noun to gradable adjective

As mentioned, my explication is based on Christiano’s definition of the term ‘democracy’ but with significant changes to be explained in this section. Here is his definition, put forth in his and Sameer Bajaj’s entry on “Democracy” in the *Stanford Encyclopedia of Philosophy*.⁸³

The term “democracy,” as we will use it in this article, refers very generally to a method of collective decision making characterized by a kind of equality among the participants at an essential stage of the collective decision making. (Christiano and Bajaj 2022)

His definition captures many important features of the ordinary concept I want to explicate, including generality and normative neutrality. In this regard I agree with Cappelen, who also notes that Christiano is one the few who emphasizes the normative neutrality of the term ‘democracy’:

One of the few texts that explicitly brings this up is Christiano’s [2008] entry on democracy in the *Stanford Encyclopedia of Philosophy*, and Christiano takes it as obvious that the term should be defined in a normatively neutral way. There’s an influential trend in the literature that denies this, however. Work on

⁸³ To repeat the point I made earlier: the definition was given in 2008 in a previous version of the *SEP* entry on “Democracy” with Christiano as the sole author. I refer to the revised entry written by Christiano and Sameer Bajaj (2022), but in the main text I will refer to the definition as “Christiano’s definition” to facilitate engagement with Cappelen’s discussion of the 2008 version.

essentially contested concepts, beginning with Gallie [1956] has held that democracy is such a concept “par excellence” and that such concepts are normative. (Cappelen 2023, 89, n. 11)

My first change of the definition is that I go from noun to adjective, and then from adjective to property concept, i.e., from term to concept. For my explication, the explicandum-term is changed from ‘democracy’ to ‘democratic’ (and as mentioned I take the property concept expressed by that term as the actual explicandum).

A second change has to do with intended scope. Christiano’s definition is modestly put forth as a stipulative definition of ‘democracy’ for the context of the Encyclopedia entry. In contrast, I will not merely give a definition of the term ‘democratic’ but pursue an explication of the concept *democratic*. In other words, I will not merely define the definiendum, i.e., the explicandum-term, but I will replace it with another definiendum, the explicatum-term, and give a definition for the new definiendum. I put forth my explicatum as a concept to be used generally, in investigations and whenever careful use of language is important. Although it would be hard to use my explication in many contexts where the D-words are used, I think that concepts closer to extant uses can be derived from my explicatum.

In the rest of this section, I explain why I go from noun to adjective, and from adjective to property concept. As I mentioned in 10.2, I have adopted from Cappelen the idea that *democratic* is more fundamental than *democracy*. Cappelen argues that: “the adjective is more important than the noun—that is [...] *being democratic* is more fundamental than *being a democracy*” (Cappelen 2023, 74). Even more strongly, he claims that “a theory of democracy is incomplete unless it is accompanied by an account of what it is to be democratic” (Cappelen 2023, 74). He goes on to admit, and indeed stress, that it is a surprising view:

I’ve found it extraordinarily hard to get people to take this idea seriously. It’s a point that’s been largely overlooked in earlier work on democracy. [...] I hope that readers will be willing to at least consider the possibility that understanding the adjective is a fundamental issue in democratic theory (even though “understanding the adjective” sounds pedestrian compared to the typically lofty goals of democratic theory). (Ibid.)

However, there is at least one predecessor, Felix Oppenheim (1971, 1981), not mentioned by Cappelen, who has argued for a very similar view. Oppenheim writes regarding the concept of democracy, that:

Democracy [...] is a property concept, i.e., the word ‘democracy’ stands for a property—namely, one that can be attributed to human organizations like trade unions, universities, or large-scale political organizations. To define the concept of democracy is to say what property it is that we attribute to an organization when we call it democratic. Not ‘democracy’ but ‘ x is democratic’ (or ‘ x is a democracy’) is the expression to be introduced into the language of political science (and later to be defined). Once ‘ x is democratic’ has been defined, we can say that the noun ‘democracy’ refers to the property x has when x is democratic. (Oppenheim 1981, 5)

Perhaps it should also be specified what the relation is between “ x is democratic” and “ x is a democracy”. I think that the difference is the following: in “ x is democratic”, the variable x can range over decision procedures as well as over groups or organizations⁸⁴ (including large political organizations such as states) while in “ x is a democracy” the variable x ranges only over countries and (arguably) other groups or organizations. Further, I think that “ x is democratic” applied to decision procedures is more basic than “ x is democratic” applied to groups or organizations. When x ranges over groups or organizations, “ x is democratic” is true when the group or organization makes (certain) decisions through a process that is democratic in the more basic sense.

Hence, as the explicandum I take *democratic* as a property which may be attributed to decision methods and *not* as a property applicable to groups, governments, countries, etc. However, in loose talk we could truly say of a group g that “ g is democratic”, but it is so in virtue of the decision methods used to make certain collective decisions in the group. As we will see, it would even be loose talk to use a qualitative concept of being democratic for a collective decision method. It should be understood as a colloquial way of saying that a collective decision method is more democratic than a contextually specified threshold. Below I will show how *democracy* can be defined in terms of democratic decision-methods.

Finally, I also agree with Cappelen regarding the uncontroversial view that ‘democratic’ is a gradable adjective. By specifying the explicandum as an adjective instead of the count noun ‘democracy’, it becomes clear that many of the most influential definitions of democracy are inadequate, because ‘democracy’ needs to be defined in terms of a gradable notion of being democratic. Since the explicandum is the gradable adjective ‘democratic’ it can naturally be used in comparative phrases such as “ x is more democratic than y ”. Hence we can easily construct the explicatum as a comparative concept, in line with Carnap’s recommendations.

⁸⁴ These two uses should be separated, and they should be explicated as different concepts.

Since most theorists of democracy focus on the noun term ‘democracy’, I will discuss extant definitions of the term “democracy” and of the concept *democracy*, but with eyes on the goal of explicating the concept *democratic*.

10.4.2 Generality requirement

Related to Cappelen’s prioritization of the adjective over the noun (or, rather, the nouns since there is both a mass noun and a count noun) is his generality requirement, which involves the rejection of what Cappelen calls the “Politics-First” approach. Here is the generality requirement:

The generality requirement: many kinds of groups, of varied sizes and varied purposes, can be democracies. Many kinds of decisions, in different settings, made by different kinds, for different purposes, can be democratic (Cappelen 2023, 85).

As pointed out by Cappelen, most theorists of democracy have been specifically interested in political democracy: “*Almost* all [work on democracy in political philosophy and political science—both theoretical and empirical] treats ‘democracy’ exclusively as meaning ‘political democracy’” (Cappelen 2023, 85). Hence, these theorists have ignored the uses of ‘democratic’ where it is not a property attributed to states and political organizations. I agree with Cappelen’s insistence that a definition of democracy needs to be general and not limited to the political domain. He emphasizes that “[t]he point is fundamental: don’t try to understand ‘democracy’ and ‘democratic’ as having new meanings when applied to a domain labeled ‘political’” (Cappelen 2023, 85). Cappelen mentions and commends Christiano as an exception among influential theorists of democracy in that Christiano does not restrict his definition to the political domain.

The morale to take away regarding ‘democracy’, I think, is not that a new theoretical concept cannot deviate heavily from ordinary usage, but that the explicandum should not be ‘political democracy’, but a more general concept.

10.4.3 Normative neutrality

Another important observation made by Cappelen, with reference to Christiano (2008), is that the D-words are normatively neutral. ‘Democracy’ and ‘democratic’ are often used to express approval, but those aspects of common usage should not be a part of the explicandum (*pace* Gallie (1956, 184), who

calls it “the appraisive political concept par excellence”). Christiano correctly emphasizes that “democracy” should be defined without normative weight so that it is compatible with the idea “that it is not desirable to have democracy in some particular contexts” (Christiano and Bajaj 2022). Thus the explicandum should be ‘democratic’ when used in a sense where it is not a contradiction to say in some contexts where we typically want expert decisions that is desirable that a collective decision method is undemocratic. As Cappelen points out, we do not want to explain disagreement between, e.g., defenders of epistocracy and defenders of democracy as due to the former’s inability to understand the words ‘democracy’ or ‘democratic’. Hence, Cappelen gives the following two principles:

Principle of lexical neutrality: The meanings of expressions ‘democratic’ and ‘democracy’ are normatively neutral—their meanings don’t encode a normative assessment.

Contextual Normativity: In some contexts, utterances of the expressions ‘democracy’ and ‘democratic’ carry significant normative weight—they are used to express assessments.

The point is that while these words are used to express approval or disapproval, the approval or disapproval is not part of the meaning of the expressions. It is only in certain kinds of settings that, e.g., the sentence “The decision was made in an undemocratic way” is used to express disapproval. It may be used by the same speaker to express disapproval regarding a government decision but not regarding decisions about how to pilot a plane or commit surgery, or, for some speakers, how to run a company.

I endorse the principle of lexical neutrality, but I reformulate it in terms of concepts instead of expressions: the concepts *democratic* and *democracy* are normatively neutral. The explicandum is the concept *democratic*, when it is used in the normatively neutral sense, although it has normative weight in some settings.

10.4.4 Logical form and domain

As I have already mentioned, the logical form of the explicandum should be clarified. When “democratic” occurs in a sentence such as “ x is democratic”, it should be specified what domain the variable x ranges over. The explicandum I am looking to replace applies to x when x is a decision procedure. When the variable in sentences such as “ x is democratic” ranges

over groups or countries, the adjective ‘democratic’ applies (if the sentence is true) to x in virtue of how certain collective decisions are made in the group or country. Hence, the basic concept of being democratic is one that ranges over decision procedures.

10.4.5 Why does it matter how the D-words are used in ordinary language?

The similarity criterion, which I have not modified in any substantive way in chapter 9, allows for a high degree of stipulative freedom. It comes, however, with the threat of conceptual proliferation and the threat of conceptual confusion and verbal disputes. For a scientific concept like *gene*, there is no big downside in specifying several gene-concepts for different theoretical purposes.⁸⁵ However, there is a larger downside to do so for concepts that are used outside science since that creates obstacles to the interactive content flow (Cappelen 2023, 189). Cappelen raises the issue of interactive content flow between scientific and theoretical contexts and the contexts of ordinary speakers. Hence, more conservatism is required in many parts of the social science and philosophy when concepts which are in common use are explicated, both to facilitate interactive content flow and to curb verbal disputes between theorists. This brings the task closer to traditional conceptual analysis, compared to concept formation in, e.g., physics. Regardless, the task is to specify in which ways the explication should be similar to the ordinary concept.

10.5 Second step: specifying the explicatum

As I have mentioned in 10.4.1, I propose to explicate the gradable property concept as a comparative concept that can be used to express how democratic different collective decision-methods are relative to each other, and derivatively, how democratic different groups, organizations and countries are relative to each other. It would be too much to require that the concept in question at the same time satisfies the similarity criterion and is incorporated into a well-connected system of scientific concepts.

⁸⁵ See for example Dupré 2016, 550–553 for a short survey of the various gene concepts.

Based on Christiano's definition of the term 'democracy', I propose the following definition of *more democratic than*, expressed by the sentence "x is more democratic than y":

x is more democratic than y iff x and y are methods of collective decision making and x is characterized by a higher degree of equality among the participants at an essential stage of the collective decision making than y is.

The definition above is to be understood as the explicit rules of use for the explicatum, expressed by "x is more democratic than y". My explicatum can easily be used to define further property concepts which can be attributed to countries, etc. A shorter and easier way to say that a certain significant amount and kind of decisions in a country *c* are made through democratic decision procedures is, of course, to say "*c* is democratic". Now that I have clarified the explicandum, specified the sense in which the explicatum should be similar to the explicandum, and specified rules of use for the explicatum, it is time to evaluate the good-making features of the explicatum, i.e., the features which make it good for inquiry and which motivate the deviations from the explicandum.

10.6 Evaluation of the explicatum

10.6.1 Defining criteria: similarity and explicit rules of use

In this section I evaluate both of the defining criteria: similarity and explicitness. Arguably, for sufficient similarity to the explicandum it is required that we can use the explicatum for decision procedures, and therefore Christiano's definition is better than other influential definitions discussed by Cappelen, such as Joseph Schumpeter's (1942), Robert Dahl's (1971), and David Estlund's (2008). In their definitions, they restrict 'democracy' to the political domain. When we classify countries or governments or political systems as democratic or as democracies, it should be because certain relevant collective decisions are made in accordance with certain decisions procedures. It seems strange to call a *method* of decision-making a democracy, but it does not seem strange to call it democratic. Hence, it is better to take as the explicandum the property concept corresponding to the adjective rather than the concepts corresponding to the nouns.

The new concept of x being more democratic than y fulfils the following roles of the pre-theoretical concept: it is generally applicable to decision procedures, it is normatively neutral, and it preserves the core notion of equality among participants. All of these requirements are fulfilled.

The second defining criterion is fulfilled since the explicatum has been provided with explicit rules of use. How good those rules are is a matter of their satisfaction of the four good-making criteria. With both defining criteria fulfilled the concept counts as a candidate explicatum, and I now evaluate how good an explicatum it is.

10.6.2 First good-making criterion: sharpness

Arguably, the comparative explicatum is sharper than the qualitative concept in Christiano's original definition, in the sense that there are fewer arbitrary borderline cases. There may still arise borderline cases when two decision methods (or, derivatively, countries) are compared, but presumably less cases than in a division of decision methods (or, derivatively, countries) into democratic and undemocratic. For example, while it is hard to judge whether Singapore is democratic it is easier to decide that Singapore is more democratic than North Korea⁸⁶ or that Singapore and The Dominican Republic are roughly equally democratic. However, the sharpness is not the main strength of the explicatum. The notion of "the essential stage of collective decision making" is also vague, but such vagueness is arguably made up for by the other criteria.

10.6.3 Second good-making criterion: connectedness

Although the explicatum cannot be incorporated into a formal or developed well-connected system of concepts, we can specify its place in a system of concepts, and we can define many important and fruitful concepts in terms of the explicatum. A viable explicatum of "democratic" should allow us to classify and rank countries. The explicatum can be used to define what it is for a country to be more democratic than another country.

⁸⁶ The statement should be regarded as loose talk for "Singapore is governed through collective decision methods characterized by a higher degree of equality among the participants at an essential stage of the collective decision than the collective decision methods through which North Korea is governed".

10.6.4 Third good-making criterion: fruitfulness

Ideally we should find examples of universal or probable statements that can be formulated with the explicatum. For such a general concept of democratic as the explicatum, it is hard to make interesting generalizations. The benefit is that more specific concepts can be constructed in terms of the same explicatum. The explicatum is sufficiently general to capture most typical uses of ‘democratic’, but we can use it to define more specific concepts, such as ‘democratic country’, which are particularly interesting to investigate (compared to, e.g., ‘democratic chess club’). Although the explicatum may not occur in many interesting generalizations, it can be used to define concepts that do. Due to the explication, we know what we are talking about, and how to specify what we are talking about, when we ask questions such as “Which factors consolidate democracy in a country?”, “Which political institutions are required for a country to be democratic?”, etc. Hence, my explication could serve as the conceptual basis for further concepts which can be used for theoretical work. Therefore it has potential to increase theoretical unity and improve communication across fields.

10.6.5 Fourth good-making criterion: intersubjectivity

Initially, the explicatum seems to be weak with regards to intersubjectivity, since it is defined in terms of ‘equality’, which itself is a contested term. That is not to say that we have to explicate every term in the definition of the explicatum-term. Oppenheim remarks regarding his explication of political concepts in terms of action concepts that:

I did not try to explicate these action concepts in turn, but considered them as given. This is perfectly in line with modern empiricism, which takes as the basis of the conceptual scheme of a given field not observation terms, but an “*antecedently available vocabulary*,” i.e., terms “which have well-established use in science and are employed by investigators with high intersubjective agreement [Hempel 1973, 372]. (Oppenheim 1981, 190)

Nevertheless, a further explication of the term ‘equality’ is in order. However, what is required for present purposes is a definition or explication of “participant equality”, and not of “equality” in general. A good starting point for a definition or explication of “participant equality” could be Dahl’s five criteria for a democratic process of governing an association. More specifically Dahl gives criteria that would have to be met for a process of governing an association “in order to satisfy the requirement that all the members are equally

entitled to participate in the association's decision about its policies" (Dahl 2021 [1998]). The five criteria are (1) effective participation, (2) voting equality, (3) enlightened understanding, (4) control of the agenda, and (5) inclusion of all adults (Dahl 2021 [1998], 37, 38). The criteria need to be modified in order to be applicable to collective decision-methods in general. Some modifications will be needed of (2) in order to uphold the generality requirement, i.e., to make the concept generally applicable for all collective decision-methods and not merely to political decision-methods. Modifications of (5) are also needed, since the criteria of equal participation need to be gradable as well. Therefore (5) should be changed so that the higher the share of legitimate participants that are included the more equal is the decision-method. With some changes, I suggest to use Dahl's five criteria for democratic processes to further explicate "equality" as it is used in the definition of "more democratic than", i.e., in the sense of "x is an equal participant in the decision procedure".

Further, it allows for a clear notion of what it is for different kinds of entities to be democratic, i.e., entities such as decision methods, decisions, laws, clubs, groups, organizations, and countries. Hence, my explication increases the intersubjectivity of the concept, without the arbitrary stipulations that are typical of many operational definitions of democracy, such as the various definitions used to measure the level of democracy across countries.

Finally, with a comparative concept there is likely a higher degree of invariance in judgement across contexts. The truth of a comparative judgement does not vary with context in the way that a qualitative judgement does. Although a contemporary observer may not want to call the ancient city state of Athens democratic due to the lack of full suffrage, the judgement that it was more democratic than Sparta does not change with the changing expectations on democratic societies. With the explicatum, we may explain the differences between Athens and modern democracies as a change in who is recognized as a participant or political subject or person, rather than a change in the concept of being democratic. Hence, we can ensure conceptual continuity with historical democratic systems.

10.6.6 Fifth good-making criterion: simplicity

The suggested explicatum is simple in the sense that it defines being more democratic in only one dimension: more equality among the participants in the decision-making process. Therefore, it is simpler than, e.g., the definitions of democracy used in the democracy indices due to their many dimensions. The minimalist definitions, such as Schumpeter's would do better, but they would

not be sufficiently similar to the explicandum, since “democracy” denotes political systems in their definitions.

10.6.7 Summary of the evaluation

I argue that my explicatum allows for a clear notion of what it is for different kinds of entities to be democratic, i.e., entities such as decision methods, decisions, laws, clubs, groups, organizations, and countries. Hence, my explication could potentially increase the intersubjectivity of the concept, without the arbitrary stipulations that are typical of many operational definitions of democracy, such as the various definitions used to measure the level of democracy across countries. Further, my explication has the potential to bring clarity to questions related to the normative status of democracy. Hence, my explication could potentially serve as the conceptual basis for normative investigations as well as empirical investigations. Therefore it has potential to increase theoretical unity and improve communication across fields.

10.7 Comparisons

To illustrate the benefits of my account over Carnap’s account I will point out the shortcomings of a strictly Carnapian explication. The main problem for an explication of ‘democracy’ on Carnap’s account would be that it is hard to satisfy both similarity and exactness to a sufficient degree. First, there is no well-connected system of scientific concepts to incorporate a concept that can replace the explicandum. If similarity is sacrificed, we may specify an exact concept that could be used to formulate empirical laws (e.g., that the democratic system in countries with market economies is more stable than the democratic system in countries with non-market economies). But the applications of ‘democratic’ are too broad to be replaced by a concept that is restricted to countries.

I now compare my explication with other definitions of ‘democratic’ and ‘democracy’. Beside Christiano’s definition, the prominent definitions are definitions of political democracy or definitions of democratic country, or democratic government, or something along those lines. Although ‘democracy’ as it is operationalized in various democracy indices may be a predictor of other properties, it is hard to see how these operational definitions

could satisfy the similarity criterion and the simplicity criterion to a sufficient degree. The democracy indices would fare very low on the simplicity criterion, due to their many dimensions. The minimalist definitions, such as Schumpeter's, would do better, but they would not be sufficiently similar to the explicandum, since 'democracy' in their definitions denotes political systems.

Now compare my procedure with the "standard approach" in measurements of democracy, as characterized by Krieger (2022):

The standard approach for producing a measure of democracy consists of three major steps [...]. The first step is to define how a non-democratic regime differs from a democratic regime (conceptualization). The next step is to compile observable regime characteristics that reflect the components of the before chosen concept of democracy (operationalization). The final step is to select a method that transforms the regime characteristics into a uni-dimensional indicator (aggregation). (Krieger 2022, 1)

On my account the explication starts, inspired by Cappelen, from the bottom, i.e., from "democratic" as applied not only to political organizations or decision procedures used in political organizations but to decisions procedures in general (used by any group), and it would start not with the classificatory concepts *democratic* and *non-democratic*, but with a comparative concept expressed by the phrase "*x* is more democratic than *y*", where the variables *x* and *y* range over decision procedures. From there, concepts applicable to groups can be derived.

10.8 Concluding remarks

To give an example of how my proposed modifications may be used in application, I have suggested an explication of the concept *democracy*. In the next and final chapter I will address these questions, and discuss the role of explication in philosophy in general, as well as the specific role of my account of explication in contrast to other accounts of explication.

11 Explication as a philosophical method

11.1 Overview

Now that I have presented a modified version of explication and given an example of how to apply it in an explication of the concept *democratic*, there remains an important question regarding my account. The question is what the methodological implications are for philosophy. Does the proposed method offer a way of making progress with regard to paradigmatic philosophical problems? I will comment on the literature on the philosophical role of both explication and conceptual engineering, and reformulate the debates and views about conceptual engineering to make them applicable to explication specifically. I begin by presenting the standard model of philosophical methodology. Then I present what I will call the standard model of explicationist philosophy. I discuss what the motivation is for applying explication in philosophy, what the scope of explication should be in philosophy, and in which way explications should be used to tackle problems in philosophy. After I have presented the current state of the debate, I reflect on the philosophical implications of my proposed modifications of explication.

As this is the final chapter of the thesis, I end it with reflections on remaining open questions, which could be the subjects of future research.

11.2 The standard model of philosophical methodology

Although many of the explications that have been mentioned and discussed in this thesis—e.g., by Frege, Russell, Tarski, Carnap, and Quine—can be regarded as philosophical projects, there is a contrast between those projects and the more traditional philosophical methods, such as the so-called standard

model of philosophical methodology (I borrow the term from Jennifer Nado (2021)). On the standard model, concepts are analysed and the analyses are tested against our intuitive verdicts in hypothetical cases. Although I expect most readers to be familiar with the standard model, I briefly present it to set things up for the ensuing discussion.

The two central elements in the methodology are conceptual analysis and the method of cases. However, not all conceptual analysis is philosophical analysis. Following Jeffrey King (2016), we may initially distinguish on the one hand between a priori and a posteriori analyses and on the other hand between interesting and trivial analyses (King 2016, 251). Philosophical analysis is traditionally taken to fall into the two former categories, as being both a priori and interesting. For examples of these distinctions, consider first an analysis of the word ‘brother’:

1. To be a brother is to be a male sibling.

Although 1 is apparently more informative than “To be a brother is to be a brother”, it is a trivial analysis. Consider, in contrast, the classic JTB-analysis of ‘knowledge’:

2. To be an instance of knowledge is to be an instance of justified true belief.

Both 1 and 2 are a priori analyses, but 1 is trivial while 2 is interesting. Further, 2 but not 1 is a *philosophical* conceptual analysis, or simply, a philosophical analysis.

Consider now the difference between 2 and the following analysis of ‘water’:

3. To be water is to be H₂O.

While both 2 and 3 are interesting and informative, 2 is a priori while 3 is a posteriori. Neither 1 nor 3 are traditionally considered philosophical analyses.⁸⁷

⁸⁷ There is, however, an alternative approach to conceptual analysis in philosophy, advocated by Frank Jackson and David Chalmers. On their account, to give a conceptual analysis is to specify the primary intension of a term (given the two-dimensional semantic framework endorsed by Jackson and Chalmers, in which expressions have two intensions). On that

In his pedagogical overview of philosophical methods, Chis Daly (2010) suggests what he calls “a working model” of philosophical analysis. Daly takes philosophical analysis to have five constituents: (1) it has the logical form of universally quantified biconditional, (2) it is necessarily true, (3) it is informative, (4) it is knowable a priori, and (5) it is testable by the method of hypothetical cases (Daly 2010, 50). A philosophical analysis, like any conceptual analysis, has the form of a universally quantified biconditional:

$$(\forall x) (Fx \leftrightarrow Gx)$$

The right-hand side of the biconditional, G, consists of individually necessary and jointly sufficient conditions for being F. Daly (2010, 50) exemplifies this with the following analysis of “x is an even number”:

$$(\forall x) (x \text{ is an even number} \leftrightarrow x \text{ is divisible without remainder by } 2).$$

To illustrate how analyses are tested by the method of cases, Daly considers the analysis that anything is an F iff it is a G:

We imagine a hypothetical case in which something is an F. We then see whether we have the intuition to describe that thing as a G. If we do, the intuition is some evidence for the analysis being true. If we have the intuition that the thing is not G, that intuition is some evidence that the analysis is false. (Daly 2010, 49)

In summary, then, on the traditional model a philosophical analysis is necessarily true, informative, knowable a priori and testable by the method of hypothetical cases (Daly 2010, 50). Now that we have the central elements of

model of conceptual analysis, we can give an a priori analysis of a term such as ‘water’. As King explains:

In the case of ‘water’, very roughly speaking it applies to the local watery stuff in any world considered as actual. That is why it yields XYZ at the XYZ world and H₂O in the actual world. We discover what the primary intension of a term is by considering various ways the world might be and asking: ‘If the world turns out that way, what would water be?’ So according to Chalmers, doing this sort of conceptual analysis is an a priori enterprise. (King 2016, 259)

King also notes, however, that the Chalmers–Jackson approach to conceptual analysis has a different purpose than traditional conceptual analysis, namely, to give reductive explanations of various phenomena in terms of microphysics (King 2016, 259).

conceptual analysis we can consider it in action, when it is used as a philosophical methodology. Nado (2021) gives a “familiar description of what philosophers do”, i.e., the standard model of philosophical methodology. Her description is worth quoting at length, since I will take it as a template when I describe what explicationist philosophers do in general and what explicationist philosophers would do if they adopted my account.

Here is Nado’s description of the standard model:

First, philosopher McA selects a philosophically interesting concept C (‘freedom’, ‘consciousness’, ‘good’, etc.), and proposes a theory which purports to delineate the conditions under which something counts as C. Ideally McA’s theory will take the form of a biconditional, the left-hand side of which contains C, and the right-hand side of which contains necessary and sufficient conditions for being C. Thus, McA’s theory (if successful) should enable us to determine, for any hypothetical case, whether it falls under C or does not.

Next, philosopher McB challenges McA’s theory by producing an imagined case (sometimes quite bizarre or complex). This should be a case which McA’s theory deems to be C, but which intuition deems to be not C (or vice versa). If McB can generate such a case, this counts strongly against McA’s theory—perhaps strongly enough to warrant its rejection. [...].

The method by which McB challenges McA’s theory is often called the ‘method of cases’; McA’s theory itself is standardly called an ‘analysis’. An analysis of what, you might ask? Well, the traditional answer would be: an analysis of the concept C. In other words, a conceptual analysis. [...] Philosophers, so the story goes, are in the business of producing and testing conceptual analyses. (Nado 2021, 1507–1508)

There are three main challenges to the standard model, namely the paradox of analysis, the critique of intuitions and the critique of the classical view of concepts. The paradox of analysis concerns how an analysis can be both correct and informative. Although the paradox was anticipated by Plato’s formulation of Meno’s paradox and was articulated in Frege’s work (Beaney 2014), the phrase ‘paradox of analysis’ was coined by C. H. Langford (1942) in relation G. E. Moore’s notion of analysis. Langford states the paradox in the following way:

Let us call what is to be analyzed the analysandum, and let us call that which does the analyzing the analysans. The analysis then states an appropriate relation of equivalence between the analysandum and the analysans. And the paradox of analysis is to the effect that, if the verbal

expression representing the analysandum has the same meaning as the verbal expression representing the analysans, the analysis states a bare identity and it is trivial; but if the two verbal expressions do not have the same meaning, the analysis is incorrect. (Langford 1942, 323)

The critique of intuitions stems from empirical studies showing both that (1) intuitions about philosophical concepts vary systematically between people and populations, and that (2) “people’s judgments about philosophical cases are sensitive to various kinds of contextual factors that seem to be philosophically irrelevant” (Knobe and Nichols 2017), such as the order of presentation. If intuitions are not reliable evidence to test our analyses against, we are left without a way of testing philosophical analyses. Hence, the standard model is not a viable methodology.

Finally, traditional conceptual analysis assumes the classical theory of concepts. On the classical theory, it is assumed (at least) that concepts have a definitional structure, and that the structure consists of simpler concepts that express necessary and sufficient conditions for falling under them (Margolis and Laurence 2023, §2.2). Alternatively put, on the classical theory a concept “encodes the conditions that are singly necessary and jointly sufficient for something to be in its extension” (Margolis and Laurence 1999, 9, n. 8). On this view, the concept *brother* consists of the simpler concepts *male* and *sibling* which are individually necessary and jointly sufficient conditions for falling under the concept *brother*. A possible explanation for the poor track record of philosophical analysis could be that concepts lack such a definitional structure (Margolis and Laurence 2023, §2.2).

In conclusion, a typical philosophical problem on the standard model is formulated in informal, ordinary language, as a question regarding some concept C, “What is C?”. The way to answer the question is to offer a definition in the form of a universally quantified biconditional with sufficient and necessary conditions for something to be C, and to test the definition against our intuitions about thought experiments. Explication as a philosophical method is motivated by the view that typical philosophical problems are misunderstood because concepts are tools which evolved for certain environments and for certain practical tasks, but expected by philosophers to do other, theoretical tasks.

11.3 The standard model of philosophical explication

In light of the challenges facing the standard model of philosophical methodology, a natural replacement is the view that philosophically interesting concepts should be explicated. Let us begin with a preliminary characterization of what I will call *the standard model of philosophical explication*:

- Philosophical problems ought not to be tackled via conceptual analysis; instead, philosophers should give explications of philosophically interesting concepts. Explications are not tested or evaluated by the method of cases, but by their usefulness for our investigative purposes.

I will now describe the methodology in action, using Nado's description of standard philosophical methodology as a template for my description, with philosophers McA as protagonist and McB as antagonist:

First, philosopher McA selects a philosophically interesting concept C ('freedom', 'consciousness', 'good', etc.), and proposes a new concept C* with rules of use which purport to capture some but not all of the most important uses of the philosophically interesting concept. Where C* deviates from the uses of the old concept C, McA points to the theoretical benefits of the new concept over the old concept. Ideally McA will give a definition of the new concept in the form of a biconditional, the left-hand side of which contains C*, and the right-hand side of which contains necessary and sufficient conditions for being C*.

Next, philosopher McB challenges McA's explication in one of the following ways: (1) McB shows that there is another concept C** that is better suited than C* for the role played by C, (2) McB argues that McA's explication fails to sufficiently satisfy McA's own criteria, (3) McB argues that McA has used the wrong criteria for their her purposes, (4) McB argues that McA has aimed for the wrong purposes. Philosophers, on the explicationist model, are in the business of constructing and using conceptual tools.

Within the standard model of philosophical explication, I discuss three extant further distinctions that have been made in the literature. The first distinction concerns the motivations for using explication as a philosophical method. The second distinction regards whether explication is seen as the only legitimate method in philosophy, or as one method among many. The third distinction regards what one thinks can be achieved when philosophical problems are

tackled by explication. The alternative views are that explications can (a) merely illuminate philosophical problems, (b) in some sense solve philosophical problems, or (c) dissolve philosophical problems. I devote a section each to these three distinctions, starting with the question of motivation.

11.4 What is the motivation for philosophical explication?

At the end of 11.2, I mentioned reasons to give up the standard model of philosophical methodology, e.g., philosophical analysis. The reasons discussed in that section are what motivated the first part of the standard model of philosophical explication, i.e., that “Philosophical problems ought not to be tackled via conceptual analysis”. In this section I will address reasons to replace philosophical analysis with philosophical *explication*. There are different ways to frame the motivation of explication as a philosophical method. It may be motivated by the view that typical philosophical concepts are defective (and hence that there is something wrong with questions posed by the use of such concepts). However, a project of explication may also be motivated by the intended improvements, regardless of any particular defects of the explicandum. This latter line of defence is available to someone who is not motivated by the idea that there is something wrong in principle with the concepts and questions of traditional philosophy, but who thinks that we can do better with other concepts. The first view I will call the *fixing language* view of explication and the second view I will call the *advancing inquiry* view of explication. Both of these views (or views close to them) are discussed under various labels in the literature on conceptual engineering. In chapter 7, I argued that the main debates in that literature on conceptual engineering lack immediate relevance for explication, but in this section I will engage with debates in that literature which *do* apply to explication: viz., debates about the role of conceptual engineering in solving philosophical problems. Therefore, I first discuss these distinctions in relation to the debates about conceptual engineering and then I apply them specifically to explication.

Under other labels, the distinction between the two views has been pointed out by Simion and Kelp (2019):

it is widely agreed in the literature that conceptual engineering is principally concerned with repairing defective concepts. [...] we propose a reorientation of

the central focus of the project away from conceptual repair and towards *conceptual innovation*. (Simion and Kelp 2019, 985)

The view that conceptual engineering is principally concerned with repairing defective concepts is what I call the *fixing language* view. Pinder describes, without endorsement, the *fixing language* view in the following way:

concepts of philosophical interest tend to be defective, and so philosophical problems ought not to be tackled via conceptual analysis; instead, philosophers should start tackling philosophical problems by improving the defective concepts. On this account, conceptual engineering is interesting insofar as both: it is an effective new methodology for solving philosophical problems; and we can make sense of concepts being defective. (Pinder 2020a, 5)⁸⁸

Simion and Kelp focus on the claim that the “literature on conceptual engineering has focused largely, if not exclusively, on conceptual repair”, and that “there is a general consensus among those working on conceptual engineering, to wit, *that it is about fixing defective concepts*.” (Simion and Kelp 2019, 987). To get an analogous account of the role of explication in philosophy, one may simply replace the words “improving” and “conceptual engineering” with “explicating” and “explication”. I will restate the account in terms of explication in this way, and as mentioned I will call it the *fixing language* view of explicationist philosophy:

- *The fixing language view of explicationist philosophy*. Concepts of philosophical interest tend to be defective, and so philosophical problems ought not to be tackled via conceptual analysis; instead, philosophers should start tackling philosophical problems by giving explications of the defective concepts in which the defects are fixed.

In contrast to that view, Simion and Kelp advocate a view of conceptual engineering as concerned with the *improvement* of concepts:

⁸⁸ Pinder summarizes what Max Deutsch calls the “standard account” of conceptual engineering (Deutsch 2020, §2). Deutsch argues that conceptual engineering is either trivial, unachievable or already commonplace, but Pinder replies that the target of his criticism is a too-narrow conception of conceptual engineering (i.e., the fixing language view), whether it is the standard account or not.

Ambitions of conceptual engineering need not be motivated by defects in our representational devices; proposals for improvements in the world of concepts will do just as well. (Simion and Kelp 2019, 988)

Pinder agrees with Simion and Kelp that conceptual engineering should not be limited to fixing language, and even more strongly he holds that such a view “is not a helpful picture of conceptual engineering in philosophy” (Pinder 2020a, 5). In Pinder’s summary:

Firstly, as Simion and Kelp [2019] have argued, conceptual engineering should not be thought of as limited to ‘fixing defective concepts’. Rather, we should think of conceptual engineering as aiming to improve upon our concepts relative to certain purposes. (Pinder 2020a, 5)

In the literature on conceptual engineering there are many different purposes motivating conceptual revisions, but I will rephrase the view restricted to explication, where the purpose is to advance inquiry. Hence, we get the following *advancing inquiry*-view of explicationist philosophy:

- *The advancing inquiry view of philosophical explication.* Philosophical problems ought not to be tackled via conceptual analysis; instead, philosophers should explicate concepts for the purpose of advancing inquiry.

Simion and Kelp recognize that repair is a form of improvement, as they note that: “[o]f course, fixing a defect of a certain sort is one way of bringing about a corresponding type of improvement” (Simion and Kelp 2019, 987). Although they claim that it is “hard to overemphasize” the significance of their reorientation from repair to improvement (Simion and Kelp 2019, 985), I think that since both conceptual defects and conceptual improvements only can be understood relative to a purpose the distinction seems to be mainly a difference in emphasis.

I understand the *fixing language* view and the *advancing inquiry* view as two expressions of the standard model of philosophical explication. Hence, the standard account should be formulated in terms that are neutral with regards to the two versions. The next distinction concerns the scope of explication as a philosophical method. Should explication exhaust the philosophical toolbox?

11.5 Should explication exhaust the philosophical toolbox?

In this section I discuss how prevalent the role of explication should be in philosophical methodology. Should explication be considered the only philosophical method or should it be considered as one philosophical method among many? At the other extreme we have the view that philosophy isn't about explication (or language engineering) at all, i.e., that such an activity falls outside the concerns of philosophy. I will not address the second extreme position, but assume that explication has at least some role in philosophy. The question is how dominant that role is. The exhaustive view of explication in philosophy is that the only legitimate activity for philosophers is explication, to be compared with the view held by Carnap in the 1930s that philosophy is to be replaced by what he called the logic of science. Scharp is a contemporary advocate of an exhaustive view (or 'large scope' view), remarking that

I've come to think that conceptual engineering can and should play a much larger role [in] philosophical theorizing. Indeed, I've come to think that most, if not all, commonly discussed philosophical concepts are inconsistent. (Scharp 2020, 397)

Among more modest views there are many possible options. Cappelen and Plunkett (2020) map the space of alternatives regarding the scope of conceptual engineering. They consider the question "What role should conceptual engineering play in philosophy?", and they break up the question into the following two components:

1. How many parts of (or sub-fields of, or issues in) philosophy should conceptual engineering play a role in?
2. For each part it should be involved in, how important should it be? (Cappelen and Plunkett 2020, 21)

Given these two components, Cappelen and Plunkett (2020, 21) list five possible answers to the question. The role of conceptual engineering in philosophy is:

1. All of All
2. All of Some

3. Some of All
4. Some of Some
5. None

The discussion is also applicable to explication specifically. Hence, the question of the scope of explication in philosophy is split into two parts: Is explication an (or the only) appropriate method in all areas of philosophy? Given the view that explication is an appropriate method in some areas of philosophy, how extensive should the role of explication be in that area? The question of how to apply explication in philosophy depends of course on further questions about the nature of philosophy and about the nature of philosophical concepts.

11.6 How can explication tackle philosophical problems?

In this section I discuss different views regarding what we can hope to achieve by explication in philosophy. Pinder distinguishes between two strategies for using conceptual engineering to tackle traditional philosophical problems, labelling them as the conservative strategy and the radical strategy. The reason for thinking of the first one as conservative is that it does not purport to solve philosophical problems, but merely illuminate them (and possibly providing reason for one solution over the other possible solutions) by showing that with the explicated concept the problem does not arise. The second strategy is radical insofar as it purports to solve philosophical problems through rearticulation. Pinder introduces the two strategies in relation to the epistemological problem of radical scepticism. He considers how the strategies would be used to tackle that problem:

A Conservative Strategy. Provide a set of technical definitions for (new or old) epistemological terms, and argue that those definitions are particularly good things to speaker-mean by those terms for the purpose of articulating and explaining epistemological phenomena. Demonstrate that a problem parallel to that of radical scepticism does not arise in this framework. Relate this result back to our ordinary concepts of KNOWLEDGE, JUSTIFICATION, etc., to obtain substantive insight into the original problem of radical scepticism. (Pinder 2020a, 7)

As examples of defenders of the conservative strategy Pinder mentions Maher (2007), Scharp (2013), Schupbach (2017), and his own work (2019) and (2020b). Maher proposes the following role for explication in philosophy:

Suppose our problem is to determine whether or not some sentence *S* of ordinary language is true. If we apply the method of explication to this problem, we will construct explicata for the concepts in *S*, formulate a corresponding sentence *S'* using these explicata, and determine whether or not *S'* is true. This does not by itself solve the original problem—that is Strawson's point—but it can greatly assist in solving the problem, in three ways. (1) The attempt to formulate *S'* often shows that the original sentence *S* was ambiguous or incomplete and needs to be stated more carefully. (Maher 2007, 333–334)

Before we move on to the other two ways in which Maher thinks that explication can assist in solving philosophical problems we should note that what makes his view an example of conservatism is that he thinks that explication can merely *assist* in solving the original problem, in contrast to radical view that explication in itself can *solve* the original problem. Maher continues:

(2) If the explicata appearing in *S'* are known to correspond well to their explicanda in other cases, that is a reason to think that they will correspond well in this case too, and hence to think that the truth value of *S* will be the same as that of *S'*. (3) We can translate the proof or disproof of *S'* into a parallel argument about the corresponding explicanda and see if this seems to be sound; if so, we obtain a direct argument for or against *S*. In these ways, explication can provide insights and lines of argument that we may not discover if we reason only in terms of the vague explicanda. (Maher 2007, 334)

Note that Maher considers a particular kind of philosophical problem, namely the problem of determining whether a sentence of ordinary language. It is not obvious how to generalize Maher's proposals to other kinds of problems, e.g., when regarding a concept *C* we ask: "What is *C*?"

Next we have the radical strategy. Once again, Pinder introduces the strategy in relation to the philosophical problem of radical scepticism:

A Radical Strategy. Provide a set of technical definitions for (new or old) epistemological terms, and argue that those definitions are particularly good things to speaker-mean by those terms for the purpose of articulating and explaining epistemological phenomena. Articulate the problem of radical scepticism within the new terminological framework, arguing that the result is a better way to understand the original problem. Show that, within the new

terminological framework, the articulated problem of radical scepticism has a solution. (Pinder 2020a, 7)

For examples of the radical strategy, Pinder mentions Carnap's RSE, Simion and Kelp (2019) and his own (2017) and (2020b). (He does not comment on the fact that he defends both strategies, even in the same paper.) A paradigmatic example of the radical strategy would be the solution of Zeno's paradoxes of motion, through mathematical explications of concepts involved in the paradoxes.

Arguably there is a third strategy not mentioned by Pinder, which I will call the abolitionist strategy. An example of the strategy can be found in Quine's conception of explication (see 6.4 for a discussion), when he writes that "In the case of the ordered pair the initial philosophical problem, summed up in the question 'What is an ordered pair?', is dissolved by showing how we can dispense with ordered pairs in any problematic sense in favour of certain clearer notions" (Quine 1960, 260). In a sense the abolitionist strategy is more radical than the radical strategy since it does not aim to solve the original problem but to dissolve it. The difference between the conservative strategy is that on the conservative strategy the original problem is left intact (but clarified) while on the abolitionist strategy the original problem is considered dissolved.

To conclude this section, I will rephrase the distinction made by Pinder in a more general form and with explicit mention of the method of explication. Here is my restatement of the distinction in an abbreviated and general formulation, not tied to any particular philosophical problem, and with the addition of the abolitionist strategy:

- *Conservative strategy.* Illuminate a philosophical problem by explicating the concept or concepts that yield the problem and showing that its problems do not arise for the explicatum, and reconsider the original problem in light of the explicatum or explicata.
- *Radical strategy.* Solve a philosophical problem by explicating the concepts that yield the problem and showing that articulated in terms of the explicatum or explicata the problem has a solution.
- *Abolitionist strategy.* Dissolve a philosophical problem by explicating the problem that yields the problem and showing that the original problem is a pseudo-problem or that it does not need a solution.

Since Carnap articulated the method of explication, there have been philosophers sympathetic to Carnap's approach and method who have applied the method to make progress in philosophy. I intentionally use the phrase "to make progress in philosophy" instead of, e.g., "to solve philosophical problems" in order to leave it open whether progress is made through illuminating, solving, or dissolving philosophical problems. All such strategies have been pursued and advocated by proponents of explication. These strategies do not appear to be mutually exclusive. In some cases we may at best achieve illumination of a philosophical problem, in some cases we may consider the problem solved, and in some cases the problem is dissolved.

However, I claim that both the conservative and the radical strategy follow a kind of business-as-usual-view of philosophical problems. It is not so much an explicit view as an assumption or expectation about how explication should contribute to philosophical progress. The abolitionist strategy, however, is motivated by a rejection of that view. Although the debate on conceptual engineering as a philosophical method is applicable to the questions about explication as a philosophical method, there is a problem with how the debate about conceptual engineering is framed. Consider the following situation, which I will call *business-as-usual*: We have a field of philosophy, with genuine problems, and with explication we get a new method to solve those problems. On this picture, it is then up to the defenders of explication to show how the new tool can be used to solve philosophical problems. The alternative is the transformative view. According to the transformative view, the remedy to philosophical problems is to switch from the activity of trying to solve them to a different activity. On the transformative view, the method of explication offers progress but it does not offer a solution to typical philosophical problems. Instead it offers an invitation to partake in different activities.

11.7 The broadened account of explication as a philosophical method

What are implications of my account for explication as a philosophical method? For Carnap, changing the conception of philosophy to the activity of language engineering was a transformation of the whole enterprise. My proposal to broaden explication is vulnerable to a bundle of worries raised by Reck (2024) in relation to the prospect of understanding explication in a broad sense, beyond formal-logical explications. Reck has interpreted Carnap's conception of philosophy as one in which "explication is the main, or even the

only, method.” He raises the possible objection that it is unfair to Carnap to interpret him so restrictively:

In [Carnap’s] defense, one might argue as follows: The tools and methods originating in modern logic and in scientific concept formation have proven fruitful in some parts of philosophy; but in other parts they do, indeed, need to be supplemented, for example, in large parts of ethics [...] In addition, from a Carnapian point of view, one can distinguish a narrower, more formal version of explication from a broader, less formal one. This is in line with the relatively broad formation of Carnap’s four desiderata of explication, as we saw; and it makes explication applicable more widely immediately. (Reck 2024, 146–147)

Against that line of argument, Reck recognizes three counterarguments:

First, Carnap’s articulation of the explication does not leave much leeway in this respect. For example, he talks about fruitfulness only in terms of formulating logical and scientific laws. Hence, even a less formal version of Carnapian explication seems narrow and restrictive. Also, formal-logical explication remains the ideal, it appears. (Reck 2024, 147)

This point has been widely recognized in the literature on explication, and taken as a reason to deviate from Carnap’s prescription (see chapter 8). The second counterargument, however, is often overlooked in contemporary literature on explication, namely that if one treats explication as a tool among others in the philosophical toolbox one breaks some of the core tenets of Carnap’s conception of philosophy. I will now pick up the thread from the end of chapter 10, quoting again the same passage from Reck:

Second, if the methodological toolbox is opened up beyond explication, in both its narrower and wider forms, this would seem to create conflicts with other core tenets of Carnap’s approach [...] More basically, Carnap says very little about what other tools might be acceptable to him, which leaves him open to this challenge. There certainly is no extended, charitable discussion of complementary tools. (Reck 2024, 147)

The other core tenets of Carnap’s approach that Reck refers to are identified earlier in Reck’s text as the goal “to erase the lines between logic, philosophy, and the sciences” and the goal “to transform philosophy into ‘mathematical’ or ‘scientific philosophy’ (Reck 2024, 143). Regarding other tools that might be acceptable, Reck points out in a footnote that there is a rare exception in Carnap’s response to Strawson where Carnap admits that the Strawsonian method of analysis might be useful for clarifying the explicandum. But Reck

concludes that “even that is done only very tentatively and half-heartedly by him” (Reck 2024, 147, n. 31). The third counterargument is related to the second. If core tenets of Carnap’s conception of philosophy are broken, the remaining position may be diluted beyond recognition:

Third, opening the toolbox further might dilute the position to such a degree that no distinctive, recognizably different conception of philosophy remains. After all, philosophers in other traditions use informal explications of concepts as well, sometimes even formal ones. [...] Suppose we interpret Carnap in a less rigid and restrictive way. How should we conceive of the proper form, goal, and reach of explication then? [...] And how far can we go until we stop philosophizing in a recognizably “Carnapian” way? Such questions seem pertinent with respect to Carnap’s legacy today. (Reck 2024, 147)

Although Reck’s discussion concerns narrow and broad interpretations of Carnap, the worries are relevant also for proposals to broaden the method further. I have recognized the limits of Carnap’s position and propose to broaden explication.

What remains of Carnap’s engineering-conception of philosophy in my modified account of explication are the following two views: (1) we are free to choose new concepts, i.e., there is an element of voluntary decision in concept choice; and (2) the decision is made and evaluated for what it can help us do (in our inquiries). The difference from Carnap lies in my view of inquiry, and in which activities I have in mind when evaluating what a concept can help us do (in our inquiries). A preliminary philosophical model for my account of explication would be:

- Philosophical problems ought not to be tackled via conceptual analysis; instead, philosophers should explicate philosophically interesting concepts. Explications are not tested or evaluated by the method of cases, but by how well they satisfy the proposed criteria. Although the criteria are general, they are “codified” or specified for each particular case.

The similarity criterion is what distinguishes explication from the mere introduction of new technical terms. However, since the similarity criterion is shared with all versions of explication, it is not a distinguishing feature of my account. For distinguishing features, we have to look at my division of criteria into defining ones and good-making ones, and my formulations of the criteria. With these modifications, can we solve typical philosophical problems better than with other accounts of explication? An equally important questions, is

whether the method can lead to better questions, and help us focus on better, more promising concepts.

11.8 Open questions and future research

In this section I address a non-exhaustive list of remaining questions and possible subjects for future research. First, there is an old tradition of revisionary philosophy, which shares the revisionary agenda of explication but not the pragmatism, conventionalism and linguistic orientation of Carnapian explication and language engineering. Should revisionary metaphysical theories of concepts such as *causation*, *persistence*, or *the self* be considered as explications? Consider an explication of “*x* persists”. What is the difference between providing a theory of persistence and of proposing an explication of “persistence”? In what sense are, e.g., David Lewis (1986) and Theodore Sider (2001) providing theories of persistence rather than giving explications of “persistence”? One suggestion is that the point of a theory of persistence is not to replace the ordinary concept of persistence (perhaps because it is not considered defective or in need of replacement). Instead, the point may be to explain the phenomenon denoted by it. Hence, the thought goes, ‘persistence’ differs from the traditional Big Words of philosophy, which are defective.

Relatedly, in the literature on explication the relation between conceptual desiderata and theoretical desiderata has not been systematically investigated. This is curious given that in many extant accounts theoretical virtues are treated in relation to the criteria of adequacy, but without any thorough discussion of how they are related. Felix Oppenheim (1981) observes the similarity between these two tasks, while holding them apart:

We have here a similarity between constructing good explicative definitions and good scientific theories. The adequacy of an empirical theory, too, must be judged by “standard criteria”, e.g., “accuracy, consistency, scope, simplicity, and fruitfulness” [(Kuhn 1977, 322)]. (Oppenheim 1981)

In the literature on Carnapian explication, the theoretical virtues are discussed by e.g., Brun and Pinder (2022b), but they are not comparing them as analogous tasks. On Pinder’s account, explication is related to theoretical virtues through his notion of fruitfulness. On Brun’s account of explication, their adequacy is measured in part by the theoretical virtues of the target theory in which the explicatum is incorporated. This is a consequence of his view that “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” (Brun 2016, 1236). Based on that view, Brun concludes that “[t]he criteria of adequacy for explications are [...] subordinate to criteria for theory choice” (ibid.). On his account, then, there is no clear distinction between conceptual virtues and theoretical virtues. On my account, they are to be held apart, but that raises questions about the relation between conceptual and theoretical virtues. It may seem natural to say that what makes concepts good for inquiry is that they allow us to formulate good theories. However, the question is further complicated when we take into account the similarity criterion and consider the role of explication in philosophical theorizing, where the analogy with scientific theorizing is contested.

The task of finding criteria for good concepts can be compared with the ongoing task of finding out what the theoretical virtues are,⁸⁹ i.e., the task of finding criteria for good theories). We look at examples of good explications, we articulate what makes them good, and in future explications we apply the criteria we have articulated. Future explications may give us reason to go back and revise our criteria.

Finally, there is the question of what the relation is between the idea of unified science and philosophical explication. What remains of Carnap’s conception of explication without the idea of unified science? Is there any aspect of it which can be upheld? As highlighted by Reck and others, a motivating idea behind Carnap’s conception of explication was to “erase the lines between logic, philosophy, and the sciences” and “thereby to transform philosophy into ‘mathematical’ or ‘scientific philosophy’” (Reck 2024, 143). Concerning my modifications, whereby the narrow, empiricist notion of unified science is replaced by purposes of inquiry more broadly construed, one may ask to what extent and in which sense an idea or ideal of unified science has been retained.

⁸⁹ See for example (Keas 2018).

References

- Achinstein, Peter. 1966. "Rudolf Carnap. The Philosophy of Rudolf Carnap." *Review of Metaphysics* 19 (3):517–549.
- Austin, John. 1956. "A plea for excuses." *Proceedings of the Aristotelian Society, New Series* 57:1–30.
- Awodey, Steve. 2012. "Explicating 'analytic'." In *Carnap's Ideal of Explication and Naturalism*, edited by Pierre Wagner. Palgrave-Macmillan.
- Babbie, Earl. 2016. *The practice of social research*. 14th ed. Boston: Cengage Learning.
- Baker, Alan. 2022. "Simplicity." In *Stanford Encyclopedia of Philosophy*.
- Beaney, Michael. 2004. "Carnap's conception of explication." In *Carnap brought home. The view from Jena*, edited by Steve Awodey and Carsten Klein, 117–150. Chicago and La Salle, IL: Open Court.
- Beaney, Michael. 2007. "Conceptions of analysis in the early analytic and phenomenological traditions." In *The Analytic Turn: Analysis in Early Analytic Philosophy and Phenomenology*, edited by Michael Beaney. London: Routledge.
- Beaney, Michael. 2013. "Analytic philosophy and history of philosophy : the development of the idea of rational reconstruction." In *The Historical Turn in Analytic Philosophy*, edited by Erich H. Reck. Palgrave-Macmillan.
- Beaney, Michael. 2014. "Analysis." In *Stanford Encyclopedia of Philosophy*.
- Blackburn, Simon. 1999. *Think: a compelling introduction to philosophy*. Oxford: Oxford University Press.
- Boniolo, Giovanni. 2003. "Kant's Explication and Carnap's Explication." *International Philosophical Quarterly* 43 (3):289–298.
- Brandt, Robert. 2001. "Modality, normativity, and intentionality." *Philosophy and Phenomenological Research* 63 (3):611–23.
- Brun, Georg. 2016. "Explication as a Method of Conceptual Re-engineering." *Erkenntnis* 81 (6):1211–1241.
- Brun, Georg. 2020. "Conceptual re-engineering: from explication to reflective equilibrium." *Synthese* 197 (3):925–954.

- Brülde, Bengt. 2000. "On how to define the concept of health: A loose comparative approach." *Medicine, Health Care and Philosophy* 3:303–306.
- Brülde, Bengt. 2010. "On defining 'mental disorder': Purposes and conditions of adequacy." *Theoretical Medicine and Bioethics* 31 (1):19–33.
- Brülde, Bengt, and Per-Anders Tengland. 2003. *Hälsa och Sjukdom: En begreppslig utredning [Health and Illness: A Conceptual Investigation]*. Lund: Studentlitteratur.
- Bryman, Alan. 2008. *Social research methods*. 3rd ed. Oxford: Oxford University Press.
- Burgess, Alexis, and David Plunkett. 2013. "Conceptual Ethics I." *Philosophy Compass* 8 (12):1091–1101.
- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- Cappelen, Herman. 2023. *The Concept of Democracy: An Essay on Conceptual Amelioration and Abandonment*. Oxford: Oxford University Press.
- Cappelen, Herman, and David Plunkett. 2020. "A Guided Tour Of Conceptual Engineering and Conceptual Ethics." In *Conceptual Engineering and Conceptual Ethics*, edited by Herman Cappelen, David Plunkett and Alexis Burgess, 1–26. Oxford: Oxford University Press.
- Carnap, Rudolf. 1935. *Philosophy and Logical Syntax*. New York: American Mathematical Society.
- Carnap, Rudolf. 1937 [1934]. *The Logical Syntax of Language*. Translated by Amethe Smeaton (Countess von Zeppelin). 4th ed. London: Routledge & Kegan Paul.
- Carnap, Rudolf. 1945a. "On inductive logic." *Philosophy of Science* 12 (2):72–97.
- Carnap, Rudolf. 1945b. "The two concepts of probability." *Philosophy and Phenomenological Research* 5 (4):513–532.
- Carnap, Rudolf. 1947. "Probability as a guide in life." *Journal of Philosophy* 44 (6):141–148.
- Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.
- Carnap, Rudolf. 1956. "The Methodological Character of Theoretical Concepts." In *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, edited by Herbert Feigl and Michael Scriven, 38–76. University of Minnesota Press.
- Carnap, Rudolf. 1956 [1947]. *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd ed. Chicago: University of Chicago Press.

- Carnap, Rudolf. 1956 [1950]. "Empiricism, Semantics and Ontology." In *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd ed., 205–221. Original edition, *Revue Internationale de Philosophie* 4: 20–40.
- Carnap, Rudolf. 1958. *Introduction to Symbolic Logic and its Applications*. Translated by William H. Meyer and John and Wilkinson. New York: Dover Publications.
- Carnap, Rudolf. 1963a. "Intellectual Autobiography." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 3–84. LaSalle, IL: Open Court.
- Carnap, Rudolf. 1963b. "Replies and Systematic Expositions." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 859–1013. LaSalle, IL: Open Court.
- Carnap, Rudolf. 1966. *Philosophical Foundations of Physics*. New York: Basic Books.
- Carnap, Rudolf. 1967 [1928]. *The Logical Structure of the World*. Translated by Rolf A. George. Original edition, Los Angeles: University of California Press. Reprint, Chicago and La Salle, IL: Open Court, 2003.
- Carnap, Rudolf. 1987 [1932]. "On protocol sentences." *Noûs* 21 (4):457–470.
- Carus, A. W. 2007. *Carnap and Twentieth-Century Thought: Explication as Enlightenment*. Cambridge: Cambridge University Press.
- Carus, A. W. 2019. "Neurath and Carnap on Semantics." In *Neurath Reconsidered: New Sources and Perspectives*, edited by Adam Tuboly and Jordi Cat, 339–361. Springer Verlag.
- Chalmers, David. 2020. "What is Conceptual Engineering and What Should it Be?" *Inquiry: An Interdisciplinary Journal of Philosophy*.
- Chalmers, David J. 2011. "Verbal Disputes." *Philosophical Review* 120 (4):515–566.
- Christiano, Thomas, and Sameer Bajaj. 2022. "Democracy." In *Stanford Encyclopedia of Philosophy*.
- Christiano, Thomas. 2008. "Democracy." In *Stanford Encyclopedia of Philosophy*.
- Collier, David, and John Gerring. 2009. *Concepts and method in social science: the tradition of Giovanni Sartori*. London: Routledge.
- Cordes, Moritz, and Geo Siegart. 2018. "Explication." In *Internet Encyclopedia of Philosophy*, edited by James Fieser and Bradley Dowden.
- Cowling, Sam. 2013. "Ideological parsimony." *Synthese* 190 (17):3889–3908.

- Creath, Richard. 1990a. *Dear Carnap, Dear Van: The Quine–Carnap Correspondence and Related Work: Edited and with an Introduction by Richard Creath*: University of California Press.
- Creath, Richard. 1990b. “The Unimportance of Semantics.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1990:405–416.
- Creath, Richard. 2012. “Before explication.” In *Carnap's ideal of explication and naturalism*, edited by Pierre Wagner. Palgrave-Macmillan.
- Dahl, Robert A. 1971. *Polyarchy: participation and opposition*. New Haven, CT: Yale University Press.
- Dahl, Robert A. 2020 [1998]. *On Democracy*. Veritas paperback edition. New Haven, CT: Yale University Press.
- Daly, Chris. 2010. *An Introduction to Philosophical Methods*: Broadview Press.
- de Fine Licht, Karl, and Bengt Brülde. 2021. “On Defining ‘Reliance’ and ‘Trust’: Purposes, Conditions of Adequacy, and New Definitions.” *Philosophia* 49 (5):1981–2001.
- de Regt, Henk W., and Dennis Dieks. 2005. “A Contextual Approach to Scientific Understanding.” *Synthese* 144 (1):137–170.
- Deutsch, Max. 2020. “Speaker’s reference, stipulation, and a dilemma for conceptual engineers.” *Philosophical Studies* 177 (12):3935–3957.
- Dipert, Randall R. 1982. “Set-Theoretical Representations of Ordered Pairs and Their Adequacy for the Logic of Relations.” *Canadian Journal of Philosophy* 12 (2):353–374.
- Dupré, John. 2016. “Social Science: City Center or Leafy Suburb.” *Philosophy of the Social Sciences* 46 (6):548–564.
- Dutilh Novaes, Catarina, and Erich Reck. 2017. “Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization.” *Synthese* 194 (1):195–215.
- Ebbs, Gary. 2017. *Carnap, Quine, and Putnam on Methods of Inquiry*. Cambridge: Cambridge University Press.
- Eklund, Matti. 2014. “Replacing Truth?” In *Metasemantics: New Essays on the Foundations of Meaning*, edited by Alexis Burgess and Brett Sherman.
- Eklund, Matti. 2015. “Intuitions, Conceptual Engineering, and Conceptual Fixed Points.” In *Palgrave Handbook on Philosophical Methods*, edited by Christopher Daly.
- Eklund, Matti. 2017. *Choosing Normative Concepts*. Oxford: Oxford University Press.

- Eklund, Matti. 2021. "Conceptual Engineering in Philosophy." In *The Routledge Handbook of Social and Political Philosophy of Language*, edited by Justin Khoo and Rachel Sterken, 15–30. New York: The Routledge.
- Eklund, Matti. 2024. "Conceptual engineering and conceptual innovation." *Inquiry: An Interdisciplinary Journal of Philosophy* doi:<https://doi.org/10.1080/0020174X.2024.2384066>.
- Ereshefsky, Marc. 2000. *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge: Cambridge University Press.
- Estlund, David. 2008. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press.
- Foster, Jen, and Jonathan Ichikawa. 2023. "Normative Inference Tickets." *Episteme* 22 (1):1–27.
- Friedman, Michael. 1991. "The Re-evaluation of Logical Positivism." *Journal of Philosophy* 88 (10):505–519.
- Friedman, Michael. 2007. "Introduction: Carnap's Revolution in Philosophy." In *The Cambridge Companion to Carnap*, edited by Michael Friedman, and Richard Creath, 1–18. Cambridge University Press: Cambridge.
- Friedman, Michael. 2008. "Wissenschaftslogik: The role of logic in the philosophy of science." *Synthese* 164 (3):385–400.
- Gallie, W. B. 1956. "IX.—Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56 (1):167–198.
- Gerring, John. 1999. "What Makes a Concept Good? A Criterial Framework for Understanding Concept Formation in the Social Sciences." *Polity* 31 (3):357–393.
- Gerring, John. 2001. *Social Science Methodology: A Criterial Framework*. New York: Cambridge University Press.
- Gerring, John. 2011. *Social Science Methodology: A Unified Framework*. Cambridge: Cambridge University Press.
- Gerring, John, and Paul A. Barresi. 2009. "Culture: Joining Minimal Definitions and Ideal Types." In *Concepts and method in social science: the tradition of Giovanni Sartori*, edited by David Collier and John Gerring. Routledge.
- Goertz, Gary. 2006. *Social Science Concepts: A User's Guide*. Princeton and Oxford: Princeton University Press.
- Goertz, Gary. 2020. *Social Science Concepts and Measurement*. Princeton: Princeton University Press.

- Goodman, Nelson. 1963. "The Significance of *Der logische Aufbau der Welt*." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 545–558. LaSalle, IL: Open Court.
- Gustafsson, Martin. 2006. "Quine on explication and elimination." *Canadian Journal of Philosophy* 36 (1):57–70.
- Gustafsson, Martin. 2007. "Carnap och Quine om explikation som filosofisk metod." *Filosofisk Tidskrift* 4.
- Gustafsson, Martin. 2014. "Quine's Conception of Explication – and Why It Isn't Carnap's." In *A Companion to W. V. O. Quine*, edited by Gilbert Harman and Ernie Lepore, 508–525. Malden: Wiley-Blackwell.
- Hanna, Joseph F. 1968. "An Explication of 'Explication'." *Philosophy of Science* 35 (1):28–44.
- Hansson, Sven Ove. 2006. How to define: a tutorial. *Principios* 13 (19):05–30.
- Hansson, Sven Ove. 2018. "Formalization." In *Introduction to Formal Philosophy*, edited by Sven Ove Hansson and Vincent F. Hendricks. Cham, Switzerland: Springer.
- Haslanger, Sally. 2000. "Gender and race: (What) are they? (What) do we want them to be?" *Noûs* 34 (1):31–55.
- Haslanger, Sally. 2006. "What good are our intuitions: Philosophical analysis and social kinds." *Aristotelian Society Supplementary Volume* 80 (1):89–118.
- Hedden, Brian, and Jacob M. Nebel. 2024. "Multidimensional Concepts and Disparate Scale Types." *Philosophical Review* 133 (3):265–308.
- Hempel, Carl Gustav. 1952. *Fundamentals of Concept Formation in Empirical Science*. Edited by Carl Gustav Hempel. Vol. 2, *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.
- Hempel, Carl Gustav. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*: The Free Press.
- Hempel, Carl Gustav. 1966. *Philosophy of natural science*. Vol. 18: Prentice-Hall.
- Hempel, Carl Gustav. 1973. "The Meaning of Theoretical Terms: A Critique of the Standard Empiricist Construal." In *Logic, Methodology and Philosophy of Science IV*, edited by Patrick Suppes et al. Amsterdam: North-Holland Publishing Co.
- IAU, (International Astronomical Union). 2006. "IAU 2006 General Assembly: Result of the IAU Resolution votes." <https://www.iau.org/news/pressreleases/detail/iau0603/>.

- Isaac, Manuel Gustavo, Steffen Koch, and Ryan Nefdt. 2022. "Conceptual Engineering: A Road Map to Practice." *Philosophy Compass* 17 (10):1–15.
- Jorem, Sigurd. 2021. "Conceptual engineering and the implementation problem." *Inquiry: An Interdisciplinary Journal of Philosophy* 64 (1–2):186–211.
- Justus, James. 2012. "Carnap on concept determination: methodology for philosophy of science." *European Journal for Philosophy of Science* 2 (2):161–179.
- Keas, Michael N. 2018. "Systematizing the theoretical virtues." *Synthese* 195:2761–2793.
- Kemeny, John G. 1959. *A Philosopher Looks at Science*: Van Nostrand Reinhold Company.
- Kemeny, John G., and Paul Oppenheim. 1952. "Degree of factual support." *Philosophy of Science* 19 (4):307–324.
- King, Jeffrey C. 2016. "Philosophical and Conceptual Analysis." In *The Oxford Handbook of Philosophical Methodology*, edited by Herman Cappelen, Tamar Gendler and John Hawthorne. Oxford: Oxford University Press.
- Kitcher, Philip. 2008. "Carnap and the Caterpillar." *Philosophical Topics* 36 (1):111–127.
- Kitsik, Eve. 2020. "Explication as a strategy for revisionary philosophy." *Synthese* 197 (3):1035–1056.
- Klein, Carsten, and Steve Awodey. 2004. *Carnap Brought Home: The View from Jena*: Open Court.
- Knobe, Joshua, and Shaun Nichols. 2017. "Experimental Philosophy." *The Stanford Encyclopedia of Philosophy*.
- Koch, Steffen. 2021. "The externalist challenge to conceptual engineering." *Synthese* 198 (1):327–348.
- Koch, Steffen. 2023. "Why Conceptual Engineers Should Not Worry About Topics." *Erkenntnis* 88 (5):2123–2143.
- Krieger, Tommy. 2022. "Measuring Democracy." *ZEW - Centre for European Economic Research Discussion Paper* 22.
- Kroeber, Alfred Louis, and Clyde Kluckhohn. 1952. *Culture: a critical review of concepts and definitions*. Cambridge, MA: Peabody Museum Press.
- Kuhn, Thomas. 1977. "Objectivity, value judgment, and theory choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, 320–39. Chicago: University of Chicago Press.
- Kuipers, Theo A. F. 2007a. "Explication in Philosophy of Science." In *Handbook of the Philosophy of Science. General Philosophy of*

- Science – Focal Issues*, edited by Theo A. F. Kuipers, vii–xxiii. Elsevier: Amsterdam.
- Kuipers, Theo A. F., ed. 2007b. *Handbook of the Philosophy of Science. General Philosophy of Science - Focal Issues*. Amsterdam: Elsevier.
- Langford, C. H. 1942. “The Notion of Analysis in Moore’s Philosophy.” In *The philosophy of G. E. Moore*, edited by Paul Arthur Schilpp. Evanston, IL: Northwestern University Press.
- Lapointe, Sandra, and Christopher Pincock. 2017. *Innovations in the History of Analytical Philosophy*: Palgrave-Macmillan.
- Leitgeb, Hannes. 2013. “Scientific Philosophy, Mathematical Philosophy, and All That.” *Metaphilosophy* 44 (3):267–275.
- Leitgeb, Hannes, and André Carus. 2021. Rudolf Carnap. *The Stanford Encyclopedia of Philosophy*.
- Lewis, David. 1986. *On the plurality of worlds*. Oxford: Blackwell Publishing Ltd.
- Limbeck-Lilienau, Christoph, and Thomas Uebel. 2022. “Introduction.” In *The Routledge Handbook of Logical Empiricism*, edited by Thomas Uebel and Christoph Limbeck-Lilienau. New York: Routledge.
- Limbeck, Christoph, and Thomas Uebel, eds. 2022. *The Routledge Handbook of Logical Empiricism*: Routledge.
- Lipton, Peter. 2004. *Inference to the Best Explanation*. 2nd ed. London: Routledge.
- Loomis, Eric, and Cory Juhl. 2006. “Explication.” In *The philosophy of science: An Encyclopedia*, edited by Sahotra Sarkar and Jessica Pfeifer. New York: Routledge.
- Ludlow, Peter. 2014. *Living Words: Meaning Underdetermination and the Dynamic Lexicon*: Oxford: Oxford University Press.
- Maher, Patrick. 2007. “Explication Defended.” *Studia Logica* 86 (2):331–341.
- Maher, Patrick. 2010. “What is Probability? MS.” <http://patrick.maher1.net/preprints/pop.pdf>.
- Margolis, Eric, and Stephen Laurence. 1999. *Concepts: Core Readings*: MIT Press.
- Margolis, Eric, and Stephen Laurence. 2023. “Concepts.” In *Stanford Encyclopedia of Philosophy*.
- Martin, Michael. 1973. “The explication of a theory.” *Philosophia* 3 (2–3):179–199.
- May, Tim. 2011. *Social Research: Issues, Methods and Research*. Maidenhead: Open University Press.
- Menger, Karl. 1943. “What is Dimension?” *The American Mathematical Monthly* 50 (1):2–7.

- Metzger et al., Philip T. 2022. "Moons Are Planets: Scientific Usefulness Versus Cultural Teleology in the Taxonomy of Planetary Science." *Icarus* 374.
- Miller, George A. 1975. "Biographical Memoirs." In. Washington, D.C: The National Academies Press.
- Millikan, R. 2001. *On Clear and Confused Ideas: Cambridge Studies in Philosophy*.
- Murphy, Dominic. 2015. "Concepts of disease and health." In *Stanford Encyclopedia of Philosophy*.
- Murzi, Mauro. 2007. "Changes in a scientific concept: What is a planet?" *PhilSci Archive*.
- Nado, Jennifer. 2021. "Conceptual engineering, truth, and efficacy." *Synthese* 198:1507–1527.
- Olsson, Erik J. 2015. "Gettier and the method of explication: a 60 year old solution to a 50 year old problem." *Philosophical Studies* 172 (1):57–72.
- Olsson, Erik J. 2017. "Explicationist Epistemology and Epistemic Pluralism." In *Epistemic Pluralism*, edited by Annalisa Coliva and Nikolaj Jang Lee Linding Pedersen, 23–46. Palgrave Macmillan.
- Olsson, Erik J. 2021. "Explicationist Epistemology and the Explanatory Role of Knowledge." *Journal for General Philosophy of Science* 53:41–60. doi: 10.1007/s10838-020-09520-8.
- Oppenheim, Felix E. 1961. *Dimensions of Freedom. An Analysis*. New York: St Martin's Press.
- Oppenheim, Felix E. 1971. "Democracy – Characteristics Included and Excluded." *The Monist* 55 (1):29–50.
- Oppenheim, Felix E. 1981. *Political Concepts: A Reconstruction*. Oxford: Basil Blackwell.
- Pinder, Mark. 2017. "The Explication Defence of Arguments from Reference." *Erkenntnis* 82 (6):1253–1276.
- Pinder, Mark. 2019. "Scharp on inconsistent concepts and their engineered replacements, or: can we mend these broken things?" *Inquiry: An Interdisciplinary Journal of Philosophy* 66 (5):863–884.
- Pinder, Mark. 2020a. "Conceptual engineering, speaker-meaning and philosophy." *Inquiry: An Interdisciplinary Journal of Philosophy* 68:224–250.
- Pinder, Mark. 2020b. "On Strawson's critique of explication as a method in philosophy." *Synthese* 197 (3):955–981.
- Pinder, Mark. 2021. "Conceptual Engineering, Metasemantic Externalism and Speaker-Meaning." *Mind* 130 (517):141–163.

- Pinder, Mark. 2022a. "Is Haslanger's ameliorative project a successful conceptual engineering project?" *Synthese* 200 (4):1–22.
- Pinder, Mark. 2022b. "What Ought a Fruitful Explicatum to be?" *Erkenntnis* 87 (2):913–932.
- Queloz, Matthieu, and Friedemann Bieber. 2021. "Conceptual Engineering and the Politics of Implementation." *Pacific Philosophical Quarterly* (3):670–691.
- Quine, Willard Van Orman. 1957. "Speaking of Objects." *Proceedings and Addresses of the American Philosophical Association* 31 (3):5–22.
- Quine, Willard Van Orman. 1948. "On What There Is." *Review of Metaphysics* 2 (5):21–38.
- Quine, Willard Van Orman. 1951a. "Ontology and ideology." *Philosophical Studies* 2 (1):11–15.
- Quine, Willard Van Orman. 1951b. "Two Dogmas of Empiricism." *Philosophical Review* 60 (1):20–43.
- Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- Raab, Jonas. 2024. "Quine on explication." *Inquiry: An Interdisciplinary Journal of Philosophy* 67 (6):2043–2072.
- Reck, Erich H. 2007. "Frege–Russell numbers: analysis or explication?" In *The Analytic Turn*, edited by Michael Beaney, 33–50. Oxford: Routledge.
- Reck, Erich H. 2012. "Carnapian Explication: A Case Study and Critique." In *Carnap's Ideal of Explication and Naturalism*, edited by Pierre Wagner, 96–116. Palgrave-Macmillan.
- Reck, Erich H., ed. 2013a. *The Historical Turn in Analytic Philosophy*. London: Palgrave-Macmillan.
- Reck, Erich H. 2013b. "Introduction: Analytic philosophy and philosophical history." In *The Historical Turn in Analytic Philosophy*, edited by Erich H. Reck, 1–36. Palgrave-Macmillan.
- Reck, Erich H. 2024. "Carnapian Explication: Origins and Shifting Goals." In *Interpreting Carnap: Critical Essays*, edited by Alan Richardson and Adam Tamas Tuboly. Cambridge: Cambridge University Press.
- Reuter, Kevin, Catherine Herfeld, and Georg Brun. 2023. "Introduction to the topical Collection: Concept formation in the natural and social sciences: empirical and normative aspects." *Synthese* 201 (3):1–10.
- Richardson, Alan. 2013. "Taking the measure of Carnap's philosophical engineering : metalogic as metrology." In *The Historical Turn in Analytic Philosophy*, edited by Erich H. Reck, 60–77. Palgrave-Macmillan.

- Richardson, Alan. 2023. *Logical Empiricism as Scientific Philosophy*. Cambridge: Cambridge University Press.
- Risjord, Mark. 2012. "Models of culture." In *The Oxford Handbook of Philosophy of Social Science*, edited by Harold Kincaid, 387–408. Oxford: Oxford University Press.
- Rorty, Richard. 1967a. "Introduction." In *The Linguistic turn: essays in philosophical method*, edited by Richard Rorty, 1–9. Chicago: University of Chicago Press.
- Rorty, Richard, ed. 1967b. *The Linguistic turn: essays in philosophical method*. Chicago: University of Chicago Press.
- Rorty, Richard. 1984. "The Historiography of Philosophy: Four Genres." In *Philosophy in History*, edited by Rorty et al. Cambridge: Cambridge University Press.
- Rorty, Richard, J. B Schneewind, and Quentin Skinner, eds. 1984. *Philosophy in History*. Cambridge: Cambridge University Press.
- Salmon, Wesley. 1989. "Four Decades of Scientific Explanation. Introduction." *Minnesota studies in the philosophy of science* 13:3–19.
- Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *The American Political Science Review* 64 (4):1033–1053.
- Sartori, Giovanni. 1984. "Guidelines for Concept Analysis." In *Social Science Concepts: A Systematic Analysis*, edited by Giovanni Sartori. Beverley Hills, CA: SAGE Publications.
- Scharp, Kevin. 2013. *Replacing Truth*. Oxford: Oxford University Press.
- Scharp, Kevin. 2020. "Philosophy as the Study of Defective Concepts." In *Conceptual Engineering and Conceptual Ethics*, edited by Herman Cappelen, Alexis Burgess and David Plunkett, 396–416. Oxford: Oxford University Press.
- Schilpp, Paul Arthur, ed. 1963. *The Philosophy of Rudolf Carnap*. LaSalle, IL: Open Court.
- Schindler, Samuel. 2018. *Theoretical Virtues in Science: Uncovering Reality Through Theory*. Cambridge: Cambridge University Press.
- Schumpeter, Joseph A. 1942. *Capitalism, Socialism, and Democracy*. New York and London: Harper & Brothers Publishers.
- Schupbach, Jonah N. 2017. "Experimental Explication." *Philosophy and Phenomenological Research* 94 (3):672–710.
- Shepherd, Joshua, and James Justus. 2015. "X-Phi and Carnapian Explication." *Erkenntnis* 80 (2):381–402.
- Sider, Theodore. 2001. *Four-dimensionalism: An ontology of persistence and time*. Oxford: Oxford University Press.

- Simion, Mona, and Christoph Kelp. 2019. "Conceptual Innovation, Function First." *Noûs* 54 (4):985–1002.
- Sjögren, Jörgen. 2011. *Concept formation in mathematics*. Gothenburg: University of Gothenburg.
- Smith, James Andrew. 2021. "Carnap and Quine on Sense and Nonsense." *Journal for the History of Analytical Philosophy* 9 (10):1–28.
- Soter, Steven. 2006. "What Is a Planet?" *The Astronomical Journal* 132:2513–2519.
- Stevens, Stanley Smith. 1946. "On the theory of scales of measurement." *Science* 103 (2684):677–680.
- Strawson, Peter F. 1963. "Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy." In *The Philosophy of Rudolf Carnap*, edited by Paul Arthur Schilpp, 503–518. Open Court: La Salle.
- Strawson, Peter F. 1992. *Analysis and metaphysics: An introduction to philosophy*. Oxford: Oxford University Press.
- Suppe, Frederick. 2017. "Definitions." In *A Companion to the Philosophy of Science*, edited by W. H. Newton-Smith, 76–78. Oxford: Blackwell.
- Tal, Eran. 2015. "Measurement in Science." In *Stanford Encyclopedia of Philosophy*.
- Tarski, Alfred. 1944. "The semantic conception of truth and the foundations of semantics." *Philosophy and Phenomenological Research* 4 (3):341–376.
- Tengland, Per-Anders. 2007. "A two-dimensional theory of health." *Theoretical Medicine and Bioethics* 28 (4):257–284.
- Tengland, Per-Anders. 2008. "Empowerment: A Conceptual Discussion." *Health Care Analysis* 16 (2):77–96.
- Tillman, Frank A. 1965. "Explication and ordinary language analysis." *Philosophy and Phenomenological Research* 25 (3):375–383.
- Tyson, Neil deGrasse. 2009. *The Pluto Files*. New York: Norton and Co.
- Vessonen, Elina. 2021. "Conceptual engineering and operationalism in psychology." *Synthese* 199 (3–4):10615–10637.
- Wagner, Pierre, ed. 2012. *Carnap's ideal of explication and naturalism*. Basingstoke: Palgrave Macmillan.
- Wilkinson, L. 1999. "Statistical methods in psychology journals: Guidelines and explanations." *American Psychologist* 54 (8):594–604.
- Williamson, Timothy. 2021. "Degrees of Freedom: Is Good Philosophy Bad Science?" *Disputatio* 13 (61):73–94.

