Markus Schrenk

METAPHYSICS OF SCIENCE

A Systematic and Historical Introduction

ROUTLEDGE

Metaphysics of Science

Metaphysics and science have a long but troubled relationship. In the twentieth century the Logical Positivists argued metaphysics was irrelevant and that philosophy should be guided by science. However, metaphysics and science attempt to answer many of the same, fundamental questions: What are laws of nature? What is causation? What are natural kinds?

In this book, Markus Schrenk examines and explains the central questions and problems in the metaphysics of science. He reviews the development of the field from the early modern period through to the latest research, systematically assessing key topics including:

- dispositions
- counterfactual conditionals
- laws of nature
- causation
- properties
- natural kinds
- essence
- necessity.

With the addition of chapter summaries and annotated further reading, *Metaphysics of Science* is a much needed, clear and informative survey of this exciting area of philosophical research. It is essential reading for students and scholars of philosophy of science and metaphysics.

Markus Schrenk is Professor for Theoretical Philosophy at Heinrich Heine University, Düsseldorf, Germany. He is the author of *The Metaphysics of Ceteris Paribus Laws* (2007), co-author of *Einführung in die Sprachphilosophie* (2nd edition 2014), and editor of *Handbuch Metaphysik* (2016).

This page intentionally left blank

Metaphysics of Science

A Systematic and Historical Introduction

Markus Schrenk

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2017
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing in Publication Data
A catalogue record for this title has been requested

Library of Congress Cataloging in Publication Data
Names: Schrenk, Markus Andreas, 1974- author.
Title: Metaphysics of science: a systematic and historical introduction/
Markus Schrenk.
Description: 1 [edition]. | New York: Routledge, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2015049235| ISBN 9781844655922 (hardback: alk. paper) |
ISBN 9781844655939 (pbk.: alk. paper) | ISBN 9781315639116 (e-book)
Subjects: LCSH: Metaphysics. | Science–Philosophy. | Philosophy and science.
Classification: LCC BD111 .S326 2016 | DDC 110–dc23LC record
available at http://lccn.loc.gov/2015049235

ISBN13: 978-1-844-65592-2 (hbk)
ISBN13: 978-1-844-65593-9 (pbk)
ISBN13: 978-1-315-63911-6 (ebk)

Typeset in Times New Roman
by Sunrise Setting Ltd, Brixham, UK

Contents

Preface

This book's primary aim is to introduce the reader to some of the key concepts within contemporary metaphysics of science: dispositions, counterfactual conditionals, laws of nature, causation, properties, natural kinds, essence and necessity.

These topics are closely connected. Consider the following relations: that something has a *disposition* to do something – for example, that these white granules are soluble – means, roughly, that the following *counterfactual conditional* is true of them: they would dissolve if they were put into water. The explanation for this potential behaviour might well be that the powder belongs to the *natural kind* salt which has *essentially* an ionic structure, *NaCl*, that figures in multiple *laws of nature*. These laws govern the *causal processes NaCl* can be involved in – for example, being torn apart by H_2O dipoles into the anionic/cationic subparts *Na*+ and *Cl*−. Finally, some laws might tell us that such causal processes happen necessarily.

The core questions Yet what exactly is a law of nature? What is a causal process? When is a sentence 'if such-and-such were the case, then this-and-that would happen' true? These are the questions philosophy of science asks and especially in its metaphysical department. Its epistemic section focuses primarily on the question how scientific knowledge – for example, about nature's laws and her regular causal processes – is accumulated. In this book we will focus on the metaphysical 'what is the nature of XYZ?' rather than the 'how do we know of XYZ?' question.

Entanglement For someone new to the subject it is a challenge that all these topics – laws, causation, etc. – are not isolated but entangled: getting to grips with one area within the metaphysics of science presupposes prior acquaintance with another which hinges on a third, etc. Yet we can turn this situation, this entanglement, into an advantage. A short historical detour will show how. It will also introduce another goal of the book.

A brief history of metaphysics of science and the plan for the book The book's secondary aim is to acquaint the reader with the historical development of metaphysics of science beginning with the early modern period but especially throughout the twentieth and at the start of the twenty-first centuries. At the beginning of the twentieth century metaphysics as a whole fell into disrepute. Or, we should say, fell *once again* into disrepute. Early in the last century, the *Logical Empiricists* revived their eighteenth-century ancestors' metaphysics critique. They, as much as the ancestors, saw no merit in non-scientific theorising about the fundamental features of reality. Metaphysics, when understood as the philosophical investigation into the most basic, ultimate, fundamental features and structures of a reality that goes beyond what can be known via (experimental) observations and sensory experiences was thought to be pointless.

The Logical Empiricists declared that the only task left for philosophy/metaphysics is the clarification of concepts as they are used in the natural sciences. Particularly, philosophy should explicate how scientific sentences can be verified by observations and how scientific concepts can be reformulated in observational terms only. (These are terms like 'is red', 'is hot', etc. that refer to sense perceptions we experience directly.) It turned out, however, that satisfactory explications of this kind are not so easily available and that it is probably not possible for a multitude of scientific terms without making assumptions that were pejoratively labelled 'nonsensical metaphysics' by the Empiricists.

Dispositional predicates were an especially hard nut to crack. Dispositional properties – we have given the example of solubility; further dispositions are inflammability, elasticity, irascibility, etc. – have in common that many objects that have them only *potentially* act or react in a certain way when in certain circumstances. As long as something soluble is not put in water it does not show its solubility. Because of their only potential manifestation, they are properties of objects that are not directly observable.

Now, because the Empiricists' aim was to base all our factual knowledge on direct sense experience, they tried to translate what it means for a thing to be disposed (to be soluble, for example) into some statement that refers to observable things, properties and events only. Yet they failed.

History tells us that each attempt to amend the shortcomings of a prior analysis led the Empiricists either to acknowledge some metaphysical assumption and/or to refer, within their analyses, to one of the other core concepts within science, like *law of nature*, *causation, natural kind* and so forth. Needless to say, these other concepts were, consecutively, themselves in need of explication.

In this book we follow the history of the analysis of dispositions. Tracing it, we not only get acquainted, bit by bit, with analyses of *counterfactual conditionals*, *laws of nature*, *causation*, *natural kinds*, *essences* and *necessity* but also with the metaphysical assumptions that have gradually been taken on board and, thus, made the philosophical scene open to metaphysics again. In other words, the (history of the) analysis of dispositional predicates will be our golden thread running through the web of the entangled core issues of metaphysics of science. Here's an overview of what awaits the readers in the individual chapters.

The prologue will offer a **brief history and critique of metaphysics**, starting with seventeenth- and eighteenth-century *Rationalism*, continuing with the classical Empiricist critique thereof, then introducing Kant's reaction to both, *Transcendental Idealism*, and ultimately arriving at twentieth-century *Logical Empiricism* and its infamous Verificationist theory of meaning. Here, the Empiricists' reasons will be given in detail for their strong anti-metaphysical stance as mentioned above. The consequences and also shortcomings of Logical Empiricisms and Verificationism will occupy a whole subsection.

The second chapter, 'Dispositions', starts with the Logical Empiricists' first attempt to reduce dispositional predicates to a vocabulary that refers to observable properties and objects only. We get acquainted with the difficulties that hide behind innocent looking 'if… then…' sentences, which lead, later, to the topics of *counterfactual conditionals.* We will also unearth how *laws of nature, causation* and the *nature of properties* become relevant. Related to properties, we will also, in the second half of the chapter, follow an ontological turn and move from semantic analysis of dispositional predicates to the ontology of categorical and dispositional properties. The grand metaphysical edifice of *Humean Supervenience* will be introduced. In this chapter we will also highlight two important aspects which are associated with dispositions: their modality (MOD) and their productive responsibility (PROD). These two features will accompany us throughout the book.

Counterfactual conditionals, the topic of Chapter 2, are if-then sentences with an antecedent that is counter to the facts: 'I keep it tight in my hands, yet, *if I were to drop this sugar cube in water then it would dissolve*'. These counterfactual conditionals seem perfect for spelling out what we mean by an attribution of a disposition to an object (here: solubility). Yet it is not so easy to say when such conditionals are true. To give their truth conditions is the topic of this chapter. In preparation, but also as a worthwhile subject in its own right, we will introduce the reader to possible worlds semantics. This also gives us the opportunity to speak about several kinds of modal properties like conceptual, metaphysical or nomological necessity and their complements, the respective possibilities.

Also: **what laws of nature say, our topic in Chapter 4**, seems to have counterfactual impact. For example, the law of gravitation claims that all massive objects are attracted by other masses – thus, if I were to let loose this sugar cube it would fall. Yet what is a law of nature? A number of theories, starting with those inspired by Empiricism and ending with those offered within the past decade, will be introduced here. The topic of *Natural Properties* will be touched in passing.

Like the chapter on laws, **Chapter 5 on causation** acquaints us with the most important and most recent theories: when is it correct to say that one event *c* causes or has caused an event *e*, as my dropping the sugar cube into water caused it to dissolve?

The sixth chapter, 'Dispositional Essentialism', turns much of the story on its head: instead of trying to analyse dispositions in terms of counterfactual conditionals, in terms of laws of nature or in terms of causation – each respective chapter has its own section on such attempts – we show how philosophers have tried to spell out these other concepts in terms of the dispositional essences of natural kinds and properties. The chapter is organised in three parts: first, reasons for holding *Dispositionalism*, the stance that dispositions are respectable, unanalysable and real properties in their own right, are introduced; second, *Essentialism*, the idea that natural kinds have their features necessarily, is presented. The third section of the chapter unites the first two and unfolds the above-mentioned dispositional essentialist theories of counterfactual conditionals, laws and causation.

An epilogue called 'Meta-metaphysics' will engage with the latest self-conscious meta-reflection questioning again the tenability of (some) metaphysical assumptions and methods that have been so freely used in the past decades. Some worries in the style of the Empiricists' will re-emerge. **The afterword** reveals those areas within the metaphysics of science we were not able to cover for reasons of space. To avoid disappointment the reader might want to consult this final section of the book before starting with the first.

Reading manual

Throughout the book we will learn – learn by doing, so to speak – what metaphysics and especially what metaphysics of science is.

- All chapters and/or subsections (except those that are themselves a résumé) end with summaries in boxes like this one.
- The Preface and Afterword, Prologue and Epilogue, and the chapters Dispositions and Dispositional Essentialism are pairs and bracket the whole book.
- The dates of famous philosophers of the past are given where their names appear first in the text.

- There are cross-references (for example, Section 10.3) to other parts of the book where the same or an adjacent subject is covered.

Some definitions of metaphysics

- 'Metaphsyics is the study of ultimate reality.' (van Inwagen 1993: 1)
- 'Metaphysics is a philosophical inquiry into the most basic and general features of reality and our place in it.' (Kim and Sosa 1999: ix)
- '[I]ts central concern is with the fundamental structure of reality as a whole.' (Lowe 2002: 3)
- '[T]o characterize the nature of reality.' (Loux 2006: 10)
- 'Metaphysics is concerned with the foundations of reality. It asks questions about the nature of the world[.]' (Chalmers *et al*. 2009: 1)
- '[E]nquiry concerning the most general questions about the nature of reality including, for example, questions about the nature of matter, abstracta, fundamentality, space and time, and causation, law, necessity and probabiltiy – that at least captures metaphysics pretty well in extension.' (Ladyman 2012: 33)
- 'The most general attempt to make sense of things.' (Moore 2012: 1)
- See also: Kristie Miller's 'Metaphysics' in (Miller 2015: 193–236)

Further literature

- Tobin, E. (Forthcoming 2017) *Philosophy of Science: An Introduction to Contemporary Problems*. London and New York City: Continuum.
  (The latest up-to-date publication on our topics but seen from a more epistemic viewpoint. An ideal companion to our book.)
- Schurz, G. (2014) *Philosophy of Science: A Unified Approach*. Abingdon: Routledge.
  (This excellent overview combines a general introduction to philosophy of science with the author's own take on the unity of the sciences.)
- Psillos, S. (2002) *Causation and Explanation*. Chesham: Acumen.
  (An outstanding book focusing on the metaphysics of science; there's a large chapter on laws, too, although the title does not indicate this.)

- Okasha, S. (2002) *Philosophy of Science: A Very Short Introduction.* Oxford: Oxford Paperbacks.
  (What it says: ‘very short’, but great for its length.)
- Ladyman, J. (2001) *Understanding Philosophy of Science.* Abingdon: Routledge.
- Bird, A. (1998) *Philosophy of Science.* Montreal and Kingston: McGill-Queen’s University Press (both highly recommended books). Ladyman and Bird have substantial parts on epistemological issues related to the sciences.
- See also Kristie Miller (2015).

Bibliography

Chalmers, D., Manley, D. and Wasserman, R. (eds.) (2009) *Metametaphysics: New Essays on the Foundations of Ontology.* Oxford: Oxford University Press.

Kim, J. and Sosa, E. (eds.) (1999) *Metaphysics: An Anthology*. Malden, MA: Blackwell.

Ladyman, J. (2012) Science, Metaphysics and Method. *Philosophical Studies.* 160. pp. 31–51.

Loux, M. (2006) *Metaphysics: A Contemporary Introduction*. New York and London: Routledge.

Lowe, E. J. (2002) *A Survey of Metaphysics.* Oxford: Oxford University Press.

Miller, K. (2015) Metaphysics. In: D. Pritchard (ed.) *What Is This Thing Called Philosophy?* Abingdon: Routledge. pp. 193–236.

Moore, A. W. (2012) *The Evolution of Modern Metaphysics: Making Sense of Things*. Cambridge: Cambridge University Press.

Van Inwagen, P. (1993) *Metaphysics.* Oxford: Oxford University Press.

Acknowledgements

I would like to thank everyone who contributed to the completion of this publication. Without your help the book would have been immeasurably more difficult to write.

I started the manuscript as a Postdoctoral Research Fellow within the AHRC-funded Metaphysics of Science Project (AH/D503833/1) at Birmingham, Bristol and Nottingham. Most parts have been written during my Postdoc years within the DFG-funded Metaphysics of Science project (FOR 1063) at Aachen, Cologne, Düsseldorf and Münster. I thank the funding bodies for their financial support. To all members of the two projects: it has been a privilege to be able to work with you!

For numerous helpful discussions of metaphysical and related matters I am grateful to David Armstrong, Helen Beebee, Alexander Bird, Dieter Birnbacher, Clandia Blöser, Elke Brendel, Carl Craver, Kristina Engelhard, Alexander Gebharter, Stuart Glennan, Rudolph Glitz, David Hommen, Peter Hucklenbroich, Andreas Hüttemann, Benedikt Kahmen, Marie I. Kaiser, Max Kistler, Charlotte Matheson, Jennifer McKitrick, Gerhard Müller-Strahl, Daniel Plenge, Nigel Sabbarton-Leary, Gerhard Schurz, Markus Stepanians, Emma Tobin and Matthew Tugby.

Amongst the many people – colleagues and students alike – who encouraged me with their comments on earlier drafts are: Alexander Anthony, Florian Boge, Alexander Christian, Mats Dietrich, Arno Goebel, Till Gallasch, Julia F. Göhner, Vera Hoff mann-Kolss, Siegfried Jaag, Daniel Koch, Marc Lange, Christian Loew, Anne Nospickel, Alexander Reutlinger, Oliver R. Scholz, Marko Tešić, Matthias Unterhuber, Barbara Vetter and Ken Waters (many of whom could also have been named in the paragraph above).

Thanks go also to two anonymous referees, not only for their enormously helpful advice but also for their kind and encouraging way of phrasing their critique. Needless to say: remaining inaccuracies are solely my own responsibility.

Bastian Walter (first and foremost) but also Sara Ipakchi helped me with all formal matters: the bibliography, formatting, layout, etc. Without their work the book could not have been printed.

The book's cover shows an artwork by the artist Susanne Britz. I am very grateful to her that she allowed the publisher to reproduce her work.

Speaking of Routledge: I am indebted to Tony Bruce, Adam Johnson, Rebecca Shillabeer, Lucy Vallance and (at Acumen) Steven Gerrard for their support and endless patience. Thanks go, too, to Tim Hyde (at Sunrise Setting) for a splendid job in improving my English and preparing the proofs.

Achim, Thomas, Julia, Elisabeth and Franz-Josef helped me with all that is non-academic.

This page intentionally left blank

1 Prologue

A brief history of metaphysics

Rumours have it that one of the founding fathers of Logical Empiricism – Rudolf Carnap (1891–1970) – never went to sleep without a copy of Friedrich Nietzsche's (1844–1900) *Thus Spoke Zarathustra* under his pillow. In this book, Nietzsche does not only have Zarathustra declare that God is dead, he also lets Zarathustra put great emphasis on the insight that all that has ever happened will recur infinitely many times. With great admiration for Nietzsche, Carnap might have read, night by night, about the eternal recurrence of the same, and other Nietzschean themes. And, yet, Carnap did not, in some sense, believe a word of what that philosopher wrote. How is this possible?

Carnap and many of his colleagues within the so called Vienna Circle, a group of scientifically minded thinkers, which held its meetings in the Austrian capital in the 1920s, believed that most philosophical problems are meaningless and especially that a certain kind of metaphysical question should be eliminated from any rational or, at least, any scientific discourse altogether. The Vienna Circle's goal, to foster the scientific world-view,[1] was perceived to be in harsh contrast to any cloudy metaphysical inquiry.

Thus, when Nietzsche presents the doctrine of the eternal recurrence of the same without basing it on prior scientific investigation, i.e. without the support of empirical observation and experiments, it must, by the standards of the Vienna Circle, be judged to be speculative metaphysical nonsense on the same level of obscurity as statements about the absolute spirit or claims concerning the essence of being and nothingness.

Interestingly, there is, in current physics, a theoretical model of an oscillatory universe that, over and over again, begins with a big bang and ends with a big crunch and, thus, goes through the same events infinitely.[2] Should this theory prove to be empirically true the Logical Empiricists could start to believe in Nietzsche's proclamation of the eternal recurrence for scientific reasons.

Their scientific-mindedness does, of course, only indicate and not justify why the Vienna Circle was so hostile to almost all metaphysical questions. Indeed, the above paragraphs make Carnap, Moritz Schlick

(1882–1936), Otto Neurath (1882–1945), Friedrich Waismann (1896–1959) and other members[3] seem to be a group of spoilsports who reject a long-standing tradition of philosophy for dubious reasons.

However, these philosophers' challenging convictions are, as we shall see, embedded into an intriguing philosophical world-view and an exciting research programme – namely, Logical Empiricism, or Neo-Positivism. In what follows in this chapter and also in the rest of the book, we will become acquainted with the Logical Empiricists' invigorating philosophical manifesto and why for Carnap and others it seemed worthy of believe and defence. We will get to know why, surprisingly, this research programme was, in the twentieth century, the origin of a discipline we today call the Metaphysics of Science, despite the fact that the name alone might have caused the Logical Empiricists a headache. As we go along, we will also find an answer to the apparent contradiction that Carnap should be an enthusiastic Nietzsche devotee even though these two thinkers' philosophical methods and aims are so dissimilar.

To appreciate Logical Empiricism's place within the development of philosophical thought we must look at the philosophical history from which it arose. I have chosen, somewhat arbitrarily and to keep the chapter shorter than it could be, to start recounting this development with the early modern period. (Some of the metaphysics of the Scholastics and also the Ancients will find its way into the book later – see *universals* and *powers* – in Sections 4.4, 6.2 and 6.4) In this introduction we will start with Rationalism, continue with the classical Empiricists' response, turn to Kantian reconciliation of the two and ultimately arrive at the Logical Empiricists, who we discuss in more detail.

In the remainder of the book we will consecutively rebuild some of the metaphysical edifices the Empiricists left in ruins. Throughout, but especially in the final chapter, we will reflect on the legitimacy of this reconstruction scheme.

1.1 Rationalism

Those philosophers with whom we start our brief historical overview, namely the so called *Rationalists* like René Descartes (1596–1650), Baruch Spinoza (1632–77) and Gottfried Wilhelm Leibniz (1646–1716), thought that *metaphysical investigations* reveal the most fundamental structure of the world, that which Goethe's Faust would describe approximately 200 years later as what 'girds the world together in its inmost being' and on which all other reality depends.[4] Equally important, the Rationalists believed that a metaphysical inquiry should deliver its insights *with absolute certainty*. It should deliver *first principles* about which there cannot be any doubt.

Note that this latter issue, *certainty*, is an *epistemic interest*, i.e. a requirement regarding our *knowledge* of metaphysical truths, whereas the former issue, the fundamental structure part, is an *ontological concern*, i.e. one focusing on *what there is* and *how it is organised*.

Concerning the epistemic part, Descartes, for example, urges us in his *Discourse on the Method*[5]

> never to accept anything for true which [we] did not clearly know to be such [. . .] and to comprise nothing more in [our] judgment than what was presented to [our] mind so clearly and distinctly as to exclude all ground of doubt.
>
> (Descartes 1637: Part II.7)

Certainty not established by the senses For the goal of achieving absolute certainty, the Rationalists found one possible source of knowledge, namely sense perception, to be wanting. It is a far too unreliable resource because our senses can be and have been frequently deceived, as, for example, optical illusions show. Two of Descartes' famous examples are a round tower that, from afar, looks square, and a huge statue that seems small from a certain distance. Descartes continues: 'In these and countless other such cases, I found that the judgements of the external senses were mistaken' (Descartes 1641: Meditation VI: 76 (53)).[6] Worse, when we are vividly dreaming or hallucinating, we only believe we perceive something real but actually do not see, hear, feel, etc. any truly existing object at all (as, again, most famously described by Descartes in his *Meditations on First Philosophy* 1641, in the dream scenario: Meditation I: 18–20 (12–14)). And, so, Descartes judges the senses negatively in that he concludes 'it is prudent never to trust completely those who have deceived us even once.' (Descartes 1641: 18 (12)).

Because of the doubtfulness of sensory perceptions, the Rationalists watched out for a different source of (metaphysical) knowledge and they thought to have found it in pure reason:

> Accordingly, if there is any certainty to be had, the only remaining alternative [to the senses, MS] is that it occurs in matters that are clearly perceived by the intellect and nowhere else.
>
> (Descartes 1641: 145 (105))

Axioms and deduced theorems The Rationalists model for such pure rational knowledge beyond doubt was mathematics and their reference point was Euclid's treatise on the *Elements*, in which Euclid (360–280 BC) axiomatised geometry. Axiomatisation means that Euclid was able to deduce mathematically a large number of geometrical truths from only a limited number (ten, to be precise) of fundamental axioms.

(In mathematics, one speaks of *axioms* when one means these first principles and of *theorems* to refer to the propositions deduced from these axioms.) A well-known example for one of the unquestioned primary statements, the axioms, is the fifth, the *parallel postulate*, which says that through a point that is not on a given line only one further line can be drawn that is parallel to the first. An example for one of the theorems, i.e. those sentences which follow from the ten fundamental axioms, is Pythagoras' theorem: if a, b and c represent the lengths of the sides of a right-angled triangle, c being the longest side, then $a^2+b^2=c^2$.

Descartes now writes:

> arithmetics, geometry and other subjects of this kind, which deal only with the simplest and most general things, regardless of whether they really exist in nature or not, contain something certain and indubitable. For whether I am asleep, two and three added together are five, and a square has no more than four sides. It seems impossible that such transparent truths should incur any suspicion of being false.
>
> (Descartes 1641: 20 (14))

Descartes, Leibniz and the other Rationalists believed that the certainty of mathematics could be achieved in philosophy, too. Metaphysical truths were thought to be logically deducible by pure reason as theorems from a few fundamental axioms. For the aim of deduction, Leibniz started to devise a logical calculus (for example, in his 1666 *Dissertation on the Art of Combinations* (in Leibniz 1969)) that should ultimately allow us to proceed with philosophical language and philosophical problems as mathematics does with numbers or geometrical figures. What is possible in mathematics should also be possible with philosophical and metaphysical concepts and words. So that, ideally, 'when there are disputes amongst persons, we can simply say: Let us calculate [lat. *calculemus*], without further ado, and see who is right' (Leibniz 1685: 51). As we shall see later, the enormous importance put on logic is a feature which unites the (in other respects very different) Logical Empiricists with Leibniz.

First principles Two questions present themselves: first, which are the metaphysically *first principles* that correspond to the mathematical axioms; and, second, where do they get their authority from? We can find five or six such metaphysically basic principles in Leibniz (there is room for dispute here about which principle shall be counted and which not), at least two of which will be of concern later: the *Principle of Contradiction* which states that 'a proposition cannot be true and false at the same time, and that therefore A is A and cannot be not A' (Leibniz 1989: 321) and the *Principle of Sufficient Reason* which states that there is no event without a sufficient cause for it, i.e. necessarily, anything that happens has been brought about by something else. The latter principle, says Leibniz, 'must be considered one of the greatest and most fruitful of all human knowledge, for upon it is built a great part of metaphysics, physics, and moral science' (Leibniz 1969: 227), thereby highlighting again the axiomatic structure of metaphysics. Indeed, the *Principle of Sufficient Reason* clearly lies at the heart of or is presupposed by all scientific endeavour.

Still, how do we know this and the other principles to be true? By and large, Leibniz and the other Rationalists simply found them to be self-evident and not in need of any further argument (this holds, by the way, also for the mathematical axioms). Descartes' way of putting this is to count as first principles only what is perceived 'clearly and distinctly' by the intellect (Descartes 1641: 35 (24)). As an example, he gives his famous *cogito* argument: 'I think therefore I am'. When confronted by Pierre Gassendi (1592–1655) with the allegation that we could be deceived even when we believe we know *clearly and distinctly,* Descartes denied this. He offered as a reliable test for clear and distinct knowledge that when we consider it we cannot doubt it (Descartes 1641: 145 (105)), and this is allegedly the case with the *cogito*. Whether this is satisfactory or a *petitio principii* might well be questioned, for wasn't indubitable truth (= certainty) our goal in the first place? Although this allegation of circularity might be right, it is not the route of critique we want to follow here.[7]

Monadology Rather, we wish to have an exemplary, brief look at what kind of overall metaphysical theory supposedly follows from, say, Leibniz's axioms. We cannot trace the whole argumentation here and Leibniz's deductions are not in all cases as transparent as they should be, even by his own standards. Apologies are nonetheless due for we will clearly do some injustice to Leibniz's grand oeuvre in presenting it here somewhat oversimplified. With this precautionary note, we nonetheless allow ourselves to say that his *Monadology* (Leibniz 1714) is an eccentric example of metaphysics that provides an easy way to unfold what later anti-metaphysicians found so doubtful in Rationalist, speculative metaphysics.

So, here it comes: for Leibniz, the ultimate building blocks of the world are what he calls *monads*: atom-like, simple substances with at least basic 'mental' capacities that allow them to perceive the world and to desire or will or have an 'appetite' for particular ends. Leibniz's reason for this stipulation is that he wanted physical matter to be itself a source of causal activity: 'A Substance is a being capable of action' (Leibniz 1989: 207) and 'we can show from the inner truths of metaphysics that what is not active is nothing' (Leibniz 1686: 64).[8] Now, because for Leibniz only something like minds can originate such activity, the world must ultimately consist in avid monads.

BOX 1.1 Rationalism

- The **Rationalists**, like Descartes (1596–1650), Spinoza (1632–77) and Leibniz (1646–1716) **distrusted the senses** as guides to metaphysical knowledge.
- This knowledge was supposed to be about the **fundamental nature of the world**, its basic building blocks and its structure.
- Moreover, it should be gained with **absolute certainty**: something the senses could not deliver – **reason** alone could.
- The Rationalists thought it was possible to **deduce from a few clear and distinct, indubitable truths** (as axioms) all the other truths about the fundamental structure of the world.
- Leibniz's ***Monadology*** is a prime example of a metaphysical system of the Rationalist kind.

1.2 Empiricism

Commit it to the flames In the preface to the first edition of his book on Leibniz, Bertrand Russell (1872–1970) writes: 'I felt – as many others have felt – that the *Monadology* was a kind of fantastic fairy tale, coherent perhaps, but wholly arbitrary' (Russell 1900: xxi).[9] This is precisely what the next philosopher we need to consider might have felt too: David Hume (1711–76), one of the greatest critics of metaphysics, enters the stage at the height of metaphysical speculation of the *Monadology* kind. He writes if not in direct reaction to Leibniz then certainly to the Rationalists' metaphysics as a whole:

> If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.
>
> (Hume 1748: Sect. XII, Part III§132:165)

If it was up to Hume, Leibniz's *Monadology* would have seen the flames. But why exactly? In order to understand the thrust of Hume's outburst – especially to apprehend the two questions he asks and answers negatively, and also why he asks precisely these two – we need to make a little detour.

Hume's impressions and ideas Early on in his take on philosophy, Hume distinguishes between *impressions* and *ideas*. Impressions are all kinds of sense experiences, i.e. the 'lively perceptions, when we hear, or see, or feel, or love, or hate, or desire, or will' (Hume 1748: Sect. II, §12: 18).

Ideas, now, fall into two categories: (1) some of them are recollections of impressions, the memory of the experience of a red spot, say, or a sour taste. Ideas are also said to subsume classes of resembling experiences. So, the idea red gathers memories of all red-impressions. As memories of sense impressions ('copies', as Hume also calls them) ideas are said to be less forceful or lively than the original experiences. Yet this aspect is less significant for us here.

Next to these elementary ideas, which were copied from simple impressions and which also subsume them, there are, (2), complex ideas that are compositions of the elementary ideas. A compound idea could, for example, be that of an apple which is composed out of the simpler ideas round, red, juicy, sweet, sour, etc. Even abstract ideas, and also those that lack a full counterpart in reality, like that of a unicorn, are still said to be composed of elementary ideas that ultimately relate via the simplest ideas to the sense impressions we actually had. In short, there is a hierarchy of ideas, at the base of which are recollections of simple impressions[10] and at the top are ideas that are composed, maybe in a rather complex way, of simple ideas.

> Our thoughts or ideas, however compound or sublime, we always find that they revolve themselves into such simple ideas as were copied from a precedent feeling or sentiment [impression, MS].
>
> (Hume 1748: Sect. II, §14: 19)

The latter need and can have no further analysis:

> These [simple] impressions are all strong and sensible. They admit not of ambiguity. They are not only placed in a full light themselves, but may throw light on their correspondent ideas, which lie in obscurity.
>
> (Hume 1748: Sect. VII, Part I, §49: 62)

Word meanings and sense experiences There is also a semantic aspect of Hume's impressions and ideas theory. For, even if Hume nowhere presents us with a fully worked out philosophy of language – that is, a theory of what words mean – he implicitly identifies *word meanings* with *ideas*: in his *Enquiries*, for example, we find formulations like 'among different languages [. . .] it is found that words expressive of ideas [. . .] do [. . .] correspond to each other' (Hume 1748: Sect. III, §18: 23).[11]

When such an identification is made, i.e. that the meaning of a word is an idea, then we can claim that any word that has a meaning is ultimately linked to some sense experience(s), namely those simple impressions the corresponding idea is ultimately connected to. This word-impression connection might be complicated, for not all words have elementary ideas associated with them. Still, eventually – maybe via complex interrelations of ideas to other, simpler ideas which do connect to simple sense impressions – there is a link from each meaningful word to some perception(s).

Now, putting together everything we have just learned about words, ideas and impressions, we arrive at the starting point of one aspect of Humean metaphysics critique: words that do not have any idea associated, that can ultimately be decomposed into the simplest ideas/impressions, are meaningless and confused. It comes as no surprise that, for Hume, many philosophical/metaphysical terms are candidates for such meaningless words. Hume states:

> When we entertain, therefore, any suspicion that a philosophical term is employed without any meaning or idea (as is but too frequently), we need but enquire, *from what impression is that supposed idea derived?* And if it is impossible to assign any, this will serve to confirm our suspicion.
>
> (Hume 1748: Sect. II, §17: 22)

Examples of meaningless terms are, for Hume, *substance*, *the self* and *soul*. None of these words correspond to an idea that is either directly copied from a sense impression (we have, for example, not yet seen a soul) or that can be analysed into simpler ideas that correspond to direct sense impressions. On these grounds alone, Leibniz's monads could be criticised, and so we have arrived at the first way in which a Humean metaphysic-critique operates.

Relations of ideas vs. matters of fact There is a second albeit related way. We initially started with Hume's famous 'commit it to the flames' quote, where he commands us to incinerate any metaphysical oeuvre if the two questions he asks about it are answered negatively. Having now introduced what Hume means by ideas and impressions, we can begin to make sense of that quote.

According to Hume, all possible human knowledge falls in precisely two categories: 'to wit, *Relations of Ideas and Matters of Fact*' (Hume 1748: Sect. IV, Part I: 25).[12] This explains already the number of his questions: any knowledge is about either of these two kinds – no more, no less. There is no other kind of knowledge to be had.

Items which fall in the first category, i.e. the Relations of Ideas, are, Hume continues, 'discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe' (Hume 1748: Sect. IV, Part I: 25). As his primary example for *Relations of Ideas* he takes mathematics ('quantity or number'), and mathematical truths are, indeed, known by mere operation of thought. Hume quotes, for instance, Pythagoras' theorem. From considerations we find elsewhere in his work it is clear that, next to mathematics, conceptual truths like 'all bachelors are unmarried' or 'my sister is female' are also to be subsumed under the heading *Relations of Ideas*.[13] Here, too, we already know from considering in our mind the concept or idea *bachelor* that bachelors are unmarried. This relates back to what we learned about ideas (and impressions) earlier: the analysis of the idea *bachelor* into simpler ideas, *man* and *unmarried*, delivers the outcome without us having to research empirically the social life of bachelors. Mere operation of thought yields the result.[14]

Matters of Fact, i.e. facts about what the world is like, are, on the contrary, known to us only by the 'testimony of our senses' (Hume 1748: Sect. IV, Part I, §21: 26): that there is an apple on the table, for example. Such knowledge cannot be gathered by mere thought: no merely mental analysis of the idea *apple* and of the idea *table* will reveal to us that there is one in front of us on the table. Only the actual impressions that we have can reveal this to us.

As already noted, the dichotomy of *Relations of Ideas* and *Matters of Fact* is exhaustive: there is no further kind of knowledge and no further way to gain it. Crystal-ball reading as much as divine revelation is not accepted as an epistemic resource.

Consequences for (Rationalist) metaphysics Now, how is all this in conflict with Descartes and Leibniz? Well, Hume vehemently denies that anything but truths of mathematics and conceptual truths are discoverable by the mere operation of thought. Yet these truths are not facts about what the world is like: 'Our reason, unassisted by experience, [cannot] ever draw any inference concerning real existence and matter of fact' (Hume 1748: Sect. IV, Part I: 27). And so, according to Hume, the Rationalists' grand project to gain metaphysical knowledge, i.e. knowledge of the fundamentals of the world, by pure thought is bound to fail.

If we want to gain knowledge about the world we need to use our senses. That, however, was unwelcome to the Rationalists, for, remember, next to wanting to get at the fundamental structure of the world they wanted to get at it with absolute certainty. However, perceptual knowledge, which does tell us about the world, as opposed to conceptual and mathematical knowledge, is, unfortunately, uncertain and fallible: 'our evidence of their truth, however great, [is not] of a like nature with the foregoing' (Hume 1748: Sect. IV, Part I, §21: 25). This is a dilemma: in the only place where certainty can be gained, in *Relations of Ideas*, there is nothing to be found about what the fundamental structure of the world is like; and where we find the latter, in *Matters of Fact,* certainty cannot be had. Thus, metaphysics – as an inquiry that, first, wishes to establish the most fundamental truths about the world and, second, to do this with absolute certainty – is not possible.

Rationalist metaphysics, so we conclude with Hume, hangs in mid-air. Instead of being inferred from indubitable axioms, metaphysical results are foggy speculation:

> But this obscurity in the profound and abstract philosophy [i.e. metaphysics, MS], is objected to, not only as painful and fatiguing, but as the inevitable source of uncertainty and error. Here indeed lies the justest and most plausible objection against a considerable part of metaphysics, that they are not properly a science; but arise either from the fruitless efforts of human vanity, which would penetrate into subjects utterly inaccessible to the understanding, or from the craft of popular superstitions, which, being unable to defend themselves on fair ground, raise these intangling brambles to cover and protect their weakness.
>
> (Hume 1748: Sect. I, §6: 11)

The Empiricist doctrine This ends our overview of the *general* way in which Hume made the metaphysical foundations of Rationalist philosophy shake and crumble. Before we see his critical instruments and his anti-metaphysical attitude at work in a concrete case – causation – a general term for Hume's philosophy has to be introduced: *Empiricism*. Empiricism quite simply contrasts with Rationalism and is a credo held not only by Hume but also by his predecessor Empiricists John Locke (1632–1704) and George Berkeley (1685–1753). Empiricism, to put it briefly, is the doctrine that all our ideas/concepts and all knowledge about the world derive from sense experience and from sense experience alone.

A prime example: causation We turn now to our concrete example, an example, actually, that will accompany us throughout the book: Hume's famous and influential views on causation. At that time, the orthodox view was that causation – say, between the event (the cause) that one moving billiard ball bumps into another and the event that the second ball starts rolling (the effect) – is a kind of necessitation: *the cause necessitates its effect*; *the effect must happen, given its cause*. Thomas Hobbes[15] (1588–1679), for example, writes 'all the effects that have been, or shall be produced, have their *necessity* in things antecedent' (Hobbes 1655: 9.5, emphasis added) and, similarly, Baruch Spinoza: 'From a given determinate cause an effect *necessarily follows* [logically or conceptually, MS]' (Spinoza 1677: Axiom 3, emphasis added).

Now, by necessity two matching things were meant. The first interpretation situates necessity in the abstract or mental realm as a matter of the inconceivability of the opposite: 'It cannot be conceived but that the effect will follow' (Hobbes 1655: 9.7). The second interpretation makes necessity a worldly connection amongst events: the first billiard ball's bumping into the other is a necessitating, driving force for the second's movement.

Although we can distinguish these two meanings of necessity, the Rationalists equated the two: 'The order and connection of ideas is the same as the order and connection of things' (Spinoza 1677: Part II, Prop. 17), i.e. abstract necessity and the driving force in nature coincide or are even seen as more or less one thing.

Note that a typical instance of the inconceivability of the opposite (the first interpretation) can be found in conceptual truths: in no way can someone be a bachelor and married, because what it means to be a bachelor is to be unmarried. Thus, it is inconceivable that there could be a married bachelor. Now, if the Rationalists were right that causal relations are of that kind then neither could there be a cause, a billiard ball bumping into another, without its effect, i.e. the second one starting to roll.

Having distinguished between *Relations of Ideas* and *Matters of Fact* Hume can now criticise this belief in causation as a necessary connection in a twofold way.

Causal necessitation is no Relation of Ideas First, he points out that a causal link is not discoverable through reason alone. There is no reasoning by which we can *deduce* effects from causes:

> The mind can never possibly find the effect in the supposed cause, by the most accurate scrutiny and examination. [. . .] A stone or piece of metal raised into the air, and left without any support, immediately falls: but to consider the matter *a priori* [i.e. merely considering the *Relations* of the *Ideas* 'matter', 'air', etc.] is there anything we discover in this situation which can beget the idea of a downward, rather than an upward, or any other motion, in the stone or metal?
>
> (Hume 1748: Sect. IV, Part I: §25: 29)[16]

Moreover, pace Hobbes, Spinoza and the other Rationalists, it *is* conceivable that something else and not the expected effect happens. Here Hume speaks of our billiard balls bumping into each other:

> May I not conceive, that a hundred different events might as well follow from that cause? May not both these balls remain at absolute rest? May not

> the first ball return in a straight line, or leap off from the second in any line or direction? All these suppositions are consistent and conceivable. Why then should we give the preference to one, which is no more consistent or conceivable than the rest? All reasoning *a priori* will never be able to show us any foundation for this preference.
>
> (Hume 1748: Sect. IV, Part I: §25: 29–30)

So, causal connections do not belong to the realm of *Relations of Ideas*, i.e. causation is after all no necessary relation of ideas. The first meaning we gave to necessity fails. Here, the Rationalists were clearly wrong.[17]

Causal necessitation is no matter of fact Are, then, causal relations at least a matter of a necessary wordly affair, i.e. can we find necessary connections amongst *Matters of Fact*? More precisely, is there an *Impression* (or are there *Impressions*) from which the (possibly complex) idea of a necessary causal connection can be extracted?

> We must consider the idea of causation, and see from what origin it is deriv'd. [. . .]. Let us therefore cast our eye on any two objects, which we call cause and effect, and turn them on all sides, in order to find that impression, which produces an idea of such prodigious consequence.
>
> (Hume 1739–40: Book I, Part III, Sect. II: 75)

Hume continues, of course, to argue that there is no such impression. A necessary connection is not discoverable by the senses: we only see one billiard ball moving, then the other, but we do not perceive the causal necessity with which that allegedly happens. There is no impression of the senses that is the impression of causal necessity or power or force:

> The scenes of the universe are continually shifting, and one object follows another in an uninterrupted succession; but the power or force, which actuates the whole machine, is entirely concealed from us, and never discovers itself in any of the sensible qualities of body. [. . .] External objects as they appear to our senses, give us no idea of power or necessary connection.
>
> (Hume 1748: Sect. IV, Part I: 63–4)

Hence, neither of the seemingly possible ways of establishing a necessitating causal link is successful: it is not discoverable within the Relations of Ideas nor are there suitable impressions that would reveal some Matter of Fact about causal necessity. This explains Hume's severe scepticism when it comes to the existence of causal necessity.[18] Before we continue with what remains of causation if we follow the Humean path, it is important to highlight two things that have been going on in the background of the argument.

1. **Conceivability and possibility.** Hume, like Hobbes, identified *conceivability* with *possibility* when he proved that causation is no relation of ideas and, thus, no matter of conceptual necessity. That is, his argument hinges on the identification of the conceivable with the possible: it is *conceivable* that the first billiard ball bumps into the second but the second does not move. Therefore, it is *possible* that the first but not the second happens; further, that what possibly does not happen cannot be *necessary*, for if it were necessary, it would definitely happen. Thus, because it is conceivable that the first happens without the second there is no necessity that the second must happen when the first does.[19]
2. **Production, causation, necessitation and necessity.** Consider the following cascade of statements: '*c causes e*', '*c necessitates e*', '*it is necessary that when c then e*'. Moving from one statement to the next seems to be a natural thing to do and, once causation is identified with necessity, it is plausible that causation cannot be observed simply because necessity cannot be observed: our senses only register *what is* but not *what must be the case, what is necessarily so*.

Next to the cascade of statements from *c causes e* to *c necessitates e* to *it is necessary that when c then e*, the Rationalists, Hume and many subsequent metaphysicians have implicitly or explicitly endorsed the following, equally plausible chain: *c* produces *e*, *c* brings about that *e*, *c* causes *e*. Linking the two chains at their common joint – *c causes e* – we smoothly move from *production* via *causation* to *necessity*.[20]

The assumption that there is a link between causal production and necessity is still prevalent. Many modern day Humeans and anti-Humeans alike believe in it. The Humeans, of course, reject causal production because of the connection to necessity, and the anti-Humeans try to prove that there is worldly necessity (and thus causation) after all.

There is a third way, though, one that became visible only much later in history. It asks us to keep only the first link from production to causation (and vice versa) but to cut the connection which production had to necessitation/necessity. This view urges us to conceive of production in a different way, one that is more akin to, say, enforcing (rather than necessitating). This is meant quite literally in the sense of Newtonian forces: that a push against the table might very likely move it forward. If it is forceful enough, it almost certainly will. Yet, this is only *almost certainly* so, for when there is a counterforce it will not. Now, think of causal production this way and not in terms of indomitable necessity, i.e. the heavy burden inherited from its Rationalist origin. Then, maybe, a causal link is observable because such a production view of causation, devoid of necessity, is immune to Hume's no-necessity attack in matters of fact (see Schrenk 2011 and 2014).

For now, however, we will go along with the production–causation–necessitation identification (we return to the alternative in Chapter 6) and make an important and revealing remark about the future development of the metaphysics of science: metaphysics critique is, due to Hume's excellent example of causation, often tied to the critique of necessities in nature. With their role model, Hume, in the background, today's anti-metaphysicians try, first and foremost, to avoid assumptions about any kind of necessity or other modality in nature. Phrased the other way round, the still ongoing assaults on necessity are, justifiably or not, thought to be a fight against bad speculative metaphysics in general (see Section 7.8).

Back to causation What remains, then, after Hume's attack, of our concept of causation when there is no necessary production? For, undoubtedly, we still make causal claims in the sciences and in everyday life and cannot simply drop it from discourse. Hume has two separate stories to tell. For one thing, alleged *causal necessity* shrinks, for him, to a mere fact about human psychology: we are accustomed to interpret certain regularities causally and *expect very strongly* that events of certain kinds that have succeeded other kinds of event in the past will also do so in future. There is, indeed, that strong expectation, that vivid feeling we have *in us*, but there is still no real necessity out there in the unfolding of events in the world:

> Either we have no idea of force or energy, and these words are altogether insignificant, or they can mean nothing but that *determination of the thought*, acquir'd by habit, to pass from the cause to its usual effect.[21]
>
> (Hume 1739-40: 657, from the later published abstract to the Treatise, emphasis added)

In other words, while there is no impression and idea of causal necessity that derives from some observations *of the world*, there is, still, the impression of the habitual transition *in our minds* from one kind of perception to another (see Moore 2012: 110 for this interpretation). So, *necessary causal connection* is not an entirely meaningless term but it means something entirely different from what we thought it means. To repeat: *necessary connection* does not refer to anything outside but to the mere habitual feeling of anticipation in us.

Here is a metaphor for what we have just said: when two objects are glued together we might come to know this because, (1), we see the glue between them or, (2), we infer from the fact that these kinds of objects usually stick together that these two exemplars are also glued. That is, in the second case we come to judge that they stick together without perceiving the glue. We are simply informed about a regularity. Now, Hume's fairly negative account of causation from above says that, when it comes to causation, there is no glue: there is no such connecting stuff in the world. At best, because of the regular observation of the co-occurrence of two events, we are trained to strongly expect them to happen together. Thus, the only glue that exists exists in our head, evoked by custom and habit.

A regularity theory of causation Hume does not end his thoughts on causation here. He still wonders what in the world rather than merely in our heads could make causal claims true. For a solution, he capitalises on the above-mentioned regular co-occurrence of alike events. Indeed, Hume ultimately offers us a tripartite definition of causation in which necessary connections do not play any role any more and in which regular co-occurrences do all the work (see also Hume 1739–40: Abstract of *A Treatise of Human Nature*: 649–50). He says:

> An actually occurring event c is a cause of an actually occurring event e if and only if:
>
> (1) **[Contiguity] c is spatially in contiguity with e:**
>
> > 'I find in the first place, that whatever objects are consider'd as causes or effects, are contiguous. [. . .] We may therefore consider the relation of contiguity as essential to that of causation.'
> >
> > (Hume 1739–40: Book I, Part III, Section II: 75)
>
> (2) **[Succession] e happens temporally after c:**
>
> > 'The second relation I shall observe as essential to causes and effects [. . .] [is] that of priority of time in the cause before the effect.'
> >
> > (Hume 1739–40: Book I, Part III, Section VI: 75–6)
>
> Again (1) and (2) together:
>
> > 'Like objects have always been plac'd in like relations of contiguity and succession.'
> >
> > (Hume 1739–40: Book I, Part III, Section VI: 88)
>
> (3) **[Regularity] all events of the same type as c are followed in spatio-temporal succession, as in (1) and (2), by events like e:**
>
> > 'a cause [is] an object, followed by another, and where all objects similar to the first are followed by objects similar to the second.'
> >
> > (Hume 1748: Sect. VII, Part II, §60: 76)

As an example, we might again think of a billiard ball *c* bumping into a billiard ball *e*, whereupon *e* starts rolling.

This concludes our brief introduction to Hume's thoughts on matters of causation. We will return to them later and also to a further Humean idea for causation, the counterfactual analysis, in Section 5.3. There, we will critically assess not only Hume's but also many other theories of causation.

We end now this very short summary of Hume's philosophy by drawing attention to the fact that he was not only extremely critical of metaphysics – he uses the term *metaphysics* often in a disapproving way (as in the final section of his *Enquiries*) – but that he was also very self-consciously cautious to avoid speculations he could not substantiate by empirical input (as his thoughts on causation attest).[22] Whether he always succeeded is controversial but the intention was clearly there.

BOX 1.2 Empiricism

- The **Empiricists** – our focus has been on David Hume (1711–76); John Locke (1632–1704) and George Berkeley (1685–1753) – believed that **all our knowledge about the world** derives from sense experience and **from sense experience alone**.
- Hume distinguished between such **matters of fact** and **relations of ideas**: the latter can be known by pure thought but only because they are merely about **word meanings**, like 'all bachelors are unmarried'. **Mathematics and logic** are also subsumed amongst relations of ideas. They too yield no knowledge about the world.
- **Metaphysics**, as the Rationalists conceived of it, namely as the pure rational inquiry which yields **certainties** about the **fundamentals of nature** is, accordingly, not possible.
- All research into **what the world is like** has to go **via the senses** but the **senses are fallible,** and **thought about words or concepts** discovers **nothing worldly** although it might deliver certainties.
- Thus, **Rationalist metaphysics** lives nowhere and is **obscure speculation**.
- Hume uses **causation** as his prime example for a metaphysics-laden concept. He shows how the **orthodox interpretation of it as necessary connection in nature** falls prey to his metaphysics critique. However, he also offers an attempt to show how **causation could become matter-of-factual**: his regularity theory.

1.3 Transcendental Idealism

We can describe *Transcendental Idealism* as a mediator between Rationalism and Empiricism. Immanuel Kant (1724–1804) was one of the first philosophers to fully realise how devastating Hume's assault on Rationalist metaphysics was. In the introduction to his *Prolegomena to Any Future Metaphysics* Kant reports appreciatively that Hume awakened him from his 'dogmatic slumber', and he continues that Hume gave his own 'investigations in the field of speculative philosophy a completely different direction' (Kant 1783: 260). Yet, rather than leaving Rationalist philosophy completely behind, Kant understood his *transcendental philosophy* as a mediator between Empiricism and Rationalism.

Empiricism claims that all our knowledge about the world derives from sense experience *and from sense experience alone*. The italicised part is an aspect we have so far neglected. It says that sensory perceptions, completely on their own, deliver insights into the world: that is, without, for example, contribution of the mind or intellect. In other words, for Empiricists, sense perceptions are not only necessary but also sufficient to acquire factual knowledge. Locke alluded to the sufficiency part with his now famous comparison of the mind to a blank slate, a *tabula rasa*, onto which sensory experiences add data without any further aid or input:

> Let us then suppose the Mind to be, as we say, white Paper void of all Characters, without any *Ideas*; How comes it to be furnished? Whence comes it by that vast store which the busy and boundless Fancy of Man has painted on it, with an almost endless variety? Whence has it all the materials of Reason and Knowledge? To this I answer, in one word, From *Experience*: in that all our knowledge is founded; and from that it ultimately derives itself.
>
> (Locke 1690: Book II, chapter I, §2: 104)

Thoughts without content are empty; intuitions without concepts are blind Now, Kant rejects this latter facet of Empiricism that the intellect plays no active role when we gain perceptual knowledge. In this respect, Kant makes concessions to the Rationalists: even in the acquisition of knowledge about *Matters of Fact*, to use that Humean phrase again, the mind is involved. Our intellect *does* make a significant contribution to what we perceive. Only if what we perceive is processed by our mind can we speak of knowledge acquisition. (We come to how precisely that is supposed to work shortly.)

Yet Kant does agree with Empiricism in the necessity part, i.e. he also believes that pure reason alone (in Hume, 'mere operation of thought') does not, unaided by the senses, have the power to accumulate knowledge about the world. Observations with our senses are indispensable.

Kant's famous slogan 'thoughts without content are empty, intuitions without concepts are blind' (Kant 1781/1787: A51/B75) summarises his position well: mere operations of the intellect ('thoughts') without empirical input are vacuous, but mere sense impressions ('intuitions') without the intellect's assistance do not accumulate knowledge either.

We return to our overall theme, metaphysics and its critique, for we can now reveal Kant's idea of what metaphysics could be. According to Kant, the one and only possible field of metaphysical exercise is to find out exactly the role the intellect plays in the formation of (perceptual) knowledge. Note that this change of perspective is a kind of 'Copernican revolution' (as Kant himself calls it) because metaphysical principles are no longer interpreted as posits about what the fundamentals of the world itself are like. (To establish those would be impossible and such claims nonsensical. Here, again, Kant is in agreement with Hume.) Rather, the task of metaphysics is to make transparent the ordering principles with which our mind structures our sense experiences. In other words, we now look inwards not outwards, plus we turn to a considerable degree towards epistemology, the theory of knowledge.

Here are some examples: 'Every event has a (deterministic) cause', 'Nature is uniform', 'Physical space is Euclidean'. These are, according to Kant, not findings about the structure of the world itself but about how our intellect organises sense perceptions. In fact, for Kant, there is no choice here: our mental set-up is such that our mind automatically and unalterably does and must pre-structure everything our senses reveal to us in the way the three exemplary principles just given have it.

Transcendental Idealism From here, a possible route for us to go further would be to focus on what Kant calls *Transcendental Idealism*, one part of which is the thought that the world, as it is in itself, is forever concealed from us: we can only see it through our native lenses. We cannot change them and pick others, nor can we get in touch with things in themselves in an unmediated way (Kant uses the term *noumenon* for the *thing in itself*). This is not Idealism in the strongest sense (the view that the external world does not exist but only our mind and its ideas): Kant agrees, there is a world. Yet the world is given to us only through appearances, never immediately, and our perceptions of it are heavily impregnated by our own ingredients. *Transcendental* is the notion Kant uses to signify that *our faculty of cognition* (Kant 1781/1787: B25) is his concern: not the world but *how we cognise it* (see Moore 2012: 121).

Categories of understanding Kant unearths twelve *categories of the understanding* (*Verstandeskategorien* in German), i.e. basic concepts our faculty of cognition operates with. These twelve fall into four groups: *quantity*, *quality*, *relation* and *modality* (see Kant 1781/1787: A79–80/

B105–6). It is less important for our purposes how Kant arrives at these categories than two other things: first, that the categories are features of our intellect (concepts it operates with) that make judgements about what we perceive to be possible in the first place instead of being extrapolated afterwards from our sense experiences. The categories belong to the preconditions of the possibility of empirical knowledge; they are not learned from perceptual experiences. Second, amongst the above-mentioned relational categories is the principle of causation. In other words, the principle of causation too belongs to the preconditions of the possibility of perceptual knowledge.

Pure intuitions of receptivity Distinct from these *categories of the understanding* (which are the ordering mechanisms of the *intellectual, conceptual side of our judgements*) are the *pure intuitions of receptivity* (*reine Formen sinnlicher Anschauung* in German). The latter pre-structure our perceptions and thus operate more directly on the *experiential side of our knowledge acquisition* than on the conceptual one. What are the pure intuitions or receptivity? Space and time! The things we perceive are spatio-temporal, i.e. in space and time, because we read space and time into our perceptions. According to Kant, space and time are not features of the world as it is in itself.

It becomes clear again why Kant calling his approach a 'Copernican revolution' is apt: he reverses the order of what is perceived of the world and what is projected into it:

> Up to now it has been assumed that all our cognition must conform to the objects [. . .] [Let] us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition [. . .] This would be just like the first thoughts of Copernicus, who, when he did not make good progress in the extrapolation of the celestial motion if he assumed that the entire celestial host revolves around the observer, tried to see if he might not have greater success if he made the observer revolve and left the stars at rest.
>
> (Kant 1781/1787: BXVI)

The transcendental method But how do we uncover these *preconditions of the possibility of cognitive achievements*? What is the path to this kind of knowledge? Kant calls the method with which we can arrive at such judgements *transcendental deduction*. Transcendental arguments (as they have also been called[23]) have the following general form: we show that A is a necessary precondition for the possibility of B. Then, because B indeed happens to be the case, we can logically deduce that A exists as well, for B could not have been without A. Here is a sketch of two examples (do not worry, for now, whether they have any validity): scientific research into planetary orbits presupposes that we think/perceive space and time to be Euclidean and absolute – that is, as a rigid container with the three coordinates (length, width and height) in which events happen in temporal succession. If our intellect were not to project these features automatically onto the perceived world our observations of the planets (through telescopes) would not deliver knowledge. So, transcendental argumentation yields judgements like 'physical space is Euclidean'. Second, going back to Descartes, we might want to reconstruct his famous *cogito* argument as a transcendental argument: Descartes claims he cannot doubt that he exists. Why so? Because if it were not true that he exists he could not think or doubt that he exists. His existence is a necessary precondition for the possibility of him thinking or doubting.

Synthetic versus analytic judgements We can put the way in which Kant exceeds Hume in another framework. We said that Hume divides things we can know into *matters of fact* (what the world is like) and *relations of ideas* (abstract mathematical and conceptual truths). Kant agrees to a high degree with this pair but he introduces a further dichotomy so that, in total, four combinations are possible. One of the four will turn out to be empty. Yet this still leaves Kant with three and not just the Humean two boxes.

First, Kant distinguishes between 'synthetic' and 'analytic' judgements. He writes:

> In all judgments in which the relation of a subject to the predicate is thought [. . .] this relation is possible in two different ways. Either the predicate B belongs to the subject A as something that is (covertly) contained in this concept A; or B lies entirely outside the concept A, though to be sure it stands in connection with it. In the first case, I call the judgment analytic, in the second synthetic.
>
> (Kant 1781/1787: A6–7)

Thus, 'All singers are musicians' is an analytic statement: the concept 'musician' belongs to the concept 'singer'. 'Ian Bostridge is a singer', however, is a synthetic statement: the concept 'singer' is not contained in the meaning of the proper name 'Ian Bostridge'. The distinction synthetic–analytic only almost corresponds to Hume's matters of fact/relations of ideas. It is very important to see that it does not do so entirely.

The *a priori* versus the *a posteriori* In order to appreciate the difference we need to turn to Kant's second dichotomy: truths known 'absolutely independently of all experience and even of all impressions of the senses' (Kant 1781/1787: B2–3) and, second, truths known *with* the help of the senses. That is, we need to distinguish further between truths known *a priori* and things known *a posteriori*.

Having done so, four possibilities principally emerge from the Kantian double dichotomy:

(1) *Synthetic a posteriori* judgements, where the predicate does not already contain the subject and where, thus, observations are necessary to know whether they are true – that there is an apple on the table, for example.
(2) *Analytic a priori* judgements: because analytic truths already emerge from the concepts they are composed of there is no need for discovery by the senses, and, so, they are *a priori* – take the 'all bachelors are unmarried' example.
(3) The third combination, *analytic* and *a posteriori*, is an empty class, for, again, what is already contained in concepts does not need to be discovered by the senses.
(4) The fourth combination is the famous Kantian *synthetic a priori*, where mathematical truths[24], amongst others, can be found. More details on that important class of judgements follow later.

Note, first, where there is agreement between Hume and Kant. The *synthetic a posteriori* corresponds fairly well to Hume's *matters of fact*. Also, when it comes to conceptual analytic truths, like 'All singers are musicians', there is concordance between the two philosophers: our intellect alone can, in an *a priori* fashion, i.e. without the aid of our senses, reveal these truths. They are known without sensory experiences because, as per Kant, their subject terms contain their predicate terms or, as per Hume, the second ideas are already contained within the first ideas. Also, there is agreement on the certainty with which the latter truths are known. As Kant writes, they are based on the logical principle of contradiction, 'for the predicate of an affirmative analytic statement is already thought in the concept of the subject, *of which it cannot be denied without contradiction*' (Kant 1783: 267, emphasis added).

The synthetic *a priori* The disagreement between Kant and Hume is in the realm of *synthetic a priori* knowledge. Sure, there are matters of fact that are known *a posteriori*, via the senses (here they agree), yet Kant recognises also synthetic ('new') knowledge that is acquired without the aid of our senses, i.e. that is acquired *a priori*, by pure reason.[25]

Kant's first example for the synthetic *a priori* is mathematics: mathematical proofs, according to Kant, go clearly beyond what numbers and mathematical functions mean. That $e^{i\pi} = -1$, is, for example, not immediately obvious by considering e, i, π, and –1. Kant's own example is the much simpler equation 7 + 5 = 12, where 12 'is by no means already thought merely by thinking of that unification of seven and five [. . .] One must go beyond these concepts' (Kant 1781/1787: B15–16). So, mathematics is synthetic (*new* or *unexpected*, if you wish) but still *a priori* because you do not need your senses.[26]

Yet if mathematics reveals some interesting synthetic truths in an *a priori* fashion, why should there not be other such kinds of knowledge? Indeed, with Kant, we know such truths already: transcendental arguments, for example, deliver such genuinely new knowledge. It is a synthetic not an analytic statement that space is Euclidean: what the words 'space' and 'time' mean does not analytically deliver that what we perceive is necessarily structured in a Euclidean way. Still, this is not something that we have learned from experience, *a posteriori*, either. Rather, seeing things through the spectacles of Euclidean space is the precondition for (sense-)experiencing the world and its objects:

> *The conditions of the possibility of experience* in general are at the same time conditions of the possibility of the objects of experience, and [. . .] for this reason they have objective validity in a *synthetic a priori judgement*.
> (Kant 1781/1787: AI58/BI97, italics added)

Metaphysics is possible in Transcendental Arguments The fundamental difference between Hume and Kant is, thus, that for Kant, but not for Hume, some synthetic judgements can be known *a priori*, and it is precisely here where Kantian metaphysics has its habitat. In the chapter *Solution of the General Question [. . .] 'How Is Metaphysics Possible as Science?*' (Kant 1783: 365–71) Kant explicitly states that all synthetic *a priori* judgements together constitute a new realm for respectable substantive metaphysics. Sometimes he also says that all metaphysics is transcendental philosophy, by which he means philosophy that concerns the synthetic *a priori*. Again, additionally to mathematics, we can enquire in a synthetic *a priori* fashion into the preconditions for the possibility of perception, i.e. the ways in which our experience of the world is prestructured, and other such aspects of our cognitive apparatus.

This completes our general take on Kantian metaphysics.[27] Just as we proceeded with Hume, we now consider Kant's views on causation in order to see his metaphysics at work. For Kant, as for Hume, causation is nothing in the world as it is in itself. In fact, any judgement about what the world or things-in-themselves really are – 'noumena', as Kant calls them – is meaningless for him. Only bad metaphysics would claim to have knowledge thereof (see, for example, Kant 1781/1787: A369).

Yet, against Hume, we do not happen to make causal judgements merely out of habit *after* we have been confronted with and become accustomed to regular occurrences like billiard balls bumping into each other. Rather, according to Kant, we can only perceive billiard-ball collisions as such because we interpret what we see in a causal manner. Possessing the concept of the connection of cause and effect is necessary to be able to make sense experiences of the billiard-ball case in the first place. Naked perceptions, not ordered by a causal structure, would be a meaningless mess for us. They would be so disorderly that no habit could even arise from them. (Note that Kant has no need to deny that causation happens with regularity. In fact, he endorses that part of Hume's theory.)

Thus, the rough picture is this: the world and its objects do affect our senses. Yet in order for us to make sense of these perceptions of the world our immanent cognitive apparatus (pre-)structures these experiences for us: we perceive events as happening in space and time in causal succession. This is a claim about our cognition, not the world itself. Knowledge about how our cognition structures experience is located within the realm of the synthetic *a priori*: it is *a priori* because we do not need experiential input to acquire it, and yet it is synthetic because it tells us something new about our cognitive apparatus.

BOX 1.3: Transcendental Idealism

- Immanuel Kant (1724–1804) is a negotiator between the **Rationalists** and the **Empiricists**. He agrees with the latter that knowledge about the empirical world can only derive from sense experiences. However, more in accordance with the former, he also points out that our mind too has a contribution to make in that **our cognitive apparatus pre-structures our experiencing.**
- The pre-structuring is done both by the **categories of the understanding**, which are the ordering mechanisms of the conceptual side of our judgements, and the **pure intuitions of receptivity**, which pre-structure our perceptions and thus operate more on the experiential side of our knowledge acquisition.
- To consider not only what is perceived but also to focus on the perceiver's cognitive set-up makes the term ***Copernican Revolution*** apt for Kant's epistemology.
- Another central pillar of Kant's philosophy is a double dichotomy. First, the semantic one: **synthetic–analytic**; and, second, the epistemic one: ***a priori–a posteriori***.
- A judgement 'A is B' is **analytic** if 'the predicate B belongs to the subject A as something that is (covertly) contained in this

concept A' ('all bachelors are unmarried' is an example); and when 'B lies entirely outside the concept A' the judgement is **synthetic**.

- Truths known 'absolutely independently of all experience and even of all impressions of the senses' are ***a priori***, and those known with the help of the senses ***a posteriori***.
- In opposition to Hume, **metaphysics** is (again) possible for Kant: all **synthetic judgements *a priori*** together constitute respectable metaphysics. **Transcendental arguments**, i.e. those that concern the conditions for the possibility of empirical knowledge, are a means to arrive at such judgements.

1.4 Logical Empiricism

1.4.1 Aufbau (Construction)

The purpose of this section is to introduce Logical Empiricism's key ideas and to highlight its relations to classical Empiricism and Kantian philosophy. In part 1.4.2 we focus on the shortcomings of Logical Empiricism[28], especially on the problems of one of its central pillars: Verificationism.

Relations to Kant Quite some time passes between Kant and the Logical Empiricists. Many important philosophical works were written in the meantime. Yet it is fair to say that Carnap, Schlick, Neurath and other twentieth-century Empiricists almost completely ignored the grand figures in the interim – for example, the German Idealists Johann Gottlieb Fichte (1762–1814), Friedrich Wilhelm Joseph Schelling (1775–1854) and Georg Wilhelm Friedrich Hegel (1770–1831). Where they did refer to figures from this school of philosophy they did so pejoratively. Kant, however, was seen as an immensely important thinker and the Logical Empiricists were highly influenced by him. For example, they agreed with Kant that metaphysics, as an inquiry into what the world in itself is like, is impossible. This is not to say that they were Kantians – on the contrary, in important respects they were not – but that they referred implicitly or explicitly to Kantian themes.[29]

Dropping the synthetic *a priori* (as a possibility for metaphysics) A major discrepancy is that the Logical Empiricists dropped the initially attractive synthetic *a priori* as a non-empty category of knowledge (see Carnap 1928/1998: §106). Remember that Kant believed that we can find out *a priori*, by mere thought, how our mind conditions our perceptions – for example, that we perceive all things happening in a Euclidean space embedded into a deterministic causal nexus and that this is a truth about us, not the world; that it belongs to the unshakeable pillars of our cognition: we cannot see anything but through that causal lens as contained in three dimensional space.

The Empiricists denied that this can be right. They did so for the following reason: the revolutionary physical theories at the turn of the twentieth century contradicted several of Kant's synthetic *a priori* principles. First, Einstein's relativity theory postulated that space-time is non-Euclidean, especially that the *parallel postulate* does not hold. Also, the latent idea of space being absolute, i.e. being a rigid container in which things are located and events happen, had to be given up. Second, as quantum mechanics tells us, some events happen not with a clear deterministic cause but spontaneously and with a certain probability only. Now, if the to-date most successful scientific theories speak against Kant's Euclidean space and against the postulate that every event has a deterministic cause then what were thought to be indubitable synthetic judgements *a priori* are, in fact, unstable, revisable assumptions. We were able to formulate these revolutionary scientific theories and to develop them in the light of experimental, observational findings. Thus, according to the Logical Empiricists, it cannot be true that the intellect or our cognitive capacities, as Kant thought of them, are fixed and inflexible. So, after all, these alleged synthetic *a priori* truths did not constitute the conditions for the possibility of empirical knowledge, especially of scientific knowledge.[30] Therefore, the Logical Empiricists again started to accept only the analytic (not the synthetic) *a priori*.

Naturalising the understanding and anti-metaphysics As a corollary to this difference with Kant, the Logical Empiricists sympathised with the naturalistic thought that can be found in the philosophy of David Hume – namely that the human brain and the workings of our mind are just as much objects for empirical research as any other entity in the world. If true then we can empirically discover why creatures like us make certain 'metaphysical' assumptions about the world. That might be, for example, because making these assumptions gives us an advantage in evolutionary fitness. Causal thinking could be a case in point. Therefore, Kant's categories of the understanding and the pure intuitions of receptivity become not only revisable but are objects of empirical research, especially of cognitive and perceptual psychology, and not of *a priori* transcendental arguments.

These departures from Kant have an immediate consequence for the possibility of metaphysics. If there is no synthetic *a priori*, where metaphysical claims could be located, then there is, after all, no place for them.

What Kant said about our example case, causation – namely that causal thinking belongs to the preconditions for the possibility of empirical knowledge – was dropped again in favour of a broadly Humean conception: first, the fact that we categorise certain events within a causal matrix is only a contingent truth about human cognition and, second, that a definition of *c* causes *e* can be given in terms of c- and e-event regularities: 'Questions about the "inner nature of the causal relations" that go beyond the discovery of certain regularities in the successions of events [are senseless]' (Carnap 1931: 237/167, my translation).

Relations to classical Empiricism After the comparison to Kant, let us now ask what the relation of logical to classical Empiricism is and why it bears the attribute *logical*. All Empiricists, old and young, share the core doctrine that knowledge about the world originates in sense perception and in sense perception alone. The major advance from seventeenth- and eighteenth-century Empiricism to twentieth-century Neo-Empiricism is an even stronger concentration on language. (Remember that we extracted already some semantic theses about word meaning from Hume and Locke.)

Logic and language As we shall see, the Logical Empiricists take these theses about language to the extreme: the Empiricist epistemic doctrine about knowledge will be fully remoulded in *semantic terms*. The possibility of that step arises mainly because of advances in formal logic and the logical analysis of language in the nineteenth and early twentieth centuries, as, for example, in Gottlob Frege's (1848–1925) *Begriffsschrift* from 1879 which bears the subtitle: *A Formal Language of Pure Thought Modelled upon that of Arithmetic*. Also extremely influential was Bertrand Russell's and Alfred North Whitehead's (1861–1947) *Principia Mathematica* from 1910–13. Bertrand Russell commented later:

> Modern analytical Empiricism [. . .] differs from that of Locke, Berkeley, and Hume by its incorporation of mathematics and *its development of a powerful logical technique*. It is thus able, in regard to certain problems, to achieve definite answers, which have the quality of science rather than of philosophy.
> (Russell 1945: 834, emphasis added)

Thus, the aspiration to clarity and simplicity of thought and the focus on language as an instrument for philosophical rigour became a driving force, even the defining criterion of Logical Empiricism (and, in fact, also of one of its grandchildren, namely *Analytical Philosophy* (see Dummett 1993: chapter 2)).

There are two main ways in which the Logical Empiricists exceeded the language affinity we found in classical Empiricism:

(1) The first is *the central pillar of Logical Empiricism*: the *verification principle of meaning*. This is a principle not about what

singular words mean (see Hume and Locke) but about the meanings of whole sentences.

(2) *Logical syntax* – that is, roughly, the grammar of sentences – is discovered as another means to criticise 'metaphysical nonsense'. We turn to Verificationism almost immediately; logical syntax has to wait a little longer.

Rationalism First, a brief final word on the relation of Logical Empiricism to Rationalism is in order: Logical Empiricism inherits its classical Empiricist ancestors' animosities against Rationalism but with one exception:[31] the above-mentioned revival and major advancement of Leibniz's *calculemus*!

Verificationism about sentence meaning The verification criterion of sentence meaning claims that a sentence – 'There is a red apple on the table', say – has meaning – that what we cognitively have to grasp in order to understand the sentence – if and only if we can specify the observation that would prove that sentence right or wrong, i.e. the observation that would verify or falsify that sentence.[32] More precisely, the criterion says something even stronger, namely not only that a sentence *has meaning* but that the method of its verification by means of observation *is its meaning*:

> The meaning of a sentence *is* the method of its verification. [. . .] A sentence that can not eventually be verified, is not verifiable at all; it then lacks meaning altogether.
>
> (Waismann 1930–1: 229, my translation and emphasis)

In the apple example above it seems fairly easy to describe an observation that would verify that sentence. 'There is a red apple on the table' is indeed a meaningful sentence: we can judge that it is true, roughly, if and only if we have a red-and-round apple impression when looking at the table. Otherwise the sentence is false. But no matter whether true or false, it is definitely meaningful because there is a method to test for it.

From epistemology to semantics Note something important: Verificationism aims to cast the Empiricists' epistemic doctrine that all factual knowledge comes from sense perception as a semantic doctrine. Indeed, if we believe that what we know is expressed (or at least expressible) in meaningful sentences then the transition from Empiricist epistemology to semantics is straightforward: all factual knowledge is expressed in meaningful sentences; only those sentences for which we are able to give a method of verification in observation are meaningful.

We cannot emphasise strongly enough that Verificationism does not simply say that it is quite useful to be able to provide a method for the verification of sentences so that we can easily observe whether they are true or not. Much more strongly, Verificationism entails that a sentence that is *in principle not verifiable* by observation *has no meaning*, i.e. no cognitive content whatsoever. Non-verifiable sentences are meaningless pseudo-sentences. The reader will already guess which (philosophical) subject is thought to be merely capable of producing meaningless pseudo-sentences like 'the Absolute enters into, but is itself incapable of, evolution and progress' (Ayer 1936/2001: 17, ascribing that sentence to F. H. Bradley). We come to this in the section on metaphysics.

Analytically true sentences We must mention a special class of sentences that need no verification in observation (it would even seem to be impossible to say which observations would falsify them): *analytic sentences* which are true in virtue of the meaning of their constitutive words alone and which are therefore knowable *a priori*. Examples are conceptual, logical or mathematical truths: 'Sisters are female siblings', '*p* or not *p*' and '2 + 2 = 4'. Neither the classical Empiricists nor the modern Empiricists had trouble accepting those sentences. Someone stating these truths, so they say, does not claim to make any factual statement whatsoever and since they are devoid of the pretence of saying something about the world they are not violating the Empiricist doctrine.

What do these statements do instead? One dominant interpretation from the debate amongst the Empiricists was that their role is to define the conceptual framework on which empirically meaningful sentences can be composed. Metaphorically speaking, they are the rules of the game, not the moves. They register conventions of language or say how we use (or ought to use) the words or symbols they contain.

Two classes of acceptable sentence In short, the Logical Empiricists, just like the old Empiricists, accept two kinds of sentences as meaningful:

> The meaningful sentences fall into two kinds: first, there are sentences which are true already because of their [logical] form ('Tautologies' after Wittgenstein; they correspond approximately to Kant's 'analytic judgements'); they do not state anything about reality. To this kind belong the formulae of logic and mathematics; they themselves are no factual statements, rather they make possible the transformation of such statements. [. . .]
>
> The truth or falsity of all the other sentences can be decided [by observations];[33] they are, therefore, (true or false) *observational sentences* and belong to the realm of the empirical sciences.
>
> (Carnap 1931: 236/166, my translation)

The hierarchy of language Verificationism about meaning says that a (non-analytic) sentence like 'There is a red apple on the table' has meaning if and only if we can describe which observation would prove it true. In this case, and similar ones like 'This bird is singing' or 'It smells of fresh coffee', our task seems fairly simple, for we can easily see the apple, hear the bird and smell the coffee. Although these sentences about observations in our daily lives cause far more trouble for Verificationism than one might prima facie think (a topic we unfortunately cannot go into in depth) we move on to more difficult statements, namely those about the imperceptible entities the sciences are involved with.

Statements about unobservables Remember that, for the Logical Empiricists, the progress of science, especially the explanatory and predictive successes of fundamental physics and the technical advances that sprang from it, became quite generally the ultimate model for any kind of intellectual endeavour.[34] Now, because the natural sciences and their theories are the paradigms for good empirical knowledge, clearly their statements should come out as perfectly meaningful in accordance with Verificationism. Yet what counts as the verifying observation for (and is thus the meaning of) sentences that contain references to unobservable, theoretically postulated entities like 'an electron passed the double slit' or 'quarks turn out to be one-dimensional oscillating strings'?

The Empiricists' answer is indirect: sentences that contain non-observational vocabulary have to be translated into sentences that do contain (only) observational terms. If such a translation succeeds the verification criterion can be fulfilled because then we can give the method for the verification of the second sentences in terms of naked eye (ears, etc.) observations.

Translations into observational vocabulary This translation method has at its core a thesis that is implicit in the Verificationist criterion for sentence meaning: our language, or at least the worthy parts of it, can be put into a hierarchical structure where terms that refer to immediately perceivable things are the basis and all further notions can be translated into or analysed in terms of this basic vocabulary. If this sounds very much like Hume's postulation that all meaningful complex ideas are compounds of the simplest ideas, which, in turn, stem from immediate impressions, then that is no coincidence. Just what Hume advised us to do with ideas, the Logical Empiricists ask us explicitly to do with words:

> For many words, and especially for almost all scientific words, it is possible to trace their meaning back to other words ('constitution', 'definition'). [. . .] In this way, every word of our language is reduced to other words and ultimately to those words figuring in [simple observational statements].
>
> (Carnap 1931: 222/152, my translation)

To indicate how the translation issue is supposed to work, we look at a slightly simpler sentence than the electron or quark example from above: 'This liquid has a temperature of a 100°C'.

While we are able to tell roughly whether something is hot, lukewarm or cold, we cannot feel temperature in a quantitative sense, i.e. we would be unable to feel/measure with our bare hands that the temperature is exactly 100°C. Hence, even for these simple sentences some intermediate steps are needed in order to fulfil the verification criterion.

Let us now see how this might work. 'Object O has temperature T' could be translated into 'If you put a mercury thermometer into O or hold it close by, then the mercury will rise (or fall) to mark T'. If we agree that all we refer to in this new sentence is directly observable (the thermometer, the mark, etc.) then we have a good candidate for the reduction/translation of the unobservable to the observable. With the help of the general reduction we can now give the method of verification for and thus the meaning of the specific liquid case. The sentence 'This liquid has a temperature of a 100°C' can be tested by the following operation: if you stick a mercury thermometer into the liquid and the mercury column rises to mark 100°C then the sentence is true, otherwise false. Thus, the temperature statement is meaningful.

The *actual* definition/reduction of all terms (or sentences) to an observational vocabulary is, of course, a utopian dream and also unnecessary for our purposes. A proof of the theoretical possibility would already be enough to support the Empiricists' credo. In fact, Rudolf Carnap, in his infamous *The Logical Construction of the World* (which we'll abbreviate as *Aufbau* from its German title *Der logische Aufbau der Welt* (Carnap 1928/1998), which explains the title of Section 1.4.1) gets down to business to prove the general possibility. It is here, in these analyses and definitions, where the advances in modern logic proved to be of indispensable help.

Sense data and the given Actually, back at the times of the *Aufbau*, Carnap's aim was even more ambitious. There, he attempted to reduce every such sentence to even more fundamental observations than those of thermometers, liquids, tables and chairs. His determined goal was to reduce everything to absolutely basic, atomic sensations like 'hot here now', 'green there', etc. which the Empiricists called 'sense data'.[35] These correspond to Hume's *impressions*, and where it was Hume's goal to show that all complex ideas can be analysed in terms of the simplest impressions, Carnap's was to 'give a rational reconstruction of the concepts of all fields of knowledge on the basis of concepts that refer to the *immediately given*' (Carnap 1928/1998: XVII, my translation, emphasis added). *The given* was Carnap's and the other Logical Empiricists' term for the sum of all simple impressions/sense data a person has.

This (over-)ambitious programme was later abandoned again and reference to observable medium-sized physical objects was allowed (see Neurath 1932–3).[36] Unfortunately, we cannot discuss the reason why the Empiricists focused on sense data and the given rather than on 'bigger' observables like tables and chairs and thermometers and liquids and why they gave up this austere programme later. However, we might occasionally speak of the Empiricists' aim to reduce everything to sense data or the given.

Anti-metaphysics We have already gathered a couple of Logical Empiricist anti-metaphysical bits and pieces: Kant's synthetic *a priori* as a possible realm for metaphysics was abandoned again. We have also mentioned the Logical Empiricists' chief weapon against sentences that allegedly express metaphysical insights: Verificationism. We can now explicitly formulate what we hinted at above when talking about Verificationism and sentence meaning.

Metaphysical pseudo-sentences In the light of a Verificationist theory of meaning, metaphysical statements such as 'humans have immortal souls', 'the laws of nature are god's will' or 'the monad is nothing but a simple substance' are allegedly revealed to be senseless pseudo-sentences. They are devoid of meaning because, according to the Empiricists, it is hard to see how they could be verified under any observation or be translated into sentences that can (which perception or, scientifically speaking, which experiment would show that monads are a simple substance, or that they are not?). Thus, metaphysics 'produces sentences which fail to conform to the conditions under which alone a sentence can be literally significant' (Ayer 1936/2001: 15).

As a consequence, metaphysical claims turn out not only to be *epistemically dubious* but, stronger, they do not even have any real *cognitive content*: they are nonsense. This verdict kept looming large within the philosophy after Logical Empiricism. All attempts to make even the most modest metaphysical claims were stigmatised.[37]

Logical syntax The significant improvements of logic (by Frege, Russell and others) allowed the Logical Empiricists to criticise metaphysics in yet another way, namely on the basis of sentence grammar/logic. This is the second advance of Logical Empiricism compared to classical Empiricism and one that exists in parallel to the Verificationist method described above. How can metaphysics be criticised on the level of grammar or syntax without the need to proceed to meaning?

Das Nichts selbst nichtet An infamous example of that method at work is Carnap's onslaught[38] on a claim Heidegger made in 1929 in his inaugural address *What is Metaphysics?* in Freiburg, namely 'Nothingness itself nothings'[39] (Heidegger 1927: 37, my translation). Carnap's critique (Carnap 1931: 230/160) is a combination of pointing out that there is no such observable event of *nothing nihilating* (Hume could have done that already) and, moreover, a syntactical/logical analysis of that sentence's form (this is the new aspect: Hume did not have Fregean/Russellian logic in his toolbox for this kind of attack).

Where, in ordinary language, we sometimes use *nothing* as a subject term, a name, or a noun, as in 'Nothing is left in the chocolate box', a logical analysis reveals that we do not mean to say that there is still *a something* in the box, namely *the nothing* or *nothingness*. Rather, we mean to utter a negative existential claim: there is no item x such that x would be in the box (= it is not the case that there is something in the box; or, in logical language: $\neg\exists x Bx$). Thus, Heidegger's *reification of nothingness* is, according to Carnap, a mere syntactical mishap.[40] Sentences like these – Carnap gives 'Caesar is and' as a further example (Carnap 1931: 227/157) – 'are effectively eliminated automatically already by grammar' (Carnap 1931: 228/157).

Metaphysics-free philosophy Frustrated, on the one hand, with the philosophical tradition of the past centuries and its grand speculative edifices, and thrilled, on the other hand, with the successes of empirical science, the Logical Empiricists declared that philosophy shall be nothing but *philosophy of science*. Carnap writes:

> What remains for *philosophy* if all sentences, that have meaning, are of an empirical nature and can be subsumed under the empirical sciences? What remains are not sentences, no theory, no system, but merely *a method*, namely logical analysis. The application of this method [. . .] serves as excision of meaningless words and senseless pseudo-sentences. In its positive use philosophy serves to clarify meaningful terms and sentences. The indicated task of logical analysis [. . .] is what we mean by '*scientific Philosophy*' in contrast to metaphysics.
>
> (Carnap 1931: 237–8/167–8, my translation)[41]

Carnap's Nietzsche admiration Carnap has something else to say about metaphysics. Interestingly, he believes that there is something valuable that metaphysicians want to get at. Yet this is not expressible by theoretical philosophical inquiry. Rather, metaphysicians secretly want to convey their

> *attitude towards life* [. . .], the mindset in which a person lives, the emotional, intentional position to his or her environment, to his or her fellow human beings, to his or her responsibilities in which s/he is engaged, to the fates s/he has to endure.
>
> (Carnap 1931: 238/168)

Carnap underlines the importance these attitudes have in our lives. He just does not think that academic theoretical metaphysics is the right means to express them. Rather, he believes that the arts are the place where these matters should prosper (Carnap 1931: 240/170). This, finally, leads us back to Carnap's Nietzsche admiration, which seemed so unlikely at the very beginning of chapter 1:

> Our guess that metaphysics is an *ersatz*, yet an inadequate one, for art seems also to be confirmed by the fact that the one metaphysician who possibly had the highest artistic talent, namely Nietzsche, made the fewest mistakes of this confusion. [. . .] In the work in which he expresses strongest what others express via metaphysics or ethics, namely in the 'Zarathustra', he chose not a misleading theoretical form, but explicitly the form of art, of poetry.
>
> (Carnap 1931: 240/170)

Where do we go from here? You hold a book on the metaphysics of science in your hands and, having read about Empiricism, classical and logical, you might be tempted to commit it to the flames. What should keep you from doing so?

BOX 1.4.1: *Aufbau* (Construction)

- Although great admirers of Kantian themes, the Logical Empiricists returned to some Humean ideas: they had good arguments for why **the category of *synthetic a priori truths* is probably empty** (see empirical findings about the nature of space and time and quantum mechanics). They added that examining **the ways in which we think and perceive is a matter of empirical research**.
- They also pushed **classical Empiricism to the extreme** in that they turned the classical Empiricist **epistemic dogma** (that all factual knowledge, i.e. knowledge about the world, has to derive from sense experience and sense experience alone) fully into the **semantic Verificationist doctrine** that the **meaning of a sentence *is* the test method through which the sentence's truth or falsity can be established by observation.**
- Together with Verificationism, the **logical analysis of the grammar or syntax** of sentences was supposed to reveal whether statements are meaningful or whether they contain sheer **metaphysical nonsense**. The only acceptable sentences

that are correct independently of sense perceptions are **analytically true sentences**.
- As well as **metaphysics ethics** and **aesthetics** were supposedly areas of philosophy that at best transport an **attitude or a feeling towards life** (*ein Lebensgefühl*) but they do **not really express meaningful propositions**.

1.4.2 Demolition

The downfall of Empiricism Logical Empiricism and specifically the Verificationist criterion of meaning is too good to be true. Philosophy of science, at least in the first two thirds of the twentieth century, was preoccupied battling the shortcomings of both, and, despite many rescue attempts, the consensus today is that Logical Empiricism and Verificationism failed, at least in their most radical forms. That means specifically that their critique of metaphysics is also untenable. Good for us, one should say, because metaphysics of science, the topic of this book, would otherwise not be possible.

We shall see in the coming chapters how philosophers regained the confidence to tackle metaphysical issues. In fact, the more they saw that the radical restrictions of Logical Empiricism were indefensible the more metaphysical territory they seem to have regained.

Here, we will sketch five of the core objections against Empiricism and Verificationism. Together, they were decisive against this prima facie attractive programme. We do not have the space to go into depth here (and there are objections other than these five) but later, especially in Chapter 2, on dispositions, we will see in more detail the problems Empiricism and especially Verificationism had to face.

(1) The myth of the given In his paper 'Empiricism and the Philosophy of Mind' (Sellars 1956)[42] Wilfrid Sellars (1912–89) attacks the Empiricist idea of raw, pre-theoretic and simple perceptions as free from any conceptualisation. Sellars coined the apt phrase 'the myth of the given' for this (untenable) presupposition, which we introduced above as Locke's blank-slate idea. Sellars argues that observation reports are contaminated by the ingredients of the observer's theoretical background assumptions. Observations are 'theory-laden': there are no impartial, neutral sense data.

Remember that the Empiricists did acknowledge some involvement of our cognition in perception: we saw this when we compared them to Kant.[43] Yet they did not realise just how much involvement there is. We cannot follow the intricacies of perception in detail but we need to keep in mind that the blank-slate idea of Empiricism was put into doubt.

(2) Provisos Earlier, we wrote confidently that a sentence like 'This liquid has a temperature of a 100°C' can easily be tested by the following operation: if we stick a mercury thermometer into the liquid and the mercury column rises (or falls) to mark 100°C then the sentence is true, otherwise false. Since all this is observable, the original temperature statement can count as meaningful. However, even such a simple sentence confronts us with a whole bunch of challenges.

For example, we have to add constraints for the right conditions, for pressure and for the workings of the thermometer and lighting, so that we do not misread the scale; also, strictly speaking, we have to exclude the case where we mistake a trick gadget for a thermometer. In other words, we have to add a whole lot of provisos (sometimes called *ceteris paribus* clauses) to the simple verification criterion: the 100°C sentence is observed to be true if and only if, under ideal conditions, while we are well awake and not hallucinating, we see a properly working, real thermometer rising to mark 100. And still we would not be at the end of our journey to get verification conditions that are watertight. We are sure the reader will find further necessary adjustments.

In other words, we are confronted with a possibly infinite number of provisos that we would have to add, some of which we are probably not aware. Therefore, it is likely that the ultimate, correct verification conditions can never be formulated. Yet, then, the true meaning of the 100°C sentence remains forever concealed from us – and that, apparently, we do not know the meaning of such a simple sentence is a strange result. It speaks against Verificationism being a correct theory of meaning for sentences.[44]

(3) Verification of universally quantified sentences The thermometer sentence form above was a singular sentence. It referred to the temperature of a particular liquid. Even more challenging are universally quantified statements. Many law statements have this form: all samples of water boil at 100°C, all electrons are negatively charged, all masses attract each other, etc. The difficulty here is that even if we have observed many single samples of, say, water boiling at 100°C, we have not done so and cannot possibly do so with *all* samples, past, present and future. Which method of verification, i.e. which meaning, could then be given for 'all such-and-such do or are this-and-this' statements? Alfred Ayer highlights this challenge for Verificationism:

> It is of the very nature of these propositions that their truth cannot be established with certainty by any finite series of observations. But if it is

> recognized that such general propositions of law are designed to cover an infinite number of cases, then it must be admitted that they cannot, even in principle, be verified conclusively.
>
> (Ayer 1936/2001: 18)

Within the Verification theory of meaning this riddle weighs heavy. Not only is it impossible to verify all cases, but within a Verificationist framework this has the consequence that universally quantified statements are meaningless nonsense! Thus, law statements would have to be banned from scientific discourse for their lack of meaning. This is, of course, absurd for they belong to the very heart of science.

Weak Verificationism. The Empiricists tried to meet these hurdles with two strategies. First, there were attempts to weaken the verification criterion to the effect that observations had only to be *somehow relevant* for the truth or falsity of a sentence in order to convey meaning instead of conclusively verifying or falsifying it. However, even these reformulations failed in the end because of further insurmountable test cases.

Meaningless law statements. The other strategy was more radical. Its proponents simply bit the bullet: they held that statements of natural law are neither true nor false and make no factual claims about the world; they are mere guidelines for scientific endeavour. Frank Ramsey (1903–30), for example, endorses the view that law statements 'are not judgments but rules for judging "If I meet a ϕ I shall regard it as a ψ"' (Ramsey 1929: 149; see also Ayer 1936/2001: 18–19, referring to Schlick 1931; for more on laws see Chapter 4).

Needless to say, those Empiricists who took this bold step owe us an explanation why exactly these statements and not others are so promising as background assumptions (that water boils at 30°C, for example). Thus, neither the weakening of the verification criterion nor the courage to accept the original's consequences did, in the end, convince.

(4) Meaning holism and the fall of the analytic–synthetic distinction

In the eyes of many, the next critique of Verificationism brought it to its knees. Take again the sentence 'This liquid has a temperature of 100°C' and its (simplified) verification conditions: if we stick a mercury thermometer into the liquid and the mercury column rises (or falls) to mark 100°C then the sentence is true, otherwise false. Suppose, now, an actual observation counts against its truth (the mercury stops at, say, 66.6°C). It will not come as a surprise that we could, in principle, nonetheless defend the claim that the liquid has a temperature of 100°C. This is possible if we revise or drop other items of our belief system. As Willard Van Orman Quine (1908–2000) writes: 'Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system' (Quine 1951: 43).

We could, for example, doubt that mercury always expands under heat, or we could doubt that the environmental pressure is normal, or we could claim the thermometer was wrongly calibrated, etc.[45] Of course, we would have to verify each of these claims in their own right. Yet, should observations count against them, we could play the same game all over again and shield also these claims from falsification: 'A recalcitrant experience can [. . .] be accommodated by any of various alternative reevaluations in various alternative quarters of the total system' (Quine 1951: 44).

The tribunal of experience. Given the possibility of revisions elsewhere in our convictions in order to save a given sentence from falsification, Quine concludes that the Empiricists' assumption that singular sentences face the *tribunal of sense experience* alone and in isolation is wrong. Singular sentences are too small a unit for a verification principle. Rather, according to Quine in his paper 'Two Dogmas of Empiricism', 'the unit of empirical significance is the whole of science' (Quine 1951: 42).[46] That is, it is always *the entire theory or the corporate body of our beliefs* that is under scrutiny. Verificationism, as a theory of isolated sentences' meaning, fails.

The second dogma that Quine demolishes had not only been supported by the Empiricists but also and especially by Kantians, namely that there is a clear distinction between analytical sentences, which are true by the meaning of their constitutive words ('Sisters are female siblings'), and synthetic truths ('My sister is a schoolteacher'), which need observations. Quine's meticulous argumentation cannot be traced here but one of his reasons to give up the synthetic–analytic distinction is related to the possibility of revision mentioned above.

The short version of the argument is this: even extreme revisions to our belief system could be considered. Quine makes this plausible by reference to quantum mechanics. There, a revision of the logical law of the excluded middle has been proposed so that, for example, light can at the same time be both a wave and not a wave (namely a particle) – and if revisions even to logic are possible then why not also revisions of analytical sentences? In the light of some observation (transsexual or transgender people may be a case in point) we might revise the proposition that sisters are female siblings.[47] If Quine is right, two central dogmas of Empiricism are untenable and the whole theory is put into doubt.

(5) The status of Verificationism itself Here is a final embarrassment for Verificationism: it does not meet its own standard, for which observation would prove that the meaning of each sentence is the method of its verification by observation? If no such method can be given then the central claim of Logical Empiricism has no semantic meaning: it would be a pseudo-sentence itself, expressing nothing.

There are two albeit fairly similar ways to deal with this quandary. Instead of treating the verification criterion as an empirical sentence, we could take it for an analytic truth which specifies the meaning of the terms it contains. To make this move more palatable we might want to rephrase the criterion in the following way: 'In every rationally conducted science the meaning of each of its empirical statements is identical with the method in which we establish the sentence's truth or falsity'. Now, the criterion defines, partially at least, what it means for an enterprise to be rightly called 'rationally conducted science'. The second alternative (Carl Gustav Hempel (1905–97) and Alfred Ayer were advocates of it) changes this reformulation from a definitional analytical statement into a prescriptive claim or a recommendation: 'In each rationally conducted science the meaning. . . *shall be identical to*. . .'.

In isolation, this final trouble for Verificationism might not weigh too heavily. One might well be willing to accept it as pragmatic advice. Yet the other shortcomings we have gathered, at least when taken in concert, make Verificationism untenable. We summarise these five reasons in Box 1.5.

BOX 1.4.2 Demolition

We were confronted with five challenges to Empiricism/Verificationism:

- **The given is a myth**: perceptions are not theory-neutral but rather **theory-laden**.
- A possibly infinite number of **proviso clauses** has to be attached to verification conditions. It might also be unclear from the outset which clauses these are.
- It is uncertain what Verificationism should say about **universally quantified sentences** like **law statements** for which there is no finite verification method. Similar to these sentences are **statements about the past or future** for which, too, there are no specifiable direct observations that could prove those sentences right or wrong.
- The **Quine–Duhem thesis of meaning holism** says that all convictions within a theory or any kind of belief system cling together. In the light of negative evidence any of a variety of sentences can be given up instead of the one allegedly under scrutiny. So it is **not singular sentences that face the tribunal of observation, as Verificationism has it, but whole theory or belief system.**

- The **application of the verification criterion to itself** reveals that it can at best be taken as **methodological advice for good science** instead of an empirically meaningful statement.

LITERATURE

- An excellent introduction to Logical Empiricism and its problems is Peter Godfrey-Smith's ***Theory and Reality*** (Godfrey-Smith 2003). The strengths and weaknesses discussed here can be found in his chapter 2.

1.4.3 Reconstruction: the road ahead

We know now that the Verificationist theory of meaning failed. This chief weapon of Logical Empiricism against metaphysics is dysfunctional: if Verificationism is false, metaphysical statements are not immediately nonsense. Their meaning or meaninglessness has to be proven on different grounds. Maybe some epistemic Empiricist doubts can be upheld against them but new arguments need to be formulated.

Still, in Chapter 2 we return to Verificationism and especially what it demands of sentences that contain dispositional predicates like 'This sugar cube is water soluble' or 'That match is inflammable'. Yet isn't going back to Verificationist ambitions an otiose enterprise? Why should we further bother and engage with some specific details of Verificationism if we have already seen decisive reasons for it being an obsolete theory?

Here's why. In having a detailed look at where exactly a Verificationist reduction of dispositional predicates to observational language fails we can see which of the metaphysical assumptions dropped by the Empiricists have to be taken on board again, gradually and cautiously. We will also see in Chapter 2 that almost all of the core concepts used within science – those of *counterfactual conditionals* (Chapter 3), *laws of nature* (Chapter 4), *causation* (Chapter 5), of *natural kinds*, of *necessity* (Chapter 6), etc. – are inseparable from theories of them. This has the welcoming side effect that the metaphysics surrounding these other concepts will also be unearthed bit by bit and, thus, the metaphysics of science will be told.

Notes

1 *The Scientific World Conception: The Vienna Circle* (Carnap *et al.* 1929).
2 For a non-formal introduction see Steinhardt and Turok (2003).
3 Other famous participants at some of the Circle's meetings in Austria were Kurt Gödel (1906–78), Karl Popper (1902–94), and Ludwig Wittgenstein (1889–1951). In Germany the Berliner Society for Empirical Philosophy met under the lead of Hans Reichenbach (1891–1953), Richard von Mises

(1883–1953) and Carl Gustav Hempel (1905–97). In the UK, young Alfred Ayer (1910–89) attracted attention with his Neo-Positivist pamphlet *Language, Truth and Logic* (Ayer 1936).

4 Legend has it, by the way, that Andronicus of Rhodes (ca. 60 BC), the first editor of Aristotle's works, placed the volume that deals with issues like being, essence, change, potentiality, cause, etc. on the shelf *behind* (meta: μετά) those volumes dealing with *physics* (φυσικά). If at all true this is a remarkable concordance of form and content. (Or, maybe, the editor intended the bibliographical sequence to match the curricular order. I owe this suggestion to Oliver R. Scholz.)

5 The full title of the book, *'Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences*', indicates that Descartes here offers one of the first treatises on correct epistemic pursuit in the natural sciences.

6 Many such examples can, by the way, be found already in the works of the ancient sceptics. See, for example, Sextus Empiricus's *Outlines of Pyrrhonism*, especially the ten tropes of Aenesidemus.

7 Descartes himself was probably aware of the fact that his answer is wanting and so also proposed as a reason that the denial of such judgements would be self-contradictory (Ayer 1936: 30–1).

8 The reader is invited to later compare Leibniz to what Dispositional Essentialists say (Chapter 6).

9 We must not omit that, immediately afterwards, we read in Russell: 'At this point I read the *Discours de Metaphysique* and the letters to Arnauld. Suddenly a flood of light was thrown on all the inmost recesses of Leibniz's philosophical edifice. I saw how its foundations were laid, and how its superstructure rose out of them'. When we come back to Leibniz's monads once in a while and use them as examples of 'bad' metaphysics we only do so from the perspective of a radical anti-metaphysician. We do not at all wish to denigrate Leibniz's philosophy, and rather side with Russell than with over-ambitious, dismissive metaphysics-critics.

10 Hume allows already for some complexity in some impressions.

11 Hume's fellow Empiricist John Locke explicitly offers such a semantic theory of word meaning (Book III in his *Essay Concerning Human Understanding* (Locke 1690)) and a similar critique of metaphysical terms as conjured up and empty.

12 Interestingly, we already find a similar distinction in Leibniz: 'truths of reason' vs. 'truths of fact' (Leibniz 1714: §§33–5), yet Leibniz puts it to a different use.

13 We have here identified *concepts* with *ideas*, a move that can be allowed for our purposes.

14 This is not in conflict with the demand that all ideas are ultimately grounded in sense experiences, for, while this might be true, the interrelation of ideas can, once their individual meanings are known, be derived without further sense data.

15 Hobbes is a bit of both – Empiricist (for example, when it comes to semantic meaning) and Rationalist (as in the above example) – and therefore hard to categorise.

16 We can already note at this point that this has more or less been the consensus ever since.

17 Much later, this insight was confronted by Donald Davidson's (1917–2003) famous critique, which, however, does not take away the gist of Hume's general point:

> Surely not every true causal statement is empirical. For suppose 'A caused B' is true. Then the *cause of B* = A; so substituting, we have 'The cause of B caused B', which is analytic. The truth of a causal statement depends on what events are described; its status as analytic or synthetic depends on how the events are described.
>
> (Davidson 1963: 14, emphasis added)

18 In the very recent literature, there has been a dispute whether Hume is merely sceptical when it comes to our epistemic access to a necessary causal connection in the world or whether he outright denies its existence. The latter has been the orthodox reading of Hume (for the mere *skepsis*, or caution, interpretation, see, for example, Strawson 1989). We need not enter into these exegetical issues here and treat, for matters of simplicity, Hume in the orthodox way as 'the greater denier of necessary connections' (Lewis 1986: ix–x).

19 Whether conceivability and possibility are identifiable is still a matter of debate (Gendler and Hawthorne 2002): there could well be things we cannot conceive of (because of the limits of our imagination) but which are nonetheless possible (think, for example, of wave–particle dualism in quantum mechanics). Also, there might be impossible things of which we believe we can conceive (we come to possible candidates in Section 6.3).

20 If we were to go with the Rationalists we would even end up with conceptual/logical necessity (not only the worldly variety we have here in mind). As noted already, ever since Hume this latter link has been irreversibly cut.

21 Note that here, again, the semantic aspect of Hume's Empiricism shines through: 'and these *words* are altogether insignificant' or '*mean* nothing but that determination of the thought'.

22 See Adrian Moore's *The Evolution of Modern Metaphysics* (Moore 2012), where he underlines this point and presents many valuable short introductions to the metaphysics of 23 philosophers, including those mentioned here, such as Descartes, Spinoza, Leibniz, Hume, Kant and, soon to follow here, Carnap.

23 Peter Strawson (1919–2006) made these arguments popular again under the 'transcendental arguments' term in his *Individuals* (Strawson 1959). We will encounter arguments akin to the Kantian/Strawsonian ones throughout the book and reflect on their validity in Chapter 7 on meta-metaphysics.

24 So says Kant, deviating from Hume who subsumed them in (2).

25 So that every analytic statement is known *a priori*, but not every *a priori* judgement is analytic: some of the latter (and of a very interesting kind) are synthetic.

26 This contradicts Hume, who, on the contrary, thought that we do not have to go beyond the concepts/ideas of 7, 5 and 12 to establish the equation.

27 To which, it should be mentioned, the *Groundwork of the Metaphysic of Morals* (Kant 1785) also belongs.

28 An almost synonymous name for Logical Empiricism is logical positivism (for a subtle difference see Wesley C. Salmon 2000). The founding father of classical positivism (and also the inventor of that name) is Auguste Comte (1798–1857).

29 Neo-Kantianism, as, for example, defended by the Marburg School (Hermann Cohen (1842–1918), Paul Natorp (1854–1924) and especially Ernst Cassirer (1874–1945)) was an influence on the Logical Empiricists (see Friedman 2000).

30 In note 28 above, we mentioned Neo-Kantianism. One way to react to the findings of modern science while still remaining within such a Kantian scheme is to say that Kant was merely wrong about the precise content of the synthetic *a priori* principles but that, still, there are such principles. The task of modern Kantians would then be to transcendentally deduce the correct principles.

31 In fact, there is another one to come: foundationalism, which we will discuss shortly.

32 A better phrase for the theory would have been *testability theory of meaning* because *to verify* literally only means *demonstrate to be true* (from the Latin *verum*, or truth) where the Empiricists indeed meant prove to be either *true or false* (see Godfrey-Smith 2003: 27).

33 Carnap writes 'protocol sentences' instead of 'observations'. We come to protocol sentences shortly.

34 This, by the way, was no less true for Hume and his self-ascribed 'experimental method'. See the subtitle to his *Treatise*: 'Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects'. Of course, for Hume, Newtonian mechanics was the role model, not relativity theory or quantum mechanics, as for the modern Empiricists.

35 Other names that circulated for the perceptually immediately given were *appearances*, *sense data*, *sensibilia*, *mental images*, *percepts*, *ideas/impressions* and *qualia*. Ernst Mach (1838–1916) (1886), Bertrand Russell (1914–19) (1986), the early Ludwig Wittgenstein (Wittgenstein 1921: §4.21ff) and the younger Rudolf Carnap (1928/1998: II C) were all friends of this sense-data atomism.

36 This happened not only because this austere form of sense-data Empiricism was hard to handle in respect of all the translations that would ideally have to be made: a further issue was that 'the given', i.e. sense data, are always only the immediate sense experiences of an individual subject so that intersubjective communicability and comparability was hard to obtain.

37 At this point one might remember that the Empiricists did accept a class of sentences, the analytically true ones, that were not in need of empirical verification. Can metaphysics ever thrive in the realm of analytic truths? Decidedly not, says Ayer (Ayer 1936: 24; Carnap 1931: 236/166), for metaphysical statements aim to reach out for factual propositions about the world. Analytic sentences, however, are about word meaning, not worldly facts.

38 As with our all too brief remarks on Leibniz earlier, it would, of course, be a great mistake to judge Heidegger's philosophy on the basis of this isolated quote. We report the Heidegger–Carnap debate because it is a striking event in the history of two grand philosophical traditions – the so

called *analytic* and *continental* traditions – and not to defame either of them. For more on the debate see Friedman (2000) and Braver (2007).

39 German: *Das Nichts selbst nichtet*. The verb *to nothing* (*to noth* or *to nihilate*) is not a neologism in English only: *nichten* does not exist in ordinary German either. Heidegger was a great inventor of philosophical terminology.

40 As Simon Blackburn sums up so aptly in his *Dictionary of Philosophy*: 'The difference between existentialists and analytical philosophers on the point is that whereas the former are afraid of Nothing, the latter think that there is nothing to be afraid of' (Blackburn 1994: 265, entry: 'Nothing').

41 All this was, by the way, seen not only as an attack on metaphysics but also on ethics and aesthetics: 'The objective validity of a value or a norm cannot [. . .] be empirically verified or inferred from empirical sentences' (Carnap 1931: 237/167, my translation).

42 For a further famous critique of sense data see John Austin (1911–60) (1962). For theory-ladenness see also Thomas Kuhn's (1922–96) *The Structure of Scientific Revolutions* (Kuhn 1962).

43 Keep in mind that when post-logical Empiricists speak of the involvement of our cognition in perception, they talk about contingent facts of human perception revealed by empirical psychology, not about the necessary preconditions of the possibility of all experience.

44 The issue of provisos and *ceteris paribus* clauses is a theme that will frequently recur in our book (especially in Sections 2.1, 5.3 and 6.2.3).

45 Compare this to problem (3), the proviso clauses: it and holism are two sides of the same coin.

46 To be fair to both Quine and Pierre Duhem (1861–1916) we should mention that Quine acknowledges in endnote 17 of *Two Dogmas* that 'this doctrine was well argued by Duhem: 303–28 [Duhem 1906]'. The doctrine is therefore known under the name *Quine–Duhem thesis*.

47 Because the analytic–synthetic distinction fell, the *a priori–a posteriori* difference was left on shaky grounds too. At best, relativised versions of it can be upheld, but we do not have the space here to argue this.

Bibliography

Austin, J. L. (1962) *Sense and Sensibilia*. Oxford: Clarendon Press.

Ayer, A. J. (1936/2001) *Language, Truth and Logic*. London: Penguin Books.

Blackburn, S. (1994) *The Oxford Dictionary of Philosophy*. Oxford: Oxford University Press.

Braver, L. (2007) *A Thing of this World. A History of Continental Anti-realism*. Evanston, Illinois: Northwestern University Press.

Carnap, R. (1928/1998) *Der logische Aufbau der Welt*. Hamburg: Meiner.

Carnap, R. (1931) Überwindung der Metaphysik durch logische Analyse der Sprache. *Erkenntnis*. 2 (1). pp. 219–41.

Carnap, R., Hahn, H. and Neurath, O. (1929) *Wissenschaftliche Weltauffassung – Der Wiener Kreis*. Wien: Veröffentlichungen des Vereines Ernst Mach.

Davidson, D. (1963) Action, Reasons, and Causes. *Journal of Philosophy*. 60. pp. 685–700.

Descartes, R. (1637) *Discourse on the Method*. Maclean, I. (trans.) (2006). Oxford: Oxford University Press.

Descartes, R. (1641) *Meditations on First Philosophy With Selections from the Objections and Replies*. Cottingham, J. (ed.) (1996 [1986]) Cambridge: Cambridge University Press (the first quoted numbers are the old pagination, the second in brackets the 1986 CUP pagination).

Duhem, P. (1906) *La Théorie physique: son objet et sa structure*. Paris: Chevalier & Rivière.

Dummett, M. (1993) *Origins of Analytical Philosophy*. London: Duckworth. Cambridge, MA: Harvard University Press.

Frege, G. (1879) HYPERLINK “https://de.wikipedia.org/wiki/Begriffsschrift” Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Louis Nebert: Halle a. S.

Friedman, M. (2000) *A Parting of Ways: Carnap, Cassirer, and Heidegger*. Chicago: Open Court.

Gendler, T. S. and Hawthorne, J. (2002) *Conceivability and Possibility*. Oxford: Oxford University Press.

Godfrey-Smith, P. (2003) *Theory and Reality*. Chicago: Chicago University Press.

Heidegger, M. (1929) *Was ist Metaphysik?* 16th edition (2007). Frankfurt am Main: Vittorio Klostermann.

Hobbes, T. (1655) De Corpore (Concerning Body). In: Molesworth, W. (ed.) (1839) *The English Works, Vol. I*. London: Scientia Aalen.

Hume, D. (1739–40) *A Treatise of Human Nature*. Selby-Bigge, L. A. (ed.), revised by Nidditch, P. H. (1978). Oxford: Clarendon Press.

Hume, D. (1748) *Enquiries concerning Human Understanding and concerning the Principles of Morals*. 17th impression of the 3rd OUP edition. Selby-Bigge, L. A. (ed.). Oxford: Clarendon Press.

Kant, I. (1781/1787 (A/B)) *Critique of Pure Reason*. Guyer, P. and Wood, A. W. (eds.) (1997). Cambridge: Cambridge University Press.

Kant, I. (1783) *Prolegomena to Any Future Metaphysics*. Hatfield, G. (trans.) (2004) Cambridge: Cambridge University Press (page references to Vol. VI of the Königlich Preussische Akademie der Wissenschaften edition as printed in the margin of the present edition).

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten (Groundwork of the Metaphysic of Morals*) In: German Academy of Sciences (eds.) (1900) *Kants Gesammelte Schriften*. Berlin: De Gruyter.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Leibniz, G. W. (1685) The Art of Discovery. In: Wiener, P. P. (ed.) (1951) *Leibniz: Selections*. New York: Charles Scribner’s Sons.

Leibniz, G. W. (1686) On the True Method in Philosophy and Theology. In: Wiener, P. P. (ed.) (1951) *Leibniz: Selections*. New York: Charles Scribner’s Sons.

Leibniz, G. W. (1714) Monadology. In: Parkinson, G. H. R. (ed., trans.), Morris, M. (trans.) (1973) *Philosophical Writings*. London: J. M. Dent & Sons.

Leibniz, G. W. (1969) *Philosophical Papers and Letters*. Loemker, L. E. (ed., trans.). Dordrecht: Reidel.

Leibniz, G. W. (1989) *Philosophical Essays*. Ariew, R. and Garber, D. (eds., trans.) Indianapolis: Hackett.

Lewis, D. (1986) *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.

Locke, J. (1690) *An Essay Concerning Human Understanding*. 4th edition. Nidditch, P. H. (ed.) (1975). Oxford: Oxford University Press.

Mach, E. (1886) *Beiträge zur Analyse der Empfindungen*. Jena: Gustav Fischer. *Contributions to the Analysis of the Sensations*. Williams, C. M. (trans.) (1897). La Salle: Open Court.

Moore, A. W. (2012) *The Evolution of Modern Metaphysics: Making Sense of Things*. Cambridge: Cambridge University Press.

Neurath, O. (1932–3) Protokollsätze. *Erkenntnis*. 3. pp. 204–14.

Newton-Smith, W. H. (ed.) (2000) *A Companion to the Philosophy of Science*. Oxford: Blackwell.

Quine, W. V. O. (1951) Two Dogmas of Empiricism. In: Quine, W. V. O. (1953) *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Ramsey, F. P. (1929) General Propositions and Causality. In: Ramsey, F. P. (1990) *Philosophical Papers*. Mellor, D. H. (ed.) Cambridge: Cambridge University Press. pp. 145–63.

Russell, B. (1900/1992) *A Critical Exposition of the Philosophy of Leibniz*. London: Routledge.

Russell, B. (1914–19) (1986) *Collected Papers, Vol. 8: The Philosophy of Logical Atomism and Other Essays*. London: George Allen and Unwin.

Russell, B. (1945) *A History of Western Philosophy*. New York: Simon & Schuster.

Salmon, Wesley C. (2000) Logical Empiricism. In: Newton-Smith, W. H. (ed.) (2000) *A Companion to the Philosophy of Science*. Oxford: Blackwell. p. 233.

Schlick, M. (1931) Causality in Contemporary Physics I & II. In: Rynin, D. (trans.) *British Journal for Philosophy of Science*. 47 (XII) (1961). pp. 177–93. 48 (XII) (1962). pp. 281–98.

Schrenk, M. (2011) Interfering with nomological necessity. *The Philosophical Quarterly*. 61. pp. 577–97.

Schrenk, M. (2014) Die Erfahrung der Widerständigkeit als Wahrnehmung kausaler Kraft! In: Spann, A.S. and Wehinger, D. (eds.) *Vermögen und Handlung*. Munster: Mentis.

Sellars, W. (1956) Empiricism and the Philosophy of Mind. In: Feigl, H., Scriven, M. and Maxwell, G. (eds.) *Minnesota Studies in the Philosophy of Science, Vol. I.* Minneapolis, MN: University of Minnesota Press.

Spinoza, B. (1677) *Ethics*. In: Curley, E. (trans.) (1985) *The Collected Writings of Spinoza*, *Vol. I*. Princeton, NJ: Princeton University Press.

Steinhardt, P. J. and Turok, N. (2003) The Cyclic Universe: An Informal Introduction. *Nuclear Physics B – Proceedings Supplements*. 124. pp. 38–49.

Strawson, G. (1989) *The Secret Connexion: Causation, Realism, and David Hume*. Oxford: Clarendon Press.

Strawson, P. (1959) *Individuals*. London: Methuen.

Waismann, F. (1930–1) Logische Analyse des Wahrscheinlichkeitsbegriffs. *Erkenntnis*. 1 (1). pp. 228–48.

Whitehead, A. and Russell, B. (1910–13), Principia mathematica 1–3. Cambridge: Cambridge University Press.

Wittgenstein, L. (1921) *Tractatus logico-philosophicus*. In: Ostwald, W. (ed.) (2001) *Annalen der Naturphilosophie, Vol. 14*. McGuiness, B. and Pears, D. (trans.). London: Routledge Classics. pp. 185–262.

2 Dispositions

2.1 Semantics

2.1.1 Core features of dispositions

A sugar cube is *disposed to dissolve* in water. We say that it is soluble or, slightly more contrived, that *solubility* belongs to its features. Likewise, we describe a match as being *inflammable* or that it has the disposition of *inflammability*. Further examples from various sciences include *being supra-conductive* and *being magnetic* from physics, *being reactive* and *being acidic* from chemistry, *being fertile* and *being predatory* from biology, *being dictatorial* and *being reform-oriented* from political science and *being market-oriented* and *being competitive* from economics. In this chapter, though, we mostly stick to everyday dispositions because the problems we must discuss arise on the level of the ordinary, and they are of the same kind as those on the scientific level.[1]

Trigger and manifestation What those objects to which we correctly ascribe a disposition have in common is that they *react or behave in certain ways* when *they are in certain conditions*. Yet until the moment of being triggered an object's disposition might well be dormant and its manifestation non-occurrent: as long as a sugar cube is not in contact with water its solubility is somewhat hidden; as long as a match is not struck or near fire its inflammability won't show itself. Thus, when we wish to explain what an object's disposition is we typically say how the disposed object *would react if it were triggered in a certain way*.[2] This has three far reaching consequences – (1) non-observability (NON-OBS), (2) modality (MOD) and (3) production (PROD) – of which, to begin with, we will only introduce non-observability. The latter two will follow in subsequent sections; all three consequences will be central to the rest of the book.

Classical and Logical Empiricism revisited Now, remember the classical Empiricist claim that knowledge about the external world can only be gained by sense experience (Section 1.2). Also recall the early twentieth-century Logical Empiricists' radicalisation of this epistemic dogma, namely to turn it into a semantic doctrine: the Verificationist criterion of sentence meaning, which says that a sentence expresses something meaningful if and only if we can spell out which phenomena we would have to observe with our five senses to call the sentence true (or false). In fact, the meaning of a sentence, according to the Logical Empiricists, *is* verification by means of observation.

We can immediately see how sentences that mention dispositional predicates are problematic for Logical Empiricists: such predicates refer to features of the world that are – unlike colours, for example – not directly observable. You see that a crystalline powder is white but you do not see straight away whether it is soluble. And even when objects react as they are allegedly disposed to do, it is questionable whether we observe *their disposition* or merely *the trigger* and then *the effect*: you see a vase sit on a shelf, then you see it being hit and fall and finally you find it broken on the floor. All this is observable, yet do you *see its fragility*?

(1) Non-Observability (NON-OBS) So, should we not, as a consequence of this non-observability, simply drop dispositional talk from all scientific discourse? After all, if it looks as if it is not amenable to Verificationism should it not simply be discarded as senseless? (Remember, from Section. 1.2, that the Empiricists thought that this is the right way to proceed with terms like *the absolute*, *monad*,[3] etc.)

Yet this won't do in the case of dispositions and the reason why this is so is analogous to the one Hume had regarding *causation* (Section 1.2): because it figures so prominently in both our ordinary and scientific talk, it is hardly possible to dismiss *causation* as meaningless. Rather, the Empiricists have to show how its meaning can be spelled out in words that refer to observable features only. Hume did offer such an analysis for causation so that his ultimate message wasn't 'there is no causation' but, rather, 'there is causation, yet it is a little different from what you think it is'.

Something perfectly analogous is true for modern Logical Empiricists and dispositional concepts. Dispositions also play essential roles in the conduct of scientific practice – remember the list of example dispositions we gave at the beginning of the chapter – and, so, Empiricists had better find a way to translate dispositional talk into a language that can fulfil their Verificationist criterion of meaning so that it becomes acceptable. This is one of the practical implementations of the Empiricists' overall aim to 'give a rational reconstruction of the concepts of all fields of knowledge on the basis of concepts that refer to the immediately given' (Carnap 1928: xvii, my translation).

Reaching this goal might not seem a major problem in the case of dispositions. Take inflammability or solubility: we have clear intuitions how to phrase an 'if. . . then. . .' sentence that specifies the observable conditions (in the antecedent if-clause) under which the entities that are soluble or inflammable show visible phenomena (in the consequent then-clause), namely *to dissolve if put in water* or *to burst into flames if struck*. These simple suggestions for a reconstruction/translation of sentences with dispositional predicates specify a method of verification: put it in water/strike it and see what happens. Does it dissolve/does it burst into flames? If yes then it is correct to ascribe solubility/inflammability to the object; if no then it is not.

Yet even if we assume that we can directly observe the respective stimuli conditions and reactions, our assumption that our goal is easily reached will prove to be wrong.

2.1.2 Void Satisfaction and modality

In order to see this, we turn to one of the classic papers in philosophy of science, namely Rudolph Carnap's *Testability and Meaning*. Together with Carnap, we

> consider the question whether the so-called *disposition-concepts* can be defined, i.e. predicates which enunciate the disposition of a point or body for reacting in such and such a way to such and such conditions, e.g. 'visible', 'smellable', 'fragile', 'tearable', 'soluble', 'indissoluble' etc.
>
> (Carnap 1936: 440)

In accordance with the informal guesses we have made above, the most straightforward formal attempt of a definition of, for example, solubility is the following *simple conditional analysis* ([SCA]):

> [SCA]:
>
> Suppose we wish to introduce the predicate 'D' meaning 'soluble in water.' Suppose further, that 'T' and 'R' are already defined in such a way that 'Tx' means 'the body x is placed into water' and 'Rx' means 'the body x dissolves'. Then one might perhaps think that we could define 'soluble in water' in the following way: 'x is soluble in water' is to mean 'whenever x is put into water, x dissolves', in symbols: $Dx =_{def.} Tx \rightarrow Rx$.
>
> (Carnap 1936: 440)[4]

The material implication Carnap is perfectly aware from the outset that [SCA] is untenable and he later continues with his own, slightly better suggestion. Yet before we can understand what [SCA] means exactly and why it fails, we need to have a brief look at the logical sentence connective $\rightarrow$, i.e. the material implication it crucially employs.

Readers who are unfamiliar with elementary formal logic should consult an introductory book before they continue reading.[5] For everyone else a reminder of the following characteristics of $\rightarrow$ will suffice: (1) The material implication belongs to the realm of extensional, truth-functional logic. (2) The truth conditions of the material implication are: $p \rightarrow q$ is true if and only if p and q are both true or p and q are both false or p is false and q is true. $p \rightarrow q$ is false if and only if p is true and q is false. (3) A consequence of these truth conditions is that the material implication is confronted with the paradox that the falsity of p alone (no matter what q) as well as the truth of q alone (no matter what p) imply the truth of $p \rightarrow q$.

Void Satisfaction (VC) Returning to Carnap and the first tentative formal analysis of dispositional predicates he discusses – namely [SCA]: $Dx =_{\text{def.}} (Tx \rightarrow Rx)$[6] – we see that features (2), and, thus, (3) threaten the definition in the form of two famous phenomena: the so called *Void Satisfaction* (VC) and the so called *Random Coincidence* (RC). We start with (VC) (and turn to (RC) later).

Carnap knew that [SCA] will not do. In *Testability and Meaning*, from which we have quoted already, he asks us to imagine a match, a, which has, as a matter of fact, never been in touch with water and never will have the chance to because, as it happens, we are about to burn it. Then, as Carnap could easily show, the sentence $Tx \rightarrow Rx$ (which defines for us what it is for something x to be soluble) comes out true for the match a. Thus, we would have to say that it is also true that the match is soluble, for this is what we defined solubility to be according to [SCA]. Yet we know that wooden things like match a are *not* soluble. An untenable result.

Let's see why it is that $Tx \rightarrow Rx$ indeed comes out true for match a. First, we replace the dummy x with the specific a. Then Ta, the first part of the resulting sentence $Ta \rightarrow Ra$, has been, is and will always be false (we could write down $\neg Ta$): the match has never been put in water and never will be because we have burned it. It has lost its chance to be ever exposed to water. Now, no matter whether the second part Ra is true or not, the whole sentence $Ta \rightarrow Ra$ is already true according to how the material implication is defined. In other words, the match *unwarrantedly satisfies* the [SCA] criterion for being soluble. We speak here of a *void satisfaction* because the match fulfils the test for solubility without really having been tested.

Some might be tempted to say that we simply have to bite the bullet and accept that the match is/was soluble, for who cares now that it has been burned? Yet one further thought reveals how devastating the (VC) really is: define the contradictory, opposite disposition *being insoluble* as $Tx \rightarrow \neg Rx$ and match a will voidly turn out to have been insoluble as well! Yet, it can hardly have been both soluble *and* insoluble.[7]

Different conditionals instead of →. The following suggestion for a solution springs immediately to mind when we are confronted with this result: if the material implication, →, has such consequences it cannot be the correct formal representation for the 'if. . . then. . .' sentence we have intuitively in mind when thinking of dispositions. In fact, the complaint continues, did we not say from the outset that what we implicitly mean with attributions of dispositions is something like this: 'if match *a were* or *had been* put into water then it *would* (*have*) dissolve(d)'? That is, we phrase the 'if. . . then. . .' sentence that belongs to disposition ascriptions in so called *counterfactual* or *subjunctive* terms.

Unlike the material implication, such conditionals do not cling to the mere facts and (also) say something about possible and maybe not actual (i.e. *counter to the facts*) states of affairs. Transcending the mere de facto truths and falsities – here, the de facto absence of the putting in water of the match – counterfactuals take into consideration what would have been the case had we done it.

The latter is precisely what the material implication does and cannot do because of its truth-functionality: the mere fact of the absence of the test fixes the truth value of the material implication (to truth) regardless of what would have happened had we really done it. In this light, the (VC) problem turns out to have been foreseeable. The material implication, so it seems, cannot be successfully used to formalise what we mean by disposition ascriptions.

(2) Modality (MOD) We can rephrase what has just been said to lead straight to the second feature – *Modality* (MOD) – of dispositions: objects that are disposed to react in certain ways when triggered seem to have some *modal connection* to possible (future) manifestations. So called *modal statements* typically say what is *possible* or what is *necessary*, or, as in the case of dispositions, what *would be the case if something else were to happen*: 'Instead of this book you could read a novel' is an example of possibility; 'necessarily, everything that is red has a colour' an example of necessity; and 'if you were to put this into water it would dissolve' for the counterfactual modality. By contrast, non-modal statements (sentences with the material implication, for example) only trace what is de facto the case (or not the case) without any claim about what is possible, necessary or would happen under certain circumstances.

As a brief interlude, consider three further facts about modality. We will get back to many of those features in due course.

First, necessity and possibility are inter-definable: if something p is necessarily the case (for example, that red is a colour) then it is not possible for p not to be the case. Or, the other way round, if it is possible that p (for example, that you hold this book in your hands) then it is not necessarily not the case that *p*. So, if we know everything about necessity we also know everything about possibility.[8]

Second, there are interesting relations between modal and non-modal statements. For example, if it is the case that this book has fewer than 500 pages then it is also possible that the book has fewer than 500 pages, for were it impossible for the book to have fewer than 500 pages it would not have fewer. We'll leave these relations in the background for now.

Third, there are many kinds of necessity (and, thus, possibility) we could distinguish: *logical necessity* (the light is switched on or it is not switched on), conceptual or *analytic necessity* (all sisters are female), *mathematical necessity* (there is no largest prime number), *metaphysical necessity* (no object's surface is both red and green all over), *moral necessity* (one ought not torture anyone) and *natural necessity*. The latter we can describe intuitively (without taking this as a proper definition) as necessity due to how the world works. So, for example, if Einstein is right and no massive object can accelerate beyond the speed of light then it is naturally necessary that nothing moves faster than light.[9]

Coming back to our core issue of dispositions, note that (NON-OBS) – non-observability, the first of our three features of dispositionality – and (MOD) are not unrelated. On the contrary, the (NON-OBS) of dispositions is (partially) a consequence of their (MOD): you can observe with your own eyes only what is factual, what is here and now in front of you. Modal facts, such as what would happen if something else had happened, transcend the directly observable.[10] Simon Blackburn expresses this fact concerning modal necessity thus:

> Observations only extend to limited periods of space and time: how could we have within our view something that essentially casts its net over the whole of space and time?
>
> (Blackburn 1990: 243)

Similarly, we do not observe what disposed objects can do or must do or what connection they might have to unactualised events or states of affairs.[11]

A different formulation from [SCA] but retaining → Carnap and the Logical Empiricists were, of course, perfectly aware of all this, i.e. of the fact that dispositionality carries some modal force, that dispositional modality is hardly observable with your bare eyes (i.e. that, here, (MOD) leads to (NON-OBS)) and that, as a result, dispositional ascriptions are not so easily paraphrased using observational terms and the material implication as interpretation of the implicit 'if. . . then. . .' statement.

Yet adherence to their Verificationist doctrine meant the Empiricists had the strong urge to continue to try to spell out dispositionality in observational terms and, also, by sentence connectives whose truth conditions do not lead us onto non-truth-functional territory.

Triumph in these endeavours would be welcome for the following reason: if it is straightforwardly observable whether Tx and Rx are true or false then the truth of $Tx \rightarrow Rx$ (or some slightly more complicated version of it) can be easily calculated: only if Tx is observed to be true but Rx observed to be false is the combined sentence false, and true in all other cases. Thus, if some (more complex) definition with a material implication turned out to be acceptable after all the whole exercise would be a paradigm case of successful Logical Empiricism at work, i.e. prove empirically (by observation) which truth value the simple sentences have, then calculate logically what truth value the complex sentence has. So, the extensional, truth-functional definition of $\rightarrow$ suits Logical Empiricism very well indeed.

Thus, Carnap and the other Logical Empiricists reacted to the (VC) by keeping the material implication after all, yet reformulating the simple $Tx \rightarrow Rx$ sentence into a slightly more complicated definition than [SCA] is.

Finally, compare the observability and modality issue regarding dispositions to Hume's view on causation (Section 1.2): Hume fought the Rationalists' conception of causation, which said that the causal relation is a *necessitating* one – where, if the cause *c* occurs so does, with *necessity*, the effect *e* – and offered a definition of causation that was in no need of reference to modal necessitating forces.

In the light of this we can see the Logical Empiricists' further endeavour to define dispositional predicates, come what may, with the material implication as an attempt not only to bypass the (VC) but also as the hope to give an analysis of dispositions without the need to refer to modal counterfactual facts.

BOX 2.1.1 and 2.1.2: Some facts about dispositions and why, for Empiricists, dispositional predicates have to be analysed. How a first attempt fails.

- Dispositions, like being soluble, manifest themselves when the object that is disposed is in certain circumstances. Dispositions are said to have a **trigger** or **stimulus,** like being put in water, and a **manifestation** or **reaction**, like dissolving.

- The dispositions of things are not directly observable. Because of their **verification theory of sentence meaning** the Empiricists saw the necessity of translating sentences that contain dispositional predicates into sentences that contain only predicates that refer to directly observable features of the world.
- The **simplest way to analyse** (sentences with) **dispositional predicates**, like 'being soluble', is possibly **[SCA]: Dx =def. $Tx \rightarrow Rx$**. The arrow $\rightarrow$ is the **material implication**.
- This definition, with **the truth-functional material implication** at its heart, **suits the Empiricist** because, if it is reliable, we simply calculate the truth of Dx as long as we can establish the truth or falsity of Tx and Rx by observation.
- The decisive reason against the definition is, however, **the problem of the Void Satisfaction (VC)**, according to which we would be obliged to attribute D (here: solubility) to anything which has never and will never be tested.
- That **(VC)** can trouble us is, on the surface, due to the **paradoxes of the material implication**. The deeper underlying problem is that dispositions comprise **modality (MOD)**: a disposition ascription to a thing is also an ascription to **what that thing *could* or *would* do if it *were* in such and such circumstances** (even when it isn't). This counter-to-the-facts assumption can hardly be encoded in the mere facts-tracing material implication.

2.1.3 Random Coincidences (RC) and Productive Responsibility

Reduction Sentences Instead of saying that the disposition solubility, Dx, is defined by [SCA], namely $Tx \rightarrow Rx$, Carnap's next idea was to propose the following explication, *Reduction Sentences Analysis [RSA]*:

> [RSA]:
>
> D cannot be defined by [SCA] [. . .]. But we can introduce it by the following sentence: Tx $\rightarrow$ (Dx $\leftrightarrow$ Rx), in words: 'if any x is put into water [. . .], then, if x is soluble in water, x dissolves [. . .], and if x is not soluble in water, it does not'. This sentence belongs to that kind of sentences which we shall call *reduction sentences*.
>
> (Carnap 1936: 440, emphasis added[12])

With this kind of definition the problem of (VC) is indeed solved. Let us see why. Remember that our match *a* has never been put in water and never will be. *Ta* is and will be false. The new definition only advises us to call something x soluble (or not) *in cases where it has been put in water*, Tx, but the definition remains silent about the case where the test has never been made (that is, there is no second sentence belonging to [RSA] starting $\neg Tx \rightarrow \ldots$). Thus, we are not forced to attribute solubility to the match as we were with [SCA]. This is, prima facie, a good result. The new kind of definition is not liable to the (VC) problem: where no test has been made we remain silent.

Partial definitions Yet the solution comes at a high cost. Sure, it is good result if the definition neither entitles nor forces us to attribute a disposition to a thing which we believe is not so disposed (solubility and the now burned match, for example). However, this also holds for untested objects which we *do* think are so disposed: take a sugar cube that has never and will never be put in water. [RSA] obliges us to withhold a judgement as to whether it is soluble or not. According to [RSA] it is undetermined whether it should be called soluble or not. Thus, [RSA] is only a *partial definition*, which does not give a decisive answer for each and every state of affairs we could encounter.

Carnap has two suggestions to upgrade the definition. First, he adds that, for one and the same dispositional predicate, we might well formulate further reduction sentences with additional test and reaction pairs: $T^*(x) \rightarrow (D(x) \leftrightarrow R^*(x))$, $T^{**}(x) \rightarrow (D(x) \leftrightarrow R^{**}(x))$, etc. For example, for being fragile one might start: 'if x is hit then if and only if it breaks it is fragile' and 'if x is dropped then if and only if it shatters it is fragile', etc. The more reduction sentences are spelled out the more likely it becomes that an object was tested by at least one of the tests T, T^*, T^{**} . . . so that we might not have to abstain from judgement after all.

Second, Carnap suggested adding informally to these reduction sentences some laws of nature that hold between the disposition and kinds of substance that generally do (or do not) possess them: in our case this would be a statement like 'Wooden objects are not soluble'. Others, for different dispositions, would be 'uranium is radioactive', 'copper is conductive', etc. From such law-like correlations (from a kind of substance to a disposition) we could deduce that the never tested but wooden matchstick is not soluble (Carnap 1936: 445). Note that with Carnap's suggestion two new concepts from the philosophy of science have entered the stage: *laws of nature* and *kinds of substance*. (We leave these topics aside for now but warn that they will cause at least as much trouble as dispositions do; see Section 4.3 and 6.3.)

Verificationism weakened Even if the trouble with judgement gaps is diminished due to Carnap's two suggestions, there remains a deeper problem. It arises when we think of the goal the Logical Empiricists had, namely to reconstruct all (scientific) concepts on the basis of concepts that refer to observables (Carnap 1928: xvii; Section 1.4.1). This was meant in the strongest way, namely that a full and *explicit definition* had to be offered such that the concept could be readily replaced by the definition in each sentence in which it occurs.

Now, however, such an exchange is no longer possible because a reduction sentence does not express a formula that is equivalent to *Dx*. Instead, *Dx* is somewhat hidden in the middle of the formula [RSA] and, thus, only implicitly defined. As a consequence, the radical Verificationist principle which demanded a full exchange of all non-observational language for observational language was weakened. The requirement for Empiricists had to become less radical:

> Descriptive predicates are [. . .] not to be admitted unless they have *some connection with possible observation*, a connection which has to be *characterized in a suitable way*.
>
> (Carnap 1937: 33, emphasis added)

'Some connection with possible observation' is much less demanding than elimination in favour of observational language! And more trouble is to come. . .

Random Coincidence (RC) [SCA] was haunted by (VC). With [RSA], there is a possible way to deal with the problem. However, when we introduced (VC) we mentioned a twin problem: Random Coincidence (RC). It's a sibling of (VC) in so far as it also arises from the shortcomings of the material implication.

Suppose that another wooden match, a^*, is in fact put in water. To our surprise it dissolves instantaneously because, unknown to us, there just so happens to be a source of some very strong radiation in the glass. This radiation is by itself forceful enough to let wood decompose (its carbon substructures are destructed by the rays, say). Now, first look at both the simple analysis of dispositional predicates, [SCA], and then at Carnap's reduction sentences, [RSA]. (RC) means trouble for both.

Both *T* and *R* are true of a^* (it is in water and it does dissolve) and so, according to [SCA], $Dx =_{def.} Tx \rightarrow Rx$, a^* would have to be called *soluble*: Da^*. The same is true for reduction sentences [RSA] like $Tx \rightarrow (Dx \leftrightarrow Rx)$, for they say that, should something be tested – and a^* was tested – then it is to be called *soluble* in cases where it dissolves – and a^* did dissolve. So, a^* has to be called *soluble* according to [RSA], too. Thus, even if they are of help against (VC), reduction sentences are no remedy for (RC), i.e. cases where objects show *T* and *R* by coincidence and not because they themselves are disposed to do so.

If we are asked what has gone wrong in such a case our answer is possibly this: the match did not dissolve *for the right reason*! In order to be correctly called *(water-)soluble* it has to dissolve *because of its inner features and the water (alone)*, not because of the accidentally, randomly present radiation.[13] From this promising intuition we can extract two aspects: (1) the first is that we might wish to specify our test conditions more precisely so that *disturbing factors* are not present; and (2) the second focuses not on the absence of unwanted factors but highlights the importance of the *right* factor: *the (allegedly) disposed object* and *water* should be *causally/productively responsible* for the dissolving (not any radiation); and if they, water and the object alone, cannot do the job the object is not so disposed.

Exclusion clauses for disturbing factors Suggestion (1) advises us to reformulate and make precise the antecedent of our conditional so that radiation and other such *interfering influences are excluded* from proper tests as unwanted external factors. For [SCA] this would be a reformulation like: $Dx =_{\text{def.}} ((Tx \wedge \neg F_1x \wedge \ldots \wedge \neg F_nx) \rightarrow Rx)$, where the F_i are factors which have to be absent for decisive tests. So, $\neg F_1x$ might say that x is not exposed to radiation, $\neg F_2x$ that it is not exposed to extreme heat, etc. (We leave the respective reformulation for [RSA] to the reader.)[14]

Note that the F_is refer both to factors that would make indisposed objects look disposed (radiation, the match and solubility) and also factors that would make disposed objects look indisposed. We have, so far, not encountered an example of the latter case. Here is one: water might be supersaturated with sugar so that an additional sugar cube would, if dropped into it, not dissolve. That sugar cube would, without any provisos, count as insoluble.[15]

Causation to the rescue We turn now to suggestion (2): what we do not want is for the test for a disposition of an object and its typical reaction to coincide only randomly – on the contrary, we want the appropriate test circumstances plus the constitution of the object to be *the crucial causal factors* of the reaction.

This suggestion was in fact advanced by philosophers working on an Empiricist analysis of dispositions.[16] However, as we shall see in Chapter 5, the enterprise to say what the causation or a causal factor is proves to be at least as difficult as spelling out what dispositions are. This is true even when we leave the strictures of Empiricism behind, which, we must not forget, make causation a very contentious issue (Section 1.2). Be that as it may, we leave aside the possibility of reformulating [SCA] and/or [RSA] via this route for now. Yet this is the place to introduce the third feature of dispositionality we mentioned at the beginning of our chapter. We have already discussed (1) non-observability (NON-OBS) and (2) modality (MOD). Now we are in a position to add and explain (3) production (PROD).

(3) Responsibility for Production, Efficacy, Causal Power (PROD) Independently from the (RC) cases just discussed, this third feature is often intuitively associated with dispositions. Consider a poison. It is lethal. It has the power to kill someone. That is, because of its power, when ingested it brings about death. It is causally responsible when the person who took it dies. Consider dynamite. It is explosive. When its powder fuse is inflamed it causes a detonation. It is within the dynamite's causal power to bring about explosions. Not all dispositions like being lethal or explosive have this connotation of conferring *active causal productive* powers to objects. Fragile things rather seem to suffer passively from being dropped. Still, one is tempted to attribute some kind of (active or passive) efficacy to the objects that have dispositions.

Now, if we could demand in our definition of dispositionality that it, the disposed object plus the trigger, were responsible for the production of the reaction, that *the object was exercising its power*, so to speak, then we would thereby exclude (RC), for these are cases where triggers and reactions are not related but only coincide accidentally (or, at least, not via the disposed object but via some other path).

Just as (NON-OBS) and (MOD) are related to each other, so is production (PROD) related to both of them. In fact, we often find a smooth transition from *x has the power to do y* or *x can be responsible for the production of y* to a modal if-then statement about non-factual events: *if x were in the right circumstances, y would result*. In history, such an alleged power has even been equated with modal necessitation or modal necessity. An exemplary example is Hobbes, who wrote 'all the effects that have been, or *shall be produced*, have *their necessity* in things antecedent' (Hobbes 1655: 9.5, emphasis added). Hume and his followers also argued that

> efficacy, agency, power, force, energy, necessity, connexion, and productive quality, are all nearly synonymous.
>
> (Hume 1739–40: Book I, Part III, Sect. XIV: 157)

only to subsequently reject all of them by pointing out that none of them belongs to the observable features of the world (NON-OBS): 'External objects as they appear to our senses, give us no idea of power or necessary connection' (Hume 1748: Sect. IV, Part I: 63–4; Section 1.2).

Summing up, we are confronted with the following tension again. On the one hand both our intuitions about dispositions and, likewise, undesirable cases of (VC) and (RC) demand (MOD) (here: if *T* were the case, R would be the case) and (PROD) (here: it is the object with its disposition that brings it about that R when T). Yet, on the other hand, because neither (MOD) nor (PROD) are directly observable (NON-OBS), the Empiricists were sceptical about them and, thus, stuck to their plan to rephrase dispositional terms in observational, truth-functional language.

Nelson Goodman, in 1955 still in the wake of (post-)Empiricism, nicely summarises the resulting agenda for (Logical) Empiricist research regarding dispositions:

> Besides the observable properties it exhibits and the actual processes it undergoes, a thing is full of threats and promises. The dispositions or capacities of a thing – its flexibility, its inflammability, its solubility – are not less important to us than its overt behavior, but they strike us by comparison as rather ethereal. And so we are moved to inquire whether we can bring them down to earth; whether, that is, we can explain disposition terms without any reference to occult powers.
>
> (Goodman 1955: 40)

BOX 2.1.3: Reduction sentences, random coincidences and causal powers

- To remedy the disadvantage of [SCA], Rudolf Carnap suggested **reducing dispositional predicates** in a different way, **[RSA]**: $\boldsymbol{Tx \rightarrow (Dx \leftrightarrow Rx)}$. These so called ***reduction sentences*** have the advantage of coping with the (VC) problem.
- Yet **implicit definitions**, like [RSA], have a couple of disadvantages, too: they **remain silent about untested objects** as to whether they are disposed or not and, in accepting reduction sentences, **the Verificationist criterion of meaning** has to be **weakened**.
- There is a further predicament for both [SCA] and [RSA] alike: **Random Coincidence (RC).** This problem is that no matter whether an object really has a disposition or not, we would have to attribute it to the object according to both [SCA] and [RSA] **even if a reaction *just so happens* without being caused by the proper test and the object's own features** (but by something else's).
- Yet we only wish to call an object *disposed* when it reacts to being triggered because ***it itself* is responsible for producing that reaction.** A randomly occurring coincidence of test and reaction does not indicate dispositionality.

- We have collected three features of disposition ascriptions that cause trouble for the Empiricists and their Verificationist criterion of meaning:

  (NON-OBS) Non-observability. Dispositional features of objects are not directly observable.
  (MOD) Modality. Disposition ascriptions are modal statements: they say what could or would happen if something else were the case and thereby transcend what is merely the case.
  (PROD) Responsibility for Production. A thing is said to have a disposition only if it itself plus the trigger are responsible for the occurence of the respective reaction (and not if the reaction occurs for some other random reason).

- These three features are not independent: **(MOD)** and **(PROD)** have, historically, been lumped together and, in Empiricist quarters, been jointly rejected because of their **(NON-OBS)**.

2.1.4 Natural properties and natural laws

So far, none of the suggestions to get rid of dispositional talk via translation into an observational, non-modal, truth-functional language have succeeded. In fact, in order to avoid future disappointment we warn the reader that no suggestion will be to the satisfaction of every philosopher. In this section we will consider a last through-and-through Empiricist effort to analyse dispositional predicates. Much can be learned from this attempt.

Carnap himself hinted at the possibility of enhancing the reduction-sentence method by supplementing it with what we generally know about the regular behaviour – which is, ideally, governed by a *law of nature* (Chapter 4) – of the material the object under concern is made of – the natural kind it belongs to (Section 6.3). For example, if a thing is made of copper we could conclude from the laws of electric conductivity without performing the respective reduction-sentence test that the untested thing is conductive.

The Finnish Empiricist philosopher Eino Kaila (1890–1958) suggested a way to incorporate Carnap's informal idea into the first simple formal definition [SCA] we considered. Kaila's aim was to meet the radical Empiricist demand for an explicit definition and to shield that definition from the problem of (VC). As we shall see, it has the potential to handle (RC), too. Also, Kaila's innovation will unearth that an analysis of dispositions and dispositional predicates can hardly succeed without also having a clear understanding of what *natural kinds* or *properties* and *laws of nature* are. This nicely sets the stage for the chapters to follow, which deal precisely with these issues; but let us now turn to Kaila's idea.

Definition Here's his suggestion for a definition of dispositional predicates:

Eino Kaila's analysis [EKA]
$Dx =_{def.} \exists F\ [Fx \wedge \exists y\ (Fy \wedge Ty) \wedge \forall z\ ((Fz \wedge Tz) \rightarrow Rz)]$

(Kaila 1939: 239)

This looks more complicated than it is. [EKA] says simply that an object *x* has the disposition *D* in cases where the following is true: (1) there is some feature *F* that *x* has (this is what $\exists F$ and Fx say; (2) there is at least one object *y* which has also *F* and which has been submitted to the test already (that's what $\exists y\ (Fy \wedge Ty)$ means); and (3) it is generally true of all objects *z* that have the feature *F* and that are or have been tested that they show or have shown reaction *R* (this is, finally, encoded in $\forall z\ ((Fz \wedge Tz) \rightarrow Rz)$).

What Kaila wishes to capture formally is exactly what Carnap had put in prose to supplement his reduction sentences: object *x* is disposed in cases where it is made of some material or has some significant feature *F* that it shares with some (other) objects *y* and it is universally (ideally, by law) the case that all other objects that have been tested have reacted positively.

Avoiding (VC) With this extension of [SCA] in place, the problem of (VC) seems to disappear, for even if *x* itself has not yet been tested (and never will be) that does not, in itself, suffice to attribute dispositionality to it. Rather, we have to check how other objects of the same kind *F* behave or behaved. Only if other objects that are *F* have all reacted positively on being tested does *x* also count as disposed; if it has not then it does not. The other objects serve, so to speak, as proxies for the untested *x*.

Take the troublesome match from above which came out soluble according to the too simple definition [SCA]. If everything goes well it should not come out soluble, according to Kaila's definition. The match is wooden (here, we naturally assume that $Fx = x$ is wooden) but no wooden thing has yet shown dissolving behaviour.[17] So, encouragingly, the final part of Kaila's definition is not fulfilled: it is not the case that $\forall z\ ((Fz \wedge Tz) \rightarrow Rz)$; on the contrary, most or all *Fs* will have reacted with $\neg R$ on *T*. In short, via [EKA], match *a* seems to be saved from being called *soluble*.

However, there are two difficulties with Kaila's formula that we need to focus on in more detail. The first concerns the feature *F*, the second the status of the universal regularity $(\forall z\ldots)$ Kaila's definition relies on. Both will lead us straight into further issues within the metaphysics of science that we will deal with later.

Nominalism: Properties and arbitrary classes of things The problem with F is that the Logical Empiricists held a very unrestricted view of what *properties* are: there is no more to a property than a set of objects and a name for it (hence *Nominalism*). Here's an example of their permissiveness: take any class of things you like and then give it a name. Choose, for example, the class which consists of some diamonds, the dark side of the moon and a random wall. Let us define, next, the (artificial) predicate *being floydy* in the following way: an object x is *floydy* if and only if x belongs to the above messy class of things, and if x does not belong to the class then saying *x is floydy* will count as false. Consequently, it would be correct to say that the dark side of the moon is *floydy* but some random animal is not, etc.

Now, the Empiricists add a radical step: take any such artificial predicate you like, take also all predicates you naturally find in your language. Each predicate, so the Empiricists claim, corresponds to a property, one property per predicate. This abundant view of properties – often called *(unrestricted) Nominalism* – amounts more or less to saying that any set of things, even the messiest one, to which a predicate might naturally or artificially refer constitutes a property. In fact, in this view properties are no more than sets of things.

We must set aside the reasons why the Empiricists defended Nominalism.[18] For now, let us accept that the Empiricists indeed had this *anything goes* concept of predicates/properties. The intuition many of us will share is that this view is weird; in fact, many philosophers today think differently about properties. We come to this is due course.

The return of (VC) Assuming that the Empiricists' abundant view of properties is correct, Carnap found that an invented predicate/property F means trouble for Kaila's definition because it enables (VC) to sneak back in. As he is allowed to do, Carnap focuses on a rather artificial class of things that contains only two objects: the match a of his own example and another random object b (a sugar cube, say). He further defines the predicate F (compare *floydy* above) as being true of something if and only if it is in the weird class of these two objects. That is, Fx holds true in cases where x is either the object a or x is the object b. Formally: $Fx \leftrightarrow (x = a \vee x = b)$.

Carnap stipulates one more thing about objects a and b. As we know, a is the already familiar match that has never and will never be put in water. So, $\neg Ta$ is true of it. For b, the random object (remember that we chose a sugar cube), Carnap assumes that it was in fact positively tested (i.e. it did dissolve in water). That is, both Tb and Rb are true.

Even if we might not see this straight away, the unwanted result that match a is to be called *disposed* (here: *soluble*) follows from Kaila's analysis and Carnap's definition of F: this is (VC) yet again!

Why is this so? Well, readers can easily convince themselves that each part of Kaila's definition is fulfilled. (If you wish to cheat then endnote[19] is for you.)

Restrictions on predicates and/or properties The most natural way to react is to dismiss Carnap's *F* as being nuts. Some restrictions, so we feel, have to be implemented as to which predicates/properties are allowed for *F*. Not anything goes. Many suggestions have indeed been made:

(1) *F* has to be a *general class*, i.e. one that has potentially infinite members and not from the outset just the two, *a* and *b* (Kaila 1942);
(2) *F* has to be an *intrinsic property* of the objects that have it, i.e. something like its chemico-physical substructure (Wedberg 1942: 44);
(3) *F* should be entrenched (or *inductive* or *projectible*), i.e. F has to be an established predicate we have successfully used in the past to make predictions (Goodman 1955: 81ff);
(4) *F* should be a *natural* property in some sense, like, for example, being made of wood or of metal (Storer 1951: 137, fn. 1).

Carnap's contrived *F* does not fulfil even one of these restrictions and, so it would seem, Kaila is safe from it. Yet all of them have their own drawbacks.

For (1), which is the most innocent of the suggestions, it could be shown that with a shrewd new Carnap-style definition (VC) returns even with a general *F*.[20]

The problem with all the other suggestions – (2) *intrinsicness*, (3) *entrenchment* and (4) *naturalness* – is the opposite: these restrictions might well work but a proper explication of what is meant by them is hard to obtain. It is also interwoven with other issues from the metaphysics of science: we might be tempted, for example, to say that *natural* properties are those that appear in the laws of nature. Yet then we are confronted with the problem of what a *law of nature* is (Chapter 4). Similarly, *entrenchment* is related to successful inductive practices. Yet how induction works and how it can be justified is also one of the persistent troubles that haunt the epistemology of science. In Section 6.3 we will have the opportunity to learn a bit more about *natural properties* and *kinds*; for now, it is enough that we have become sensitive to the issue: something more has to be said about properties than what the Logical Empiricists said.

(RC) revisited We have not forgotten that, next to (VC) (which Kaila's definition was originally supposed to solve), there is the issue of (RC). Let us briefly see how Kaila's [EKA] fares with (RC).

In [EKA] the allegedly disposed object has to belong to a class of things, *F*, for which the test-reaction occurrences are a *universal* *regularity*: $\forall z((Fz \wedge Tz) \rightarrow Rz)$. That is, with Kaila's definition, it would be a rather big coincidence if all the many tested objects that are *F* had randomly, just by chance, shown the reaction *R*. Such global coincidences are unlikely to happen. So, this looks promising.

Laws of nature Of course, it would be better still if Kaila had demanded that that regularity is not just true but that it is a *law of nature* or, at least, that it follows from natural laws. This is so because without this additional requirement there is still, in principle (even if it is unlikely) the chance that all instances occurred by chance. Think of classes *F* that do not have very many members, as is the case with, for example, newly synthesised materials or superheavy atoms like ununoctium. There, only a few tests for certain dispositions might have been made and reactions might indeed have occurred only by chance as artefacts of the experimental set-up instead of the material's own features and the proper test. Yet if we knew that the correlation from test to reaction is *a law of nature* for the relevant material then we could be sure that the few positive instances occurred for the right reasons.

Thus, a law-like, so called *nomological* connection between being *F*, being tested and reacting would provide us with the resources to avoid the problem of (RC). It should be noted, however, that a nomological link would be somewhat akin to (PROD), the demand for a causal responsibility between disposed object, trigger and manifestation. Also, laws of nature might not be free of (MOD): if it is a law that all things *F* react with *R* when in *T* then this entails that if something were an *F* and *T*-ed it would react with *R*.

The important part of this section is not so much the addition of another twist in the analysis of dispositional predicates (here: [EKA]). Rather, it is that the issue is very closely related to the topics of *natural properties* (Sections 4.2, 4.3 and 6.3) and *laws of nature* (Chapter 4).

BOX 2.1.4: Eino Kaila's definition: a crucial reference to natural properties and laws of nature

- **Kaila** incorporates formally into an explicit definition what Carnap has already suggested in prose as an enhancement of his own implicit definition (reduction sentences): **reference to a class of like objects** and **a statement about the general behaviour of members of that class**.
- With his **explicit definition**, he would also fulfil the Logical Empiricists' desideratum to make it possible to exchange all

terms that do not refer to directly observable features of things (like their dispositions) for those that do.

- However, it turns out that, for the definition to have a chance to work, the mentioned class cannot be any odd bunch of objects but that it has to be a somehow **general class of things** that share some **intrinsic, entrenched or natural properties**. We hinted at the fact that the precise explications of what these demands come down to are complicated and in any case hardly available for the Empiricist.
- Also, the reference to the **universal, regular behaviour** of the class's objects probably does not suffice. Instead, the stronger demand for **laws of nature** seems to be necessary.

2.1.5 Summary: (NON-OBS), (MOD) and (PROD)

Dispositions are features of things that are not immediately observable (NON-OBS). Thus, the Empiricists' aim was to translate sentences that contain dispositional predicates into sentences that refer to observational terms only. Two major stumbling blocks were in their way: cases of (VC) and (RC). Superficially based in the paradoxes of the material conditional, (VC) and (RC) reveal deeper concerns. What (VC) unearths is that dispositional ascriptions are hidden (MOD): they implicitly say what could happen to a thing even if it does not now happen. What (RC) uncovers is that we are attributing some responsibility, some productive link, instead of mere loose, coincidental occurrences of events when we ascribe dispositions to things (PROD).

Unfortunately, both the supposed modal features of allegedly disposed objects and their assumed causal powers are unobservable and, thus, the Empiricists had all the more reasons to be sceptical of dispositions. Being confronted with the problems of (VC) and (RC) caused great tension: on the one hand these difficulties indicated a possible failure of the Empiricists' enterprise. On the other hand they increased the Empiricists' incentive to continue with their quest for a reformulation of sentences with dispositional predicates in terms of pure observational non-modal language. This is to get rid not only of (unobservable) dispositions but also of the modality and causal powers associated with them.

As we have seen, the attempts [RSA] and [EKA] were at best partial successes. Certainly, the ideas to incorporate *causation*, *laws of nature* or *natural kinds,* or to use, instead of the material implication, *counterfactual conditionals* would be of great help. However, these would be pyrrhic victories because the issues of (MOD) and (PROD) will in further analyses resurface as major stumbling blocks. We will see this when we turn to these topics in their own right (Chapters 3 to 6).

Before this we have to comprehend a major change in focus within the philosophy of science that took place in the literature on dispositions after Logical Empiricism: the turn from observations and sense data, semantics and language to ontology.

2.2 Ontology

2.2.1 Farewell to Empiricism

Remember Logical Empiricism and its core: the Logical Empiricists pushed classical Empiricism to the extreme. They turned the classical Empiricist epistemic dogma – that all factual knowledge, i.e. knowledge about the world, has to derive from sense experience and sense experience alone – into their semantic Verificationist doctrine: the meaning of a sentence is nothing but the test method through which the sentence's truth or falsity can be established by sense perceptions. If there is no such test method there is no meaning to the sentence. It then expresses nothing or, at best, nonsense.

We have twice seen how hard it is to defend this doctrine: in Chapter 1, where we encountered several general difficulties for Verificationism (Section 1.4.2), some of which were problems for the analysis of dispositions, too; and in Section 2.1 we were confronted with more trouble, i.e. (VC), (RC) and the related issues regarding (MOD) and (PROD).

These difficulties contributed to the downfall of Empiricism because they are not limited to dispositions. In fact, we need to understand that the debate around dispositions has been of such a central importance not because fragility, solubility and other ordinary dispositions were deemed so very interesting. Rather, it was because sentences that contain non-observational scientific vocabulary, i.e. theoretical terms like *electron* or *quark* that do not refer to entities or properties that can be directly seen, heard, etc., also had to be translated into sentences that contain merely observational terms. These translations for theoretical terms in general have exactly the form the analyses of dispositional predicates exemplify: both share an *if such and such test is performed then this and that observation can be made* structure. And, so, problems that arise for dispositions arise for scientific theoretical terms of all kinds, too. This is why Empiricism as a whole was shaken so much by the trouble dispositional predicates caused.

2.2.2 An ontological turn

With these severe difficulties for Empiricism/Verificationism becoming more and more obvious, philosophers became less and less interested in pursuing the Logical Empiricists' programme. They turned away from the notorious focus on language and observation and dared to move towards ontological questions.[21]

Indeed, from the late 1960s till the early 1980s a debate arose surrounding the question of whether dispositions have categorical bases. Here is David Armstrong (1926–2014), one of the first to tackle these issues:

> To speak of an object's having a dispositional property entails that the object is in some non-dispositional state or that it has some property (there exists a 'categorical basis') which is responsible for the object manifesting certain behaviour in certain circumstances, manifestations whose nature makes the dispositional property the particular dispositional property it is [. . .]. In asserting that a certain piece of glass is brittle, for instance, we are ipso facto asserting that it is in a certain non-dispositional state which disposes it to shatter and fly apart in a wide variety of circumstances.
>
> (Armstrong 1968: 86)

This quote can be found in Armstrong's seminal *A Materialist Theory of the Mind*, a book that comes, as the title indicates, not from general metaphysics but rather from the philosophy of mind.

Body and Soul One motivation for Armstrong to write his book was that he found some of Gilbert Ryle's (1900–76) talk about *mental dispositions* dissatisfying. Ryle had made progress towards a kind of *Monism* when it comes to the mental and the physical world. Saying of a person that they know, feel, wish, etc. something is, for Ryle, not to attribute some property to that person's soul or inner mind which would exist separately from and additional to their physical body. (This latter position is called *Dualism*, for there would be two things, a body and a soul.) Rather, these are simply attributions of dispositions to that person. Claiming that a person wishes to drink a beer is not to say that their soul is in that longing state but simply that they are disposed to open the fridge when they know there is beer inside, or that they are inclined to buy a bottle as soon as they have the chance, etc. (Ryle 1949).

So far so good. Armstrong accepts this. Yet he points out that there must be something in the brain of a person who wants a beer that is different from that of a person who does not. In other words, the Rylean (mental) disposition must, according to Armstrong, have some (physical) basis or realiser in the brain. Otherwise it would be mysterious why some people have the disposition and some do not. Again, Armstrong's thesis is that mental dispositions have (or are even identical to) physical bases, in this case brain states.

Unfortunately, we must leave these intriguing questions about the relation between mind and body aside here. Suffice it to say that they are but one, albeit intriguing, example of the general question 'Is a dispositional property identical with whatever in the object explains its presence?' (O'Shaughnessy 1970: 2).

Much more innocently (rather than loaded with complicated issues surrounding mind and brain), we might ask whether sugar is *soluble in water* because of its $C_{12}H_{22}O_{11}$ substructure and the dipole nature of water, H_2O. Remember another place where reference to some other property of the supposedly disposed entity was seen as essential for an analysis of dispositions (Section 2.1.5): Eino Kaila saw in the postulation of such an internal property (plus a law of nature that governs the behaviour of it) a way to handle the problems of analysis with the material implication.

Two things are noteworthy at this point. First, that the issue (as Armstrong presents it here) is phrased wholly in realistic or physicalistic terms – that is, it is a direct question about how things stand in reality. It is neither about predicates nor about observations (see below some caveats about this remark). Second, a distinction has been made on the ontological level – *dispositions vs. categorical properties* – and an intimate relation between these distinct kinds of property is assumed, namely that whenever an object is disposed it has some (non-dispositional) categorical property which is somehow responsible for the object being disposed.[22]

2.2.3 Categorical properties

What are these so called *categorical* or *occurrent properties* as opposed to dispositional properties? Typically, categorical properties concern the structure or form of something or its spatio-temporal location. Examples include the above-mentioned *molecular structure* (of, for example, sugar), *hanging on a wall* (of a picture) or *being pentagonal* (of a building). Contrary to dispositional features, those properties are *alive* as soon as an object has them. A good working hypothesis for the distinction of these two important kinds of property – dispositions vs. categorical properties – comes from Andreas Hüttemann:

> A *dispositional property* is a property that, if instantiated by an object, *is manifest under specific conditions only*. A *categorical property* by contrast is a property that, if instantiated by an object, *is manifest under all conditions*.
> (Hüttemann 2009: 225)[23]

Solubility, a disposition instantiated by sugar, manifests itself only when the sugar is in water; being $C_{12}H_{22}O_{11}$, also instantiated by sugar, is manifestly there, here and now, under any condition. We will, in the next section, unearth further features of categorical properties. Here, we have to insert a quick remark about observability in order to avoid misunderstandings.

Observability revisited Remember that we have left behind the Empiricists' obsession with verification via sense perception. Still, you might object that in order to find out whether an object has a categorical property you sometimes have to observe how it reacts in certain test conditions. This is especially the case with categorical properties like chemical structures of substances or, worse, the features of quarks. Thus, so this objection continues, categorical properties have this trigger-manifestation make-up just as dispositions do and, so, there is no difference between these allegedly different kinds of property after all.

Now, this is indeed true if we stick to the dogmas of Empiricism, and we have already mentioned that for Logical Empiricists there really is not much difference between dispositions and theoretically, scientifically postulated categorical features like features of quarks or chemical substructures.[24] At best, the macroscopic features of medium-sized objects, like something's being white, being rectangular, being in the room, etc., which can be observed with the naked eye without particular tests, could truly count as categorical properties.

Yet having left the Empiricists' perceivability issue behind, the blurred difference becomes more distinct. This is because the *if trigger then manifestation* core of dispositions is no longer meant to be about observations but merely about *what happens when*, whether these happenings are observable or not. This *what happens when* structure does not belong to the heart – to the essence – of categorical properties. The latter ones are just what they are. Observable or not, a sugar cube instantiates and manifests being $C_{12}H_{22}O_{11}$ as it sits there in the jug. Its solubility, however, is merely instantiated and not yet manifest (if ever).

Thus, one standard way to make the distinction between dispositions and categorical properties is that where dispositions are essentially related to what could or would have to happen if something else were the case, categorical properties are non-modal properties. (They are, if at all, only related to such modal facts via laws.)

Predicates revisited Another complication emerges when we reconsider the distinction between properties, i.e. the features of things in the world, and predicates, i.e. those linguistic terms language users employ to refer to such features. We said that we are now firmly on ontological grounds, dealing with properties, and we will stick to this assumption throughout major parts of the book. However, we must mention in passing that some philosophers – for example, Hugh Mellor – focus on predicates rather than properties to spell out the dispositional–categorical distinction. They claim that the difference is not an ontological matter: that is, not a difference in the properties objects have. Rather, they say that the difference is semantic, i.e. that it is a difference regarding what the words we use mean or entail:

> Dispositionality is a feature not of properties but of predicates [. . .]. Properties in our sense [. . .] need not in themselves be either dispositional or categorical: those that exist can just be.
>
> (Mellor 2000: 768–9)

When we use a dispositional predicate to describe objects, as in 'This sugar cube is soluble', we *mean* that sugar will dissolve when in water. This conditional belongs to the semantic meaning of a dispositional predicate. By contrast, when competent speakers use a categorical description, like 'Sugar has macromolecular structure $C_{12}H_{22}O_{11}$', they do not mean to imply any potential behaviour of sugar. Of course, they might also tacitly associate some behaviour with that molecular structure, but they do not *mean* this behaviour when using 'being of structure $C_{12}H_{22}O_{11}$'. So, the distinction dispositional–categorical has been made on the level of language, not on the level of what there is in the world.

Even if one buys this language-centricity, one might still wish to ask ontological questions about the objects and their features. For example: if the distinction is purely semantic, i.e. only on the level of language, what, then, are properties themselves? Are they neutral: that is, neither categorical nor dispositional? This is a position called *Neutral Monism*. Or are they both simultaneously? This could be called *dual-sided,* or *identity, view*.[25] Or is this a question that makes no sense? Further, we might wonder: do some dispositional and categorical predicates refer to one and the same neutral or dual-sided property? For example, does *solubility* refer to the same property as $C_{12}H_{22}O_{11}$?

As you have seen in our reference to Mellor, the insistence that the difference between the categorical and the dispositional is in language and not in the world is still an option (his article is from 2000). However, to keep things simple, from now on we adopt the majority opinion that the distinction is meant to be about properties in the world.

Pan-Dispositionalism vs. Categoricalism Whether, even if the principal distinction is meant to be about the world, both kinds of property really exist is, of course, another question. (Compare: we can distinguish unicorns from horses but only one of the two really exists.) And, so, some people claim that there is really only one type of property in the world. This *Property Monist* camp of philosophers divides into two antagonistic factions: the *Pan-Dispositionalists* claim that all properties are, after all, dispositional; the others, sometimes called *Categoricalists*, suppose that all properties are categorical (at least, at the most fundamental level of reality).[26] We will hear about Categoricalism in the next section. Dispositionalism is our topic in Chapter 6.

BOX 2.2.1–2.2.3: Dispositions and categorical properties

- Turning away from **Empiricism**, which, as we know, **saw insurmountable problems,** philosophers turned from semantic considerations of dispositional predicates towards the **ontology of dispositional properties**.
- A dispositional property, like being soluble or being inflammable, manifests itself when the object that has the disposition is in certain circumstances. Thus, dispositions are said to have a **trigger** (or **stimulus**) and a **manifestation** (or **reaction**).
- An object's **categorical properties**, like having a rectangular form or having a specific chemical structure, are, by contrast, manifest as soon as the object has them. Categorical properties need no trigger, they are always right there. (What you see is what you get!)
- As opposed to dispositional properties, **categorical properties are non-modal** properties: **dispositions are related to what could or would have to happen if something else were the case**, whereas categorical properties mediate to such modal facts only, if at all, via laws. In fact, the latter difference is the basis for what has become the **standard distinction between dispositions and categorical properties.**
- Some philosophers believe that, even if this distinction can be made conceptually, all properties that really exist in the world are of one of the two kinds. This position is called Property Monism. It divides into the two opposing groups of **Dispositionalists** and **Categoricalists** (as well as a third, **Neutral Monists**, who think that the division exists only in language). We also find the position of **Property Dualism**, the position that both kinds of property exist. (In this book, we focus by and large on the division between Dispositionalists and Categoricalists.)

Literature

- Two well-known introductions to the topic of dispositions, their features and the history of the controversy surrounding them are **Stephen Mumford's *Dispositions*** (1998) and **George Molnar's *Powers*** (2003). A concise history can be found in **Wolfgang Malzkorn's** (1999).

- For the discussion that followed in the twenty-first century see the edited volumes **Kistler and Gnassounou** (2007), **Handfield** (2009), **Damschen** *et al*. (2009), **Marmodoro** (2010); see also **Friebe** (2014).

2.2.4 Categoricalism

Categoricalists' dismissive claim about dispositional *properties* is very similar in spirit to Logical Empiricists on dispositional *predicates*.[27] The Categoricalists' belief is that alleged dispositional properties of objects are reducible to these objects' categorical features plus, maybe, laws of nature or causation or other addenda. Hugh Mellor has summarised this stance poignantly:

> Dispositions are as shameful in many eyes as pregnant spinsters used to be – ideally to be explained away, or entitled by a shotgun wedding to take the name of some decently real categorical property.
>
> (Mellor 1974: 157)[28]

A Categoricalist's first shot could, for example, be this: sugar's solubility is nothing but sugar's molecular structure $C_{12}H_{22}O_{11}$ plus the laws of nature that demand that $C_{12}H_{22}O_{11}$ disintegrates in water.

Multiple realisability Of course, Categoricalists are aware of the phenomenon of multiple realisability of dispositions: sugar is soluble and so is salt. However, salt has that property because of a totally different substructure: salt is an ionic compound, *Na+Cl-*, with which water dipoles can causally interact, whereas sugar is the macro-molecule $C_{12}H_{22}O_{11}$, which also disintegrates in the vicinity of water dipoles. Thus, solubility is, if at all, realisable by or reducible to different underlying structures. Ambitious attempts to reduce dispositions to categorical properties which can easily deal with multiple realisability will be introduced in the next chapter.

Micro–Macro Note something else of importance: for the Categoricalist, the question is not (or not only) the general one of whether macro objects have some of their macroscopic features because of the way they are composed from smaller units with specific micro properties. Put in such broad terms, the answer is fairly uncontroversial: almost all properties of macroscopic objects depend in some way on the micro properties of the objects' parts. A painting's overall visual properties, for example, depend, amongst other things, on the quality of the paint and its distribution. If, as a painter, you wish to change how your painting looks (under the same observation conditions) you have to change the distribution of colours, at least in some parts of your canvas. There is no macro change (what the canvas shows) without a micro change (how the oil or acrylic, etc. is applied).

The more specific question that can be asked is whether the (macro) *dispositional properties* of an object are ultimately there due to no properties other than the (micro) *categorical properties* of the object (and, as we said, probably the laws of nature involving these properties) or whether some of those micro properties are themselves dispositional. Categoricalists answer decisively that the most basic properties of the most fundamental entities in nature are all categorical. (Think, as a close approximation to the *most fundamental properties*, of those fundamental physics postulates.) All the rest, and especially the dispositional properties of things, somehow arise out of these fundamental categorical ones. This Categoricalist position deserves the label *Reductionist*: at its core is the belief that dispositional properties ultimately reduce to categorical ones.

Hume's Dictum So far so good, but what is the Categoricalists' motivation for their credo? For the Empiricists and their dogmas the issue was clear: to fulfil the Empiricist doctrine of Verificationism at the heart of which lies the conviction that all knowledge about the world comes via the senses. But why would anyone insist that the world's only (or fundamental) properties that ground everything else are of a categorical nature?

The short answer is that Categoricalists inherit from Hume (and the Logical Empiricists) not the obsession with language and observation but the sceptical attitude towards (MOD) and/or (PROD). Dispositions, if they existed as real properties in their own right, would bring (MOD) and/or (PROD) to the world (Sections 2.1.2, 2.1.3) and, so, Categoricalists wish to show that dispositions do not exist or that they reduce to some underlying categorical basis. For this enterprise to succeed it is, of course, of primary importance to make sure that the basis to which dispositions are reduced is not itself modally loaded or powerful. For this reason, categorical properties are assumed to be free from any modal commitment.

Before we come to this important characteristic of categorical properties and its consequences we need to pause for a second and ask why anyone would keep the anti-modality, anti-power attitude when the old reason for sustaining it, unobservability, is no longer judged to be of importance. This is astonishing – and it is even more surprising that people starting to be puzzled about it is a recent development:

> The positivists [Empiricists, MS] had either to rework the notion of a law [or causation, or dispositions, MS] or abandon it altogether. But no one is a positivist any more, and the shortcomings of Verificationist theories of

> meaning are so well known as not to bear repeating. It is odd, then, that contemporary philosophers should flock to the banner of Hume. Lewis does not announce himself a positivist, and presumably would be embarrassed at the association. The semantic theory that underlies Hume's own views has been thoroughly discredited. Why should one have 'Humean scruples' any more?
>
> (Maudlin 2007: 70–1)

We will come back to this in more detail in Section 7.8. For now, we simply assume that Humean Categoricalists (David Lewis (1941–2001) was mentioned by Maudlin) have (other) good reasons for their sceptical attitude towards (MOD) and (PROD). A term that has been established recently for the anti-modality, anti-powers dogma that there are no necessary causal connections in nature is *Hume's Dictum* (Wilson 2010). We will now unearth what it means for a property to be free of any such modal/causal import.

Quiddities Suppose *being NaCl* is a categorical property. We might, of course, associate certain potential behaviours with *NaCl* because we have observed that *NaCl* has figured in some causal chains in the past – that it dissolved when put in water, for example. Yet, so Categoricalists claim, this potential behaviour, these possible causal interactions, these roles *NaCl* can play in the world, neither belong to the chemical definition of what it is to be *NaCl* nor to this property's essential nature, i.e. to what it is to be that property.

Categoricalists sometimes prompt this latter intuition by asking us to imagine the laws of nature to be slightly different than they actually are. With different laws, things (here, especially those things that have the categorical property *NaCl*) would behave differently. In other words, it seems conceivable that, with the laws changed, *NaCl* does not dissolve when in water, that it does not engage in that causal process. The possibility of a scenario where *NaCl* behaves differently under different laws shows, so the Categoricalists would continue to argue, that *NaCl*'s current regular behaviour does not belong to its very nature, for that behaviour could change while *NaCl* is still what it is.[29]

This is very different from dispositional properties like solubility. Sure, if the laws were different salt might not dissolve anymore when put in water. Yet, if that is so, then salt has also lost its solubility (while it is still *NaCl*!), for dissolving when put in water is simply what it is to be soluble. It belongs to solubility's essential core to endow its bearer with the capacity to dissolve.

Back to categorical properties: we might start to wonder what such a property's essential core is if it is not the potential behaviour things can display when they have that property. The answer Categoricalists are going to give might come as surprise. Before we turn to it, consider the following analogy.

Interlude: Haecceities What can change about you without you losing your identity? A new haircut does not change your very nature. A sex change might be more fundamental. Still, would you say that you ceased to exist and a new person came into being should you undertake one? Is it not still you, albeit with a different sex? But can you turn into a poached egg and still be you? That is contentious to say the least. Still, some people think you can because what it is to be you is to have a certain *brute identity*, which is untouched no matter what qualities you otherwise have or in what relations you stand to other objects. This primitive identity is called a *haecceity*, from the Latin demonstrative pronoun *haec*: this. (Maybe souls going through different incarnations are haecceities?).[30]

Back to Quiddities Be that as it may, this idea of a primitive identity can be assumed about properties as well – this position is called *Quidditism* (from the Latin *quid*: what) – and 'Quidditism is to properties as haecceitism is to individuals' (Lewis 2008: 209). It says that quiddistic properties have a brute identity, which is untouched even if the laws of nature were to change massively and objects that have such a property behaved utterly differently from how they do in the world as it is. Now, you won't be surprised to hear that Categoricalists believe just that: each categorical property has a primitive identity that distinguishes it from other categorical properties, but these brute identities are independent of the behavioural potentials these properties endow their bearers with.

So, the answer to our question above – what is a categorical property's essential core (if it is not the roles that things that have it can play)? – is simple: their essential core is nothing but a (naked) primitive identity about which nothing more can be said. (From now on we will use the terms *Quiddity*, *quiddistic property* and *categorical property* interchangeably.)

Anti-modality This might sound very dissatisfying (and we will come back to it), yet Categoricalists hail this core feature of quiddistic properties. The reason is quite obvious: for what, again, would be the alternative? Suppose a certain role belonged to a property's identity, to its very nature, to what it is to be this property. This, for example, could be the role that makes it the case that an object that has this property dissolves if put into water.[31] If this is so then certain arrangements in the world are impossible: if an object has this property then the object *cannot* be both in water and not dissolve, i.e. this would be an impossibility. Another way of saying that something is impossible is to say that it is necessarily not the case. Thus, properties to whose identity such roles belong have some necessities inscribed in themselves; and necessities – think of Hume's Dictum – are *persona non grata* for the Categoricalist. So, Categoricalists embrace the primitive identity of Quiddities, which does not commit to any necessities: Quiddities by themselves are naked and do not prescribe anything.

The reader will long have noticed that we are back again to the categorical properties–dispositions divide: if your properties have more than a brute identity, i.e. if they are what they are because they bestow the objects which have them with the potential to react in certain ways in certain circumstances, then you have dispositional properties. Because of this potential or power that would come with them, Categoricalists wish to avoid dispositions and embrace Quiddities.

Consequences of Categoricalism/Quidditism Categoricalists are confronted with surprising consequences, and anti-Categoricalists such as the friends of dispositions regard them as a good reason to reject Categoricalism/Quidditism. Here's one: if the behaviour it enables or enforces an object to perform does not belong to the nature of a property – if the very core of what it is to be that property is brute – then how could we ever experience such a property? Our experiences of things in the world have to be causal: we get to know which objects have which features either via direct sense experience (a causal chain directly from the object to, say, our eyes via light waves) or indirectly via what these objects cause in other objects which are then, maybe via further links in the chain, causally connected to our senses.

Yet if, as a Categoricalist assumes, the causal routes the properties might initiate do not belong to their identity then we do not get to know the properties themselves. What we get to know – the causal interaction – comes, so to speak, later. So, the causal role the object plays is, on the one hand, the closest we can get towards the property in itself; on the other hand, because that role does not belong to the property's identity, it somehow veils its identity from us.

If the reader is now reminded of old acquaintances then this is quite right: remember Kant, who, for very different reasons and via quite a different argumentational route, ended up postulating unreachable *things in themselves* that lie hidden behind the appearances we have access to. Categoricalists have noticed and embraced this parallel between Quiddities and Kantian things in themselves. Influenced by Rae Langton's work *Kantian Humility* (Langton 1998), David Lewis, for example, agrees and welcomes that Categoricalism forces us to be humble regarding the prospects of knowing *categorical properties in themselves* (Lewis 2008).

Think about some further radical consequences this 'neo-transcendental idealism'[32] regarding properties might have: a categorical property, call it 'being electronegatively charged', which is, today, responsible for objects repelling each other if they both have it, could, tomorrow, be replaced by the property, call it 'being massive', which is, today, responsible for objects that possess it attracting each other. That is, the two primitive identities behind being electronegatively charged and being massive could swap their roles and this would happen unnoticed by us (Black 2000: 94; Lewis 2008: 212).

Or think of this further, similarly weird possibility: objects might have a million more quiddistic properties – Lewis calls them 'idlers' – than we ever would have dreamt of and that we'd never be able to notice: 'Idlers are those [. . .] properties [. . .] that are instantiated [. . .], but play no active role in the workings of nature' (Lewis 2008: 205). They just sit there (on objects), so to speak, and do nothing.

Now, most Categoricalists, including Lewis, have an affinity with Empiricism and science,[33] i.e. they believe that science is an epistemic route to knowledge par excellence. Yet empirical science, based on observations, delivers causal roles only, and, so, if Quidditism is the right view about properties, science can never convey the true nature of things/properties.

All these are the costs of quiddistic Categoricalism,[34] a view that wishes to avoid properties being modally loaded or charged with causal power. As we said earlier, such a view is not uncommon. In fact, it is one of the two big movements in contemporary metaphysics (to be more precise: in properties ontology) of which Dispositionalism is the other. Lewis's defense of Categoricalism has been central to contemporary metaphysics; it is a view that has been adopted by many philosophers and heavily criticised by others. Because of its importance, especially to modern considerations regarding dispositions, we will dedicate a section to the basics of Lewis's metaphysics and compare it to aspects of Logical Empiricism. The outline we will give of it now is sketchy but refinements will be offered in Sections 3.4 and 4.3.

2.2.5 Humean Supervenience

Today's metaphysical community is, roughly, divided into two camps: the Humeans, who adhere to Hume's Dictum, as introduced above, and the anti-Humeans', who have abandoned it together with Empiricism.[35]

David Lewis is a famous defender of Humeanism. He has, to honour Hume, the great denier of necessary connections, dubbed the specific dogma he believes in *Humean Supervenience*. He describes it in the following way:

> It is the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, *just one little thing and then another.* [. . .] For short: we

> have an arrangement of qualities. And that is all there is. There is no difference without a difference in the arrangement of qualities. All else supervenes on that.
>
> (Lewis 1986a: ix–x, emphasis added)

Compare Lewis's 'just one little thing and then another' with Hume's 'loose and separate' remark and you will immediately see why the epithet *Humean* is apt for Lewis's view:

> There appears not, throughout all nature, any one instance of connexion which is conceivable by us. All events seem entirely *loose and separate*. One event follows another; but we never can observe any tie between them. They seem conjoined, but never connected.
>
> (Hume 1748: Sect. VII, Part II, §58: 74, my italics)

But let us turn back to Lewis's own quote. With 'arrangement of qualities' he means a mosaic of categorical properties, of Quiddities, as introduced in Section 2.2.4.

The computer-screens metaphor We can imagine Lewis's world to be like our computer screens (Lewis 1986b: 14). Such screens consist of small pixels (on good screens they are, individually, invisible to the naked eye). They each display one of 16.7 million distinct colours and, as a whole, compose a picture (or a movie if we consider the development through time). Just which colour is shown at which specific pixel at a certain time is totally independent of which colours are shown at other pixels, i.e. each pixel can be autonomously programmed/coloured. Note that this latter fact is analogous to what we have said about categorical properties: they too can be rearranged freely. The instantiation of one categorical property at one place and time does not prescribe the categorical properties instantiated elsewhere.

Supervenience With the computer-screen pixel metaphor we can also indicate what is meant by *supervenience* in Humean Supervenience. The pictures we can see on the monitor *supervene* on the arrangement of colours located at the individual pixels, i.e. loosely speaking, the overall pictures depend on or arise from the colour distribution on the pixels. Whenever, on two different screens of the same product line, the colours on the individual pixels are distributed exactly alike then these screens will definitely show the very same picture. Phrased the other way round, no difference is possible in the overall picture if there are no differences in the underlying structure.[36]

The idea of supervenience can be expressed more formally in different and non-equivalent ways. A good starting point is this:

> A set of properties A supervenes upon another set B just in case no two things can differ with respect to A-properties without also differing with respect to their B-properties. In slogan form, 'there cannot be an A-difference without a B-difference'.
>
> (McLaughlin and Bennett 2014)

Needless to say, just as the overall picture depends on the pixels so do the individual objects depicted within the picture. Suppose we have, as our screensaver, Van Gogh's painting *Bedroom in Arles* (with its two chairs, the table, the bed, etc.) and we want the bed to appear elsewhere or be replaced by a futon then we ultimately have to re-colour the small atomic pixels. Just as the picture as a whole (the bedroom) is nothing over and above the distribution of colours on the smallest pixels, so are the appearances of objects within it (chairs, table, bed, etc.).

Let us – as a summary and a very rough outline of Lewis's metaphysics, i.e. of Humean Supervenience[37] – extract three features from the screen metaphor, which Lewis holds true, *mutatis mutandis,* about the universe, its objects and its properties:

(1) The world is a three-dimensional grid (comparable to the two-dimensional screen). A fourth dimension, additional to the first three spatial ones, is time. Instead of colours, fundamental properties are instantiated at the space-time points of the grid and thereby compose the mosaic that is the world.[38] The world (or parts of it) changes when the fundamental properties at specific points, at a specific time, change.

Think, as an approximation of what Lewis has in mind when he asks for the *fundamentality* of his properties, of the most advanced physical theories and the properties they postulate (like *being electrically charged* or *having gravitational mass*). Of course, physics might err and the world's real fundamental properties might be altogether different but, as we said, they are a good estimate of what is meant here. (We later have to add *perfect naturalness* to the attributes of Lewis's fundamental properties. This will become crucial but it can wait till Section 4.3).

(2) The fundamental properties from (1) are categorical properties, i.e. like the colours of the screen they can be freely (re-)arranged: that is, a property instantiated at a single point does not depend on (other) properties instantiated elsewhere. This is the *free recombination*

which *categorical properties*, but not dispositions, allow (Section 2.2.4). Remember, too, that the possibility of free recombination entails the Humeans' creed that there are no necessary connections in the world: nothing necessitates anything else – no property instantiation here enforces another there.

By the way, Lewis speaks of a mosaic of *local* matters of particular fact, i.e. localised at single space-time points (we will criticise locality in Section 7.8). This locality also implies that these properties are meant to be *intrinsic, non-relational*. Being the brother of my sister Julia is a relational property of mine (I have it only because something extrinsic to myself exists, namely my sister); having a mass of x kilograms, however, is my very own local intrinsic problem (it does not depend or relate to anything external).[39]

Now, there is, after all, one single type of relational, extrinsic property Lewis allows the points in his mosaic to have, namely the spatio-temporal relations between them. For example point $\langle x, y, z, t\rangle$ might have the extrinsic, relational property of being 1 cm below and 1 second after point $\langle x^*, y^*, z^*, t^*\rangle$. (All this is again directly analogous to the pixels on our metaphorical screen.)

(3) Everything else – macro objects and their macro properties (even living beings, persons, societies, etc.) and, importantly, also causality or the laws of nature, etc. – supervene on this arrangement, i.e. the spatio-temporal relations between those points with fundamental properties' instantiations. That is, the macroscopic objects of the world, their properties and causal or nomological features depend on the arrangement of fundamental properties of the three-dimensional (or, with time, four-dimensional) grid just as much as the screen-chairs, -tables, -beds, etc. arise from the individual colours of the pixels.[40]

Note that (2) comprises Hume's Dictum as mentioned above, i.e. the anti-modality, anti-power attitude that assumes there are no necessary or other modal connections ((MOD): Section 2.1.2) or productive forces ((PROD): Section 2.1.3) as fundamental features in the world. The combination of (1) and (3) we could label *supervenience*, so that Humean Supervenience turns out to be the combo of Hume's Dictum and supervenience.

Anti-Humeans and supervenience Note, now, that supervenience alone, i.e. (1) plus (3) but without (2), could also be upheld by Dispositionalists. In fact, many non-Humean philosophers believe that macro objects and their properties (a table's colour, surface structure, hardness, etc.) depend or supervene on or are reducible to more fundamental (micro) objects and their properties (the material the table is made of and its features). It is just that those fundamental properties are, by their lights, not (only) categorical but (also) dispositional (Section 2.2.4). That is, they deny one of the two central aspects of Humean Supervenience: they deny the no-necessary-connections verdict but affirm the belief in supervenience on more fundamental entities.

So, when Lewis writes:

> The point of defending Humean Supervenience is to resist philosophical arguments that there are more things in heaven and earth than physics has dreamt of.
>
> (Lewis 1999: 226)

then many non-Humean philosophers could join in. They would just deny that the full package of Humean Supervenience is the (or the only) means to that end.[41]

Further comparisons to Empiricism As we just said, Lewis's business to trace back every commonplace feature of the world to the mosaic of fundamental categorical properties parallels the Empiricists urge to rationally construct everything out of the flux of elementary sense data (Section 1.4.1) – and there is another respect in which Lewis's project is similar to that of the Logical Empiricists: their respective reduction bases – the mosaic of categorical properties and the flux of elementary sense data – share a core feature:

Just as sense data are said to appear and disappear as they please, without any perceivable necessitating force between them (the Logical Empiricists inherited this belief from Hume), so do the categorical properties which make up the four-dimensional mosaic that is the world without any necessitating link between their individual instances: no sense datum and, likewise, no categorical property here and now necessitates some other sense datum/categorical property there and later. That is, categorical properties are, in this respect, modelled after the Empiricists' sense data (or qualia, if you wish).

However, also note a grave difference between categorical properties and sense data: where we have no experiential access at all to the quiddistic primitive identity of a categorical property, sense data *just are* what is immediately given. Their nature and identity is precisely what is experienced, what we are directly acquainted with. (This leads – remember the Kantian considerations in Section 2.2.3 – to the consequence that Lewis's mosaic of categorical properties, realistically construed as the pattern that makes the world, is just as inscrutable as Kant's things in themselves.)

Ontology vs. semantics Here is a reminder of another difference: Logical Empiricism was occupied with the task of confirming an epistemic credo by means of a semantic dogma: all factual knowledge comes from sense experience and it can be shown that this is true by proving that the Verificationist theory of meaning is adequate. Humean Supervenience, on the other hand, is a metaphysical doctrine (think of the ontological shift!): all there is is a huge pattern of point-size, loose and separate property instantiations.

Post-Empiricist metaphysics Lewis is just one defender of a Humean metaphysics of science, an exemplar for the agenda of various anti-dispositional ontologies. The Categoricalisms of other philosophers, like, for example, David Armstrong, differ in significant ways (compare, for example, Armstrong's theory of laws of nature to Lewis's in Chapter 4). Yet they have in common the urge to show that dispositions are, really, only categorical properties plus, maybe, the laws. They often also share the conviction that the world as a whole, with its macroscopic objects and their properties, is, in the end, nothing but what goes on at a fundamental level.

The other post-Empiricist group of metaphysicians of science, the anti-Humeans, often share this latter belief in a fundamental level out of which everything else is built, but they insist that (some or most or all) fundamental properties are dispositional. We focus on this group's ideas in Chapter 6. Yet when we turn to the other issues within the metaphysics of science – counterfactual conditionals (Chapter 3), laws of nature (Chapter 4) and causation (Chapter 5) – we will always take both perspectives: the Humeans' and the anti-Humeans', where the first fight against necessities, causal powers and dispositions, and the second welcome these features. This frontier is characteristic of twentieth and twenty-first-century metaphysics of science.

BOX 2.2.4 and 2.2.5 Quiddities and Humean Supervenience

- **Categoricalists** share the belief that (all) **dispositions** have a **categorical underlying basis**, which is causally responsible for the possible dispositional behaviour. If so, dispositions might be **multiply realisable**. This is the case when they have more than one possible categorical basis.
- The debate about dispositions moved from the Empiricists' venture to analyse dispositional predicates in terms of observational language to a discussion of **whether dispositions supervene on categorical properties**.
- Categorical properties, or **Quiddities**, as they are often called, have the questionable feature of being **inscrutable**: like

Kantian things in themselves their intrinsic nature remains forever concealed from us.

- Yet philosophers who subscribe to **Hume's Dictum**, that there are **no necessary connections in nature**, still prefer them to dispositions because they, unlike dispositions, do not bring any such modal link into the world.
- David Lewis's **Humean Supervenience**, being in some (yet by no means all) respects heir to Empiricism, adheres to this dictum and is, moreover, built on the idea that **everything** – macro objects and their macro properties (even living beings, persons, societies, etc.) – **supervenes on a spatio-temporal arrangement of categorical fundamental properties**.

Notes

1 Synonyms for *disposition* are *capacity*, *tendency* and *power*. Philosophers make subtle distinctions between those terms occasionally. We will treat them alike unless we explicitly say so.

2 Recently, the requirement of a trigger condition has been contested by Barbara Vetter (Vetter 2015). However, we stick to the orthodox reading, where both trigger and manifestation belong to the disposition. We will, by the way, use *trigger*, *stimulation*, *stimulus conditions* and, likewise, *manifestation*, *reaction* and *effect* synonymously.

3 As we'll see in Section 6.4, monads are not such a bad comparison for dispositions despite many dissimilarities: both could be conceived of as little powerful agents.

4 I have taken the liberty of simplifying Carnap's original definition by deleting explicit references to time. I have also used modern logical symbols and easier-to-remember predicate letters: *D* for disposition, *T* for trigger and *R* for reaction.

5 For our purposes, even a very short introduction suffices. We recommend Graham Priest's (2001), especially chapters 2 and 7.

6 Some readers might wonder why there is no $\forall x$ in front of the whole formula [SCA] and, later, [RSA] and [EKA] to say that they hold for any object *x*. This is exactly right. Carnap (and later Kaila) omitted this first universal quantifier only to make their definitions more legible (Carnap 1936: 434).

7 Yet, see Clarke (2010), who argues for the possibility of things having opposing powers. Also, note that we could further define *being not-soluble* as $\neg(Tx \rightarrow Rx)$ and compare it to *being insoluble* $Tx \rightarrow \neg Rx$. This is a delicate matter once we exchange the material implication for a counterfactual conditional (Section 2.4).

8 For more on this interrelation see Section 3.3 on possible worlds semantics. By the way, the semantics of counterfactual conditionals was formulated by Robert Stalnaker (Stalnaker 1968) and independently by David Lewis (Lewis 1973).

9 I have taken most of the above examples of necessities from Lange (2009: 45).

10 Not all modal facts transcend the observable. For instance, if you see a red rose right in front of you you thereby see that it is *possible* for roses to be red (I owe this reminder to Barbara Vetter).

11 Yet see the psychological literature on *affordances*: (Gibson 1977).

12 [RSA] is my abbreviation; also, I have again used the letters *D*, *T*, *R*, exchanged the logical symbols for modern ones and omitted time.

13 Note something about the descriptive name *Random Coincidence*: even in such cases, there will clearly be *some cause* determining how the reaction happens. It is just that, seen from the perspective of the disposition under concern, the cause is not the right one – and this is the sense in which the attribute *random* has been used. If you wish to have a case which is *random* in the sense of happening undetermined, simply suppose that the radiation was emitted *by pure chance*

due to a *probabilistic quantum process* related to a coincidental atom decay when we dropped the match into the water.

14 Readers of Chapter 1 will of course be reminded of the section on provisos in (Section 1.4.2). Note, too, that we will encounter them again and again: in the case of dispositions in the guise of *antidotes* and *masks* (Section 3.5) and as *ceteris paribus conditions* in Section 6.2.

15 Note furthermore that even if we succeed in getting rid of the challenge of (RC) with these amendments we would, in the case of [SCA], still be faced with the problem of (VC) – maybe even more so because the more specific we make the testing conditions the harder it is for them to obtain. Thus, the instrument by which we avoid (RC) for [SCA] is a mechanism to raise the probability of (VC). For [RSA], which avoids the problem of (VC), the disadvantage of being only a partial definition is aggravated, for the dispositional predicate is now only defined for circumscribed cases.

16 See Burks (1951, 1955) and Sellars (1958).

17 Sure, wood rots if it is too long in water and eventually it disintegrates. We mean here a more or less instantaneous dissolving.

18 Here's one: assuming that properties are something over and above the set of things that have them might lead to the postulation of non-verifiable entities: what are these things, properties, which exist next to the objects? Do they live in Platonic heaven? (Or so Empiricist scruples might start…) Note that Nominalism is a view of properties already proposed by some scholastics. Famously, William of Ockham (1285–1347) defended it. Yet the question of the status of properties dates back at least to Plato's theory of forms, which contradicts Nominalism.

We will be occupied with the question of predicates and properties throughout the book, especially in the discussions of David Lewis's perfectly natural properties in (Section 4.3) and of natural kinds (Section 6.3).

19 Remember, his formula asks whether there is an *F* for *a* that fulfils three desiderata. (1) That the object under concern, *a*, is *F*. This is true because *a* belongs to the weird class for which *F* stands. (2) The part $\exists y(Fy \wedge Ty)$ demands that there is at least one thing *y* which is *F* and which has been tested. There is such a thing: *b*! It is, we have assumed, both in class *F* and it has been tested. (3) The final part $\forall z\ ((Fz \wedge Tz) \rightarrow Rz)$ asks that all objects that are *F* and that have been tested have responded positively with reaction *R*. Now, there are only two objects in *F*, namely *a* and *b*. Of these, only *b* was tested so we only have to check whether it responded positively. It did! This is something we assumed of *b*, the sugar cube, and, so, all three requirements for *a* to be *D* are met. *a*, the match, has to count as soluble yet again. The match *voidly* fulfils even Kaila's definition.

20 We leave these trickeries aside here but mention, for logic buffs, that this wicked general F is: $Fx =_{def.} (Tx \rightarrow Rx)$ (Wedberg 1944: 237). It has also again to be assumed that there is/was at least one *F* that had been positively tested.

21 An apology: this ontological turn deserves far more attention and happened much more gradually than outlined here.

22 Many philosophers, Armstrong himself but also, for example, Elisabeth Prior (Prior *et al*. 1982), think that all dispositions have categorical bases, i.e. when an object has a disposition there is some underlying categorical structure which is responsible for that disposition.

23 There have been many other attempts to spell out the distinction. Michael Esfeld, for example, links dispositions directly to causation and says that what they are consists crucially in what they might cause. This is a line of thought we will follow in Section 6.2. Categorical properties, on the contrary, are what they are independently of what they can causally do (Esfeld 2007: 140). For further variants of the distinction between dispositional and categorical properties see, for example, Prior (1982) and Mumford (1998: chapter 4).

24 There are very subtle distinctions between dispositions and theoretical terms, which need not concern us here. See Carnap's 'The Methodological Character of Theoretical Concepts' (1956a).

25 Charlie Martin (1997), John Heil (2005) and Galen Strawson (2008) have sympathised with Neutral Monism or dual-sided, or identity, views. Stephen Mumford has changed his opinion since 1998 (Mumford 1998: 116–17) and is now a Pan-Dispositionalist (2004; Mumford and Anjum 2011: 3). Other sympathisers with Pan-Dispositionalism (at least when it comes to natural properties) are Popper (1959), Harré and Madden (1975), Shoemaker (1979), Bird (2007), Chakravartty (2008) and Bostock (2008).

26 For a defence of Pan-Dispositionalism see, for example, Bird (2007: 3–4, 44ff.) and (Mumford and Anjum 2011: 3ff.). Mellor's ontological views are, we should mention, closer to Pan-Dispositionality than to Categoricalism, after all. David Lewis and David Armstrong are known as the arch-defenders of Categoricalism. See especially the foreword of Lewis (1986a: ix–xi) and chapters 3–5 of Armstrong (1997).

27 The difference is the move from language and observations (dispositional predicates and observational vocabulary for the Empiricists) to ontology (dispositions to categorical properties in the world for the post-Empiricist Categoricalists).

28 Mellor, a dispositional realist, later continues as wittily as he started:

> It is time to remove this lingering Victorian prejudice. Dispositions, like unmarried mothers, can manage on their own. [. . .] My strategy will be to show the offending features of dispositions to be either mythical or common to other properties of things; just as loose living is no prerogative of the unmarried and so is no proper basis for discriminating against them.
>
> (Mellor 1974: 157)

29 Note that we have, again (cf. Section 1.2), moved smoothly from conceivability to possibility. This is a contentious move. In Sections 6.3 and 6.4 we will see how some philosophers, the Dispositionalist Essentialists, doubt that it is possible for *NaCl* to obey different laws of nature (Bird 2001).

30 To learn more about *how individuals might have been* see Mackie (2009).

31 Assume, for the sake of the argument, that it would definitely do so and disregard the possibility of supersaturation or some other preventers.

32 This label is probably too strong. There are very many and deep differences between Kant and Lewis.

33 Lewis expresses his science-dedicated views thus: 'Most likely, if Humean supervenience is true at all, it is true in more or less the way that present physics would suggest [. . .] If physics itself were to teach me that it is false, I wouldn't grieve' (Lewis 1986a: xi).

34 For more see Bird (2007: 76–9).

35 This division overlaps a lot (if not entirely) with the Categoricalism–Dispositionalism dichotomy.

36 For discussions of supervenience and its most common variants see Kim (1993).

37 In Sections 2.4, 2.5, 3.3 and 4.3 we refine this outline. For an in-depth descriptions of Lewis's Humean Supervenience see, for example, Weatherson (2015).

38 Here's a difference between screen and world, colours and fundamental properties: Lewis allows more than one property to be instantiated per space-time point, while there can be only and precisely one colour per pixel on a screen.

39 For precise definitions of intrinsicality and extrinsicality see Lewis (1983), Lewis and Langton (1998) and (Hoffmann-Kolss 2010).

40 We restrict ourselves in this section to the supervenience of macro objects and properties on the mosaic. How causation, laws, the truth of counterfactual conditionals and, in fact, dispositions arise from the distribution of fundamental categorical properties is the issue of Chapters 3, 4 and 5.

41 As a preview of Chapters 3 to 5 note the following: when the anti-Humeans allow some modality to be included within the fundamental supervenience base they contradict the Hume's-dictum part of Humean Supervenience. The addition of anti-Humean connections to the supervenience base might come in different guises: one might add dispositional properties to the set of fundamental intrinsic properties (Chapter 6), laws might be said to be parts of the world's fundamental building blocks (see Armstrong: Section 4.4, and Maudlin: Section 4.6) or counterfactual facts might be said to belong to nature's basic equipment (see Lange 2009: chapter 4), etc. Whatever the anti-Humeans enhance the basis with, what remains is said to supervene on it. So, you might, as an anti-Humean, have categorical properties plus counterfactual facts as your fundamental elements and claim that the rest – macro properties as well as laws and causation, etc. – supervene on those (Lange 2009). Or you have categorical properties and laws as fundamental (Armstrong 1997) and have counterfactuals, causation, etc. supervene. Or. . . the possibilities are numerous, and we can find a defender for each.

Bibliography

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*. London: Routledge.
Armstrong, D. M. (1997) *A World of States of Affairs*. Cambridge: Cambridge University Press.
Bird, A. (2001) Necessarily, Salt Dissolves in Water. *Analysis* 61 (4). pp. 267–74.
Bird, A. (2007) *Nature's Metaphysics*. Oxford: Clarendon Press.
Black, R. (2000) Against Quidditism. *Australasian Journal of Philosophy*. 78 (1). pp. 87–104.
Blackburn, S. (1990) Hume and Thick Connexions. *Philosophy and Phenomenological Research*. 50, Supplement. pp. 237–50.
Bostock, S. (2008) In Defence of Pan-Dispositionalism. *Metaphysica*. 9 (2). pp. 139–57.
Burks, A. (1951) The Logic of Causal Propositions. *Mind*. 60 (239). pp. 363–82.
Burks, A. (1955) Dispositional Statements. *Philosophy of Science*. 22. pp. 175–93.
Carnap, R. (1928) *Der logische Aufbau der Welt*. Berlin: Weltkreis.
Carnap, R. (1936) Testability and Meaning I. *Philosophy of Science*. 3 (4). pp. 419–71.
Carnap, R. (1937) Testability and Meaning II. *Philosophy of Science*. 4. pp. 1–40.
Carnap, R. (1956a) The Methodological Character of Theoretical Concepts. *Minnesota Studies in the Philosophy of Science*. 1. pp. 38–76.
Chakravartty, A. (2008) Inessential Aristotle: Powers without Essences. In: Groff, R. (ed.) *Revitalizing Causality: Realism about Causality in Philosophy and Social Science*. New York: Routledge. pp. 152–62.
Clarke, R. (2010) Opposing Powers. *Philosophical Studies*. 49 (2). pp. 153–60.
Damschen, G., Schnepf, R. and Stüber, K. R. (eds.) (2009) *Debating Dispositions – Issues in Metaphysics, Epistemology and Philosophy of Mind*. Berlin/New York: De Gruyter.
Esfeld, M. (2007) *Naturphilosophie als Metaphysik der Natur*. Frankfurt: Suhrkamp.
Friebe, C. (2014) Categoricalism Versus Dispositionalism: A Case Study in Metametaphysics. *Journal of General Philosophy of Science*. 45 (1). pp. 5–15.
Gibson, J.J. (1977) The Theory of Affordances. In: *Perceiving, Acting, and Knowing*, Shaw, R. and Bramsford, J. (eds.). Hoboken, NJ: John Wiley and Sons Inc.
Goodman, N. (1955) *Fact, Fiction and Forecast*. 4th edition (1983). Cambridge, MA: Harvard University Press.
Handfield, T. (ed.) (2009) *Dispositions and Causes*. Oxford: Oxford University Press.
Harré, R. and Madden, E. H. (1975) *Causal Powers: A Theory of Natural Necessity*. Oxford: Blackwell.
Heil, J. (2005) Dispositions. *Synthese*. 144 (3). pp. 343–56.
Hobbes, T. (1655) De Corpore (Concerning Body). In: Molesworth, W. (ed.) (1839) *The English Works, Vol. I*. London: Scientia Aalen.
Hoffmann–Kolss, V. (2010) *The Metaphysics of Extrinsic Properties*. Frankfurt: Ontos.

Hume, D. (1739–40) *A Treatise of Human Nature*. Selby-Bigge, L. A. (ed.), revised by Nidditch, P. H. (1978). Oxford: Clarendon Press.
Hume, D. (1748) *Enquiries concerning Human Understanding and concerning the Principles of Morals*. 17th impression of the 3rd OUP edition. Selby-Bigge, L. A. (ed.). Oxford: Clarendon Press.
Hüttemann, A. (2009) Dispositions in Physics. In: Damschen, G., Schnepf, R. and Stüber, K. R. (eds.) *Debating Dispositions – Issues in Metaphysics, Epistemology and Philosophy of Mind*. Berlin/New York: DeGruyter.
Kaila, E. (1939) *Den mänskliga kunskapen*. Stockholm.
Kaila, E. (1942) Der physikalische Realitätsbegriff: Zweiter Beitrag zum logischen Empirismus. *Acta Philosophica Fennica*. 4. p. 191.
Kim, J. (1993) *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
Kistler, M. and Gnassounou, B. (eds.) (2007) *Dispositions and Causal Powers*. New York: Routledge.
Lange, M. (2009) *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford: Oxford University Press.
Langton, R. (1998) *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Clarendon Press.
Lewis, D. (1983). "Extrinsic Properties". Philosophical Studies, 44: 197–200.
Lewis, D. 1973 *Counterfactuals*. Oxford: Oxford University Press.
Lewis, D. (1986a) *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.
Lewis, D. (1986b) *On the Plurality of Worlds*. Oxford: Blackwell's.
Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
Lewis, D. (2008) Ramseyan Humility. In: Braddon-Mitchell, D. and Nola, R. (eds.) *Conceptual Analysis and Philosophical Naturalism*. Cambridge, MA: MIT Press. pp. 203–22.
Lewis, D. and Langton, R. (1998) Defining 'Intrinsic'. *Philosophy and Phenomenological Research*. 58 (2). pp. 333–45.
Mackie, P. (2009) *How Things Might Have Been: Individuals, Kinds, and Essential Properties*. Oxford: Oxford University Press.
McLaughlin, B. and Bennett, K. (2014) Supervenience. In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition (online). Available from: http://plato.stanford.edu/archives/spr2014/entries/supervenience/ (accessed 10 March 2015).
Malzkorn, W. (1999) On the Conditional Analysis of Dispositions. In: Meixner, U., Simons, P. (eds.) *Akten des 22. Internationalen Wittgenstein–Symposiums*. Kirchberg/W., 15–21 August 1999.
Marmodoro, A. (ed.) (2010) *The Metaphysics of Powers*. New York: Routledge.
Martin, C. B. (1997) On the Need for Properties. *Synthese*. 112 (2). pp. 193–231.
Maudlin, T. (2007) *The Metaphysics within Physics*. Oxford: Oxford University Press.
Mellor, D. H. (1974) In Defense of Dispositions. *The Philosophical Review*. 83 (2). pp. 157–81.

Mellor, D. H. (2000) The Semantics and Ontology of Dispositions. *Mind*. 109 (436). pp. 757–80.
Molnar, G. (2003) *Powers: A Study in Metaphysics*. Mumford, S. (ed.). Oxford: Oxford University Press.
Mumford, S. (1998) *Dispositions*. Oxford: Oxford University Press.
Mumford, S. (2004) *Laws in Nature*. London and New York: Routledge.
Mumford, S. and Anjum, R. L. (2011) *Getting Causes from Powers*. Oxford: Oxford University Press.
O'Shaughnessy, B. (1970) The Powerlessness of Dispositions. *Analysis* 31. pp. 1–15.
Popper, K. R. (1959) *The Logic of Scientific Discovery*. London: Hutchinson.
Priest, G. (2001) *Logic: A Very Short Introduction*. Oxford: Oxford University Press.
Prior, E. W. (1982) The Dispositional/Categorical Distinction. *Analysis*. 42 (2). pp. 93–6.
Prior, E. W., Pargetter, R. and Jackson, F. (1982) Three Theses about Dispositions. *American Philosophical Quarterly*. 19 (3). pp. 251–57.
Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson.
Sellars, W. (1958) Counterfactuals, Dispositions, and the Causal Modalities. In: Feigl, H., Scriven, M. and Maxwell, G. (eds.) *Minnesota Studies in the Philosophy of Science, Vol. II*. Minneapolis, MN: University of Minnesota Press. pp. 225–308.
Shoemaker, S. (1979) Identity, Properties and Causality. In: Shoemaker, S. (1984) *Identity, Cause and Mind*. Cambridge: Cambridge University Press. pp. 234–60.
Stalnaker, R. (1968) A Theory of Conditionals. *Studies in Logical Theory, American Philosophical Quarterly Monographs*. 2. pp. 98–112.
Storer, T. (1951) On Defining 'Soluble'. *Analysis*. 11 (6). pp. 134–37.
Strawson, G. (2008) The Identity of the Categorical and the Dispositional. *Analysis*. 68 (4). pp. 271–82.
Vetter, B. (2014) Dispositions without Conditionals. *Mind*. 123 (489). pp. 129–56.
Vetter, B. (2015) *Potentiality: From Dispositions to Modality*. Oxford: Oxford University Press.
Weatherson, B. (2015) Humean Supervenience. In: Loewer, B. and Schaffer, J. (eds.) *Blackwell Companion to David Lewis*. Chichester, West Sussex: Wiley-Blackwell, pp. 101–15.
Wedberg, A. (1942) Review of Kaila 1939. *Journal of Symbolic Logic*. 7. pp. 43–4.
Wedberg, A. (1944) The logical construction of the world. *Theoria*. 10 (3). pp. 216–46.
Wilson, J. (2010) What is Hume's Dictum, and Why Believe It? *Philosophy and Phenomenological Research*. 80 (3). pp. 595–637.

3 Counterfactuals

3.1 Motivation. Where we come from

In Section 2.1 the Logical Empiricists offered us two steps towards an analysis of disposition ascriptions. First, they considered that an example such as ‘*x* is soluble’ is equivalent to a simple ‘if. . . then . . .’ sentence: ‘if *x* is put in water then it dissolves’. Second, they said that this if-then clause can be interpreted as the material implication, with its specific truth conditions known from elementary logic: $Dx =_{def.} Tx \rightarrow Rx$.[1]

The latter step is problematic. This becomes most visible when we reconsider the so called *Void Satisfaction* (VC) problem: if we really interpreted the if-then in the above simple analysis of disposition ascriptions as a material implication we would have to attribute the disposition, here solubility, to each and every object that has never and will never be tested, be it made of sugar, wood or iron. These objects would *voidly fulfil* the definition. This is because the material implication is defined to be true as a whole (no matter what material *x* consists of) already when its antecedent ‘*x* is put in water’ is false, i.e. if it is not (now) the case that *x* is in water.

This is surely an untenable result. Yet rather than to drop the problematic second step – that is, rather than to conclude that the material implication with its truth conditions is not the right way to interpret the if-then claim inherent in dispositional ascriptions – the Logical Empiricists tried to temper the first step. That is, in order to circumvent the (VC) problem they tried to complicate and/or rephrase the simple *if trigger then reaction* (Sections 2.1.3 and 2.1.4.) However, they continued to interpret the $\rightarrow$ that still appeared in the midst of their more complex analyses as material implications. (In Section 2.1 we made plausible why this strategy seemed, back then, to be right; but we also saw why it was not entirely compelling.)

Enter: counterfactual conditionals Therefore we try, in this chapter, something else: we leave step one intact, i.e. we keep the simple *if trigger then reaction* structure (*if Tx then Rx*), but turn instead to the other possibility for improvement, namely to alter step two and give the if-then clause an interpretation and different truth conditions that deviate from the material implication. Only later we will see that we probably have to complicate the sentence as well. We mentioned the (VC) problem above. It is the other problem we discussed in Section 2.1, the Random Coincidence (RC), which will eventually force us to make both alterations.

One main desideratum for the new interpretation of if-then is, of course, that it gets right what the material implication got wrong, and this concerns specifically the cases where the antecedent of the if-then clause is, as a matter of fact, false.

That these cases in particular are important for dispositional ascriptions and deserve our focused attention is already obvious on an intuitive level, independently of the technical (VC) problem: when we say that something is disposed to do something we first and foremost have in mind what it would do in *circumstances that do not now obtain*, i.e. that are counter to actually existing facts. We might, for example, say: 'Take care, this vase is fragile', when the cause for its breaking – dropping – has not yet occurred but threatens to occur. In saying that it is fragile, we warn the carrier of the vase what would happen if those circumstance came about.[2]

Even the die-hard defender of extensional logic, Willard Van Orman Quine, agrees (here speaking of solubility):

> To say of a particular lump of stuff that it is soluble in water is [. . .] to say more about it than just that whenever it is in water it dissolves. [. . .] For a lump to be soluble we must be able to say of it that if it were in water it would dissolve; we need an 'if-then' connection governed by necessity, and by a necessity that goes beyond mere generality over time. To say merely that the lump dissolves at each time that it is in water is too weak.
> (Quine 1966: 71)[3]

The change of the grammatical mood of the antecedent and the consequent (here: from *is put* to *were put* and *dissolves* to *would dissolve*) is, of course, only a superficial move, merely indicating the next substantial change: to give truth conditions for these counterfactual conditionals.

Notation Let us, for ease of presentation, introduce a new symbol, an arrow preceded by a box, $\Box\!\!\rightarrow$, when we mean counterfactual conditionals together with their own to-be-introduced truth conditions. ($\Box\!\!\rightarrow$ has today become the standard notation for counterfactual conditionals.) This box-arrow will be flanked by an antecedent p and consequent q just as in the material conditional $\rightarrow$. So, $p \Box\!\!\rightarrow q$ reads: 'if it were the case that p then it would (also) be the case that q'.

As we know, the material implication, $p \rightarrow q$, is extensionally defined, i.e. its truth value is solely dependent on/already determined by the truth values of its component sentences. This is not the case for counterfactual conditionals: the truth (or falsity) of $p \square\!\!\rightarrow q$ has to depend on more than just the truth values of its antecedent, p, and consequent, q, if it depends on these truth values at all. When truth values (alone) are not decisive, i.e. when additional or totally different information is needed, we say that we are dealing with an *intensional* sentence connective.

Specifying the truth conditions for counterfactual conditionals and, specifically, what kind of additional information is needed next to or instead of the truths or falsities of the component sentences is the topic of this chapter.[4]

We will first (Sections 3.2–3.4) consider counterfactual conditionals independently of dispositions, for they occur in many other contexts as well: 'If kangaroos had no tails, they would topple over' is the one with which David Lewis (1941–2001) starts his book *Counterfactuals* (Lewis 1973a) – but once we have a good grasp on their truth conditions we turn back to an analysis of dispositions with the help of counterfactual conditionals (Section 3.5).

BOX 3.1: Conditionals within dispositions: material implication versus counterfactual conditionals

We dissected the analysis of dispositional predicates into **two steps**:

- First, we claimed that **an *if. . . then. . .* sentence is tacitly meant** when we attribute dispositions to things: this vase is fragile, i.e. if you drop it then it breaks.
- Second, we saw that the ***if. . . then. . .*** can be interpreted in different ways: as **material implication** or as **counterfactual conditional**, i.e. different truth conditions can be given for it.
- If you read it as **material implication** you do not do justice to the so analysed dispositional predicate (e.g. Void Satisfaction (VC) and Random Coincidence (RC): Sections 2.1.2 and 2.1.3). You could try to complicate the analysis while sticking to the material implication, but we saw that these attempts fail too.
- Thus, we make it our business to see whether an interpretation of the ***if. . . then. . .*** as **counterfactual conditional** succeeds better to capture what we mean by disposition attributions.
- Yet for this task to succeed we first need to give **truth conditions for counterfactual conditionals** (Sections 3.2–3.4) before we insert them into our analysis of dispositional predicates (this is our task in Section 3.5).

3.2 Post-Empiricist truth conditions for counterfactual conditionals

In the wake of the Logical Empiricists, Roderick Chisholm (1916–99) and Nelson Goodman (1906–98) were two of the first philosophers to give accounts of subjunctive or counterfactual conditionals. Both underline their importance for the analyses of dispositional predicates and how closely related these two subjects are, but their treatment of counterfactuals is independent of dispositions (Chisholm 1946: 291; Goodman 1947: 3; 1955: 40–9).[5]

The core idea Their basic idea is simple. A counterfactual conditional $p \,\Box\!\!\rightarrow q$ is true iff:

> there is a valid argument of the following form: it follows logically from $p, L_1, \ldots, L_n, C_1, \ldots, C_m$ that q,[6]

where p (the antecedent of the counterfactual) belongs to the set of premises and q (the counterfactual's consequent) is the conclusion of the argument. What are the additional $L_1, \ldots, L_n, C_1, \ldots, C_m$, though?

The *L*s are not unfamiliar entities (see Carnap and Kaila: sections 2.1.2 and 2.1.3, who also used them for their purposes): the *L*s are statements of *laws of nature*. Needless to say, if we make free use of laws we need a proper account of them, too, and laws will be our focus in Chapter 4. For now, however, we operate with an intuitive notion of laws[7] and spell out, in due course, what it is that makes them particularly appropriate to help define the truth conditions for counterfactual conditionals.

So what is the idea behind taking laws on board? This is best seen when we consider a counterfactual from the sciences: 'if this electron were to move into that magnetic field it would follow a curved trajectory'. This seems right *because it is a law of nature* that charged particles move in curves through magnetic fields. That is, the presence of laws in the set of premises in the above argument is really not so surprising. We can already grasp why it will often not be enough to quote just one law but many: in the electron example, in order to be able to deduce the consequent from the antecedent we need to add that it is a law that electrons are negatively charged. We then have the following valid argument:

Premise 1:	This electron moves into that magnetic field.	(p)
Premise 2:	All electrons are negatively charged.	(L_1)
Premise 3:	All negatively charged particles that move into magnetic fields follow curved trajectories.	(L_2)
Conclusion:	This electron follows a curved trajectory.	(q)

This makes plausible the $L_1, \ldots, L_n$ in those arguments which are meant to formulate the truth conditions for the counterfactual. Before we come to what the $C_1, \ldots, C_m$ are (and which might, occasionally, be unnecessary), let us remain with the laws for a while.

Clearly, the law statements in the premises *have to be true*. We cannot just take any made up 'law' and formulate a valid argument. That would enable us to make true whichever counterfactual we please. This would be wishful thinking.

Why laws? A slightly more difficult question is why, instead of laws, not any general truths (Chisholm 1946: 301–2)? Suppose it was true that all the coins in my pocket are Euro coins. Would the counterfactual 'if this coin (referring to one on the table) were in my pocket it would be a Euro coin' be true? There is, after all, a valid argument of the required form: 'This coin is now in my pocket. All coins in my pocket are Euro coins. Therefore, this coin is a Euro coin'. Yet it doesn't seem right that this valid argument, with a *merely true* but *not nomological* general statement, makes the counterfactual true, for if that coin is, as a matter of fact, a Pound coin it will not (miraculously) turn into a Euro coin. Rather, the true-so-far general sentence that all coins in my pocket are Euro coins would become false.

That is, in order to give truth conditions for counterfactual conditionals the general statements in the valid arguments have to be stably true and force everything that falls within their scope to comply with them. In other words, they have to be a law of nature. Consider, instead of the sentence about my pocket, the sentence 'if this coin were made of copper it would conduct electricity'. Now, this seems right, for 'All copper conducts electricity' is a stable law of nature and if we plug it into the respective valid argument we indeed get support for the truth of the counterfactual conditional.[8]

Background conditions We said that a counterfactual $p \Box\!\!\rightarrow q$ is true if and only if there is a valid argument of the form 'from p, $L_1, \ldots, L_n$, $C_1, \ldots, C_m$ follows logically that q', with the Ls being true laws of nature. What are the $C_1, \ldots, C_m$, which we have so far neglected?

Let us turn to an example from everyday life: 'Had the match been scratched, it would have lighted' (Goodman 1947: 115). Let us follow an example of Carnap's (Section 2.1.3) and assume that the match has just been destroyed (dissolved in acid, for example) so that the antecedent is forever false. Still we claim that, p, had it been scratched (a minute ago) then, q, it would have lighted: that is, if we take our truth conditions seriously, we claim that there is a valid argument from p, some laws L_i and . . . to q. Let us see whether there is such a valid argument.

We quickly realise two things. First, there probably isn't a law of nature that directly correlates match scratchings and lighting. Sciences are concerned with subtler underlying phenomena – for example, how friction (scratching) causes heat and heat causes some chemical (in the match's head) to burst into flames. But let that pass and pretend, for the sake of the argument, that there are macro laws relating matches to lighting. (This is innocent for we do have good grounds to believe that such macro behaviours are indeed governed by laws operating on the respective micro structures.)

Still, that macro law won't just say that all matches light when scratched. Rather, matches that are dry, well made, of wood and sulphur, surrounded by enough oxygen, etc. light when scratched. (Let us suppose that we have such a law with all the necessary conditions listed in its antecedent.) Now, what we claim when we claim that it is true that had the match been scratched, p, it would have lighted, q, is that there is a valid argument from p, the macro law (with the lengthy antecedent that the match is dry, the match is well made, the match is of wood and sulphur, there is enough oxygen, etc.), to q. These additional background conditions that are necessary for that argument to go through (because the law(s) demand for these conditions to be in place) are our $C_1, \ldots, C_m$!

Note that these circumstances have to actually obtain for the counterfactual claim to be true, for if you were to make that claim in empty space (where there's no oxygen) you would be wrong: the match would not light even when struck. In other words, someone who makes the counterfactual claim presupposes tacitly that the background conditions are in place.[9]

Let us, for the sake of brevity, lump all the laws together into the placeholder L, and all the background circumstances into the placeholder C. Then we have arrived at the following truth conditions (using symbol $\models$ for *logically implies*):[10]

$p \mathrel{\Box\!\!\rightarrow} q$ is true if there are true L and true C so that: $p, L, C \models q$.

Unfortunately, we are not yet done with this attempt. Actually, the real difficulties arise now and, as we will see, the answers are not entirely satisfactory.

Logical compatibility There have to be restrictions on what we allow into C because if we allow any actual truth to be eligible to figure in the background conditions we get unwanted results. Take, for example, $\neg p$. When we consider the counterfactual 'had p been the case. . .' we assume that the opposite, $\neg p$, is indeed the case. So, we might think that it cannot be harmful to put it into *C*, for it cannot be bad to add as many truths as possible, can it (Goodman 1947: 118)?

Unfortunately, it can, for if we were allowed to do this we could make any counterfactual true. Why is this? Remember, we want a valid argument from *p*, *L*, *C* to *q*. Yet if we put $\neg p$ into *C* we have *p* and $\neg p$ in our set of premises and from contradictions anything, any *q*, follows. Consequently, only those truths that are logically compatible with *p* may figure in *C*.[11]

Nomological compatibility Unfortunately, logical compatibility is not enough. We need two further kinds of compatibility: *nomological compatibility* and *cotenability*. Both are fairly complex issues and we can only roughly indicate what they are supposed to save us from.[12] First, here's a short version of nomological compatibility: there might be some facts *C** which, although they do not contradict *p* logically (as $\neg p$ would), speak *nomologically* against *p*. That is, it is not logic but some law of nature that says *C** and *p* cannot go together or, to put it another way, *C** and some law implies that $\neg p$. Thus, we had better keep *C** away from our antecedent or contradictions could threaten again. *C** and *p* are not nomologically compatible – hence, we disallow such nomologically incompatible *C*s.

Cotenability Logical compatibility together with nomological compatibility of all the *C*s with *p* is still not enough. We need yet another requirement for the *C*s. Worse, this last one threatens to make our entire analysis of counterfactuals circular. Before we throw in the towel, let us first see what cotenability is (Goodman 1947: 115–18).

Suppose we intuitively believe 'Had that match been scratched (*p*), it would have lighted (q): $p \mathrel{\Box\!\!\rightarrow} q$' to be true, i.e. we believe that the necessary supportive background conditions *C* are realised – the match was dry, well made, of wood, etc. – and also that there are true laws *L* supporting that kind of reaction – here, for example, that all dry, well made, etc. matches that are scratched do indeed light. In other words, we may assume that we have *L* and *C* so that from *p*, *L*, *C* follows *q*.

Unfortunately, if all this is the case, a second, unwelcome conditional turns out to be true too: 'Had that match been scratched (*p*), it would not have been dry ($\neg r$): $p \mathrel{\Box\!\!\rightarrow} \neg r$'. To see why this weird counterfactual is also true we would have to go through some knotty and tedious arguments. Let us spare ourselves the details except one (we come to this shortly) and focus on why we should want $p \mathrel{\Box\!\!\rightarrow} \neg r$ to be false: were it true then, together with our original, $p \mathrel{\Box\!\!\rightarrow} q$, we would have to believe both that had the match really been scratched it would light (*q*) and that it would be wet (i.e. not dry: $\neg r$). However, we can assume that this is physically impossible: wet matches do not light. We would end up in an unwanted nomological contradiction. Thus, we should see to it that $p \mathrel{\Box\!\!\rightarrow} \neg r$ does, after all, not come out true.

To grasp the strategy Goodman proposed we need to know one detail from the argument why, so far, $p \mathrel{\Box\!\!\rightarrow} \neg r$ comes out true. For that argument to work, $\neg q$ has to be allowed to be one part of C, the background conditions within the truth conditions for $p \mathrel{\Box\!\!\rightarrow} \neg r$. (This has not been forbidden yet: q itself does not even figure in that second conditional.) If we could disallow its presence in C everything would be fine. (We must ask the reader to trust us that $\neg q$ is the troublemaker.)

Here's a way to protest against using $\neg q$ when evaluating $p \mathrel{\Box\!\!\rightarrow} \neg r$: because we firmly believe in our first conditional, $p \mathrel{\Box\!\!\rightarrow} q$, we would want to say that $\neg q$ is in some way not compatible with p. Because of this first conditional we want to claim that $\neg q$ would not be the case if p were (rather, q would!). p and $\neg q$ are, in this sense, not *cotenable*, where *not being cotenable* is defined thus:

> Two sentences S_1 and S_2 are not *cotenable* iff S_2 *wouldn't be true* if S_1 *were*.

Let us, then, add cotenability with p to logical and nomological compatibility with p to our list of restrictions. The unwelcome counterfactual $p \mathrel{\Box\!\!\rightarrow} \neg r$ cannot come out true anymore because, to prove its truth, we would need to use $\neg q$. Yet $\neg q$ is not cotenable with p and thus forbidden for usage.

The failing ultimate definition Although the reader will surely have massive doubts already that this is a kosher way to modify our truth conditions (here's a hint: circularity!), we want to write down the overall definition we have now arrived at:

> $p \mathrel{\Box\!\!\rightarrow} q$ is true iff:
>
> there is a valid argument $p, L, C \models q$ with L being actual laws of nature and C those actually true background conditions that are logically and nomologically compatible and cotenable with p.

Now, here's the ultimate letdown. Our aim was to spell out circumstances under which we could say that a given counterfactual conditional is true. We hope, thereby, to illuminate the meaning of such counterfactuals. Now, however, it turns out that for this enterprise we need the notion of *cotenability*. Yet cotenability is defined via a counterfactual conditional: two sentences S_1 and S_2 are not cotenable iff S_2 *would* not be true if S_1 *were*. Thus, the truth conditions for counterfactuals are spelled out in terms of other counterfactuals. An unwelcome result. Goodman comments:

> We can never explain a counterfactual except in terms of others, so that the problem of counterfactuals must remain unsolved. Though unwilling to accept this conclusion, I do not at present see any way of meeting the difficulty.
>
> (Goodman 1947: 121)

There is, however, some consolation: even though we ended up in a definition circle we have learned a good deal about counterfactuals: their closeness to laws, for example, and how much their truth still depends on actual facts, even though they are about events that are not factual.

Note one further thing about the evaluation of counterfactuals here proposed. We have indeed given truth conditions for $p \mathrel{\Box\!\!\rightarrow} q$ which do not mention the actual truth values of p and q. That is, we have given a truly non-extensional, i.e. intensional, definition. This means, too, that no direct observation of the truth or falsity of p and/or q could establish the truth or falsity of $p \mathrel{\Box\!\!\rightarrow} q$. This is very different from the truth-functional material implication $p \rightarrow q$.

What we still would need to test, of course, is whether this interpretation of counterfactual conditionals could help with the analysis of dispositional predicates. If we put Goodman's and Chisholm's counterfactuals into the analysis is, for example, the (VC) problem solved? Random Coincidences (RC)? We will answer this question not for Goodman's and Chisholm's theory (and why should we, since we know it is circular?) but for Lewis's and Stalnaker's theory (to be introduced in the next but one section). Their theory has become the orthodox way to interpret counterfactual conditionals.

Before we turn to Lewis and Stalnaker we must learn some core features of so called *possible worlds semantics*. This will be helpful in its own right for it allows us to get a better grip on modalities like necessity and possibility (we touched on this issue in Section 2.1).

BOX 3.2: Goodman's and Chisholm's attempt

- When **Goodman** and **Chisholm** tried to spell out truth conditions for counterfactual conditionals they made use of an idea which Frank Ramsey attributed to John Stuart Mill: 'In general we can say with Mill that "If p then q" means that q is inferable from p, that is, of course, from p together with certain facts [C] and laws [L] not stated but in some way indicated by the context' (Ramsey 1931: 248, my additions in brackets).

- In more detail Goodman and Chisholm proposed that $p \mathbin{\Box\!\!\rightarrow} q$ **is true iff there is a valid argument from *p, L*, and *C* to *q*, where *L* is a set of laws (or one single law) and *C* is a set of conditions that in fact hold true** where and when the counterfactual is uttered (or held true at the time for which the counterfactual is meant to be true).
- The idea behind the *C* is that we hold **counterfactuals true under some but not other conditions**: if it is a quarter to twelve it might be right to say that had we caught that leaving bus we would just have been in time for the movies, yet if it is already five to twelve that might be wrong.

Certain additional conditions must hold, especially for the Cs:

- *C* as a whole and also any individual member of *C* have to be **logically and nomologically compatible with *p***, for otherwise anything would follow and any counterfactual be true.
- Finally, *C* can only contain statements that are **cotenable with *p***, i.e. be such that **had *p* occurred they would still be true**. Unfortunately, the necessity of this requirement brought our definition attempt to its knees, for cotenability is defined in terms of a counterfactual: **our truth conditions become circular** (or an infinite regress looms).
- Disregarding the failure of the attempt, we need to be aware of the following fact: even if we had succeeded with the above definition we would still have to **define what a law of nature is**, for we have made crucial use of this notion.

3.3 Interlude: *Possible Worlds Semantics*: the world is not enough

Our aim in this section is to introduce today's standard way to speak about modalities like necessity and possibility. We insert this excursus as it will lay the fundamentals for the theory of counterfactuals we discuss next (Section 3.4). In other words, understanding how counterfactuals work presupposes the possible worlds semantics for modality we unfold now. Our path will lead us through a multiverse of possible worlds. At first glance, this seems to be an enormous detour for our purposes, but, as will become clear fairly quickly, it isn't.

The best of all possible worlds In his *Théodicée* (Leibniz 1709), Leibniz tried to show how it is possible that a benevolent, omniscient and omnipotent god can allow evil to exist. Leibniz's line of argument is, roughly, this: being omniscient and omnipotent, god could create just about any world he likes, including good ones and evil ones. Yet, being also benevolent, he surely created the best of all these possible worlds. That there is some residue of evil in the one he finally created means only that no better world, with less pain and suffering, was possible. In other words, despite all the evil in the world, it still is the best of all possible worlds. One reason why god had to risk some evil is to endow human beings with free will.

We won't indulge theological disputes further, and we use future references to god only as a metaphor. What is important is to note that the idea of *possible worlds* was born: god could have created a huge variety of worlds and, in his mind, he might have considered them one by one. In a more profane vocabulary we could simply say that the world could have been otherwise and there are multiple different ways in which it could have been. Each such divergent way the world could have been we call a *possible world*. To make this idea even more palatable consider a bag full of Lego pieces. From these pieces we can build a whole variety of objects, and so could god build worlds from some fundamental elements. This is the general picture.

What these fundamental elements are that god could play with is controversial, and ideas change from philosopher to philosopher. Leibniz speaks of collections of *compossibles* or of individual substances. He also mentions that laws to which those elements conform belong to the basics of such a possible world. Regarding the laws within our actual world, Leibniz claims that 'God chose the most perfect order, that is, the order that is at once simplest in general rules and richest in phenomena' (Leibniz 1686: §6). (This perfect order is actually one feature which, according to Leibniz, makes ours the best of all possible worlds.)

Carnap's state descriptions It might come as a surprise to the reader of our historical introduction (Sections 1.1. and 1.4) that Carnap of all people was one of the first philosophers, after Wittgenstein (1889–1951) in his *Tractatus*, to make use of Leibnizian possible worlds. However, as we would expect, Carnap brings Leibniz's idea down to earth. First and foremost, he does not want to solve the theodicy problem but rather to illuminate the notions of *necessity* and *possibility*. Second, he rephrases the concept of a possible world (or, as he calls it, more modestly, 'state description') in terms of sets of true atomic sentences (Carnap 1956: 9). That is, instead of speaking about concrete fundamental elements of some sort ('Lego pieces') that are imagined to be rearranged in some way, Carnap asks us to consider sets of simple sentences written in some favourable vocabulary (for example, in the vocabulary of fundamental physics). Here, *simple* means more precisely *atomic*: the sentences shall not contain any sentence connectives like *and*, *or*, etc. and there should be no redundancies ('there's an electron at space time point <*x*, *y*, *z*, *t*>' would be a good candidate for one of many such sentences).[13] Here's Carnap's idea:

> A class of sentences in S_I [written in a certain favourable language with its basic vocabulary, MS] which contains for every atomic sentence either this sentence or its negation, but not both, and no other sentences, is called a *state-description* in S_I, because it obviously gives a complete description of a possible state of the universe of individuals with respect to all properties and relations expressed by predicates of the system. Thus the state-descriptions represent Leibniz' possible worlds or Wittgenstein's possible states of affairs.
>
> (Carnap 1956: 9)

With Wittgenstein, we can also capture the idea in terms of truth tables: write down, in the top row of a truth table with multiple columns, all the atomic sentences of your chosen language, one per box. We suppose that the top row is in some sense complete, i.e. that it contains all such atomic sentences. Anyway, suppose that all the other rows contain together all the possible different permutations of truth values there are for these atomic sentences. Then each row with its specific distribution of truth values, starting from the one with only Ts in its boxes and ending with the one that has only Fs, presents a possible world (Wittgenstein 1921: 4.21–4.431, especially 4.26 and 4.31).

Here's a miniature model of what's going on. Suppose the top row lists atomic sentences about attendance at a party: 'Alex is at the party', 'Siggi is at the party', and so on for other potential guests. Then the other rows of the table, distributing truth values for these sentences differently, each specify a different party. For example, the first row (below the top one): true for the first, true for the second, etc.; n^{th} row: false for the first, true for the second, etc. The first row specifies the party where everyone came, the n^{th}, maybe, one where everyone but Alex turned up, etc.

It could turn out that, in regard to some crowd of people, some logically possible distribution of truth values does still not appear in that table – maybe, Alex and Siggi promised each other never to go alone to parties. Then the row (i.e. party, i.e. miniature world) where only Alex (only Siggi) attended is not a possibility (if they stick to their promises). We come back to this potential restriction on the totality of logically possible distributions.

Here is a hint of what we are going to use it for: take the promises made by Siggi and Alex as metaphors for the laws of nature and consider a more realistic list of atomic sentences (descriptions of the subatomic world, for example). It might then be, due to the laws of nature, that some rows, although logically possible, do not appear or are written in thin grey rather than in bold black letters because the laws (promises) do not allow, say, electrons (Alex) to appear without unit negative charge (Siggi). We return shortly to these considerations.

Necessity, possibility, impossibility and contingency defined Leibniz's possible world talk is an excellent tool to explicate what we mean by our central modal terms, like necessity, possibility, impossibility and contingency, and Carnap is the first to make use of that tool (Carnap 1956: §41).[14] Suppose you have *all* possible worlds in front of you. (Imagine them either as different concrete objects or as sets of propositions or as rows in a truth table: whatever your favourite way is to think of them. We won't, here, discuss which way should be preferred.) Imagine further that you observe that in some of these possible worlds donkeys speak. Then it would be natural to say that it is possible that donkeys speak. Or watch out for worlds where there is an object that is both red and green all over. If you do not find such a world it seems to be correct to say that it is impossible for objects to be red and green all over – and we can treat necessity and contingency similarly. Here's a summary:

> A sentence or proposition *p is necessary* (or necessarily true) if and only if *p* is true in all possible worlds. We write: $\Box\, p$
> A sentence or proposition *p is possible* (or possibly true) if and only if *p* is true in at least one possible world. We write: $\Diamond\, p$
> A sentence or proposition *p is impossible* (or impossibly true) if and only if *p* is true in no possible world.
> *p is contingent* (or contingently true) if and only if *p* is in fact true, yet false in some other possible world.
> (No symbol has been established for impossibility or contingency because they are definable in terms of the widely used box, $\Box$, for necessity, and the diamond, $\Diamond$, for possibility.)

Some readers might wonder whether these definitions are not circular, for the word 'possible' sometimes appears in the *definiendum* as well as in the *definiens*. This is indeed unfortunate but merely due to a misnomer. Take 'possible world' to be 'a name for a certain collection of things – big, universe-size things' (Carroll and Markosian 2010: 10), or sets of sentences or rows in a truth table. 'We could have just called them the worlds, and left the word 'possible' out of the explanatory parts of the accounts altogether' (2010: 10). Thus, where it could cause confusion, we will just speak of 'worlds'.

Anyway, defined as above, the necessity operator, $\Box$, 'is very similar to the universal quantifier and "$\Diamond$" to the existential quantifier' (Carnap 1956: 186). It's just that they quantify not things but worlds. Indeed, we also get the same kind of inter-definability: $\Box p$ iff $\neg\Diamond\neg p$, and $\Diamond p$ iff $\neg\Box\neg p$. Translated into prose we see that this interdependence coheres very well with our intuitions: p is necessary should it not be possible that p does not obtain, and p is possible should it not be that case that non-p is necessary. (With these clues, the reader is invited to define contingency and impossibility on the basis of necessity and possibility.)[15]

Scope restrictions The possible worlds interpretation of modality (necessity, possibility – also counterfactuality, as we will soon see, etc.) is a very handy tool indeed. Next to the above inter-definability, which matches so well our intuitions, it allows us to model the different kinds of necessity (and possibility) we might have in mind (Section 2.1.2): logical necessity (the light is switched on or it is not switched on), conceptual or analytic necessity (all sisters are female), mathematical necessity (there is no largest prime number), metaphysical necessity (no object's surface is both red and green all over), nomological necessity (no massive object accelerates beyond the speed of light).

The key to interpreting these necessities within possible world talk is *restriction*. Here's how: above we said that necessity, by which we tacitly meant necessity *tout court*, operates like a universal quantifier: it is truth *in all worlds*. Now, if we wish to define different necessities of lesser strength, we can *restrict the scope* of the *quantifier* to fewer than absolutely all worlds. We could, for example, single out all and only those worlds that have the same laws of nature as our own world. We call these worlds the *nomological worlds*.

We have now the resources to define *nomological necessity*: we restrict this necessity's range to the nomological worlds. A statement p is then *nomologically necessary* if it is *true in all the law respecting worlds*. (A note on vocabulary: nomological necessity is also often called *natural* or *physical* necessity. We use those terms synonymously even though *physical* has an unsolicited Reductionist air to it, as if the laws of physics are real laws but not the ones of chemistry, etc.)

Here we have tacitly (and maybe presumptuously) presupposed that our own world is at the centre of the multiverse of all possible worlds and we have defined physical necessity in terms of the laws that obtain *in our world*. If you put a different world at the centre of your considerations, a world with deviating laws, then your nomological necessity will range over different possible worlds. In other words, there is a certain relativity regarding your starting point. Yet unless we make it explicit we will take our own world as point of departure.

How does nomological necessity compare to, for example, logical necessity? Intuitively we'd like to say that whatever is nomologically possible is also logically possible: it is nomologically possible that humans travel to Mars in spaceships and, so, it is logically possible. The converse is not true, some logical possibilities are not physical possibilities: that something moves faster than light, for example. In terms of possible worlds we get that *the physically possible worlds are a subset of the logically possible worlds*. There are, thus, fewer physically possible worlds than logically possible worlds. Going back to the truth tables from above: there are fewer physically possible rows than logically possible ones. The latter are all of the rows, the former just a subset thereof.

We have, as exemplars, picked out physical/nomological and logical possibility. What we said about their relation extends, *mutatis mutandis*, to the other necessities and possibilities: what is physically possible is also metaphysically possible; what is metaphysically possible is also conceptually possible, and the later is also logically possible. In terms of possible worlds this translates into the following picture: the physically possible worlds are a proper subset of the metaphysically possible worlds, which are a proper subset of the conceptually possible worlds, which are a proper subset of the logically possible worlds.[16]

What possibilities and/or necessities are there? The picture we have just drawn makes some heavyweight philosophical assumptions: most prominently, that there are all these necessities (and, translated, these subsets of possible worlds) and also that they are so ordered. Needless to say, not everyone shares these presuppositions: some will say that metaphysical necessity and possibility do not exist at all (so that there is no distinct realm of possible worlds that could be singled out here; Humean philosophers have an affinity with this view). Others will say that physical and metaphysical possibility are the same so that the physically possible worlds and metaphysically possible worlds form one and the same set; Saul Kripke flirts with this option: 'a good deal of what contemporary philosophy regards as mere physical necessity is actually necessity *tout court* [i.e. metaphysical necessity, MS]' (Kripke 1980: 164).[17]

There is also controversy about whether the logically possible worlds are a subset of a still wider set in which illogically possible worlds could also be found, or whether the logically possible worlds set the limit and exhaust all possible worlds (Lewis believes the latter (Lewis 1973a: 7)). Yet since there are different logics, with different axioms, it is not so absurd to speak of worlds that are illogical in respect of some but not another axiomatic system.

For now our goal has merely been to spell out broadly *in terms of possible worlds* what we mean when we say that something is necessary or possible. The philosophical intricacies we spare for the chapters to follow.

Especially, we leave aside here any confession as to what possible worlds are, ontologically speaking: they might be, fairly innocently, sets of propositions or sentences (see Carnap), imagined scenarios in ours or god's mind (for the latter, see Leibniz), concretely existing real entities in parallel universes (see Lewis) or yet something else.

For us, possible worlds semantics for necessity and possibility serves first and foremost as a manner of speaking, a handy tool for precise expression. That is, when we say things like 'There is a possible world in which the laws are different' we simply mean to say that it is in some sense possible that the laws could be different. We do not wish to imply any existential commitment as to whether such a world 'really' exists.

BOX 3.3: Possible worlds

- The idea of **possible worlds** goes back to **Leibniz. Wittgenstein** and **Carnap** have refined the idea and found it to be a handy tool **to formalise** what we mean when we use the **modal operators 'it is necessary that. . .' or 'it is possible that . . .'**
- A sentence or proposition p is **necessary** (or necessarily true), $\Box p$, if and only if p is **true in all possible worlds**.
- A sentence or proposition p is **possible** (or possibly true), $\Diamond p$, if and only if p is **true in at least one possible world**.
- Necessities (and thus possibilities) of **different strengths** can be introduced in that we **restrict the scope of 'all' to certain subsets of worlds**. For example, a proposition p is *physically* necessary if and only if p is true in all *physically* possible worlds.
- **Necessities come in a hierarchy**: logical necessity, being the strongest, is truth in absolutely all possible worlds, metaphysical necessity in slightly fewer, physical in even fewer, etc.

3.4 The present orthodoxy for counterfactuals

In the previous section we saw how we can reformulate the claim that a sentence (or a proposition, or the fact it expresses) is necessarily true in terms of possible worlds:

> p is necessarily true (we wrote in symbols: $\Box p$) *iff* p is true in all possible worlds.

We saw also that we can restrict the range of the necessity under concern: for *physical* necessity, for example, we could say instead that it is truth in all *physically* possible worlds.

Necessary material implications On the basis of these thoughts we might form the following idea to give truth conditions for counterfactual conditionals. Just say that 'if p were the case q would be the case' equals the claim that 'it is necessary that if p then q', i.e. that 'if p then q' is true in all possible worlds:

$$p \Box\!\!\rightarrow q \quad \textit{iff} \quad \Box(p \rightarrow q), \text{ that is}$$
$$\textit{iff} \quad (p \rightarrow q) \text{ is true in all possible worlds.}$$

In yet other words this would mean that all worlds where p is the case q is also the case or, in short, that all p-worlds are also q-worlds.

Prima facie, this looks like an excellent solution. With the necessity of the conditional link we would make the connection between p and q much tighter than it used to be when we expressed it by the pure material implication. In fact, we would make the link between p and q as strong as it can get in demanding that it holds in every possible world.

If we were to interpret $p \Box\!\!\rightarrow q$ as $\Box(p \rightarrow q)$ we would even solve the (VC) problem we encountered in Section 2.1, when we started analysing dispositional predicates. For suppose we were to say that 'x is soluble' means, necessarily, if x is put in water it dissolves. Then the fact that we do not actually put it in water, here in our world, does not cause the trouble it did for the naked material implication. Remember, the trouble with the latter implication was that it would be already true if its antecedent is false (the respective x is not actually put in water).

If, however, we are allowed to attribute 'soluble' to x if and only if *in all possible worlds where it is indeed put in water it dissolves* we would have to check these other worlds first. That is, the facts in our own world alone are not decisive. The missing test – it is not put into water *here* – is delegated to these other worlds and, depending on how x behaves *there*, it counts *here* as soluble or not. Let us suppose, for now, that x does dissolve in all those other worlds where it is indeed put in water. It then seems we are truly justified in saying, here and now, that it is soluble, i.e. that it would dissolve were we to put it in water. This looks like a positive result.

Unfortunately, it isn't. The strength of $\Box(p \rightarrow q)$ is not only a blessing. In fact, we shall see that it is over the top. Here's what we mean. When we utter counterfactual conditionals we want them to be flexible in a certain sense – more flexible than the strict necessary conditional.

Consider the following counterfactuals which we might want to hold true simultaneously:

> If Otto had come [p_1], it would have been a lively party [q]; but if both Otto and Anna had come [$p_1 \wedge p_2$], it would have been a dreary party [¬q]; but if Waldo had come as well [$p_1 \wedge p_2 \wedge p_3$], it would have been lively [q]; but. . .
> (Lewis 1973a: 10, my additions in brackets)

In other words, we get:

> $p_1 \Box\!\!\rightarrow q$ *and* $p_1 \wedge p_2 \Box\!\!\rightarrow \neg q$ *and* $p_1 \wedge p_2 \wedge p_3 \Box\!\!\rightarrow q$ *and* so on with alternating $q\,/\neg q$.[18]

This feature of counterfactuals cannot be realised by the strict necessary conditional. Take any two consecutive sentences from this cascade of statements, say: $\Box(p_1 \rightarrow q)$ and $\Box(p_1 \wedge p_2 \rightarrow \neg q)$. Suppose we hold them true. Then $(p_1 \rightarrow q)$ and $(p_1 \wedge p_2 \rightarrow \neg q)$ are true in all possible worlds. Yet from this we can conclude that, in any such world (for example, in ours), $p_1 \wedge p_2$ cannot be true, i.e. Otto and Anna definitely cannot come together. For if p_1 and p_2 were true then we could (per *modus ponens*) deduce q from $(p_1 \rightarrow q)$ and $\neg q$ from $(p_1 \wedge p_2 \rightarrow \neg q)$: contradiction! That is, holding true $\Box(p_1 \rightarrow q)$ and $\Box(p_1 \wedge p_2 \rightarrow \neg q)$ together commits us to also holding true that it cannot be the case that $p_1 \wedge p_2$, i.e. that it is impossible for Anna and Otto to come together to the party. This is surely not what we had in mind when uttering the cascade of party-counterfactuals (or you are willing to allow logically impossible worlds, an option we leave aside).

Let us sum up what we have got: from the (VC) problem we know that we need an interpretation of the counterfactual conditional that is stronger than the material implication; and from the argument given here we know that the strict necessary conditional would be too strong. Thus, counterfactuals, $(p \Box\!\!\rightarrow q)$ must be something in between the too loose $(p \rightarrow q)$ and the too strong $\Box(p \rightarrow q)$. In possible-words talk, we must look at more than just our world, yet looking at all of them (which necessity demands) would be too much.[19]

Fine-grained resemblance ordering David Lewis and Robert Stalnaker had, independently of each other, an ingenious idea how to get the balance right (Stalnaker 1968; Lewis 1973a). For each individual counterfactual conditional, roughly speaking, look at just the relevant worlds and leave out those that are too remote from ours. To understand this strategy we first need to sort or align possible worlds in a more finely grained manner than we did when we lumped them together into the vast sets of the logically possible worlds, the nomologically possible worlds, etc. (Section 3.3).

Let us, for illustration, remain within the set of physically possible worlds, i.e. those where our laws of nature hold. There are many different such worlds and they vary in their overall similarity. Take the world as it actually is. Now, take another one which is exactly like this one, yet where you have one hair less on your head. That world is very similar to ours indeed. Another world where everything remains the same but where you have *much* less hair is more different. Yet even that world is much more similar to ours than the one in which you weren't born (everything else being the same), and so on. You can easily imagine worlds being ordered in abstract space in respect of their similarity, where those that are more similar are located closer to ours.

Various respects of resemblance Sometimes different aspects under which worlds can resemble each other can stand in conflict or result in different orderings of worlds: just as you can sort all your books by aligning them to the left and right of your favourite novel according to page number or size or colour or content or. . . you can imagine worlds to be sorted either according to this or that criterion, and it would be surprising if all these sorting criteria yielded the same arrangement. Note that this does not make things totally relative or subjective: once the criteria are fixed – and the counterfactual often suggests such criteria – the ordering of worlds is an objective matter.

Just which criterion for resemblance is intended or important depends on which question you are interested in and in which context you ask it. If someone says that if Caesar had been in command in North Korea he would have used the atom bomb they seem to be thinking of possible worlds in which Caesar, with his character traits, is in command in twentieth-century Korea. If, however, they utter instead that if Caesar had been in command in Korea he would have used catapults, what they have in mind is holding the technology of Caesar's time fixed and looking at worlds in which Caesar's Korean battlefield resembles that of ancient Rome. So, for the evaluation of the first counterfactual we should consider those worlds closer to ours where Korea's armament resembles today's weaponry. For the second counterfactual those worlds are closer to ours where Korea's weaponry resembles ancient Rome's technology (see Lewis 1973a: 15, but also Chisholm 1946: 303–4).

Let us leave the difficulties of context dependence and different-world orderings aside and suppose that it is perfectly clear for any specific utterance of a counterfactual what the resemblance criteria are. Consider, for example, Lewis's kangaroo counterfactual conditional: 'If kangaroos had no tails, they would topple over'. Let us look at a couple of sets of possible worlds. There are, for example, those which resemble ours in almost every respect except that kangaroos are on average slightly smaller but otherwise pretty much the same (including having tails). A little more dissimilar are those worlds where kangaroos have no tails (and, thus, also leave no tail traces in the sand, etc.). Yet more dissimilar are those where kangaroos have no tails but use crutches to walk around, etc.

Now, when we say that if kangaroos had no tails they would topple over we certainly do not mean that in every possible world where kangaroos have no tails they topple over. When we utter that counterfactual we are simply not interested in worlds where kangaroos use crutches. That is, if someone were to reply, ‘But what if they had no tails and used crutches, then they would not topple over, would they?’, we would answer, ‘Oh come on, who cares about such fantasy worlds? I meant they’d topple over if everything else is as much as it is’.

The formal definition Slightly more abstractly put, this leads to the following recipe: you decide whether a counterfactual $p \Box\!\!\rightarrow q$ is true in that you start looking one by one at possible worlds in the order of their similarity (in our geometric metaphor: in order of closeness to the actual world). Some very close ones will be so similar that kangaroos have tails there, too. They are not so interesting. Yet sooner or later you visit possible worlds where kangaroos have no tails though these worlds still resemble ours closely in many other respects. What happens to these kangaroos there? That is, what happens where kangaroos have no tails, everything else being pretty much equal? Do they topple over in these worlds? If *yes* then the counterfactual is correctly uttered in our world, i.e. if in worlds that are most similar to the actual in all other respects but in which p is true q is also true.

One thing is for sure: if we were to move further out we would find worlds where p is true but q is not. These are the remote worlds, fairly dissimilar to ours, where kangaroos do not have tails but use crutches (in order not to topple). Do these worlds matter? As we argued above, they don’t. It is enough for a counterfactual $p \Box\!\!\rightarrow q$ to be true that *in the closest worlds where p is true, q is true*, no matter what happens in less close, i.e. very dissimilar, worlds.

Here’s a formal expression of this recipe (the first entry of which will have to be clarified afterwards):

> At a world @[20] the counterfactual conditional $p \Box\!\!\rightarrow q$ is
> (1) *vacuously true* if there are no possible worlds where p is true;
> (2 *non-vacuously true* if, amongst the worlds where p is true, some worlds where q is true are closer to @ than any world where q is not true;[21] or
> (3) *false otherwise*.
>
> (Lewis 1973a: 16)

The first entry becomes immediately clear when we remember that the only worlds we have excluded are logically impossible worlds where, say, sentences s and $\neg s$ are true together. Thus, there being no worlds where p is true (as in (1)) means that p is internally contradictory. So, for example, 'if the light were switched on and off at the same time I could see in the dark' is vacuously true because it is logically impossible for the light to be both on and off. We can motivate this move, i.e. the move to say that these counterfactuals are true, by saying: if you're willing to go that far (i.e. to be illogical) then everything can be true.[22]

The second and third entries express exactly what we said more informally above: there are many p-worlds where, unlike our own world, the antecedent, p, of the counterfactual under consideration is true. If, amongst those worlds, those where q is also the case are closer to ours than those where it is not the case, then the counterfactual is true.[23]

Test cases We started this section with two puzzles. For conditionals associated with dispositions we saw that the material implication is too weak (remember the (VC) and the (RC) (Sections 2.1.2, 2.1.3). We need a stronger conditional that does not only consider actual truths or falsities of the antecedent and consequent. A necessary material implication would deliver this strength, for it would consider truth values throughout every possible world. It would, in fact, solve the (VC) issue. Yet we saw that, as a general interpretation of counterfactuals, it would be too strong: remember the puzzle with the cascading party-counterfactuals.

Our task is now twofold: we need to check whether the Lewis/Stalnaker interpretation of counterfactuals can cope with such cascading conditionals and we need to see whether Lewis/Stalnaker counterfactuals are appropriate for an analysis of dispositions and especially whether they solve the (VC) and (RC) problem. We will do the latter in Section 3.5, after ending the present section with the party we looked at above and, thereafter, a brief comparison of Goodman/Chisholm and Lewis/Stalnaker.

Cascading counterfactuals So, let us focus on a very small region of our world, namely the place where the party took place. Let us agree that the rest of the world remains exactly as it is. That is, in what follows, we look successively at worlds more and more dissimilar to ours, but only in respect of the party. In some such party worlds the host offered vodka, in others gin, in some she played drum and bass in others dubstep, etc., etc. Be that as it may, suppose that at the party, as it actually happened, neither Otto nor Anna nor Waldo, etc. came. Consider now two sets of other possible worlds: one set, call it O (for Otto), to which Otto but neither Anna nor Waldo came. There are many such worlds: for example, some where Otto drinks a lot, others where he drinks little, some where he is very talkative, others where he is silent, etc. Consider a second further cluster of worlds, call it *O&A*, to which both Otto and Anna (but not Waldo) came. Again, there are many such worlds with varying progressions of the party. Of course, all the *O&A* worlds are also *O* worlds (but not vice versa).

This will be important for evaluating the first counterfactual in our cascade: 'if Otto had come, it would have been a lively party'. We need to consider whether, amongst the *O* worlds that are closest to ours, the party is lively. Note something important here: for our context and purposes, those *O* worlds where Anna is not also present (i.e. those *O* worlds that are not also *O&A* worlds) are closer to the actual world than those where she is also present because the latter are more dissimilar to our present party: you have to make two major changes to arrive at them, i.e. you need to have Otto *and* Anna present. For the mere *O* worlds, however, the only change you have to make is to let Otto in. Back to the evaluation of the first counterfactual: if in these *O-only* worlds the party is lively then the counterfactual is true, if not then not.

This no longer contradicts the second counterfactual: 'if Otto and Anna had come, it would have been a dreary party'. Here, we need to go to the first worlds where Anna is also present. If amongst those worlds the ones which are closest to ours have miserable parties the second counterfactual is true (and for Otto-Anna-Waldo worlds. . . you can play this game as often as you like). Thus, holding many such counterfactual conditionals with the consequence's truth and falsity alternating does not commit us to untenable contradictions. The key for this solution is the variable strength of counterfactuals. Every counterfactual brings its own (restricted) sphere of worlds with it: counterfactuals do not need the consequent to be true in *all* possible worlds where the antecedent is. Those closest, most similar to ours will suffice.

Comparisons to Goodman and Chisholm Remember Goodman and Chisholm's theory of counterfactuals (Section 3.2). There, set *C* of actually obtaining background conditions and true laws of nature, *L*, were relevant for the truth or falsity of the counterfactual. Where have the *L*s and *C*s gone in Lewis/Stalnaker? Into the similarity ordering of worlds, of course! We evaluate counterfactuals in that we look at those that are similar to ours, i.e. at those in which enough circumstances *C* equal those of the actual world or situation.

What about the laws of nature, though? Weren't they a crucial justification within Goodman and Chisholm's theory for when and whether counterfactuals are true? Moreover, laws are even important when we intuitively (as non-philosophers) evaluate counterfactuals: 'Sure, if you were to jump out of the window you'd fall! It's the law of gravitation, stupid.' Like the circumstances *C*, the laws *L* have not disappeared from Lewis's account. The key is again the similarity of worlds: worlds with different laws are dissimilar to ours. When we considered the party worlds we did not consider a party where, because of the lack of gravitation, people were floating around but only worlds where the law of gravitation holds.[24]

Similarity measures Lewis actually gives a list of 'weights or priorities' for similarity measurements of worlds, of which the first three are:

> (1) It is of the first importance to avoid big, widespread, miracles, i.e. diverse violations of law.
> (2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.[25]
> (3) It is of the third importance to avoid even small, localized, simple violations of law [i.e. miracles].
>
> (Lewis 1979, quoted in Lewis 1986a: 47–8)[26]

The second entry mirrors what we said about *C*: where Goodman needs to add statements to the premises set of his derivations that record particular obtaining facts, Lewis needs to focus on worlds in which these facts obtain just as in our actual world. The first entry confirms what we just said about laws: worlds can only be considered as similar when they do not massively diverge in their laws of nature.

However, in Lewis's formulation some violations of laws seem to be allowed. The third entry liberalises nomic concordance in that it demands only *importance* but not *necessity* for the avoidance of small, localised, simple violations of law. Why is this?

Miracles: violations of laws Well, what does it take, quite generally, to conceive of a possible world where, as opposed to our world, some event has not happened, yet where almost everything else is the same as in our world, i.e. where we have the least overall departure from actuality?

Suppose you play roulette. The ball lands on a black number but you placed a fortune on red. You say: 'Had the ball landed on red, I'd be rich'. That's true, we want to say intuitively. Yet in order to evaluate the correctness of that counterfactual statement within Lewis's framework we need to see what happens in worlds that are closest to ours but in which the ball in fact landed on red. Now, which are the closest worlds, i.e. those most similar to ours, in which the ball did land on red but where everything or at least most other things are equal?

If you do not allow miracles, i.e. small violations of laws, then the past history of such a world must have been different from ours: the croupier must have thrown the ball slightly harder or softer or the friction of the wheel must have been different. Yet for the latter to be true the table must have had a different past: it must have been used more or less or have been produced slightly differently, etc., etc. The further back you go the more differences to the actual world will accumulate. Thus, in total, that world and our world will be fairly different after all: to alter that past a little means to alter its past a little and so on. In other words, altering the immediate past (without miracles) means altering the whole past and thereby we make the (past) histories of our two comparative worlds very different.

This is not so if we allow for a tiny miracle: consider a world which shares our past exactly. How can the ball, there, have landed on red? Well, let a miracle occur where the ball jumps just that millimetre further and then let the world take its normal course again, just as law-governed as ours is. Aren't the two worlds until after the miraculous divergence very similar despite the miracle? Especially when compared to the world where the ball falls on red because the past is so very different? In the currency of similarity, so the intuition might go, tiny miracles seem a cheaper option than altering the whole past.

Here's a further thought about tiny miracles: when, along the history of the possible world, should we surgically intervene with such a miracle (to make the ball fall on red)? Well,

> it will normally be best not to diverge at all from the actual course of events until just before the time of e [here: the ball falling on red]. The longer we wait, the more we prolong the spatiotemporal region of perfect match between our actual world and the selected alternative.
>
> (Lewis 1973b: 171)

Note also that the possibility to alter laws without thereby making worlds so dissimilar that they cease to be of interest allows us the possibility to utter and evaluate counterfactuals that consider violated laws in their antecedent: 'if the gravitational force between two massive bodies were to decrease not with the square but the cube of the distance then. . .'. Such counterfactuals are central to science and metaphysics of science.

These were some of the motivations and intuitions for Lewis's proposed ordering of similarity aspects. Lewis had other reasons, which have to do with so called backtracking counterfactuals and time's arrow (Lewis 1979), which we come to in Section 5.3.

Possible worlds: useful fictions? Here's a warning: the advantage of possible worlds talk is deceptive and can lead us into the temptation to engage in it without reflection. We then talk as if possible worlds were planets that could be visited and where we could make observations and experiments that cannot or have not been done on earth. When, after experimenting in these foreign lands, we find that such and such is the case, the counterfactual, uttered here on earth, is true: 'because – lo! – in that other world it actually happened as it says!'.

Now, even though a few philosophers do indeed believe that possible worlds exist just as our own does, the above is clearly not what happens.[27] For even if these worlds really existed (which most philosophers do not believe) the only epistemic access we would have to them is not via space travel or observation through telescopes but via thinking of actual facts and actual laws and then measuring the odds whether, were the antecedent really true, that would lead to the consequent. In other words, Goodman's and Chisholm's thoughts on counterfactuals (Section 3.2), with explicit reference to laws, are by no means obsolete once we have possible worlds talk at hand.

Possible worlds are a handy semantic tool – a helpful quick way of speaking – but not a *carte blanche* to disregard actual facts, actual laws and what we could derive from them if we take different premises on board. If possible worlds semantics gives correct truth conditions then the way *we know* whether these conditions are fulfilled is still via recombining facts in imaginative experiments where we consider what events the laws permit (Lewis 1986b: §2.4: 113).

In a way, both accounts of counterfactual conditionals in this and the previous section resemble the idea of a computer simulation to help find out whether they are true. It is just that Goodman and Chisholm give you the program code for that simulation whereas Lewis and Stalnaker actually let you play the game.

BOX 3.4: Lewis's and Stalnaker's account

- Within so called '**possible worlds semantics**' a counterfactual conditional $p \mathbin{\Box\!\!\rightarrow} q$ is true iff amongst the worlds where p is true, some worlds where q is also true are closer to the actual world than those where q is not true. In other words, some $p \wedge q$ worlds are closer to our world than any $p \wedge \neg q$ worlds.
- This formulation allows counterfactuals to have **variable strictness**: they need not range over all possible worlds as a necessity operator does when prefixed to the material implication.
- The metaphor of closeness of worlds demands **similarity measures**. Worlds are closer to one another, i.e. they are more similar to each other, when they share as many facts and as many laws as possible. Sometimes facts need to vary (p, for example, has to be true at a world to be considered) and sometimes laws can be violated (see miracles).

- There are **accounts of (counterfactual) conditionals** other than those discussed in Sections 3.2 and 3.4, some of which say that these conditionals lack truth conditions and have instead assertability conditions only. There are yet others that turn to conditional probabilities to analyse conditionals: 'if *p* were the case then *q* would be' is saying that the probability of *q* is high(er), given *p*.

LITERATURE

- A prominent overview can be found in Dorothy Edgington's articles 'On Conditionals' (1995) where, amongst others, Goodman's (§3) and Lewis's (§4) accounts are critically evaluated. Jonathan Bennett (2003) has *A Philosophical Guide to Conditionals* and Frank Jackson (1991) has edited a collection of seminal papers on *Conditionals*.

3.5 Counterfactuals and dispositions

Counterfactual or subjunctive conditionals are important in all areas of metaphysics of science. Yet our own reason for considering them in more detail was to find an adequate analysis of dispositions or dispositional predicates. The suggestions we encountered in Section 2.1 (with a material implication at heart) were troubled by the problem of (VC) (and (RC), to which we come later). Now, if we use, say, Lewis's analysis of counterfactual conditionals in order to interpret the if-then statement that is inherent in dispositional ascriptions would we then solve that problem? Yes we would!

(VC) Consider an old acquaintance: Carnap's match that has never been put to the test and never will be and thus would count as soluble in water under the primitive analysis of solubility (with the material implication as interpretation of the if-then). Here's a new counterfactual analysis of solubility:

> Something *x* is soluble *iff* if *x* were to be put in water it would dissolve. That is, the closest, most similar worlds to ours in which *x* is indeed put in water are worlds in which *x* dissolves. There might be remoter, more dissimilar worlds in which it doesn't but they are of no immediate concern.

Now, taking this new analysis for granted, would we have to attribute solubility to Carnap's match? Luckily, we would not, and this is because we have good reasons to believe that the counterfactual is false because worlds which are similar to ours in all respects except (more or less) that the match is there put in water are worlds in which it doesn't dissolve. Why? Because those worlds share not only the vast majority of actual circumstances but also our laws of nature and there are no laws that have wooden material dissolve in ordinary H_2O. In fact, the laws will assure wood's integrity in water (at least for quite some time, until it rots). Thus, the match does not count as soluble and the (VC) problem is out of the way. Compare with Kaila's analysis (Section 2.1), where laws are explicitly mentioned and, so to speak, visible on the surface.[28]

New (RC). ***Finks*** **and** ***reverse finks*** Counterfactual conditionals can solve the (VC) problem. Can they also handle (RC), the second class of difficulties the Logical Empiricist were plagued with, where trigger and reaction of a disposition co-occur, not because of the disposition but by accident, randomly? After Stalnaker (1968) and Lewis (1973a) had published their theories of counterfactual conditionals and maybe after Chisholm's (1946) and Goodman's (1947) approaches, people seemed to have tacitly assumed that they can.[29] Lewis, for example, writes:

> All of us used to think, and many of us still think, that statements about how a thing is disposed to respond to stimuli can be analysed straightforwardly in terms of counterfactual conditionals.
>
> (Lewis 1997: 143)

This folklore was shaken when, in 1994, Charles Martin (1924–2008) revived the dispositions debate with his seminal text 'Dispositions and Conditionals' (Martin 1994).[30] In fact, after his publication new papers on the conditional analysis of dispositions mushroomed.

In his paper Martin shows convincingly that the bare counterfactual conditional fails:

> x has the disposition $D \leftrightarrow$ if x *were exposed* to the test T, x *would show* the reaction R.

His counterexample is of a live wire (*being live* is the disposition in question)[31] to which a machine – Martin calls it an *electro-fink* – is connected. This machine is imagined to be built in such a way that it stops the power supply immediately if something touches the wire (the machine (fink) is more or less just a simple fuse).

Consider now the following straightforward conditional analysis of the disposition *being live*:

> x is live *iff* if x were touched by a conductor, then electric current would flow from x to the conductor.

This turns out to be an inadequate analysis. Why? The wire is live *ex hypothesi*, i.e. it does have the disposition yet the conditional is not true because of the fink. Go to the closest worlds to ours where everything is more or less like here – there's the wire, it is live, there's the electro-fink, there are the laws of nature – yet where there is this one difference: the wire *is indeed touched* there. What happens in those worlds? Well, the moment the wire is touched (by a conductor) the wire is live no longer due to the electro-fink and thus no current flows to the conductor. In our possible worlds talk this reads: the closest *p*-worlds are non-*q* worlds and, thus, the counterfactual conditional is false. Therefore, it can't be a proper analysis of dispositions because the disposition ascription is true (as we assume) while the counterfactual comes out wrong: something can have the disposition without the conditional being true of it.

Worse is to come. We can rephrase the story *mutatis mutandis* such that a reverse-electro-fink is operating on a non-live wire and thereby show that something can fail to have the disposition despite the respective conditional being true of it (see Martin 1994: 3; the reader is invited to spell out Martin's claim in terms of possible worlds semantics).

(RC) and finks Note the following similarity, especially of this reverse-fink, to the (RC) we know from Section 2.1.3: here, the wire would show the appropriate reaction on being tested although it was not so disposed; there, a match did surprisingly dissolve in water because some radiation, strong enough to make wood disintegrate, was also present. What happens in both cases is that an *indisposed* object fulfils the (therefore obviously inadequate) *conditions for dispositionality* and it does so because it shows the reaction on being triggered for some but not the right reasons.[32] Relate this back to the intuitive feature (PROD) of dispositions: *production*. If a fink is operative to bring about or produce the manifestation rather than a true disposition of the object then the object can hardly be called disposed.

Just as with the simple *material conditional* analysis of dispositions, the question now arises of how we can enhance the simple *counterfactual conditional* analysis so that we avoid being forced to attribute dispositionality to indisposed objects and vice versa.

Proviso clauses Our immediate reply to Martin is likely to be this: the peculiar intervention of a fink is not what normally happens and what we mean to say is something like this: 'The wire is live iff if it were touched by a conductor *under normal circumstances* (to which the presence of weird finks does not belong), then electrical current would flow from the wire to the conductor'. In other words, we would want to add a *proviso clause* to the conditional. We encountered these clauses, which are often also called *all else equal*, or *ceteris paribus,* clauses, or a *saving clause*

(Martin 1994: 5), a couple of times (as early as in our historical chapter on Empiricism (Section 1.4.2)), and they will reappear in Sections 6.2 (on *ceteris paribus* laws) and 6.4 (on dispositional essentialism).

Martin anticipates this reaction (Martin 1994: 5). Yet he insists that simply adding such proviso clauses generates a further problem, namely to say precisely what we mean by 'normal conditions', or, to put it the other way round, which circumstances or interferences (like those of finks) we wish to exclude.

There seem to be only two ways to exclude the unwanted cases. (1) We itemise each and every possible one ('If the wire is touched and there is no fink, no fuse, there is no one cutting the wire nor does the power station break down, etc.'). The problem with this solution is that we will most probably have to leave a gap in the antecedent, indicated by the *etc.*: there might well be an endless number of possible interferers and we might not know them all.

Or, (2), we try to find a general feature shared by all unwanted interferers. Then we could exclude them in one sweep: things or events that share this feature, i.e. that are similar in this respect, should be absent. But similarity in which respect? Martin answers tentatively 'that each of these events brings it about that it is not the case that the wire is live at a certain time' (Martin 1994: 6). However, he is aware of the fact that this solution leads to an intolerable result, for, if we use this suggestion to enhance the original definition of the dispositional predicate, we end up in a circle:

> If the wire is touched by a conductor, and nothing happens to make it false that the wire is live, and yet other things are equal, then electrical current flows from the wire to the conductor.
>
> (Martin 1994: 6)

In abstract terms we have arrived at the following impasse. Either we get a gappy definition:

> x has the disposition D iff if x were exposed to the test T in the absence of interfering factors F_1, F_2, F_3 . . . x would show the reaction R.

Or we get a circular one:

> x has the disposition D iff in conditions where no event brings it about that it is not the case that x has the disposition D: if x were exposed to the test T, x would show the reaction R.

Martin's conclusion regarding this negative result is radical and pessimistic. We can neither extensionally list all the cases to be excluded (they are too various and numerous) nor can we define a common criterion of those cases without ending up in a definition circle. Hence, according to Martin, counterfactual conditionals are not the appropriate tools to define dispositional predicates. For him, dispositional predicates are irreducible:

> The central argument of this paper has been against any attempt to account for dispositions that makes essential use of strong conditionals and counterfactuals. [Those] counterfactuals have a place, of course, only as clumsy and inexact linguistic gestures to dispositions and they should be kept in that place.
>
> (Martin 1994: 7–8)

Martin's paper in *Philosophical Quarterly* was the starting point for a fresh debate about dispositions and their potential analysis. The prevention of finks (and other counteracting devices) was the main concern in this discussion for over a decade. The goal has either been to enhance the conditional analysis with additional clauses to make it work or to show that this is not possible and, so, that dispositions are irreducible real properties. We will sketch two such further approaches: David Lewis's, which argues for the conditional analysis, and Alexander Bird's, which argues against Lewis's attempt.[33] We will not go into much depth here. Yet we will extract from the discussion what relates to the story as we have told it so far.

Lewis's analysis In 'Finkish Dispositions' (Lewis 1997) Lewis did not want to follow Martin's bold step, i.e. the step to accept dispositions as real, irreducible properties in their own right. This comes as no surprise if we remember Lewis's affinity with Hume and, thus, his metaphysical anti-dispositional attitude, which we discussed at length in Section 2.2.5 and which manifests in his credo of Humean Supervenience. In short, Lewis's aim is to show that, ultimately, all dispositional properties arise from categorical properties (Sections 2.2.3, 2.2.4) plus laws and/or causation, i.e. their existence is wholly dependent on these other entities (Lewis 1997).

Lewis offers a sophisticated reformulation of the primitive counterfactual version. His reformulation manages to circumvent reference to normal or ideal or *ceteris paribus* conditions, which looked so troublesome to Martin. Instead of such proviso clauses, Lewis crucially adds that each object that is disposed has an intrinsic (categorical) property *B*, which is causally responsible for the dispositional behaviour of the object. In the case of solubility, for example, this property *B* may be a molecular structure or an ionic compound (in the case of salt: $Na^{+}Cl^{-}$) with which water dipoles can causally interact.

In ‘Finkish Dispositions’ Lewis remains neutral on the character of the basis property *B* (Lewis 1997: 144). This is certainly wise when it comes to ordinary dispositional properties like fragility or solubility because the underlying structures might themselves be of a dispositional nature. However, ultimately, i.e. at the most fundamental level of the world, Lewis would want the final bases to be categorical (Section 2.2.5).[34]

Finkish base Having revived the idea to incorporate reference to a base property into the analysis of dispositions, Lewis can indicate his starting point for a solution to Martin’s fink cases: ‘A finkish disposition is a disposition with a finkish base’ (Lewis 1997: 149), i.e. an object with a finkish disposition is an object that loses the base property of the disposition once the trigger for the disposition is pulled. Again, in other words, the diagnosis of what exactly happens in fink cases is that the trigger for the disposition is also a cause for the basis of the disposition to disappear. In fact, it disappears faster than it could react appropriately with the expected manifestation. It is lost before it does what it is supposed to do.

Here, then, is Lewis’s definition:

> Something x is disposed at time t to give response r to stimulus s, iff, for some intrinsic property B that x has at t, for some time t’ after t, if x were to undergo stimulus s at time t and retain property B until t’, [then] s and x’s having of B would jointly be an x-complete cause of x’s giving response r.
>
> (Lewis 1997: 157, my addition in brackets)

Let us explain this ‘unlovely mouthful’, as Lewis himself called his analysis (Lewis 1997: 157), by first showing why the definition is immune to finks. The crucial phrase is: ‘retain property *B* from *t* to *t*’’.[35]

To evaluate whether something is disposed, i.e. whether the according counterfactual is true, we have to consult the closest possible worlds to ours where the whole antecedent – here, ‘if *x* were to undergo stimulus *s* at time *t* and retain property *B* until *t*’’ – is true. If, in these worlds, the consequent is also true then the counterfactual as a whole is true and thus the disposition ascription. Now, in order to look for closest worlds in which this antecedent is true we have to make sure both that the trigger has occurred there and that *B* is retained for long enough to cause the reaction (this is what the time interval *t to t’* demands). Now, the latter means basically that no fink operates (successfully) to destroy the causally active basis. If in such finkless worlds, the response comes about (response *r* is caused) our object may justifiably be called disposed; if not then not. All seems well.

There are two further things to be clarified: why does Lewis add that *B* should be intrinsic? What does he mean by ‘x-complete *cause*’ and why does *causation* appear here at all?

Intrinsicality Lewis asks for the property *B* to be an intrinsic[36] property, i.e. one that the object has just by itself. This is because he has the intuition that we attribute dispositions only to things that are themselves responsible for the dispositional behaviour. If, for example, Willie is a pacifist but his older brother is the town's infamous rowdy it might be true of Willie that if you were to provoke him you would be beaten up. Still, according to Lewis, this is not because of a disposition of Willie's but of his brother's (Lewis 1997: 147). Willie has no intrinsic disposition for violence. Thus, since in Lewis's definition the disposition depends so crucially on *B*, *B* had better be intrinsic so that *D* is too.[37]

Causation Together *B* and the stimulus cause the reaction. If everything is in its right place then *the stimulus, s,* and *the object x's having the base property B* 'would jointly be an *x*-complete cause of *x*'s giving response *r*'. Why does causation have to enter the picture?[38] Why not simply have 'then *x* would give response *r*' as the consequent of the counterfactual? Well, what would be the use of *x* having some intrinsic property *B* if *B* were not crucial causally involved in making the reaction come about?

Here is another, more concrete reason to demand that an intrinsic *B* is causally active: if we were not to ask for this then our enemy, the (RC), could strike again. Let *x* be our old acquaintance, the insoluble match.[39] It has some intrinsic properties – for example, it is made of wood. Now, let it be put in water and the random strong radiation occur again so that it does dissolve. Nearly all the desiderata of Lewis's definition for solubility would be met: *x* has some intrinsic property *B* and it retains it. *x* does undergo stimulus *s* and *x* does give response *r*. Luckily, the match, *x*, would still not count as soluble: it is not any of the intrinsic *B*s that, together with the stimulus, is *a cause* of the reaction. The actual cause for the match dissolving is the extrinsic radiation. Thus, because of Lewis's demand that it is an *intrinsic basis* which *causes* the reaction, it is clear that, here, the match dissolves for the wrong reasons (or, better, wrong causes). Thus, we need not attribute solubility to it.

Responsibility for production Here's a final, important thing to notice. When we encountered the (RC) for the first time we extracted from the discussion a feature we intuitively think dispositions have: a responsibility for production, a kind of efficacy, some causal power, i.e. feature (PROD) in Section 2.1.4. This is precisely what we can extract again from the considerations about finks and what, indeed, Lewis encodes in demanding *B* (together with *s*) to be the cause of *r*.[40]

Bird's antidotes Lewis seems to have saved the counterfactual conditional analysis of dispositional predicates. The (VC), and thus (MOD), is taken care of by the mere fact that counterfactual conditionals are employed, and the (RC), here especially in the guise of finks and reverse-finks, is ineffective too because Lewis's definition also demands causal responsibility of the disposition's basis for the reaction (cf. (PROD)). In other words, the reaction must not merely coincidentally occur with the trigger, it must have been caused by the basis.

Unfortunately, there is, next to finks, a whole zoo of little machines, interferers and unwanted processes – namely *antidotes*, *prodotes*, *masking* and *mimicking*, to name but a few – which still make life difficult for those who aim to provide a counterfactual analysis of dispositions. As we will indicate now, not even Lewis's promising analysis is immune to them. We will briefly focus on antidotes as introduced by Alexander Bird in his 'Dispositions and Antidotes' (Bird 1998).[41]

Bird's derivative of finks, namely *antidotes*, do not destroy the intrinsic basis, *B*, of a disposed object (and thus Lewis's demand that *B* be retained till *t*' is still met). Yet antidotes interfere with the further causal process extrinsic to the object. Here is one of Bird's examples (Bird 1998: 229). A uranium pile has the following disposition: when it gets beyond critical mass it chain-reacts catastrophically. Luckily, in nuclear power stations there is a safety mechanism which lets boron-moderating rods absorb the fatal surplus radiation in case radioactivity increases dangerously. These boron rods thus prevent a chain reaction. Although the intrinsic structure, *B*, of the uranium pile is not altered, its present disposition will not manifest itself. Hence, Lewis's demand for retaining the intrinsic basis is futile: interferences that happen extrinsic to the disposed object can still prevent the reaction from coming about, although the trigger is there and the object is definitely disposed. Thus, for the truly disposed uranium pile, Lewis's analysis would wrongly say that it is not disposed.

One can react to Bird's counterexamples in two general ways: one could either argue against their success or, being convinced of their defeating power, suggest another revision of the counterfactual analysis.[42] We do not have the space to discuss either way but rather wish to come to concluding remarks on counterfactuals and the counterfactual analysis of dispositions.[43]

The big picture Carnap and the other Logical Empiricists tried to define dispositional predicates in terms of observation vocabulary in order, first, to make disposition ascriptions verifiable, in line with the Verificationist theory of meaning. Second, they sought thereby to get rid of the potential danger to attribute (occult) powers or other unwanted modal connections to the world. Being Hume disciples, this was a major desideratum for them. Lewis and other Post-Empiricist Humeans gave up Verificationism but they continued to dismiss powers from their ontology. As we have seen in Sections 2.2.4 and 2.2.5, the goal thereby shifted from analysing dispositional predicates in terms of *observational vocabulary* to showing that dispositions can be reduced to *categorical properties*.

Lewis and his followers' analyses are more sophisticated in that they allow counterfactual conditionals instead of material implications. This chapter's purpose has mainly been to say what a counterfactual conditional is and especially to introduce truth-conditions of counterfactual conditionals. However, we have also turned to the attempts to define dispositions in terms of counterfactual conditionals.

We have seen reformulations and drawbacks. The way in which we presented this back and forth might have given the impression that philosophers played with the definition attempts for dispositions for the sheer joy of a good puzzle. Some might well do, but it is important to keep in mind that there has been an agenda behind the efforts and the attacks: as we said, those who are keen on an analysis of dispositions in non-dispositional terms are likely to be Humean-minded philosophers (like Lewis) and those who give counterexamples are most probably anti-Humeans (like Martin). The counterfactual analysis of dispositions has been one of the major battlegrounds for Humeans against anti-Humeans.[44]

Whether or not some counterfactual analysis has succeeded or some future one will succeed we cannot answer here. If we apply a pessimistic induction we might be tempted to say that, since so far all analyses have had counterexamples and thus failed, the next ones will too. If this is true it seems that anti-Humean Dispositionalists will win the battle.

Yet even if a watertight characterisation can be given Dispositionalists might have a trump card up their sleeve. Here is how John Heil expresses this possibility:

> Even if you could concoct a conditional analysis of dispositionality impervious to counter-examples, it is not clear what you would have accomplished. You would still be faced with the question, What are the truth makers for dispositional claims? Suppose you decide that 'object *o* is fragile' implies and is implied by '*o* would shatter if struck in circumstances *C*'. You are not excused from the task of saying what the truth maker might be for this conditional. Presumably, if the conditional is an analysis, its truth maker will be whatever the truth maker is for the original dispositional assertion. This is *progress*?
>
> (Heil 2005: 345)

In other words, it is not so clear which side of the alleged definitions, the left or the right of the *iff*, is ontologically prior to the other. If this is right, then the proxy war Humeans and anti-Humeans have fought on the battleground of the conditional analysis of dispositions was futile. Other arguments would have to be given for the ontological priority of either dispositions or non-dispositional properties. We have given some already (Section 2.2.4) and we will look for some again later (Section 7.8).

In any case, many philosophers have – also in the light of the failures of the counterfactual analysis – accepted dispositions as being primitive, non-reducible, real features of the world. Rather then analysing them away, their aim is to analyse everything else: laws (Chapter 4), causation (Chapter 5), counterfactuals (this chapter), etc. with the help of dispositional powers. We see how this is supposed to work in Chapter 6, on Dispositional Essentialism.[45]

BOX 3.5: The counterfactual analysis of dispositions

- Carnap expressed his hope that once proper **truth conditions for counterfactual conditionals** are available a **conditional explication of dispositions** can be given (Carnap 1963: 951). However, even if we believe that puzzles about the logic of counterfactuals are uncontroversially solved, a straightforward analysis of dispositions is still not easily had.
- **Charles Martin** showed with his **fink and reverse fink cases** (which are, in fact, derivatives of our good old (RC)) that at least a **simple counterfactual conditional analysis** – for example, x has the disposition D iff if x were in stimulus circumstances s it would react with reaction r – is not successful. A fink, remember, is a device that reacts to the very same trigger as the disposition. Yet, viciously, the fink's reaction is to destroy the disposition before it can show its reaction. Thus, although an object has that (finkish) disposition, it would not show the reaction were it triggered.
- Reactions to Martin come from, for example, **David Lewis**. Lewis's idea is, in short, to say this: x is disposed to show reaction r on stimulus s iff if s came about and ***x*'s intrinsic character were unchanged** then the stimulus and x's intrinsic setup would cause the reaction r. In other words, no fink will mess with the object's intrinsic underlying basis for the disposition.
- Analogously to Martin's finkish assault on the simple analysis, **Alexander Bird** attacks Lewis's definition with **antidotes**. Antidotes are interferers that leave x's intrinsic character intact but **mess with the causal process towards reaction *r* extrinsic to the object *x***.

- Whether Lewis's, Bird's or any conditional analyses of dispositions with the help of counterfactuals succeeds remains a matter of debate. In fact, the publication output on the counterfactual analysis is not unlike the Gettier industry within epistemology.

Recent Publications on the Analysis of Dispositions include:

- Barbara Vetter (2014) 'Dispositions without Conditionals'
- Troy Cross (2012) 'Recent Work on Dispositions'
- David Manley and Ryan Wasserman (2008) 'On Linking Dispositions and Conditionals'
- Choi, Sungho (2008) 'Dispositional Properties and Counterfactual Conditionals'
- Michael Fara (2005) 'Dispositions and Habituals'

Notes

1 For the semantics/logic of the material implication consult an introductory logic book – for example, Priest (2001).

2 We said above that we focus on *counterfactual* conditionals which are uttered when their antecedent is (supposed to be) false. Note that there is also a type of natural language conditional, more neutrally called *subjunctive conditional*, where the presupposition of the falsity of the antecedent is not made. Consider the two following conditionals: 'Were there tea in this cup then there would be H_2O in the cup' and 'Had there been tea in this cup then there would have been H_2O in the cup'. Someone who utters the first (*merely* subjunctive) seems to be neutral on the issue ('I don't know whether there is tea or not, but were there tea…'), while someone who utters the second (counterfactual) seems to presuppose the falsity of the antecedent ('Had there been tea… but I believe there isn't').

Be that as it may, following (Chisholm 1946: 290), (Lewis 1973a: 3) and many other philosophers, we will assume that this difference is one of the pragmatics of language use only: we tend to utter the first if we presuppose nothing about the truth of the antecedent, and the second if we believe it to be false. Yet, the truth conditions, i.e. the circumstances under which the two utterances are true, are the same. In other words, we treat counterfactuals and subjunctive conditionals alike and use the term 'counterfactuals' to mean both.

3 Of course, Quine sees this kind of necessity ultimately analysed in terms of 'Hume's regularities' (Quine 1966: 76).

4 Some philosophers believe that no matter how much information we add to the truth or falsity of the two sentences a counterfactual conditional is composed of, we cannot determinate a truth value for the conditional as a whole when the antecedent is false. Rather than to claim that it is true (or false) we should say that, under certain circumstances, we are *justified to believe* in a counterfactual or that we *can be more or less certain* of it (rather than to say it is true). Such theories often operate with degrees of belief and probabilities: the *if* is interpreted as *given that* or *on the supposition that* and asserting 'if *p* were the case *q* would be the case' is interpreted as saying something along the lines of 'When I assume that *p* is given then my degree of belief that *q* is the case is (much) higher than my belief that not-*q*'. See Adams (1975) and consider Edgington (1995) for a very helpful (if deliberately opinionated) overview of all kinds of theories for conditionals.

5 We will by and large follow Goodman's ideas but note that Chisholm goes through very similar steps. With Goodman and Chisholm we limit our inquiry to 'those [conditionals] which are not analytic or logically true' (Chisholm 1946: 294).

6 We actually have to add: 'and there's no such kind of argument with which $\neg q$ could be deduced', but we leave this aside for now.

7 Both Goodman and Chisholm acknowledge the problem of laws. While Goodman favours an anti-Realist account of what laws are (see the last chapters of Goodman 1955), Chisholm flirts at one point with Charles Sanders Peirce's (1839–1914) idea that nomicity is 'an irreducible ontological category' (Chisholm 1946: 306). This is an idea which we will later find in Dretske, Tooley, Armstrong (Section 4.4) and, differently, in Maudlin (Section 4.6).

8 Note that some philosophers hold that laws of nature themselves have counterfactual form. Goodman must (and does) deny this for otherwise his definition of counterfactuals is circular. Laws on his account are no more than a special, more stable kind of regularity.

9 Note that this is not to say that someone who has made a counterfactual claim made an elliptical claim and that they rather meant 'had p and C_1 and … and C_m been the case, then q'. That would be a much weaker claim, for then $C_1, \ldots, C_m$ need not be true (as p need not be true). Yet the original claim presupposes implicitly the truth of $C_1, \ldots, C_m$. $p \mathrel{\Box\!\!\rightarrow} q$ can only be true if $C_1, \ldots, C_m$ are.

10 This end product might remind some readers of the so called *Ramsey Test* for counterfactuals, which has been an inspiration not only for Goodman and Chisholm but for anyone working on counterfactual conditionals:

> If two people are arguing 'If p, then q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge [including some other facts and knowledge about laws of nature, MS] and arguing on that basis about q; so that in a sense 'If p, q' and 'If p, $\neg q$' are contradictories. We can say that they are fixing their degree of belief in q given p.
>
> (Ramsey 1990: 154–5)

It should be noted, however, that Ramsey's aim was not to give truth conditions for 'If p, then q?' (subjunctively construed) but only something like acceptability conditions.

Also, readers who have knowledge of Hempel's theory of explanation will find similarities between $p, L, C \models q$ and the DN model.

11 A quick note on q, i.e. our consequent and conclusion: we have assumed that q is de facto false. Since we only allow truths in C we could not have the false q in it (besides, this would make it all too easy to deduce q from it). We should also not take $\neg q$ into C for then we could only deduce q if the premises were internally contradictory. In other words, C shall be neutral regarding q or $\neg q$. Neutral here means logically compatible with both.

12 Goodman (1947) discusses in more depth the issue of nomological compatibility and further, less important blemishes of the definition.

13 Less technically, think of a novel. Although hopefully not written in atomic sentences it also contains a kind of (state) description, i.e. a description of a possible world or, at least, a small section thereof (possible worlds are meant to describe *everything* – novels restrict themselves on what is essential for the plot).

14 In Leibniz's works we seem to find only hints of the possibility to define modalities in terms of possible worlds.

15 For historical completeness, we should add that it was Saul Kripke in the late 1950s and early 60s who developed the formalities further (see, for example, Kripke 1959, 1963). Readers who are interested in the formal aspects of modal logic should, however, first consult the excellent textbook by Hughes and Cresswell (1996).

This is almost all we want to say here about formal logic. We just want to add briefly that some but not all modal logical systems contain the following axioms $\Box p \rightarrow p$ (what is necessary is also the case), $\Box p \rightarrow \Box\Box p$ (what is necessary is necessarily necessary, or (roughly) what is necessary in our world, is necessary in every world), $\Diamond p \rightarrow \Box\Diamond p$ (what is possible is necessarily possible, or (roughly) what is possible in our world, is possible in every world).

16 In the section on Marc Lange's theory (Section 4.5) we recapitulate this hierarchy of necessities from a different angle.
17 Having spoken of the multitude of commonly recognised necessities, it becomes clear that we can actually sort and arrange possible worlds in any respect we like. We might want to speak of legal necessity and means to capture all possible worlds (or a small fragment thereof: some particular country) that have the same legal system as the actual one. More abstractly speaking, we can take any respect R in which worlds can look alike, then restrict our attention only to those possible worlds that R-resemble each other and define R-necessity as truth in all R-resembling worlds: 'Necessity of a certain sort is truth at all possible worlds that satisfy a certain restriction' (Lewis 1973a: 4).
18 We could also do this with Goodman's match: if match *m* had been scratched it would have lighted; but if match *m* had been wet and scratched it would not have lighted; but if match *m* had been wet and scratched and the surrounding temperature had been extremely high it would have lighted...
19 You might wonder, just out of curiosity, whether the problem of (RC) would be solved with a necessary conditional even if we have rejected this solution already. This might be so because a mere coincidence of *p* and *q* here, in our world, does not make ($p \rightarrow q$) true in all possible worlds. Yet as the (RC) is closely related to feature (PROD) of dispositions, namely to causal productivity (Section 2.1), we need to ask further whether causal productivity is also captured by a necessary conditional. Here, the answer is not so straightforward. We come back to this in Section 6.4.3.
20 The '@' sign is often used to indicate that we are talking about our actual world. This is also intended here. Yet, note that the definition can take any world at its centre.
21 For Stalnaker, entry (2) is different: (2') non-vacuously true if at *the* world closest to @ where *p* is true, *q* is also true (Stalnaker 1968). Yet we need not go into the details of the consequences of this difference here.
22 We must not conceal that we actually do sometimes engage into counterfactual reasoning with illogical antecedents. In 'Disposition Impossible' Carrie Jenkins and Daniel Nolan offer, for example, the case 'if Jane were to encounter a round square object Jane would be surprised' (Jenkins and Nolan 2012: 738). Yet we will leave these impossible dispositions aside here.
23 In Section 5.3 there is a paragraph on backtracking which further highlights the importance of these similarity considerations. Interested readers might want to make a little detour.
24 We do not wish to go much deeper into comparisons, but see how Lewis (1973a: §2.6: 69) deals with cotenability and the threatening circularity problem.
25 Many authors believe that the future of the worlds (from the moment *e* occurred) is, as opposed to the past, of lesser or no interest for similarity between worlds. This sounds intuitively right: if we consider a counterfactual (say, had *c* not happened *e* would not have happened) we allow for the then unfolding world to be different from the actual one. In fact, we possibly expect that the worlds will diverge, maybe massively, from that moment on. Yet, Lewis makes no such assumption that past similarity is more important than future similarity.
26 We disregard a fourth entry he gives: 'It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly' (Lewis 1979, quoted in Lewis 1986a: 48).
27 David Lewis is a defender of such a modal realism (Lewis 1986b).
28 Note as an aside that the truth conditions we have given for counterfactuals depend crucially on the laws of nature plus some actually obtaining facts. That is, while they are somewhat hidden, laws still play a crucial role for dispositional analyses. Dispositions, counterfactuals and laws are obviously very intimately related.
29 Although no one ever explicitly defended it in the literature.
30 John Heil (personal correspondence) reports that Martin had written the paper a couple of decades earlier.
31 Other examples can be found easily if it is in doubt whether being live should count as a disposition.
32 As a matter of fact, our original matchstick (RC) case already means trouble for a counterfactual analysis of dispositions and Martin would not have had to invent a new case. Suppose *p* and *q* are in fact both true. Now, the most similar *p*-world to our own is... our world! Is *q* there true? Yes, incidentally it is and, thus, the counterfactual comes out true (Lewis 1973a: 26). Apply this to the dissolving match: it would be attributed 'soluble' according to the respective counterfactual definition.
33 There are answers to Martin's challenge that differ from Bird and Lewis's. Lewis takes Martin's attack seriously, and he reformulates the too simple-looking conditional. Lars Bo Gundersen (2002) and,

independently, Sungho Choi (2006), who is probably the world's leading expert on the conditional analysis of dispositions, argue differently. They give convincing reasons that the semantics of counterfactual conditionals which allows conditionals to be variably strict is already by itself rich enough to buffer the dispositional conditional against finks and the like so that reformulations are not needed.

34 Note that Lewis's idea to refer to some (underlying) property is hardly a new idea: remember what, for example, David Armstrong had to say about dispositions and their basis (Section 2.2.2) and, also, remember that Eino Kaila saw in the postulation of something like such a property a way to handle the problems of the analysis with material implication.

35 Note as an aside that *B* is not the dispositional predicate *D* itself so that there is no danger of circularity in Lewis's account. It is also important to see that Lewis writes 'for some intrinsic property *B* that *x* has'. The formulation is such that for a different object, *x**, some other basis, *B**, could play our first *x*'s *B*'s role. In other words, the disposition *D* is not confined to one and only one basis property *B*. Instead: *D* is multiply realisable (Section 2.2.4).

36 We considered intrinsicality very briefly in Section 2.1.4. Further reading on the issue of intrinsic versus extrinsic properties can, for example, be found in Lewis (1983), Lewis and Langton (1998) and Hoffmann-Kolss (2010).

37 Not everyone shares the intuition that all dispositions are an object-intrinsic affair. For an influential argument for the existence of extrinsic dispositions see McKitrick (2003).

38 We ignore that Lewis writes 'x-complete cause' instead of merely 'cause'. The completeness demand is there simply to ensure that everything about the object *x* that contributes to the reaction is retained till *t*'. If Lewis had merely spoken of 'a cause' that has to be retained then some other crucially relevant causal factors could still be finkishly lost (Lewis 1997: 149).

39 See Lewis (1999: 148), where he discusses fragility, but the gist of the argument is the same.

40 You might protest that Lewis did clearly not think of any occult power or even necessitating force when he thought of the *B* being a cause. Indeed, nothing would be more alien to a neo-Humean and, thus, Lewis gives a Humean definition of what causation is which is entirely devoid of reference to productive powers. We'll get to know his theory, amongst many others, in Chapter 5, on causation.

41 For masking and mimicking see Johnston (1992). Our notion of a *prodote* cannot be found in the literature. Just as an antidote is the equivalent of a fink, let us call a *prodote* the analogue to a reverse fink.

42 In fact, Bird has his own interesting suggestion for a revision: Bird (1998: 231ff). See Spohn (1997: 337) for a similar answer.

43 For further discussions see, for example, Molnar (1999), Malzkorn (2000), Gundersen (2000, 2002, the latter with a reply by Bird (2000)) and Choi (2003).

44 Alexander Bird is special: he first attacks the (reductive) counterfactual analysis, then gives a solution to his own counterexample, but still becomes later a defender of Dispositionalism. How this is possible is hinted at in Section 6.4.2.

45 The back and forth with definition attempts, counterexamples and redefinitions, etc., which ultimately leads to turning things on their head, is no single case in philosophy. Readers familiar with basic epistemological questions will have heard of Gettier's (and similar) cases that make life hard for those who aim to define what it means for a person to know some fact, *p*, in terms of true justified belief plus some other features. Partially in the light of the continuing definition failures some, and notably Tim Williamson, declare knowledge to be primitive and unanalysable and many of the other concepts which were thought to figure in the analysans (belief, for example) to be a derivative of knowledge (Williamson 2000).

Bibliography

Adams, E. W. (1975) *The Logic of Conditionals*. Dordrecht: Reidel.

Bennett, J. (2003) *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Bird, A. (1998) Dispositions and Antidotes. *Philosophical Quarterly*. 48 (191). pp. 227–34.

Bird, A. (2000) Further Antidotes: A Response to Gundersen. *Philosophical Quarterly*. 50 (199). pp. 229–33.

Carnap, R. (1956) *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press. 2nd, enlarged edition.

Carnap, R. (1963) Arthur Pap on Dispositions. In: Schlipp, P.A. (ed.) *The Library of Living Philosophers, Vol. XI*: La Salle, IL.: Open Court.

Carroll, J. W. and Markosian, N. (2010) *An Introduction to Metaphysics*. Cambridge: Cambridge University Press.

Chisholm, R. (1946) The Contrary-to-Fact Conditional. *Mind*. 55 (220). pp. 289–307.

Choi, S. (2003) Improving Bird's Antidotes. *Australasian Journal of Philosophy*. 81 (4). pp. 573–80.

Choi, S. (2006) The Simple vs. Reformed Conditional Analysis of Dispositions. *Synthese*. 148 (2). pp. 369–79.

Choi, S. (2008) Dispositional Properties and Counterfactual Conditionals. *Mind*. 117 (468). pp. 795–841.

Cross, T. (2012) Recent Work on Dispositions. *Analysis Reviews*. 72 (1). pp. 115–24.

Edgington, D. (1995) On Conditionals. Mind. 104. pp. 235–329.

Fara, M. (2005) Dispositions and Habituals. Noûs. 39(1). pp. 43–82.

Goodman, N. (1947) The Problem of Counterfactual Conditionals. *Journal of Philosophy*. 44 (5). pp. 113–28.

Goodman, N. (1955) *Fact, Fiction and Forecast*. 4th edition (1983). Cambridge, MA: Harvard University Press.

Gundersen, L. B. (2002) In Defence of the Conditional Account of Dispositions. *Synthese*. 130 (3). pp. 389–411.

Gundersen, L. (2000) Bird on Dispositions and Antidotes. *Philosophical Quarterly*. 50. pp. 227–229.

Heil, J. (2005) Dispositions. *Synthese*. 144 (3). pp. 343–56.

Hoffmann-Kolss, V. (2010) *The Metaphysics of Extrinsic Properties*. Frankfurt: Ontos.

Hughes, G. E. and Cresswell, M. J. (1996) *A New Introduction to Modal Logic*. London and New York: Routledge.

Jackson, F. (ed.) (1991) *Conditionals*. Oxford: Clarendon Press.

Jenkins, C. and Nolan, D. (2012) Disposition Impossible. *Noûs*. 46 (4). pp. 732–53.

Johnston, M. (1992) How to Speak of the Colors. *Philosophical Studies*. 68 (3). pp. 221–63.

Kripke, S. (1959) A Completeness Theorem in Modal Logic. *The Journal of Symbolic Logic*. 24 (1). pp. 1–14.

Kripke, S. (1963) Semantical Considerations on Modal Logic. *Acta Philosophica Fennica*. 16. pp. 83–94.

Kripke, S. (1980) *Naming and Necessity*. Oxford: Blackwell.

Langton, R. (1998) *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Clarendon Press.

Leibniz, G. W. (1686) On the True Method in Philosophy and Theology. In: Wiener, P. P. (ed.) (1951) *Leibniz: Selections*. New York: Charles Scribner's Sons.

Leibniz, G. W. (1709) *Theodicy – Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*. Farrer, A. (ed.), Huggard, E. M. (trans.) (1952). New Haven, CT: Yale University Press.

Lewis, D. (1973a) *Counterfactuals*. Oxford: Oxford University Press.
Lewis, D. (1973b) Causation. *Journal of Philosophy*. 70 (17). pp. 556–67. Quoted in Lewis 1986a. pp. 159–213.
Lewis, D. (1979) Counterfactual Dependence and Time's Arrow. *Noûs*. 13. pp. 455–76. Quoted in Lewis 1986a: 32–51.
Lewis, D. (1983) Extrinsic Properties. In: Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press. pp. 111–15.
Lewis, D. (1986a) *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.
Lewis, D. (1986b) *On the Plurality of Worlds*. Oxford: Blackwell's.
Lewis, D. (1997) Finkish Dispositions. *Philosophical Quarterly* 47 (187). pp. 143–58.
Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
Lewis, D. and Langton, R. (1998) Defining 'Intrinsic'. *Philosophy and Phenomenological Research*. 58 (2). pp. 333–45.
McKitrick, J. (2003) A Case for Extrinsic Dispositions. *Australasian Journal of Philosophy*. 81 (2). pp. 155–74.
Malzkorn, W. (2000) Realism, Functionalism and the Conditional Analysis of Dispositions. *Philosophical Quarterly* 50 (201): pp. 452–69.
Manley, D. and Wasserman, R. (2008) On Linking Dispositions and Conditionals. *Mind*. 117 (465). pp. 59–84.
Martin, C. B. (1994) Dispositions and Conditionals. *The Philosophical Quarterly*. 44 (174). pp. 1–8.
Molnar, G. (1999) Are Dispositions Reducible? *Philosophical Quarterly*. 49 (194). pp. 1–17.
Nolan, D. (2005) *David Lewis*. Chesham: Acumen.
Priest, G. (2001) *Logic: A Very Short Introduction*. Oxford: Oxford University Press.
Quine, W. V. O. (1966) Necessary Truth. In: *The Ways of Paradox and Other Essays*. Revised and enlarged edition (1976). New Haven, CT: Harvard University Press. pp. 68–76.
Ramsey, F. P. (1931) *The Foundations of Mathematics and Other Logical Essays*. Braithwaite, R. B. (ed.) London: Routledge & Kegan Paul.
Ramsey, F. P. (1990) *Philosophical Papers*. Mellor, D. H. (ed.). Cambridge: Cambridge University Press.
Spohn, W. (1997) Begründungen apriori-oder: Ein frischer Blick auf Dispositionsprädikate. In: Lenzen, W. (ed.) *Das weite Spektrum der analytischen Philosophie -Festschrift für Franz von Kutschera*. Berlin: De Gruyter. pp. 323–45.
Stalnaker, R. (1968) A Theory of Conditionals. *Studies in Logical Theory, American Philosophical Quarterly Monographs*. 2. pp. 98–112.
Vetter, B. (2014) Dispositions without Conditionals. *Mind*. 123 (489). pp. 129–56.
Williamson, T. (2000) *Knowledge and Its Limits*. Oxford: Oxford University Press.
Wittgenstein, L. (1921) *Tractatus logico-philosophicus*. In: Ostwald, W. (ed.) *Annalen der Naturphilosophie, Vol. 14*. pp. 185–262. McGuinness, B. and Pears, D. (trans.) (2001). London: Routledge Classics.

4 Laws of nature

4.1 Basic intuitions about laws

Setting the stage Finding the laws of nature is one of the central tasks of the natural sciences. To explicate what a law of nature is, is one of the central projects within the metaphysics of science. This is an intriguing task not only in its own right. We have already encountered several places within metaphysics where reference to laws became necessary.

Carnap suggested to list laws alongside his reduction sentences in order to bridge the definitional gaps they leave behind (Section 2.1.3), Kaila's definition of dispositional predicates incorporates reference to universal regularities that are best interpreted as laws (Section 2.1.4), Goodman and Lewis need them for their theories of counterfactual conditionals (Sections 3.2 and 3.4).

Since we have made it our strategy throughout the whole book to start with Empiricism affine theories we also start this chapter with Empiricist efforts to define what laws are (Section 4.2). Suppose we succeeded with such a definition. Then we could plug it into Kaila's, Carnap's, and Goodman's efforts to define dispositional predicates or counterfactuals and open up the possibility for these enterprises to succeed. Unfortunately, we will once again see the radical strictures of Empiricism fail.

General intuitions Here are, before we plunge into Empiricism again, some basic characteristics of laws of nature or law statements philosophers have defended: truth, objectivity, generality/universality, contingency, support of counterfactual conditionals, confirmability by induction, being explanatory valuable, aiding predictions and retrodictions.

A little scattered in the chapter we will encounter them all again but here are already some preliminary thoughts on some of these features: *truth* can be problematic when we consider that some laws are idealized or hold, at best, other things being equal (Section 6.2.1). Does *objectivity* mean total independence of human opinions, goals and needs (Section 4.3)? Some philosophers advertise a more pragmatic attitude.

Generality/Universality may mean four different things: that, (i), laws are valid everywhere in space and time (although there might be some biological laws, say, concerning just what happens on earth), or, (ii), that they hold for any object, or, (iii), under any circumstances (although, again, some laws have proviso clauses or are idealised), or, (iv), that they hold for the widest range of variables (in case they have a mathematical encoding).[1] Being epistemic issues, the laws' role in *explanations*, *predictions*, and *retrodictions*, their *confirmability by induction* and how they are *falsified* will not be a topic of our chapter. Yet we will target laws' relation to *counterfactuals* in more detail.

Necessitation Additional to the just mentioned features of laws there is one central issue about them that has divided the community for centuries. Some philosophers have the feeling that natural laws govern the events in the world: what a law says *must* happen, or what a law forbids *can not* happen. This intuition might in part originate in our actual day-to-day experiences when we feel resistance against some of our actions. Some goals are not merely difficult to achieve, they seem *impossible*: we cannot walk through a concrete wall, run a marathon in under one hour, or jump five metres high unaided. In concert with the facts about our current body mass, leg muscles, and the earth's gravitational field, the laws of nature seem to prohibit this latter kind of leap. (This will remind the reader of two central features we intuitively attributed to dispositions: modality (MOD), and productive power (PROD) (Sections 2.1.2, 2.1.3; see also the beginning of 6.4.2)).

For other philosophers, laws have more of a descriptive character: the laws are (merely) accurate reports of what regularly happens or is universally the case. This intuition comes from the observation that nature seems to be uniform. Alleged laws like Boyle's law (which says that for a fixed amount of an ideal gas kept at a fixed temperature, its pressure and volume are inversely proportional: $p = k/V$) or Einstein's famous mass–energy equivalence ($E = mc^2$) record these universal regularities.

Those who hold the first intuition (that the laws necessitate what happens and prohibit what does not happen) do not think the second intuition is wrong. In fact, if, what the laws say, *must* happen, then it also always does happen and we get the regularities for free. The necessities in nature supposedly *produce* (PROD) the regularities and thus explain why they are there. Yet those who subscribe to a pure regularity view (that's the second intuition above) deny that laws have any productive force.[2] With these features of laws in mind we will now take a little historical detour.

Historical background The two positions we have here described are still very much alive in contemporary metaphysics and we will get to know proponents of both sides. Historically, we know, of course, one prominent place where they collided: recall David Hume's argumentation that the postulation of necessary connections in nature is so suspect that we should rather remove it from our ontological inventory. We also know that the Logical Empiricists adopted this scepticism towards any necessity or necessitation in nature. Hume's opponents, remember, were the Rationalists. Although we cannot get into the details here we should at least mention that some of them were, at least tacitly, religiously motivated when it comes to laws: some divine creature wills what is happening in the world and, surely, what god wills must happen.

Descartes, for example, explained the regular behaviour of the inactive matter the universe is filled with by reference to god's decreeing the laws. This view had rivals of a similar nature: God could, instead, himself act upon nature. The laws then would be regularities in god's actions. Or there could be lesser divine beings that push and pull matter around in accordance with god's will. These three suggestions, the latter of which he endorses, come from Ralph Cudworth (1617–88) (see Cudworth 1678).

In the eighteenth century, forces join passive matter. In Kant's *Metaphysical Foundations of Natural Science* we even find that the concept of matter is unthinkable unless we do perceive it as attracting and repelling other bodies (see Kant 1786: IV. 510 and 514). Matter, now being active itself, needs no external laws anymore to govern it.[3] (Compare this to the Dispositionalism we develop in Chapter 6.)

The view was short lived, though. Nineteenth-century science thought forces obsolete. The physicist Gustav Kirchhoff (1824–87), for example, writes that the sole task of mechanics is to describe the phenomena (here: movements of bodies or particles) but not the (alleged) causes, i.e. forces, behind them. Forces, according to Kirchhoff, were helpful to conceive mechanical theories but in advanced science reference to such striving or pushing should be avoided (see Kirchhoff 1876).

Kirchhoff leads us to the starting point for this chapter: being suspect of speculative metaphysics the Logical Empiricists set out to formulate godless, necessity-free, forceless approaches to laws. However, we will see how hard it is to abstain from necessity altogether and, thus, also get to know reasons why twentieth-century metaphysicians started to believe in it again.

Before we start, let us agree to call sentences *accidentally true* if they are true *just as it so happens* rather than because of a natural law (see Goodman 1955: 73). For example, that there are two pencils on my desk is such an *accident*. That they both gravitationally attract each other is, however, no accident: that's what Newton's law says about all masses. Also note that we use the terms *nomic* or *nomological* (from Greek *nómos* for law) synonymously for *according to law* or *lawlike*.

4.2 Early regularity accounts

Language first! Unsurprisingly, for the Logical Empiricists the starting point to theorise about the nature of laws is to focus on their linguistic representation. To them, this approach via language suggests itself for two reasons.

The first is independent of Empiricism and is simply the observation that prototypical *law statements* as they occur in the actual practice of science seem to share a linguistic feature, namely, that they all are *universal generalisations*, i.e. sentences of roughly the following form 'All such and such are so and so': 'All changes in hereditary dispositions are caused by gene mutation', 'Gases expand when heated at constant pressure', 'Each metal has a specific melting point', etc. It will turn out that this characteristic is considered by many a necessary feature for law statements but on its own it is not sufficient. (Note that the universality of these statements captures the respective regularity in nature).

The second reason to focus first on the linguistic expression of alleged laws is specific to the Logical Empiricists' general affinity to logic and language. Remember, that this is nowhere more prominent than in their Verificationist criterion of (sentence) meaning (Section 1.4.1).[4] In accordance with it, Empiricists attempted a definition of lawhood that could in principle be cast into a form like this: *L is a law if and only if . . .*, where the dots have to be filled with innocent bits and pieces that do not violate Verificationist restrictions like observability and also refrain from reference to necessity.

Lawlikeness and truth Here's a start: from paradigm cases of assumed law statements ('All electrons are charged') the Empiricists extracted the linguistic feature of universality (the idea just mentioned above) and decreed that a sentence is to be called *lawlike* if it has the form of a universal statement. (In logical form these are those sentences starting with a universal quantifier: $\forall x\ (Ex \rightarrow Cx)$; E might stand for *being an electron*, and C for *being charged*).

Unfortunately, it is pretty clear that not all statements that can be called *lawlike* according to this definition are also laws: some such sentences, like 'all cows are reptiles' that have that form, are simply *false* and should therefore surely have no chance to count as laws. Thus, as a second ingredient, *truth* must be added to lawlikeness in order to make a law. Hence, a first extension of the above starting point is this: for any statement *L*, *L* is a law of nature if and only if *L* is *lawlike* and *true*.

A division of labour This is in principle a neat idea for note that two tasks are now separated: the further explication of what it means for a statement to be *lawlike* can be delegated to the philosopher and whether it is *true* to the empirically working scientist. If both is the case, i.e. the statement is *lawlike* and *true*, then it expresses a law.

Analyticity So far so good, but this cannot be the last word. First of all, if lawlikeness were just generality then universal logical truths like 'All ravens are ravens' would count as laws according to this criterion. Yet these 'laws' have no empirical content. We can know them to be true without a single experiment and surely without sense perception. Thus, such logical truths, even if in universally quantified form, should not qualify as laws of nature for laws should be about matters of fact and have empirical content. So, let us exclude analytically true statements.

True generality Yet there is an even trickier issue. The true statement 'All coins in my pocket are silver' is, although universally quantified not really general enough in its scope to count as a law: not only does it refer to some particular individual (me) but also to a specific region in space and time (my pocket). A law statement, so is our intuition, shall not have such reference to individuals and limited areas in space-time. Rather, we want laws to hold everywhere at any time and to hold for all (and potentially infinite) members of a sufficiently general class of things (all electrons, for example, and not coins in specific pockets). As a consequence, we have to demand further (in addition to being non-logically true and universal) that no proper name or other singular term (like the indexical *my pocket*) should appear in the statement and no reference to particular space-time points should be made. (Time and space can, of course, appear as variables and/or intervals but not as specific coordinates).

Unfortunately, this still doesn't do the job for we can hide reference to individuals in pseudo-general predicates. (We learnt a very similar trick already: remember Carnap's response to Kaila (Section 2.1.4)) If we define the class F as the class of things that are identical to a coin in my pocket then we can, afterwards, translate the above pocket-statement into one that looks on its syntactical surface like a perfectly general law-like statement: 'All Fs are silver'.

Natural kinds, again In order to get rid of these unwanted pseudo-universal sentences we have to transcend further the criteria of syntactic form and logic alone, i.e. universal quantification, no tautology, no singular but only general terms, and add in semantic features: we might, for example, want to say that the general terms in a law statement should stand for *natural properties* or *natural kinds* like electrons, tigers, water, gold, photons, etc., and not for artificial classes like the above F. Then we could try to define laws as non-logically true statements that are universally quantified *and contain kosher predicates only* that stand for natural kinds and/or properties.

This – the demand for naturalness (or intrinsicality or projectibility or some such) – is, of course, the very same reaction we came across when trying to rescue Kaila from Carnap (Section 2.1.4). Naturalness of properties will also figure prominently in Lewis's account of laws, which we discuss in Section 4.3.

Reichenbach's golden spheres However, even if we wholeheartedly embrace the existence of natural properties and/or kinds – that is, properties or kinds nature herself comes equipped with – our present class of alleged law statements will still be too inclusive. Consider the statement 'All golden spheres have a diameter of less than 10m'. We can safely assume that this is, as a matter of fact, true in our universe (if not, just increase the diameter until it is true; the example is from (Reichenbach 1947: 368) who, there, spoke of golden *cubes*). Then, according to our current theory, it ought to count as a law because it is a universal statement that contains natural property terms only: if *gold*, referring to a chemical element, does not count as natural then what does? (The other general terms like *sphere* and *diameter* are allowed on the grounds that mathematics and mathematical terms are, especially in the physical sciences, common if not essential ingredients in law statements).

Yet we have the strong intuition that even if universally true it is a mere accident, not a law, that no lump of gold of that size exists. We feel that it easily could have existed and, thus, that this statement can't be lawlike. This becomes especially clear when we contrast the gold-sentence with an exactly parallel one about uranium-235: 'All uranium spheres have a diameter of less than 10m'. The latter, in contrast, should count as lawlike because the free electrons in uranium-235 would kick off a chain reaction that would make any sphere of that size (precisely: one which is above critical mass) blast. However, on the basis of the syntactic and semantic criteria so far given we cannot distinguish between the gold and the uranium sentences: both sentences are true universal non-analytic statements that mention only natural and mathematical properties.

Note that, here, the temptation is strong to say: '*Necessarily*, there are no such uranium spheres. There can't be! Yet nothing *necessitates* that there are no such golden spheres'. Reacting this way would be very anti-Humean and anti-Empiricist, of course. It would postulate, over and above the mere regularity, a factor that makes sure that things stand as the law statement but not the accidental truth states. We will see soon how some philosophers dared to oppose Hume and the Logical Empiricists and chose this option after it had been considered untenable for 200 years (Sections 4.4–4.6).

No objectivity in lawhood Other people were willing to throw in the towel and give up the attempt to find objective criteria to distinguish between accidentally true and nomologically true universal sentences. If one makes this pessimistic anti-Realist step it is of course best taken before the rescue attempt via natural kinds so that one does not have to carry that metaphysical baggage. We did mention such views already in our historical outline when engaging with Logical Empiricism (Section 1.4.2). Similarly, Gilbert Ryle called laws 'inference tickets':

> A law is used as, so to speak, an inference-ticket (a season ticket) which licenses its possessors to move from asserting factual statements to asserting other factual statements. It also licenses them to provide explanations of given facts and to bring about desired states of affairs by manipulating what is found existing or happening.
>
> (Ryle 1949: 117)[5]

Goodman concurs:

> I want to emphasize the Humean idea that rather than a sentence being used for predictions because it is a law, it is called a law because it is used for predications.
>
> (Goodman 1955: 21)

Within these approaches to lawhood, the distinction between accidentally true sentences and lawlike statements becomes a more pragmatic issue: we might want to call those universal statements that we find more useful for predictions, explanations, counterfactual reasoning, etc. *laws*. Objectively, however, these *laws* are really not separated from the accidents by any profound ontological gap (see Ayer 1963: 230).

BOX 4.1 and 4.2 Early regularity accounts

- Logical Empiricists had an interest in spelling out what a **law of nature** is without reference to any kind of **necessity**.
- Focusing first and foremost on **statements** and trying to give **linguistic criteria** to distinguish **law statements** from those that state **accidents** they split the task to say what a law is into two: say what **lawlikeness** is (of a statement), then define *law*

in saying that a ***true*** **lawlike statement** signifies a law. The analysis of *lawlikeness* is a task for the philosopher, the finding of truth one for the scientist.

- We discussed four attempts to refine lawlikeness. None of them, not even in combination, sufficed.
  (1) Pure universality is not enough: there are universal but **accidental truths**.
  (2) Also, the **logically true universal statements** have to be excluded.
  (3) Moreover, generality should not only be a **syntactic feature** on the surface of the statement. The predicates in it should be true of **possibly infinite entities** (i.e. the predicates should be **projectible**). One way to secure this would be to demand that they refer to **natural properties.**
  (4) Not even taking step (3) is enough to weed out the accidents: **Reichenbach's golden** and, respectively, **uranium spheres** attest to this.
- We introduced an **anti-Realist approach to laws**: there is **no objective difference** within the class of true universal statements. Rather, for **pragmatic considerations** regarding their **usefulness for predictions** we might classify some as laws, others not.

4.3 A sophisticated regularity account

There is a more sophisticated account of laws of nature that respects Regularist intuitions and yet has the potential to handle cases like the golden sphere (Section 4.2). This account is holistic in that it does not judge individual statements whether they are laws but rather whole systems of sentences.

Although the idea has been around for a while – John Stuart Mill (1806–73) had already endorsed it in 1843 and Frank Ramsey in 1928[6] – some time passes before a philosopher picks it up again to develop it into a full blooded theory but, first, here's Mill with a foretaste of what awaits us:

> According to one mode of expression, the question, *What are laws of nature?* may be stated thus:—What are the *fewest* and *simplest* assumptions, which being granted, the whole existing order of nature would result? Another mode of stating the question would be thus: *What are the fewest general propositions from which all uniformities which exist in the universe might be deductively inferred?*
>
> (Mill 1843: 317, emphasis added)

Systematising the world The refinement of this Mill–Ramsey theory in post-Empiricist times comes from David Lewis. It is therefore often referred to as the *Mill–Ramsey–Lewis view of laws*.[7] Another name is *Best System Account (BSA)* for reasons that will become clear immediately.

Lewis presents his ideas first in his 1973 book *Counterfactuals*, which has been central for us already, in Section 3.4, on this very issue. Also remember that we got to know Lewis in our sections on the ontology of dispositions endorsing what he named 'Humean Supervenience' (Sections 2.2.3–2.2.5). That name should be telling: Lewis's theory of laws will shun reference to necessity or any other secret connection in nature.

Lewis makes a profound Realist assumption: the world is, fundamentally, a four dimensional space-time mosaic of instantiations of point size categorical properties (Section 2.2.5). You, me, tables and chairs are all ultimately composed of such pixels.

With this assumption in the background let us see how Lewis defines lawhood. He starts: suppose you knew everything about the past, present and future, all facts, all events, just everything. Further, assume that you wish to note down some or even all of your knowledge in an economical way. That is, you don't just want to list the vast amount of facts (there's an electron there, here was a dinosaur once, there moved a star, . . .) but you want a good summary. For example, instead of saying about each and every electron that it is negatively charged (e_1 is charged, e_2 is charged, etc.) you'd say, only once, that *all electrons are charged*. Since it simply follows logically from the general statement that an individual electron you encounter is charged you do not need to repeat this: your summary has become much simpler. We imagine that you repeat this successful strategy with other true generalisations (about tigers, maybe, or uranium or gold). In other words, you systematise your knowledge.

We can imagine that if we have a whole bunch of generalisations we can deduce further ones from them. For example, if we also include that like charged objects repel each other we can deduce, with the help of the already mentioned regularity, that electrons repel each other.

Now, here's a further way in which you might be more economical: if there is a way to state exactly how, say, water is composed out of smaller units (for brevity, let's say it is two hydrogen (H) atoms and an oxygen atom (O) in a specific arrangement) you might not have to say much about water anymore if your summary contains information about its composition and all the info about the building blocks (such information is encoded in 'bridge laws').

System competitions We can assume that such systematisations – Lewis speaks of *deductive* systems because of the many just mentioned internal interrelations – can be done in various ways and with differing elegance. If you merely talk about medium-sized objects you will leave out a lot about micro-entities. If, on the contrary, you say everything about the microstructure then, maybe, what can be known about macro stuff follows already from summaries of the micro world because the macro consists of the micro. (We are aware that this might be wishful thinking but let us go along with the assumption).

Lewis is only interested in the best of these possible systems. He further specifies what he means by *best*: it's the one system that achieves a far better combination of simplicity and strength than any of the other competing systems. Sounds good, but what is simplicity, what strength? To have *strength* is to bear a great deal of informational content about the world, and to be *simple* is to state everything in a concise way, not to be redundant, etc.[8]

Unfortunately, strength and simplicity tend to pull in different directions. Suppose you had not said anything yet about protons. If you add a general statement about these subatomic particles your system gets stronger. It has more informational content. Yet at the same time your system is less simple for it lists that additional sentence. Still, if you can buy a lot of strength for a small complication you should do it. This is why Lewis does not ask to maximise strength and simplicity separately but to balance them:

> What we value in a deductive system is a properly balanced combination of simplicity and strength – as much of both as truth and our way of balancing permit.
>
> (Lewis 1973: 73)

Laws of nature Now, what has all this to do with the laws of nature? Here's Lewis's definition:

> A contingent generalization is a *law of nature* if and only if it appears as a theorem (or axiom) in [. . .] the true deductive systems that achieves a best combination of simplicity and strength.
>
> (Lewis 1973: 73; also see Lewis 1983a, 1994)

Fundamental and derived laws; different sample systems A few remarks about this definition are in order: by *contingent* Lewis means to exclude analytically true statements so as to make sure that laws are only statements with empirical content (Section 4.2). The *axioms* of a system are those general statements that cannot be deduced from other statements in the system. They count as the *fundamental laws*. Theorems are deduced statements. They are less fundamental but count, according to

Lewis, also as laws. Remember our water example from above: in axiom/theorem terms we can now speculate that there are water theorems deduced from axioms about *H* and *O*. In this way, the best system might include laws that correspond to our special science laws from chemistry, biology, psychology, etc., if they follow as theorems from the fundamental axioms/laws (whatever they are).

Note, however, that, for the possibility of the deduction of water laws from subatomic laws, the system would have to say how water is related to its subparts.

To get a still better grasp on Lewis's best-system idea let us have a look at three examples for possible competing systems. We could have, (i), have a gigantic lookup-table: for each space-time point the table lists the properties that are instantiated there (charge at $<x_1, y_1, z_1, t_1>$, mass at $<x_2, y_2, z_2, t_2>$, etc.) This would be as strong as it gets but it would not be a very simple system and hence probably lose against the others (it would anyway not contain any *general statement* and therefore no laws). We could, (ii), have a single line: 'all electrons have unit charge'. This is a very simple but indeed very weak system for which it is not possible to win any competition. We could, (iii), have present day physics. If we are lucky it is not too bad a system when it comes to a convincing, balanced combination of strength and simplicity. Yet it will probably be superseded at some point by a stronger and simpler arrangement and, if we are very lucky, we might approach the system that, like in Lewis's assumption, is the best seen from an omniscient god's-eye perspective. That is, if we are lucky, we, as scientists, arrive at *what the laws really are* (but what they are is, of course, independent of our inquiry).

Back to gold and uranium Remember that we introduced the Mill–Ramsey–Lewis account of lawhood in reaction to Reichenbach's golden-spheres problem. Historically, it was not developed in direct reaction to the problem, but Lewis does underline that it has the potential to solve it (see Lewis 1973: 74). Here's how: while a general statement like 'All golden spheres have a diameter of less than 10m' might be true of the world (past, present, and future) it is too marginal a truth, unconnected to anything else, to belong to an elite class of axioms and theorems of the best system. This is unlike the similar generalisation about uranium. That such spheres do not exist is connected to, even follows from, true general statements about sufficient uranium concentrations that lead to nuclear chain reactions. Therefore, the uranium-sphere theorem is probably derivable from these axioms. Thus it, but not the gold statement for which this is not the case, is a law.

Natural properties Recall how we arrived at the gold–uranium problem: Empiricist attempts to define what a law is did not seem possible, unlike when we declare that the properties the law statement refers to are *natural properties*.

In 1973, Lewis thought that he, unlike these earlier regularity theories, needs no reference to naturalness. He believed that best-system competitions sort out, just by themselves, what the laws of nature are and, in passing, thereby which properties are natural ones: *unnatural* properties would be weeded out because systems that refer to them could, so the idea and hope, not possibly win the competition.

This turned out to be wrong (we'll come to the reason shortly) and in reaction Lewis became a defender of the idea that nature herself is divided into her very own classes of entities. In fact, the idea of mind-independent natural properties became a primitive fundamental building block in Lewis's metaphysical system:

> Formerly I had been persuaded by [Nelson] Goodman and others that all properties were equal: it was hopeless to try to distinguish 'natural' properties from gruesome gerrymandered, disjunctive properties. Eventually I was persuaded, largely by D. M. Armstrong, that the distinction I had rejected was so commonsensical and so serviceable – indeed, was so often indispensable – that it was foolish to try to get on without it.
>
> (Lewis 1999: 1–2; see also Lewis 1983a)

Here's why it is indispensable for Lewis to take naturalness on board as an essential feature of the properties under concern:

Gruesome properties If any kind of predicate (referring to the weirdest classes of things (Section 2.1.4)) were allowed in the languages of competing systems then competitions/comparisons would not be possible for the lack of a proper inter-system-simplicity-measure. We explain this with the help of a toy example. (Be warned, despite it being a toy example, the argumentation is complex.)

Suppose we talk about a real mosaic with its coloured tesserae. Now, let the competing systems of one set of systems *S1* describe patterns in the mosaic in terms of our normal colour terms, including *green* and *blue*. Let the systems of another set of systems *S2* use the terms *grue* and *bleen* instead, where, translated into our normal language, '*x* is grue' means: either *x* is green and on the lower half of the mosaic *or x* is blue and on the upper half of the mosaic (and vice versa with *bleen*).

Again, *S1* and *S2* are meant to be *sets* of many competing systems, say, for *S1*: *s1*, s1**, s1****, . . . and, respectively, for *S2: s2*, s2**, s2***,* . . . Each, *S1* and *S2*, have, intrinsically, their winning system. Call their local winners: *s1* and *s2*. Remember, *s1* is written in the green/blue language and *s2* in the grue/bleen language.[9]

Suppose now that, as a matter of fact, it is true of the mosaic that all square tesserae in the lower half are green and all square tesserae in the upper half are blue. That's quite a complicated thing to say in the language of *S1*. Now, look at how we'd state this very same fact in the language of *S2*: 'All square tesserae are *grue*'. This is a very simple sentence. It is informative, and, thus, strong, too. Hence, we may assume that the best system *s2* of *S2* includes that sentence as one of its axioms. (Systems within *S2* that do not include it lose in strength considerably and can hardly compete.) How about the systems in *S1*? Well, complicated as the sentence is when expressed in the language of *S1* it might not find its way into *S1*'s best system s*1*. It might add too little strength for its complicated form.

The best system in S1 might however register other regularities as axioms. Let's stipulate additionally, again as a matter of fact, that all *round* tesserae of the entire, whole mosaic are green. Now, that fact is easy to state in terms of *S1* (it is simply to say 'all *round* tesserae are green') and, thus, the best system *s1* of *S1* might well have it as an axiom. Not so in *S2's* best system, *s2*. For try to express that all round tesserae (everywhere) are green in a *grue* language. That might well be too complicated a sentence to be in the winning team, i.e. we may assume that, for its lack of simplicity, it is not in *s2*.

What have we got? *s1* and *s2* are two local best systems, one for *S1* in a green/blue language, one for *S2* in a grue/bleen language. Yet they have registered different general statements as axioms. *s1* has 'all round tesserae are green' as one of its axioms, yet *s2* lacks in its axioms a corresponding sentence that would express that fact in its own language; similarly, s_2 has 'All square tesserae are *grue*', yet *s1* lacks in its axioms a sentence that expresses that generality. That, however, means that the two systems differ in what they say the laws are. The one has 'all round tesserae are green' as a law which the other does not include. It is extremely important to understand that *s1* and *s2* not only speak a different language they single out different generalisations as laws, that is, even after a possible translation from one in the respective other language there is no match in laws!

Which system, *s1* or *s2*, of these two local best ones is, now, the ultimate best? That is, which one states the true laws of nature? This turns out to be a perspectival issue if, indeed, we believe that all predicates/properties are alike. We might choose *s1* or *s2* depending which language we like best. Yet who are we to say whether *green* is any better than *grue*?

This is exactly the place where naturalness helps: although we, ourselves, do not have authority to decide nature does. If she really has her own properties, the natural ones, then, clearly, we should only look at systems whose predicates refer one to one to those properties. That's the only natural way to describe the world! Any other description might have its best system but why consider such an unnatural description if the *natural* laws are what we wish to unearth. Nature might say: '*my true colours* are green and blue. So, if you want to see what *my laws* are don't describe my objects as grue or bleen!'.

Refining naturalness Let us have a closer look at naturalness. At the bottom there is perfect naturalness. This is *naturalness proper*: nothing else really is. Is *being water* such a property? Well, maybe not. Maybe *being a quark* or *being an electron* is and *being water* is only derivative thereof. If the derivation (via bridge laws, see above) is not too complicated and contrived we could allow ourselves to call *being water* a 'fairly natural property'. (Not so with *being a smartphone* or *being a burned match*, etc.)[10]

(Perfectly) natural properties are those that make objects that have them objectively resemble each other (all other resemblances would be artificial). Also, if you have characterised an object completely by the natural properties it has, there is no more to say about it. Everything else would be redundant (see Lewis 1983b: 13–4; 1986: 50–69).

The Better Best Systems Account (BBSA) Recently, some philosophers have given the Lewisean best-system account a new twist (amongst others: Schrenk 2007, 2008, 2014; Cohen and Callender 2009, 2010)[11]. They still do not believe in natural properties or at least remain agnostic on that matter. Unlike Lewis these philosophers weren't interested in the most fundamental laws but rather in those of various actual sciences: the laws of biology, those of chemistry, the physical ones, etc. Therefore, they suggested fixing a set of predicates separately for each of these sciences: the set of biological predicates for biology, the set of chemical predicates for chemistry, etc. Best-system analyses then have to be conducted per science and its very own fixed vocabulary. The advantage of this *better best systems account* (BBSA) (as Cohen and Callender aptly call the approach) is that it can remain neutral on the topic of naturalness.

Back to the merely best: simplicity and strength Even if we have natural properties and the predicates that refer to them at hand Lewis's idea is still not independent of an anthropocentric factor: the standards of simplicity and strength might well be human-made standards.

Lewis's answer to this allegation (strength and simplicity for whom?) is to say that if nature is very orderly and uniform indeed, then there will be a system (mirroring this natural order) which is so good that it is far ahead of the others no matter what measure of simplicity and strength we apply.

There's a related issue where he gives the same optimistic answer: what if the world is a messy place and there are not that many true regularities? This could have the effect that there really isn't a best system but only ties of very mediocre ones. For Lewis's theory, however, it would not matter, for, if the world is such a mess, we would anyway also intuitively deny that there are laws. If the theory comes to the same conclusion that's all the better. Thus, Lewis writes:

> If nature is kind, the best system will robustly be the best—so far ahead of its rivals that it will come out first under any standards of simplicity and strength and balance. We have no guarantee that nature is kind in this way, but no evidence that it isn't. It's a reasonable hope. Perhaps we presuppose it in our thinking about law. [. . .] I can admit that if nature were unkind [. . .] then lawhood might be a psychological matter. [. . .] But I'd blame the trouble on unkind nature, not on the analysis.
>
> (Lewis 1994: 233)[12]

Modality We come to an important, as yet untouched issue: (nomological) necessity. Lewis's sophisticated regularity account gets rid of unwanted accidentally true statements like 'All golden spheres have a diameter of less than 10m': they do not belong to the theorems of the best system. So far so good, yet remember our initial reaction to the golden-spheres *law* that threatened the Empiricists' regularity account of lawhood. It was: *nothing really prevents us* from building such a sphere. *We could do it*. It is for this reason that the golden-spheres statement cannot be a law. Compare: if Einstein is right it is not only true that nothing moves faster than light, *nothing ever can move faster than light* and any attempt to accelerate an object to higher speeds *must fail*'.

Because of these intuitions, some philosophers thought that while Lewis could get rid of the golden-spheres pseudo-law he did so *by chance*, or, at least, not for the right reasons. Lack of simplicity and strength is a good argument against the golden-spheres *law*, yet ultimately it is its lack of a modal punch that really disqualifies it from law status.

A Lewisean can react to this allegation – the lack of modal impact – in the following way. Define *nomological*, i.e. *law generated necessity* thus:

> A statement, *p*, is nomologically necessary if and only if *p* follows from the laws of nature (being the axioms of the best system). Since every statement follows from itself the laws themselves are also nomologically necessary.
>
> (van Fraassen 1989: 44–5)

This manoeuvre indeed has as a result that the golden-sphere regularity isn't; yet the one about superluminal speed is nomologically necessary. Yet defenders of a more robust sense of lawhood and necessity still won't be satisfied. Closing the gap between necessity and lawhood by fiat does not convince them (Section 4.4).

Laws transcend the mere facts Anti-Humeans complain about another issue. Lewis's laws are very much tied to the facts. Yet so opponents argue, laws might transcend the mere facts. We simplified an example from Michael Tooley (1977: 669) to make this point:

Imagine a miniature universe with 10 sorts of fundamental particles. All kinds of reactions between those particles happen all the time and the outcome of these reactions is perfectly regular: for example, whenever *A*-kind particles react with *B*-kind particles a *C*-particle emerges. Suppose, however, that there are two sorts of particles, *X* and *Y*, which, as it so happens, never come close to each other throughout the whole history of that universe. It seems, now, that a Lewis-style best system would not contain any generalisations regarding *X*-*Y* interactions because such a generalisation would complicate the system while not adding any strength: since *X*-*Y* collisions do not in fact happen the informational content of the respective statement would be zero. And yet we might feel that if the particles were to interact they would do so in some predetermined law-governed way. Neither that law nor the corresponding modal fact about *X* and *Y* particles would be captured by Lewis's best system.

In other words, Lewis's theory of lawhood would not be extensionally adequate because it would leave out some of the laws: the *X*-*Y*-particles scenario is in one respect the complement to the original golden-spheres problem. The latter originally caused the difficulty that too many statements would count as laws which are only accidentally true. Here, the contrary is the case: some laws that (intuitively) exist slip through the net and are not registered as such.

There are two ways Lewiseans can answer (see, for example, Helen Beebee 2000). Either they argue that, despite prima facie appearance to the contrary, it is simply not the case that vacuous generalisations like the *X*-*Y law* do not find their way into best systems. One reason why such empty regularities might after all figure in the best system is that if such a generalisation is interwoven with other generalities about *X*s and *Y*s it might well provide deductive strength (costing only little complication for what it delivers). Another reason might be that that empty *law* is deducible from the others. Think of the periodic table of elements: some laws for the so far unsynthesised elements in the bottom row can be deduced from the laws of elements from above. Thus, although there is no actual regularity, they, being theorems of the best system, count as laws.

This first defence strategy tries to satisfy the anti-Lewisean intuitions in that it argues that Lewis's theory is extensionally adequate after all. Yet a good Lewisean should probably choose a different route because this first defence does make some concessions to the opponent in so far as it buys from them that the *X*-*Y* regularity is indeed a law.

The other way to react is simply to criticise the *X*-*Y* particles example as begging the question. That is, the Lewiseans should simply say that they do not share the intuition that the empty generalisation is a law. Or, even stronger, that the idea that laws are more than regularities was secretly smuggled into the scenario from the outset: assuming that there is an *X*-*Y* law beyond what is regularly happening *is* assuming, from the start, that laws are more than regularities. This is what Lewiseans deny. So, they can simply stick to their guns and continue to claim that there is no *X*-*Y* law.

Explanatory force Note one final, related allegation that has been put forward against the Mill–Ramsey–Lewis view (and *a forteriori* to simpler regularity views). On these accounts, laws seem not to have explanatory power. Someone might ask: 'Why is this electron moving on that trajectory?' and a scientist might answer: 'Because it is a law that all electrons behave that way in electromagnetic fields'. Now, if laws are but summaries of everything that in fact happens then this single event is (partially) explained by itself because it is one of the many single events the law is a mere summary of.[13]

BOX 4.3 A sophisticated regularity account

- The **Mill–Ramsey–Lewis** view of laws, also known as the **Best System Account,** asks us to do the following: suppose you knew everything and organised it as simply as possible in various competing deductive systems that mention **perfectly natural properties** only.
- A **contingent generalisation** is a **law of nature** iff it appears as an **axiom** or **theorem** in the one true deductive system that achieves a best combination of **simplicity** and **strength**,
- where to be **simple** is to state everything in a concise way, not being redundant, etc.; to have **strength** is to bear a great deal of informational content about the world.
- Lewis has to assume that the world brings with it its **very own natural properties**. Omitting this assumption leads into the same sort of trouble we know from early analyses of dispositional predicates.
- Lewis's analysis has to face the allegation that it is **anthropocentric** because the standards of simplicity and strength are.
- There are three further standard critiques of Lewis's theory: (i) **laws lack modal impact**; (ii) some **laws transcend the facts** and are, thus, not captured by best-system analyses; (iii) Lewis laws have **no explanatory value**.

4.4 The necessitation account

Some philosophers, notably David Armstrong, who, Lewis reports, has convinced him of the existence of natural properties, were dissatisfied with the absence of any necessitating modal push within laws, and so Armstrong – along with Fred Dretske (1932–2013) and Michael Tooley (see Dretske 1977; Tooley 1977)[14] – developed a theory of laws in which necessitation plays a crucial role.

Materialism Armstrong fits especially well into our overall plot: although he is convinced that necessitation of some anti-Humean kind has to be assumed, he still aims to remain as close to the Empiricist ideal as possible. In what respects? At least in their science affinity and especially also in their later *Materialism* that everything, i.e. chemical, biological, psychological, mental facts, are ultimately constituted by the physical stuff the world is made off (plus its internal relations).[15]

Universals Armstrong and Lewis agree that we have to assume that nature comes, by itself, equipped with natural properties which 'carve nature at its joints' (Plato's *Phaedrus*: 265d–266a). Without this assumption, both Davids claim, it is impossible to construct a theory of lawhood. Armstrong also shares with Lewis the view that properties are non-dispositional Quiddities (Sections 2.2.3, 2.2.4). There is a difference that remains between them, though. Lewis still sympathises with the Empiricist idea that properties are no more than classes of things that might resemble each other in a certain respect. Sure, some such classes are perfectly natural while others are not, yet other than being perfectly natural there need not exist anything over and above these classes.[16]

Armstrong, however, believes that there is more to a property than classes of (alike) entities. He thinks that properties like *being an electron*, *being charged*, etc. are entities in their own right, over and above sets of individual objects that have or instantiate them. So, two kinds of things exist, the property and the class of things that have this property. We might want to think of the property as the respect in which the objects resemble each other and then reify: the respects of resemblance exist as much as the things that resemble each other in these respects.

This is, by the way, the old idea of *universals* as we find it, for example, in disputes amongst Plato (428/427–348/347 BC) and Aristotle (384–22 BC). *Universals* is the name philosophers have given properties when they are conceived of as being more than just classes of things.[17] Trying to be as down to earth (i.e. as Naturalistic or Materialistic) as possible with his assumption of universals, Armstrong favours Aristotle's theory: properties, although being more than classes of things, still only exist when things have them, i.e. in so far as they are instantiated *in re*, in these things. Plato was metaphysically more daring in this respect and had universals exist *ante rem*, i.e. prior to or independently of the things that can have them. Shunning an empirically unreachable, abstract Platonic heaven, where these universals would have their habitat, Armstrong goes broadly with Aristotle's conception of them being located in our concrete space-time. Universals 'have no existence, except in [their] instantiations' (Armstrong 1997: 506).[18]

Laws of nature For Armstrong, a *law of nature* is, now, a special relation, called *nomological necessitation* (abbreviated, *N*) that holds between universals. Saying that '*electrons are negatively charged* is a law' is saying that nomological necessitation holds between the universal *being an electron* (*E*, say) and the universal *being charged* (*C*, say): *N* (*E*, *C*). *N* is a second-order relational universal, connecting universals rather than things. Still, *N* belongs as much to the inventory of the world as other universals do. *N* belongs, just as the other natural properties, to nature's basic inventory.

Before we go further into Armstrong's theory, note the entire shift from the Empiricist's (and also Lewis's) focus on lawlike *statements* to worldly (second-order) states of affairs. We speak directly about how things are, sparing the detour via language.

Advantages What's the advantage of Armstrong's view? First of all, it aims to meet the intuition that what the laws demand must be the case: if *N* holds between, say, the two universals *F* and *G* i.e. *N(F, G)*, then any object, *a*, that has property *F*, must have *G* as well. Or, if we are dealing with a causal succession law, that *a* is *F* will *necessitate* or *bring about* that it (or something else) is or becomes *G*.

Second, we can see how the golden-spheres statement does not express a law: the property of being a golden sphere of that diameter, so we are asked to assume, does simply not stand in the nomological relation *N* to the property of bursting apart.

Third, why could the uninstantiated reaction of *X*- and *Y-types* particles be law governed even if no actual *X* ever meets an actual *Y* (the example we adopted from Tooley above)? Because the universals *X* and *Y* might well stand in nomological necessitation to some kind of reaction *R*, *N(X and Y, R)*, so that, if an *x* were to meet a *y*, *r* would be necessitated.

Problems for the Armstrongian N. What's the catch? For one thing, we have certainly entered anti-Humean, anti-Empiricist territory: a necessitating connection in nature is granted existence and this goes strongly against Humeanism.

There's a practical and a deep problem here. The practical one is: granted *N*'s existence, how do we know between which universals nomological necessitation holds? Armstrong's answer is simple: the normal conduct of science is a very good path. Scientific laws, i.e. those laws science postulates and works with, have a good chance to coincide (ultimately) with the true laws of nature. This is not at all unlike Lewis's view on the epistemology of law discovery.

The deeper problem is to establish the existence and nature of *N* in the first place. Armstrong agrees entirely with the Empiricists that there is no *a priori* route here and he shares with them the suspicion that no direct sense perception can be had of the second order universal *N*. (There is a twist, however, on observability to which we come soon).

Inference to the best explanation (IBE) Is there a third way, another method to establish or, at least, justify the belief in the existence of *N*? Armstrong thinks there is:

> I hold [. . .] that the law involves an extra thing, some further state of affairs. [. . .] The postulation of that extra thing [i.e. N, nomological necessitation; MS] is a case of *inference to the best explanation*.
>
> (Armstrong 1983: 55; also see 52–3, emphasis added)

What's that, an *inference to the best explanation*?[19] Here is, first, an example from the practice of the natural sciences of such an inference: when, in 1920, Adams and Leverrier discovered that Uranus's orbit diverged from what Newtonian theory predicted they postulated the existence of a further planet (which they called 'Neptune'), the gravitation of which was supposed to disturb Uranus. And, indeed, this planet was subsequently made visible by stronger telescopes. The postulation of Uranus was, before the empirical verification of its existence, simply *the best explanation* for the disturbance of Uranus's orbit and, therefore, its existence was assumed ('inferred').

The general argument structure for such a kind of IBE is roughly this: if a phenomenon *P* needs an explanation and the best (available) explanation of *P* needs to presuppose or postulate that some entity exists (an entity which crucially figures in the explanation) then accept the explanation and assume the reality of the respective entity.

Now, as seen in the quote above, Armstrong invokes an IBE argument within metaphysics. For what is Armstrong's *N* a better explanation than any other attempt? The answer brings us back to the alleged advantages of *N* we have listed above. Armstrong believes that his theory of laws *explains best* our intuition that what the laws demand must be the case. Also, as a byproduct, Reichenbach's golden-spheres problem is solved and, finally, *N* can make plausible how the uninstantiated reaction *R* of *X* and *Y* particles can be law governed even if no *X* ever meets a *Y*. Thus, the best theory of laws assumes the existence of nomological necessitation, so, via IBE, we should believe that N does indeed exist. Thus, we have found a potential answer to the deeper problem from above how we can argue for the existence of N even if it cannot be observed and there is no *a priori* access.

Let us grant, for now, that Armstrong's inference indeed goes through but keep in mind that, in Chapter 7, on meta-metaphysics, especially in Section 7.3, we will show how IBEs can fail in science and why they are even more contentious within metaphysics.

The identification and the inference problem Now that we (provisionally) believe in its existence, we can still ask further what exactly the nature of nomological necessitation is. This has been called the *identification problem*. Also, we might want to know how it manages to do what it allegedly does, namely bring about the actual happenings in the world: how $N(F, G)$ makes it the case that *all Fs are Gs*. This is known under the name *inference problem*.[20]

Here's a way to spell out the inference problem[21]. Let us grand that $N(F, G)$ is the case and that something, a, is F. How does $N(F, G)$ manage to make a also G? One way to approach an answer is to parallel it to logical inferences: we get from Fa and $N(F, G)$ to Ga just as we get, per *modus ponens*, from p and $p \rightarrow q$ to q.

Yet *modus ponens* 'works' because the material implication in $p \rightarrow q$ is defined in such a way that, in case of the truth of p and also $p \rightarrow q$, q must also be true. Therefore, if you want 'F, $N(F, G)$, so G' to be like *modus ponens* you ought to provide us with a definition or semantics of $N(F, G)$ that makes G follow *modus-ponens-esque* in a similar way from $N(F, G)$ and F. The most natural way to try to do that is to stipulate that $N(F, G)$ itself *logically implies* that $\forall x\ (Fx \rightarrow Gx)$, so that we can indeed logically derive Ga from Fa via $\forall x(Fx \rightarrow Gx)$ which is contained in $N(F, G)$.

Yet this is problematic. First, you would beg the initial question. You would explain why N(F, G) produces the regularities by pure stipulation. Second, even if we granted you this move you would be in danger of collapsing the theory into a regularity theory for now it seems as if $\forall x(Fx \rightarrow Gx)$ would do the job to make Ga true when Fa is true. Thus, $N(F, G)$ is in danger of turning into an idle wheel.

Instead of (logical) inference similar to *modus ponens* we could postulate a worldly mechanism: for any individual F, $N(F, G)$ instantiates as a kind of mechanism which operates on that particular F in order to reliably produce a G. Note aside that by identifying such a mechanism the *identification problem* – what's the true nature of N? – could also be answered. Yet at least three questions ask themselves about such a postulated mechanism: (i) Why does it operate on every *F*? We'd need an explanation for this regularity. (ii) Why is that mechanism, once it's in operation, so reliably producing *G*? (iii) What's the secret inner life of that mechanism?

Nomological necessitation and causation Armstrong's own suggestion for such a mechanism leads us to the concept of causation. His idea is, roughly, this: nomological necessitation, this law making second-order relation between universals, manifests itself in the world as causation and, also, causation is, contra Hume, occasionally observable. Thus, we sometimes have experiential contact to nomological necessitation via causation and it is via causation that nomological necessitation drives the goings on in the world. All this needs further explanation.

If nomological necessitation, *N*, holds between two universals, *F* and *G*, say, and an actual object instantiates *F* then *N(F, G)* is also instantiated, namely, as the causing of *G*. In other words, while the two universals *F* and *G* are instantiated as the particulars' actual property tokens, the nomological necessitation between *F* and *G*, *N(F, G)*, is instantiated as causation: 'causation is nothing but the instantiation of a law' (Armstrong 1997: 506–7; see also Armstrong and Heathcote 1991). Causation is, so to speak, the incarnation of nomological necessitation, father and son.[22]

What has just been said is clearly very anti-Humean. Remember how Hume argued that causation is not a real power, not some force in the world that would be producing the goings on in nature. Here we have Armstrong saying just this and he pushes this line even further. Causation is, according to Armstrong, occasionally observable:

> The dyadic predicate 'causes' is as much an observational predicate as any other predicate in our language, especially in such cases as our awareness of pressure on our own body.
>
> (Armstrong 1997: 228)

So, especially when we ourselves (our bodies, that is) are involved in some causal chain and feel certain physical constraints we feel, as causation, the instantiation of nomological necessitation.

We can interpret Armstrong's 'nomological necessitation is instantiated as causation' theory as a partial answer to (i) to (iii): the mechanism is causation and we are directly acquainted with causation when it actuates as pushes and pulls on our body. Bodily experiences are meant to throw some light on the secret behind (iii) and to solve the identification problem, too. Also, when we are pushed hard enough we know well that we inevitably fall. Maybe this answers (ii). Only (i), why the mechanism/causation reliably kicks in (for any *F*) remains questionable.[23]

In summary, there are two lines of argument Armstrong offers for *nomological necessitation* which are mutually supporting: first, an inference to the best explanation for the existence of the second order universal *N* and, second, an observability thesis regarding *N*'s instantiation as causation. This, too, is also meant to answer both problems van Fraassen raised: the *identification problem – N* is, in the end, causation and with causation we have direct experiential acquaintance; the *inference problem – N(F, G)* brings about that any *F* is a *G* because with each *F*, *N(F, G)* is instantiated as the causation of *G*.

Necessitation: production or modality? Remember a distinction we made in Section 2.1.3 regarding dispositions: there, we highlighted the difference between, (PROD), responsibility for production, and (MOD), modality. We might guess that Armstrong's *N* has little to do with the second and everything with the first. The closeness of *N* to (PROD) is obvious in that Armstrong identifies instances of *N* with causation.[24] In order to see how it differs from (MOD) it helps to point out that Armstrong assumes it to be a contingent matter which universals stand in the nomological relation *N*: other possible worlds might have totally different laws than ours or even none at all. That is *N* might, there, bind different universals than here – $N(H, J)$, for example, instead of $N(F, G)$. In some lawless worlds, no universals might be *N* related whatsoever (see Armstrong 1983: 158–71). In other words, in accordance with the general intuition we mentioned at the beginning of this chapter (Section 4.1), it is not (metaphysically) necessary that *N(F, G)* holds: $\neg\Box\, N(F, G)$.[25]

Whether you go with Armstrong or not, *N*'s affiliation with causal production (PROD) and its distance to (metaphysical) necessity (MOD) reveals that within the cascade of concepts Hume rejected while thinking they were synonymous ('efficacy, agency, power, force, energy, necessity, connexion, and productive quality' (Hume 1739–40: Book I, Part III, Sect. XIV: 157)) there is much leeway for distinctions: here between efficacy, force, production on the one, necessity on the other side. Be that as it may, one should not confuse Armstrong's necessitation, *N*, with (metaphysical) necessity.

Necessitation and nomological necessity Despite what we have just said about the differences between necessitation and (metaphysical) necessity we can, similarly as for Lewis's laws (see van Fraassen's suggestion (Section 4.3)), define a weaker kind of (modal) necessity such that Armstrong's laws are necessary in this respect. Restrict the set of possible worlds to be considered in the following way: look at all and only those worlds that share their laws with ours, i.e. our *N* relations. Then define the weaker *nomological necessity* as truth in all and only these possible worlds. Consequently, the laws themselves come out as *nomologically necessary* (they hold in these worlds by definition) while they are still *metaphysical contingent.*[26]

General critique Although Armstrong is in many respects fairly close to Humeanism (see his defence of Quidditism and the contingency of laws) some radical Humeans will be sceptical about Armstrong's *N*.

Whether one is a Humean or not, there are aspects of Armstrong's lawmaker *N* one can criticise. (i) There are non-causal laws next to the causal ones: Pauli's exclusion principle, the conservation laws or simply that electrons are negatively charged, for example. Yet if *N* always instantiates as causation we have to conceive also of those laws as causal. (ii) There are laws (or derived truths) which seem to be causal, yet they do not, it seems, necessitate their effects but only make them likely: 'Smoking causes cancer' or 'Unrestricted speculations on the stock market cause financial crises' being examples.

Armstrong has reacted to (i) in saying that he construes causation broadly so that any kind of *making it the case that* counts as causal even if there is no, now narrowly conceived, causal event happening (like in the electron/charge example). Concerning (ii) he has a story to offer about both probabilistic laws (see, for example, Armstrong 1997: 237) and also about so called *proviso* laws (or 'oaken' or 'defeasible laws' as he called them (see, for example, Armstrong 1997: 231) that hold only when the circumstances are exactly right. We cannot get further into these issues here (but see our section on *ceteris paribus* laws in Section 6.2); for a critique of Armstrong's solution see Schrenk (2010).

BOX 4.4 The necessitation account

- While Lewis throws out a best-systems net in order to capture the laws, Armstrong (also Tooley and Dretske) seeks to anchor those regularities which deserve law status in deeper grounds: in **nomological necessitation**, *N*.
- ***N* is a second-order relational universal that binds first order universals**. 'Electrons are negatively charged' is a correct law statement if and only if being an electron and being negatively charged stand in the *N* relation.
- When, say, *N* holds between the universals *F* and *G* then, whenever something *x* has *F*, ***N* is instantiated as well, namely, as the causation of** *G*, so that it is true of all *x* that are *F* that they are also *G*.

- In other words, on Armstrong's nomological necessitation account of laws the lawlike **regularities are produced/necessitated** by the *N* relations.
- Armstrong neither ignores **probabilistic laws** nor so called **ceteris paribus laws** where *G* might not come about despite *F*'s being instantiated but we have skipped this part of his theory for the sake of brevity.

4.5 The counterfactuals account

Laws and counterfactuals stand in a close relation. A law, so we have suggested so far, supports counterfactuals whereas an accidentally true regularity does not: that *Fs are Gs* is a law makes true the counterfactual that if this *x* were an *F* it would also be a *G*. However, should all *Fs* so far have been *Gs* only accidentally, by pure chance, then there is no reason to believe that if something else were an *F* it would also be a *G*.

The relation in the direction from laws to counterfactuals – laws first, then counterfactuals – has been endorsed by all protagonists we encountered so far: Goodman in Section 3.2, Lewis in Sections 3.4 and 4.3 and Armstrong in Section 4.4. It can be turned on its head, though: instead of saying that laws support counterfactuals we may reverse the order and say that the laws (but not the accidents) amongst the regularities are those generalities that are *stable or invariant under counterfactual perturbations*.

'Come what may', we might say, 'the regularities that are laws will remain the same, they won't be false, ever'. Try as we might, we won't accelerate anything to a velocity higher than the speed of light. This is clearly not correct of accidentally true regularities. Those, it seems, we could (easily) make false. If you put some Pound coins into your pocket where, so far, have been only Euro coins, this, regrettably, won't turn Pounds into Euros, but would instead make the generalisation 'all coins in my pocket are Euros' false.

Marc Lange, who has introduced and developed such an account of laws (Lange 2000, 2005, 2009),[27] offers a nice example that captures this intuition:

> The laws of nature would surely have been no different had Jones missed his bus to work this morning. [. . .] But had Jones missed his bus to work, then various accidental truths, such as his perfect attendance record, would not still have held.
>
> (Lange 2005: 415)

More abstractly, Lange says about the laws (as opposed to the accidents) that they 'would still have held had q been the case, for [almost] any q' (Lange 2005: 415). Admittedly, the *almost* I have inserted here is vague. One of our main tasks will be to make it more precise.

Counterfactuals first Let's get into the details. Take the set of all facts about our world including all general regularities, lawful and accidental, and other facts, like that you are reading this book right now, that I have written it, etc. Now, Lange wishes to find a method to single out within that large set, call it Σ, those smaller subsets which contain only laws. The set Σ is, thus, the basic material to work on and the law-finding mechanism is our next topic. (We are, like gold diggers, panning for the nuggets in Σ).

What is second, the right mechanism, criterion, or testing schema, *TS*, to separate the laws in Σ from the accidents (see Lange 2009a: ix)? We have already introduced the basic idea: subsets of Σ that contain only law statements (as opposed to subsets that contain accidents) 'would still have held had q been the case, for [almost] any q' (Lange 2005: 415).

Before we continue, note something important: we have spoken of a *testing schema*. This phrasing sounds very epistemic, as if it were us searching for the laws, but this is a mere *façon de parler*: Lange is very clear that this *testing* is an ontological matter. It's not how we ascertain the laws but what *makes them laws*. However, since speaking of *testing* is a convenient way to put things, we continue to allow ourselves this language.

Of course, in order to perform that *test* on our material Σ, we have to know from the outset which counterfactuals of the general kind 'p would still have held, had q been the case' are true. Thus, third, we have to suppose with Lange that, *CT*, all counterfactual truths are fixed from the outset. (In fact, these counterfactual truths are fundamental building blocks in Lange's ontology. We will highlight this point later again). That is, we assume that there is a complete predetermined list of all true counterfactual claims: 'if the earth had been slightly more massive the moon would have a different orbit', 'if you had jumped you'd have fallen', and so forth for any q and p and corresponding $q \mathrel{\Box\!\!\rightarrow} p$.[28] Abstractly speaking, for all p_i and all q_j we either have the counterfactual $q_j \rightarrow p_i$ or $q_j \rightarrow p_i$ depending which of the two is true. Later, those p_i will be said to be stable for which there is no q_j and related true counterfactual with a negated p_i in the consequent.

In sum, we have three basic ingredients to start with: the set Σ of all facts whatsoever amongst which we want to find the laws. This is done via, *TS*, the testing schema: for any p, see whether there is a q and a true counterfactual $q \mathrel{\Box\!\!\rightarrow} \neg p$. If so, p is unstable because, as that counterfactual tells us, were q the case p would no longer be the case. Thus, p would not be a law. Yet if there is no such true counterfactual with a negated p as consequent then *p* is a law for come what may (for any *q*) *p* remains true. Now, clearly, for *TS* to be possible, it has to be clear from the outset which counterfactuals, *CT*, are true.

The set of all sub-nomic truths On each of Σ, *TS*, and *CT* we need to be more precise. Let us start with Σ. What exactly is Σ's content? For a start: nothing but the truth. Σ contains no falsehoods. Also, all of Σ's entries are written in pure factual form ('all uranium spheres are. . . ', '$E = mc^2$', etc.). Lange speaks here of 'sub-nomic claims' (Lange 2009a: 17–18).[29] That is, its statements do not say things like: '*It is a law that* all uranium spheres are. . .', '*It is nomologically necessary* that $E = mc^2$', etc. Again: '$E = mc^2$' is a mere fact stating the sub-nomic claim that is in Σ; '*It is nomologically necessary* that $E = mc^2$' is a meta-claim about the nomic status of the former sub-nomic claim which is not in Σ.[30]

We will now have a closer look at *TS*, the testing schema. Which *q*s are eligible for this test? (As we will see, there is a type of *q*s for which the test outcome would not tell the laws from the accidents but would treat them alike).

(i) Suppose, we had no restrictions whatsoever on *q*. Suppose especially that we were allowed to take even those *q*s that *logically contradict* the general statements in our world description. Would these unrestricted counterfactual variations unearth a difference between laws and accidents? Surely not.

Take, for example, the (intuitively accidental) general statement *s* that all of the apples on a tree in Marc Lange's garden are ripe (my knowledge of Lange's garden stems from himself: (Lange 2005: 415)). Clearly, if we take *q* to be 'there is an unripe apple on Lange's tree' statement *s* fails the test, for 'had there been an unripe apple on Lange's tree still all apples had been ripe' is clearly false. The apple regularity is not stable under these kinds of directly contradictory counterfactual perturbation. Yet neither are stable what we intuitively believe to be laws. They would share the accidents' fate: if we allow a counterfactual assumption, *q*, that *logically contradicts* (potential) laws then even the firmest laws break: 'if this specific mass m_0 corresponded to an energy of m_0c^4 then still $E = mc^2$ would hold for any mass' is false. Even the (potential) laws do not withstand such a counterfactual test.

In other words, the stableness we are interested in is not really stability *come absolutely whatever may be* but only stability under a wide range of counterfactual/subjunctive suppositions. Those *q*s (as counterfactual assumptions) that logically contradict members of the set of statements to be tested clearly make both accidents and laws within that set break alike. We would not gain a discriminating factor between the two. So we exclude these kind of *q*s from our testing.

(ii) Let's look at a second class of possible *q*s (a weird one, admittedly): those states of affairs Bill Gates desires (see Lange 2009a: 8). Now, consider two counterfactual tests (one for an alleged law, one for an accident): '(Even) if Bill Gates wanted a uranium sphere bigger than 10m diameter, all uranium spheres would still be smaller' and 'If Bill Gates wanted a sphere of gold with a diameter bigger than 10m, all lumps of gold would still be smaller'. The first, it seems, is true. The *law* in its consequent part is stable: no matter how rich or skilful Gates is there would not be such a sphere. This is unlike the second counterfactual: Gates is rich enough to buy that much gold and nothing would prevent him from welding his aureate ball. So, the second counterfactual, being false, shows that the golden-sphere regularity would break.

So far so good. There seems to be some difference between laws and accidents if we choose the Bill-Gates-desires class of *q*s. Yet this weird class of *q*s is not sufficient to tell all accidents from laws. This is so, simply because there are some accidents even Bill Gates could not change. Let, for the sake of the argument, 'all galaxies contain at least one gold atom' be a true accident. Rich as Gates might be, he cannot do anything about this. And, so, under our current test, this accident – stable as it is under whatever Gates desires – comes wrongfully out as a law. Thus, we still do not have a really satisfactory means to discriminate laws and accidents. (Besides, it would be strange indeed if the metaphysics of laws depended on Microsoft.)

(iii) We could now be tempted to exchange Bill Gates for future generations of scientists which have all knowledge there is to be had about the world and whose technology is so potent that they can really do anything within the limits of the physical laws. Then, it seems, whatever they do, the laws could still not be broken while all accidents are now unstable (even the galaxy/gold atom regularity). Good as the suggestion sounds, it is not feasible: it is, unfortunately, plainly circular! It might kick out all the accidents but only because we smuggled in reference to something we want to explain, namely the laws, for saying 'that the future generation *can really do anything within the limits of the physical laws*' amounts to saying that the laws are such facts that would still have held had *q* been the case for 'any q that is [logically] consistent with the *laws*' (Lange 2005: 415, emphasis added).

How, then, can we keep the general intuition – laws are those regularities that are stable under counterfactual perturbations, accidents are not – and yet avoid the circularity? Let us go back to suggestion (i). It had no restrictions whatsoever on *q*. Especially also those *q*s that *logically contradict* general statements were allowed. This was not successful. Yet maybe all we need to do in order to turn this into a success story is to say that all *q*s shall be allowed as counterfactual probes that are *logically consistent* with our set. So, the only restriction on *q*s is *logical consistency* with the entries we wish to test. This avoids the shortcoming of (i) and it widens the too narrow (and anthropocentric) (ii). Yet it does so in such a way that no circularity occurs like in (iii). This is in fact Lange's solution!

We have figured out now which counterfactual variations, q, are admissible: those that are *logically consistent* with every member of our set of sub-nomic truths to be tested, and we ultimately want to say that sets of laws but not of accidents are stable under these counterfactual perturbations. Another way to put this is to say that when a subset of sub-nomic facts is stable then we have justification to call its members laws. Being unstable, on the contrary, is an indication of the accidentality of at least some member of that subset.[31]

Let us, thus, define *sub-nomic stability* for subsets Γ of Σ in a more formal manner and then see the definition at work:

> A non-empty set of sub-nomic truths Γ is 'stable' exactly when for each member p of Γ the following is true: take any q that is *logically consistent* with every member of Γ, then even if q had been the case, p would still have been true, i.e.: q $\Box\!\!\rightarrow$ p.
>
> (Lange 2005: 420; 2009a: 29)

Lange's orchard Let's consider another example. Take (again) a couple of sentences that describe Marc Lange's garden, 'All of the apples are ripe' being one of them (and maybe also that the apple tree is to the left of a cherry tree, etc). All else we ask of this non-empty set is that it states only sub-nomic truths, i.e. it does not include judgmental statements like '*It is a biological law* that plants produce oxygen' but it might include the merely descriptive 'Plants produce oxygen'.

How stable is this toy example set, Γ (which is a very small subset of all the sub-nomic truths as in Σ), under the allowed counterfactual suppositions? Consider an assumption, q (where q or $\neg q$ is in $\Sigma\backslash\Gamma$), which is *logically consistent* with all entries in our toy set, namely: 'Had there been fewer sunny days last summer then . . .'. Had this q indeed been the case, i.e. had there indeed been fewer sunny days then we can be sure that not all apples would have ripened. Thus, taking 'there were fewer sunny days last summer' for q and 'all apples on the apple tree are ripe' for p we know that $q \Box\!\!\rightarrow p$ *is false* (instead: $q \Box\!\!\rightarrow \neg p$ is true) (see Lange 2005: 420). Therefore, the set of truths about Marc's garden, Γ, including this unstable p, has failed the examination. The set of sub-nomic facts about his orchard isn't stable (according to the definition of stability above) and is, thus, no set of laws (see Lange 2005: 419).

Two trivially stable sets We have just seen a stability-test in action, namely on a set that proved to be unstable. Are there stable sets at all?

Yes! Here are at least two of them: (i), the set of *all* sub-nomic truths together, i.e. Σ itself (see Lange 2005: 422) and, (ii), the set of all logical truths.

(i) Why is Σ stable? Because any counter to the facts q does logically contradict some entry of that set for it does contain absolutely all facts. Again, it being a complete set of facts any counter to the facts q must logically contradict one member. So, here, stability is voidly satisfied: there simply is no counterfactual q that is logically consistent with the set of all sub-nomic truths.[32]

(ii) And why is the set of all logical truths stable? This is equally trivial: for any q that is consistent with the logical truths it holds that were it the case then any logical truth would still hold. Here, too, the stability test is fulfilled. The elite class of logical truths, we should highlight, belongs as a subset to all truths, Σ, together.

Here's an interesting third observation, (iii), about some of the intermediate sets (somewhere in the middle between Σ, as in (i), and the set of logical truths, as in (ii)) that contain both accidents and laws. Clearly, the accidents in such a set are not stable. However, you might want to try to stabilise them, i.e. to immunise the accidents within by taking other truths on board. Lange (2009a: 32–6) gives the following example: suppose a set contains the accidental truth, g, that whenever the gas pedal of a certain car is depressed by x inches and the car is on a dry, flat road, then the car's acceleration is given by $a(x)$. This is, per se, unstable because g might not still have held had the engine contained six cylinders instead of four (this is a counterfactual test which g fails). Now, here's a trick to make g stable after all: add to the set a description of the car's engine. Now, you are no longer allowed to launch the above counterfactual against g because its antecedent would logically contradict that newly added fact.

The problem with this stabilisation attempt of statement g is that it can only be gained on the costs of a new unstable member within the set, namely the very description of the car's engine:

> But now to be stable, the set must also include a description of the engine factory, since had the factory been different, the engine might have been different.
>
> (Lange 2009a: 32)

Lange believes that this cascade would go on and on – you stabilise one statement but get another unstable one – until you reach the maximal set Σ. This leads to the following consequence regarding sets with accidents:

> no set containing an accidental truth is stable except for the set of all truths, laws and accidents, which is trivially stable [see above].
>
> (Lange 2005: 420)

We can turn this the other way round: if there should exit stable non-maximal sets Γ (i.e. proper subsets of Σ) then we can be sure there's no accident in there.

Λ for law The common-or-garden orchard subset of truths wasn't a stable set (even if it had some laws in it). This was because a change of some fact (the sunlight hours) outside that set lead to changes inside (the ripeness of apples) – and so it was no set of laws. Σ itself (as in (i)) and the set of logical truths, (ii), from above are trivially stable sets. Finally, (iii), we know that no set containing an accident can be stabilised unless you make the maximal set out of it.

Here comes the crucial question: are there any stable sets somewhere in the middle of the trivially stable stets Σ and the logical truths? Stable sets which contain laws only? Yes, there are, says Lange: there is one largest non-maximal set, call it Λ (*lambda* for *law*), and there is, within Λ, a single nested hierarchy of Λ subsets. That is, we have a (smallest) such stable set Λ_0 which is contained in a bigger stable set Λ_1, which is contained in a yet bigger set Λ_2, etc. till we reach Λ. We will leave the proof (see Lange 2009a 37–42) that Λ and the nested hierarchy Λ_1, Λ_2 etc. exist aside and concentrate on the benefits of the result.

Lange can make plausible that the hierarchy of these nested sets corresponds to a hierarchy of strata or levels within the natural laws we intuitively expect from them: the highest level (corresponding to the smallest set, Λ_0) contains the laws of logic only (this is actually the set (ii) from above), one level down in the hierarchy (here: Λ_1) contains also meta-laws of nature like the composition of forces, the conservation laws, etc., and in stratum Λ_2, we have also the individual force laws concerning weak and strong subatomic forces, the electromagnetic force, and gravitation (see Lange 2009a: 49), etc. till we reach some Λ (which is still a proper subset of the maximal set Σ, our set of all subnomic facts).

Here's an example to make this hierarchy plausible: we can imagine that there were more or different individual forces than there actually are. That is, we can imagine, for example, gravitation not to exist but some psychic force instead. Yet at the same time, we could remain faithful to the meta-law of force composition, i.e. we could say that whatever the specific forces are they add up to a resultant force according to that meta-law. In other words, we could counterfactually toy around with Λ_2's first order laws – they might have been different – and leave, at the same time, the meta-laws, Λ_1, intact. Thus, lawhood comes hierarchically ordered in strata with different degrees of stability.

Interlude: modality, especially necessity This hierarchy of sets comes in handy in another respect. Consider *necessity* rather than nomicity. As we know, necessities share aspects with laws. Here are two such features: first, 'Intuitively, "necessity" is an especially strong sort of persistence under counterfactual perturbation' (Lange 2009a: 75); second, laws and necessities also have the hierarchical structure in common:

> The traditional picture of the genuine modalities involves a series of concentric circles: the (narrowly) logical necessities are a proper subset of the conceptual necessities, which are a proper subset of the metaphysical necessities, which are a proper subset of the natural necessities.
>
> (Lange 2009a: 87; Section 2.3)

Thus, Lange moves swiftly from necessities to laws (and vice versa): '*p* is a law in virtue of its necessity, and its necessity is constituted by its membership in a sub-nomically stable set' (Lange 2009a: 89).

Defining what a law is Let us focus on that latter quote for it (almost) defines what lawhood ultimately is in Lange's theory and, after all this preparation work, it is about time that we explicitly write down his final definition. Here are, then, the ultimate fruits of our labour:

> it is a law that *m* if and only if m belongs to at least one nonmaximally sub-nomically stable set [i.e. to one of the Λ_0, Λ_1 . . .].
>
> (Lange 2009a: 42, my addition in brackets)

Laws of logic Per construction, Lange's system is such that the smallest subset Λ_0 of the hierarchy of laws is the set of logical truth. Lange bites this bullet: logical truths 'are "by courtesy" counted amongst the laws' (Lange 2009a: 15). If you hesitate to accept that logic and nomicity should be on a par think of the hierarchy as the hierarchy of necessities (see above): the nomological, the metaphysical, the logical. Necessities – while varying in strength – are one of a kind and Lange's assimilation of logic to nomicity is more palatable.

Dispositions We know that the sub-nomic fundament must not already contain nomological statements like '*It is a law that*. . .' (otherwise the account would be circular). We learn further from Lange that asking of a set to contain only sub-nomic truths means demanding the absence of 'other species of modality', too, including logical ones like 'it is logically necessary that' (see Lange 2009a: 18).

Yet what about *dispositional statements*? Is 'salt is *water soluble*' sub-nomic? Is 'smoking *causes* cancer'? That depends. If 'salt is water soluble' equals saying that salt has a *NaCl* substructure and that it is a *law* that water dipoles tear it apart then such a dispositional statement is *not* sub-nomic because it (tacitly) refers to a law (Sections 2.1.4 and 2.2.2). Thus, it mustn't be contained in Σ, the set of sub-nomic truths.[33]

If, on the other hand, a dispositional statement is shorthand for a counterfactual conditional with sub-nomic antecedent and consequent (here: if salt were put in water then it would dissolve) then dispositional facts might be on a par with counterfactual truths. This, then, amounts more or less to saying that dispositional facts belong to *CT*, the counterfactual truths about our world and not to Σ.

If this is right then the laws would somewhat depend on counterfactual truths and/or dispositions. This would, of course, be an outcome which is remote from, in fact, *contrary* to our very early starting point in this chapter on laws: to find a definition of lawhood that could be used in definitions of dispositions.

The counterfactual facts again Lange's account depends on the prior truths of certain counterfactual/subjunctive facts (see *CT*) and, in fact, Lange believes that these subjunctive truths belong to what the world is made of at its very basis. They (like the facts in Σ)[34] belong to nature's ontological fundament:

> I propose that with these subjunctive facts we have reached ontological bedrock. They are [. . .] primitive, lying at the bottom of the world. They are the lawmakers.
>
> (Lange 2009a: 136)

We might still have the intuitive urge to ask what makes these counterfactuals true. Worse, we might have the feeling that it is the laws that do: 'Were I to try to jump 10m high unaided I still would not succeed' is true, we might wish to say, because of some facts about my body *and, essentially, the law of gravitation*. Yet Lange denies this and he also must deny it: his theory would be circular otherwise. To make this more acceptable Lange points out that epistemic priority does not have to mean ontological priority (see Lange 2009a: 136), i.e. while epistemic subjects might indeed first find out the laws and infer from these the counterfactuals, still, ontologically, the counterfactuals are prior.

It is for future metaphysicians to decide whether this ingenious approach can be defended. Note, that it is certainly an anti-Humean account: the counterfactuals, and maybe dispositions, on which the hierarchy of laws and necessities depend are themselves modal. They exceed the merely factual (and especially what is directly observable).

Woodward's and Roberts's theories We end this section hinting at two theories that have considerable overlap with Lange's. James Woodward had already sketched a similar theory in 1992:

> [I]n actual scientific practice, the notion of lawfulness is closely connected with the notions of stability and invariance. As a rough approximation, we

> may say that a law, in contrast to an accidentally true generalisation, expresses a relationship which not only holds in the actual circumstances but which will remain stable or invariant under some fairly wide range of changes or interventions.
>
> (Woodward 1992: 202; see also his 2003: 236 ff.)

A difference between Lange and Woodward is that for the latter there is no sharp distinction between laws and accidents: the difference is only a matter of degree, namely of the degree of stability. As we have seen, Lange's theory sharply tells laws and accidents apart. Also, where Lange focuses on laws, Woodward, especially in 2003, centres more on causation (Section 5.5).[35]

Also John Roberts has recently presented an intriguing view on laws that has elements not only of a subjunctives theory but also other here introduced theories. Roberts calls his the 'measurability account of laws' (2008: 28–30). It shares, (i), with the Empiricists a closeness to *what can be observed*, here, what can be measured in experimental setups. It has, (ii), a flavour of laws as *inference tickets*: laws being those statements in our theories that entitle to measurement predictions. And, (iii), it takes from Lange the *stability requirement*: '$\forall$Q (if Q is logically consistent with P, then Q $\Box\!\!\rightarrow$ (still) P' (Roberts 2008: 179). The following quote displays all these three elements:

> What it is for someone to regard something as a law of nature [ii] is for it to be one of the principles that guarantee the reliability [iii] of the methods they regard as good measurement methods [i].
>
> (Roberts 2008: 37, my additions in brackets)

Roberts's full account is complex and, with regret, we can't give a proper introduction here.

BOX 4.5 The counterfactuals account

For Marc Lange there are two sets of basic facts, *Σ* and *CT*, and a crucial test schema, *TS*:

- Σ: All **sub-nomic facts** taken together: 'There are n planets in the universe.', 'Everything moves at most as fast as the speed of light', 'The apples in Lange's garden are ripe', etc.
- *CT*: All **counterfactual/subjunctive facts**: 'If I had jumped I'd have fallen', 'If this had been an electron (instead of a photon) it would have moved slower', 'If there had been less sunny days, the apples would not have been ripe', etc.

- It is a **law of nature** that *m* if and only if *m* belongs to stable subsets (of the set Σ of all subnomic facts).
  When is a set stable?
- *TS*: We test whether **a set of sub-nomic facts is stable** in that we ask whether its entries break, i.e. become false **under counterfactual perturbations**. That is, we test for any entry *p* in the set whether it 'would still have held had q been the case, for [almost] any q' (Lange 2005: 415).
- There are many stable sets, Lange claims, of which two are **trivially stable**: the **set of all sub-nomic truths**, Σ, and the **set of logical truths**, Λ_0. Right in the middle is a nested hierarchy of stable sets of more and less fundamental sets of laws, Λ_1, Λ_2. . .
- Differently strong **modalities**, specifically **necessities**, (natural, metaphysical, logical) correspond to these nested sets. Still, because of the hierarchy and, correspondingly, because stability comes in degrees, (most) laws could have been different. Only the laws of logic are super-stable.
- Lange's theory is **anti-Humean** because it depends on modally non-innocent **counterfactuals** which are said to belong to the fundamental building blocks of the world.
- **Dispositions** could either turn out to be on a par with the fundamental counterfactual facts (in which case laws depend on them) or they are analysable in terms of laws. Lange does not commit himself to either possibility.

4.6 Laws as unanalysably fundamental

There is a philosophical position that takes *laws of nature* to be *ontological primitives*. Accordingly, laws cannot be reduced to anything that is more fundamental. John W. Carroll (Carroll 1994) and Tim Maudlin (Maudlin 2007: 5–49) endorse such a view. This seems, at first glance, to be a lazy option, especially in comparison to the intense labour that has gone into the attempts sketched in Sections 4.2–4.5. Maudlin says to his defence:

> Taking laws as primitive may appear to be simple surrender in the face of a philosophical puzzle. But every account must have primitives. The account must be judged on the clarity of the inferences that the primitives warrant and on the degree of systematization they reveal among our pre-analytic inferences.
>
> (Maudlin 2007: 15)

This sounds right and, indeed, the attraction of Maudlin's account lies not so much in the characterisation of what a law is – and how could it when that notion is primitive – but rather in what can be done with such an unanalysed concept when we make use of it for other purposes – for example, for accounts of counterfactual conditionals, causation, dispositions, etc.

The starting point for Maudlin is a provocative view[36] about how metaphysics of science ought to be done in general: 'Metaphysics', so he tells us, 'can do no better than reflect on physics', at least 'in so far as it is concerned with the natural world. [. . .] The philosopher's proper task is the interpretation and elucidation of those theories' (Maudlin 2007: 1); for more on such meta-philosophical considerations see Chapter 7).[37]

Laws as bedrock Now, nothing in these scientific theories and practices, Maudlin continues, suggests that laws of nature should be analysed any further. Rather, (i), '[laws] ought to be posited as ontological bedrock' (Maudlin 2007: 1) and, (ii), 'the notion of a law cannot be reduced to other more primitive notions' (Maudlin 2007: 18).

Maudlin does, of course, not stop here. Despite the ontological and also conceptual fundamentality of laws something more can be said about their nature. One of the first notable observations Maudlin makes is that laws (or the corresponding statements) as they appear in the sciences are hardly ever in the syntactical form $\forall x\ (Fx \supset Gx)$ philosophers so much love. He concedes, for example, that Schrödinger's equation $i\hbar(\partial\psi/\partial t) = H\psi$ could be 'tortured' into the former form (see Maudlin 2007: 11) but he doubts that we'd gain anything much by doing so.

Remember that the universally quantified form captures the belief that laws hold everywhere under all conditions. This, however, is something Maudlin thinks is no requirement of lawhood (see Maudlin 2007: 12). It is imaginable that laws hold in some areas of space at some time but not at others or that they hold only under very specific circumstances. The latter is certainly true of many special science laws like the biological laws of evolution or even of physical laws that are restricted to certain conditions (as the pendulum law $T = 2\pi\ (L/g)^{1/2}$ is which holds only when the amplitude is limited to small swings).

Flotes, slotes and lotes Rather than universality (which he rejects), Maudlin takes a different feature a huge class of laws share as central: they all govern or generate a system's or particle's evolution through time. Or, if we do not focus on physics, how objects in a broader sense, e.g. plants, ecosystems, etc., evolve, at least under normal conditions. The former, physical laws, Maudlin calls 'FLOTEs': '*F*undamental *L*aws *o*f *T*emporal *E*volution'; the latter, i.e. the non-fundamental variant, 'LOTEs' (or 'SLOTEs' when non-fundamental *S*pecial sciences are meant). These three descriptive names indicate a further point Maudlin thinks can be made about laws: (S)LOTEs are parasitic on or true in virtue of the former *fundamental* laws, FLOTEs, of temporal evolution and (S)LOTEs are, also, more likely to meet exceptions should something interfere (see Maudlin 2007: 12–15).

We can, as an intermediate summary, collect three overall attitudes Maudlin has regarding laws: (i) he is a *Realist* (laws objectively exist, independent of us), (ii) he is *Primitivist* (laws are ontologically basic and unanalysable) and (iii) he is a *Physicalist*: 'The only objective primitive laws I believe in are the laws of physics' (Maudlin 2007: 157–8).

Here's a small critical point: Maudlin seems to have in mind only diachronic developments – temporal evolution – when he theorises about laws. What, though, with *synchronic* laws of coexistence, such as the ideal gas laws, Pauli's exclusion principle or laws about how properties cluster (for example, that electrons always have negative unit charge). Maudlin's brief answer is that they can, somehow, be construed as consequences of dynamical laws. Yet he leaves this task to others (see Maudlin 2007: 13).

Related to this diachronic business just mentioned and, moreover, to the fact that Maudlin's laws are meant to be the governing or generating force behind a system's or particle's evolution through time we can, once again, make a connection to what we have extracted from dispositions in our early discussions: (MOD) and (PROD) (Sections 2.1.2, 2.1.3) Maudlin's laws are clearly on the *responsibility for productivity* (PROD) side!

Given that the temporal laws are realistically conceived as generators or producers of later states, Maudlin is decisively on anti-Humean and anti-Lewisean grounds. Indeed, he explicitly writes: 'The "necessary connection" that Hume sought at the heart of causation is nomic necessity' (Maudlin 2007: 145), i.e. the laws' generating and producing force (PROD).

Laws, causation, counterfactual conditionals and dispositions We have made it an implicit rule in this book not to mix topics extensively within our sections. When we talk about laws we talk about laws, when about causation about causation, etc. The nature of Maudlin's Primitivism about laws makes it, however, desirable to indicate, if briefly, how accounts of other central notions of the metaphysics of science arise from his theory. It is only this way that we can appreciate the merits of his endeavour.

Clearly, causation, counterfactual conditionals, and dispositions belong together. For example, that something caused something else suggests that the counterfactual 'had the first not happened so would not have the second' is true. Yet according to Maudlin, despite the correlation of the two (here causation and counterfactuals) they cannot be reduced to one another. Rather, their correlation is explained by a common third factor: the laws! Laws are, so to speak, 'a common cause' of counterfactuals, causation, and dispositions (see Maudlin 2007: 144–5). So, let's see how they all individually depend on laws.

Causation We start with *causation*. For Maudlin, Newton's first and second laws of motion are important blueprints for his theory of causation:

> The First Law, the so-called *Law of Inertia*, states that a body at rest will remain at rest and a body in motion will continue in motion at a uniform speed in a straight line, unless some force is put on it. [. . .]
> The second law then specifies how the state of motion of an object will *change* if a force is put on it.
>
> (Maudlin 2007: 155)

Maudlin believes that there are other such pairs of laws. He calls them *quasi*-Newtonian. They come divided into *inertial laws* which describe how a system behaves when nothing acts upon them, and *laws of deviation* which specify how the systems behaves when disturbed.

In terms of Newtonian or *quasi*-Newtonian laws, a *cause of an effect* is then analysable as the disturbance of an inertial system. Suppose two particles, *P* and *Q*, collide. *P* pushes the formerly resting *Q* away. The cause, here *P* colliding with *Q*, is what stirred up, what changed the inactivity (or uniform motion) of *Q*, i.e. the effect. Causes are, thus, law governed disturbances of inertial behaviour and effects the resulting deviations thereof.[38]

Here's a difficulty for this theory of causation: not all spheres of the world are governed by laws that come in these neat and useful (quasi-)Newtonian pairs. Only some laws (or classes of laws) are *quasi*-Newtonian. Which ones are? Well, maybe surprisingly, Maudlin believes that the laws of the special sciences, i.e. biology, chemistry, etc. (including classical mechanics) are:

> The special sciences and plain common sense as well, will seek to carve up the physical world into parts which can, fairly reliably, be described as having inertial states (or inertial motions) which can be expected to obtain unless some specifiable sort of interference or interaction occurs. [. . .]
>
> We [. . .] understand much of human biology in quasi-Newtonian terms. The inertial state of a living body is, in our usual conception of things, to remain living: that is why coroners are supposed to find a 'cause of death' to put on a death certificate.
>
> (Maudlin 2007: 159)

Suppose that this is right. Then we also have a first theory of causation for special science events and, too, for the goings on in our common environment.

Yet, second, what about non-Newtonian fundamental physics such as quantum mechanics? Regarding these fundamental physical laws, Maudlin confesses, he has little faith that they could turn out to be *quasi*-Newtonian, i.e. divisible into pairs of inertial laws and laws of deviation. What's a cause, what an effect there then? Maudlin's fall-back option is to say the following: for fundamental micro-events on the fundamental level of the world,

> the entire antecedent state of the world is the cause [. . .]. That being the minimum information from which, together with the laws of physics, the event can be predicted.
>
> (Maudlin 2007: 168)

Ontologically speaking, the picture Maudlin draws here is this: given a certain antecedent state of the universe, 'the laws of temporal evolution operate, whether deterministically or stochastically, from that initial state to generate or produce later states' (Maudlin 2007: 174) and amongst the later states the micro event we have in mind as effect can be found.

Is Maudlin's theory of causation dualistic, then? For there are two *conceptual ways* in which he explicates causation: one for the special sciences and our everyday world, and another for fundamental physics and its laws. Well, remember his Reductionism regarding LOTEs to FLOTEs: 'The only objective primitive laws I believe in are the laws of physics' (Maudlin 2007: 157–8). This suggests that also what causation *ontologically really is*, is ultimately what was described for the fundamental level: FLOTEs generating world states from prior world states. This is so, even though the former *quasi*-Newtonian conception is closer to our intuitive notion of causation where not world states but separable individual events count as causes (see Maudlin 2007: 157, 168). Thus, the quasi-Newtonian picture of causation turns out to be only a working hypothesis or a tool for the special sciences or everyday life. Yet causation really is what FLOTEs do.[39]

Counterfactuals Let us, next, see how *counterfactual conditionals* depend on (F/S)LOTEs. Take, as an example: 'were I to drop this glass on that stony floor, it would break'. How is it determined whether this conditional is true or false? Maudlin advises a three step procedure. First, take the antecedent of the conditional to be a kind of command which issues the following order: fix a relevant time, let's say *now*, and the state of the world at that time. Second, alter that state minimally so that the antecedent is true (the glass does drop). Third, let the laws (the FLOTEs) do their job, i.e. let them generate or produce the next state of the world. Now, the counterfactual is true if and only if its consequent is true regarding that second state of the world (see Maudlin 2007: 23).[40] Maudlin's theory is the ontological equivalent to Goodman's epistemic/logical account. Instead of premises and deductive consequences Maudlin speaks of initial world states and production of later ones (Section 3.2).

Dispositions What do we do, next to counterfactuals and causation, with *dispositions*? They are neither reduced to counterfactuals nor causation, but they, too, can be directly analysed in terms of laws. This holds at least for non-fundamental dispositions:

> Fragility is not as 'real' a property as mass for the simple reason that we take it to supervene on fundamental physical and chemical properties. Once you have specified the exact physical composition of the window, *and given the physical laws*, it follows by analysis that the window will break if a rock is thrown at it with sufficient force.
>
> (Maudlin 2007: 72)

The reader will see similarities to Kaila's theory (Section 2.1.4). Maudlin's is in essence the same – slightly revised, of course, and with a non-Empiricist conception of laws. Yet what about fundamental properties like being massive? Are they dispositional? Although Maudlin is certainly not a Lewisean believer in Quiddities, we get no clear answer here (see Maudlin 2007: 72–3).

BOX 4.6 Laws as unanalysably fundamental

- Tim Maudlin (and also John Carroll whose thoughts we could, regrettably, not introduce here), claim that laws belong to the fundamental fabric of the universe. They are **ontologically basic** and **not grounded in anything else**. Thus, **an analysis** or explication of laws in terms of some other concepts **is not possible**.
- However, a little more can be said about these laws. For Maudlin, they come in two varieties: in **FLOTE**s: ***F*****undamental *L*****aws of *T*****emporal *E*****volution**; and in a non-fundamental variety: **LOTE**s. Strictly speaking, the second are not real, they are parasitic on or true in virtue of the former fundamental laws only.
- **Instead of universality** – which LOTEs anyway do not possess because of possible interferers but which FLOTEs also might lack because they could differ at other times or places – Maudlin takes the following as the laws' common important characteristic: to be **producers or generators of future states of the universe given some initial state**. Compare this to feature **(PROD)** dispositions intuitively have.
- This is at the heart of Maudlin's **analysis of causation**, too. Here, he distinguishes again between fundamental and non-fundamental causation. For the latter, **causes are LOTE**

governed disturbances of LOTE governed inertial behaviour of systems/objects and effects the resulting deviations thereof. In the realm of FLOTEs there are, according to Maudlin, no such isolatable systems and the picture is rather that, holistically, all **FLOTEs together causally produce a whole subsequent world state from a temporally previous one**.

- Maudlin's analysis of **counterfactuals** is *Goodman ontologised*: a counterfactual **'if p were the case, q would be the case' is true** *iff* **FLOTEs produce a world state in which *q* is the case from a world state in which *p*** together with other relevant circumstances is the case.
- Macroscopic **dispositions**, finally, can be analysed à la Kaila in terms of (underlying) basic properties and laws/(F)LOTEs.

4.7 Laws and dispositions

Remember our starting point in Chapter 2 on dispositions: we wanted to know whether it is possible to define dispositional predicates in terms of non-dispositional vocabulary. Alternatively, or additionally, whether dispositional properties exist in their own right or whether they are ontologically parasitic on more fundamental material like causation, counterfactuals, or laws of nature.

One such early (Empiricist) account was Eino Kaila's. It says, slightly modernised, that an object *x* has a disposition *D* to react with reaction *R* when in triggering circumstances *T iff* the following is true: *x* has some intrinsic structural, categorical property *S* and, moreover, it is a law of nature that objects that have/are *S* react with *R* when *T*-ed.

Of the philosophers (together with their law theories) we got to know in this chapter, Tim Maudlin and David Armstrong have explicitly put forward reductive theories along Kaila's lines. Maudlin's we have just described above, and Armstrong's doesn't differ much (Section 2.2).

We discussed three more theories of laws in this chapter: Lange's (Section 4.5), Lewis's (Section 4.3) and the early Empiricists' (Section 4.2). The latter's take on dispositions are actually the ones we encountered first in Chapter 2. Here we only added their so far missing (albeit failing) theory of laws.

Lewis sees a closer immediate relation of dispositions to counterfactuals rather than to laws. However, as we have seen in Section 3.5, laws play an important indirect role: the place where they come into play is in the truth conditions of the counterfactuals.

As we have seen, Lange sympathises at some places with a theory for dispositions along the Kaila–Armstrong–Maudlin line. Yet he also finds plausibility in the option that dispositional statements are directly equivalent to counterfactual conditionals. If so dispositional facts and counterfactual truths are on a par. In the latter case, it could turn out that laws depend on dispositions rather than the other way round. This would turn our whole endeavour so far on its head.

In fact, other philosophers have explicitly done this. A pioneer of a theory that has laws depend on dispositions rather than the other way round is Nancy Cartwright, who, in *How the Laws of Physics Lie* (1983) and in *Nature's Capacities and their Measurement* (1989), has courageously put forward such ideas when the majority of philosophers were hostile towards dispositions, capacities, powers and the like (Section 6.2).

Only in the late 1990s and early 2000s did beliefs in the reality of dispositions return to the philosophical mainstream. In Chapter 6 we will get to know the main protagonists of the so called *Dispositional Essentialism*.

Notes

1 The four interpretations of universality and parts of the list of common attributes to laws are taken from Hüttemann (2007: 140–1).

2 Note that (i), whether laws necessitate the coming about of events in the world and (ii), whether the laws themselves could not have been different, i.e. are necessarily as they are. Actually, even those philosophers who think that what the laws say must happen often hold that what these laws are is a contingent matter. (For a further discussion see Sections 4.4 and 6.4.2.)

3 I owe all my knowledge about seventeenth- and eighteenth-century conceptions of laws to Andreas Hüttemann (Hüttemann 2007). A very informative chapter on the history of the concept of laws of nature, including ancient and medieval thought, can be found in Mauro Dorato's *The Software of the Universe* (Dorato 2005: 1–30). Dorato's book on laws also attempts an answer to the rarely asked question why statements of laws of nature are often in mathematical form (Dorato 2005: 31–65).

4 It is one thing to say what laws or law statements are and another to spell out how we can verify them, even if these issues are, for the Empiricist, closely related: if you cannot say how to verify (or falsify) candidate laws by some kind of sense experience they are nothing but bad metaphysical speculation. This complex and interesting task of the verification of law candidates can not, however, be a topic for us. Nor will we deal with the tricky question what the appropriate scientific, experimental methods are to discover natural laws in the first place. This is rather a subject for a book on the epistemology of science. Rather, we remain faithful to metaphysical issues and simply ask how to characterise what laws of nature are. If you are interested in the epistemology of laws Carl Gustav Hempel's *Aspects of Scientific Explanation* (1965, especially parts I and IV) and Karl Raimund Popper's *The Logic of Scientific Discovery* (1959) are the classical starting points. A currently very popular theory of confirmation is Bayesianism; see, for example, Howson and Urbach (1993). A recent original approach is Wolfgang Spohn's ranking theory (Spohn 2012: 271–303). Textbooks that deal with the epistemology of science include Bird (1998a: part II), Ladyman (2002: part I) and Tobin (forthcoming).

5 A few pages later, Ryle too advertises this view regarding dispositions (Ryle 1949: 119).

6 See Mill (1843: 317) and Ramsey (1928: 140–4). Further above, we have Ramsey give assent to a different, pragmatist conception of lawhood. That was indeed the view he endorsed later in his short life.

7 Note, however, that we find similar ideas also in Leibniz who, remember (Chapter 2), claimed that when god created the best of possible worlds, namely ours, he 'chose the most perfect order, that is, the order that is at once simplest in general rules and richest in phenomena' (Leibniz 1686: §6).

From a Leibnizean standpoint, Mill–Ramsey–Lewis aim to retrace god's order. There are also traces of what became the Mill–Ramsey–Lewisean view in Kant, see his *Critique of Judgement* (Kant 1790: 179f, 183f, 185). Note, however, that both Leibniz and Kant are anti-Humeans in that they believe in the necessitating force of laws.

8 We leave out, at this place, a third important parameter: *fit*, which concerns probabilistic laws. *Fit* is a measure of how well the probabilities a system postulates match the actually occurring respective frequencies. In other words, one could say that *fit* measures how likely the actual world is according to the system: the likelier, the higher the fit-score.

9 Note aside: the predicates *grue* and *bleen*, as originally introduced by Nelson Goodman, were defined slightly differently (Goodman 1955: 74ff). Also, for Goodman, his artificial predicates serve a different issue, namely to unearth a new riddle of induction.

10 Lewis does not give us a clear decision procedure when a property is natural to which degree. Also, how derived laws are derived remains unexplained (see Lewis 1983a/1999: 42).

11 BBSA-style ideas can also be found in Halpin (2003), Taylor (1993: 97), Roberts (1998), Loewer (2007) and Albert (2000).

12 Lewis has changed his view on this issue between 1973 (Lewis 1973: 73; see also Lewis 1983a: 41) and 1994 (see above).

13 See Jaag (2015) for more on the explanation relation between laws and singular facts. Actually, Jaag discusses this issue for all theories of laws, not just for Lewis's. See Loewer (1996; 2012) for a possible solution to the explanation problem.

14 For reasons of brevity we will restrict our section to Armstrong's variant of the theory.

15 See Armstrong's *A Materialist Theory of the Mind* (1968) and our brief discussion of some of its core theses in the dispositions chapter (Section 2.2).

16 In fact, Lewis tells us that he likes to refrain from judgement whether universals exist or not (see Lewis 1983a). He can allow himself to remain agnostic because his metaphysics does not depend on this issue at all. Only naturalness (either of classes or universals) is crucial.

17 The discussion whether and how universals exist was repeated more than a thousand years later within scholastic philosophy, Thomas Aquinas (1225–74) being the most famous proponent of universals, William of Ockham (1287–1347) the most famous Nominalist.

18 Note that there are reasons why Dispositional Realists (Chapter 6) should, in this respect, be Platonists rather than Aristoteleans (Tugby 2013).

19 Inferences to the Best Explanation are discussed at large by Peter Lipton in his like named book (Lipton 1991). In Section 7.3 we will have the chance to evaluate this scientific and philosophical method.

20 See (van Fraassen 1989: 96–109) for these allegations and their two names, see also Lewis (1983a), who utters similar kinds of worries.

21 Do not confuse the *inference* in *inference problem* with the *inference* in *inference to the best explanation*. The first is the specific inference from $N(F, G)$ to $\forall x\ (Fx \rightarrow Gx)$, the second are general inferences to the existence of whatever best explains some phenomenon.

22 For comparisons with religious matters see also note 31 in Chapter 6.

23 By explicitly pointing out that the latter remains unsolved I am not implicitly assuming that the answers to (i) and (ii) are really plausible.

24 There are philosophers who suggest making the productive character of Armstrong's *N* (or a similar connection) even more dominant. Lewis Creary, Richard Corry, and myself have all suggested to characterise N more in terms of Newtonian forces rather than to allude to necessity/necessitation (Creary 1981; Corry 2006; Schrenk 2011).

25 For more on the difference see Schrenk (2005). There are, however, authors who have suggested in one way or another that the identification of nomological necessitation, N, with full blown metaphysical necessity, $\Box$, should be made: $N(F, G)$ equals $\Box\forall x(Fx \rightarrow Gx)$ (see Kneale 1949, Elder 1994 and Swoyer 1982). Actually, Armstrong, influenced by Donald Baxter (2001), flirted once with a similar move for a short while (Armstrong 2004a, 2004b): 'The connections between universals involved in laws of nature are necessary rather than contingent' (Armstrong 2004a: 136). Armstrong had rejected this view again (personal communication 2008).

Things might get even more contingent: like all universals, *N* itself is, for Armstrong, a quiddistic property (Sections 2.2.3 and 2.2.4). So, it too can lose the lawmaking role it has in our world. Thus, not only do we get the above contingency of laws – that *N*, being the lawmaker, creates different laws in different worlds – but also that *N*, in yet other worlds, ceases to be a lawmaker. Some philosophers (see Bird 2007: 93; Section 6.4.2) find this unacceptable and suggest to demand that necessarily, for all *F* and *G*, if *N* (*F*, *G*) then $\forall x(Fx \rightarrow Gx)$, i.e. $\Box\ (\forall \mathrm{F, G}\ (N\ (\mathrm{F, G}) \rightarrow \forall \mathrm{x}\ (\mathrm{Fx} \rightarrow \mathrm{Gx}))$. This would say, more or less, that *N* can't lose its lawmaking (regularity generating) force.

26 Robert Pargetter (1984) has formulated a theory of laws where the just defined nomological necessity is the fundamental, not further analysable notion (instead of Armstrong's necessitation).

27 Here, we will sketch Lange's theory as he developed it from his 2005 paper onwards. For the sake of brevity we leave Lange (2000) aside.

28 We introduced the symbol $\Box\!\!\rightarrow$ in Section 3.1.

29 Compare this to deductive systems that describe the world's mosaic of perfectly natural properties in Lewis's best-system analysis (2.2.5 and 4.3) where, also, the regularities are first and foremost stated in a completely factual and non-modal way. Only later, law status might be conferred to them. Lange's account also shares with Lewis's that it is holistic: we do not get the laws individually but all of them together as members of stable sets. (This is unlike Armstrong where, even in isolation, *N* (*F*, *G*) can be a law).

30 Some readers might now ask whether dispositional truths may be in Σ. That's a very good question and we will return to it later.

31 Here's a quick reminder that, although we speak of *justification* and *indication,* we do not mean anything epistemic. Better: a subset of sub-nomic facts being stable is what *makes* its members laws.

32 The factual material for forming counter to the facts assumptions to any proper subset Γ of Σ is always to be found in $\Sigma\backslash\Gamma$, i.e. the remainder without Γ.

33 This is, by the way, a position Lange sympathised with in his book on the philosophy of physics (2002: 76). In his 2009a book he seems much more neutral.

34 Here's a gem for those of our readers who are familiar with modal logic: subjunctive facts can entail the actual facts so that the members of Σ need not be a separate basic ontological category within Lange's ontology. Here's why: take any fact, *p*. It is entailed by any counterfactual of the form $T \Box\!\!\rightarrow p$ (with *T* being some logical truth). So, if, for all actual facts p_i, we have a corresponding counterfactual fact, $T \Box\!\!\rightarrow p_i$, *amongst the true subjunctives* in *CT* then we need no separate set Σ because it, Σ, can be gained from a subset of the subjunctive truths plus logic (see Lange 2009b: 304, fn.11). In other words, the world's fundamental fabric is made of subjunctive facts and logic alone!

35 By the way, Woodward explicitly endorses the dispositions-first view: 'An account of laws which relies on notions like invariance and stability seems to fit naturally with the idea [. . .] that laws are abstract [. . .] descriptions of capacities and dispositions characteristic of particular objects or systems' (Woodward 1992: 205). Compare this to Cartwright and other Dispositionalists in Chapter 6.

36 For reasons of space we must put our focus on Maudlin's recent ideas. Carroll's precedential theory has found its way into textbooks already (see, for example, Psillos 2002: 175–7).

37 This echoes the Logical Empiricists in so far as they, too, thought that the only task left for philosophy is to clarify the concepts that occur within science. Needless to say, Maudlin is neither a Verificationist nor even a Humean.

38 Andreas Hüttemann suggests a similar analysis of causation (Section 6.4.3). Instead of laws he speaks, however, of the inertial behaviour a system is *disposed* to. A disturbance is, for Hüttemann, an interference into a disposition's manifestation: 'a cause is an actual disturbing factor (antidote) to the default behaviour that a system is disposed to display' (see Hüttemann 2013: 116).

39 Remember Armstrong's nomological account of causation. It is similar in so far as he, too, has laws (here: nomological necessitation) causally produce the ongoings in the world. Yet, Armstrong's N can also operate locally where universals are instantiated in individual separated objects. Maudlin's fundamental story about causation is, on the contrary, holistic.

40 All this is more complicated than depicted here. For example, probabilistic laws and also LOTEs with exceptions do not determine strictly what the outcome of such an evolution is. Of course, Maudlin is aware of this and he has ways to deal with these complications (Maudlin 2007: 21–34). More on laws with exceptions, *ceteris paribus* laws, can be found on pages 45–7 of that book, and there is more on counterfactuals (and causation) on pages 143–69.

Bibliography

Albert, D. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*. London: Routledge.

Armstrong, D. M. (1983) *What is a Law of Nature*? Cambridge: Cambridge University Press.

Armstrong, D. M. (1997) *A World of States of Affairs*. Cambridge: Cambridge University Press.

Armstrong, D. M. (2004a) How Do Particulars Stand to Universals? In: Zimmerman, D. W. (ed.) *Oxford Studies in Metaphysics, Vol. 1*. Oxford: Clarendon Press. pp. 139–54.

Armstrong, D. M. (2004b) *Truth and Truthmakers*. Cambridge: Cambridge University Press.

Armstrong, D. M. and Heathcote, A. (1991) Causes and Laws. *Noûs*. 25 (1). pp. 63–73.

Ayer, A. (1963) *The Concept of a Person and other Essays*. London: Macmillan.

Bartels, A. and Stöckler, M. (eds.) (2007) *Wissenschaftstheorie. Texte zur Einführung*. Paderborn: Mentis.

Baxter, D. (2001) Instantiation as Partial Identity. *Australasian Journal of Philosophy*. 79 (4). pp. 449–64.

Beebee, H. (2000) The Non-Governing Conception of Laws of Nature. *Philosophy and Phenomenological Research*. 61 (3). pp. 571–94.

Bird, A. (1998a) *Philosophy of Science*. Montreal and Kingston: McGill and Queen's University Press.

Bird, A. (2007) *Nature's Metaphysics*. Oxford: Clarendon Press.

Carroll, J. W. (1994) *Laws of Nature*. Cambridge: Cambridge University Press.

Cartwright, N. (1983) *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Cartwright, N. (1989) *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Cohen, J. and Callender, C. (2009) A Better Best System Account of Lawhood. *Philosophical Studies*. 145. pp. 1–34.

Cohen, J. and Callender, C. (2010) Special Sciences, Conspiracy and the Better Best System Account of Lawhood. *Erkenntnis*. 73 (3). pp. 427–47.

Corry, R. (2006) Causal Realism and the Laws of Nature. *Philosophy of Science*. 73 (3). pp. 261–76.

Creary, L. (1981) Causal Explanation and the Reality of Natural Component Forces. *Pacific Philosophical Quarterly*. 62 (2). pp. 148–57.

Cudworth, R. (1678) *The True Intellectual System of the Universe*. London.

Dorato, M. (2005) *The Software of the Universe: An Introduction to the History and Philosophy of the Laws of Nature*. Aldershot: Ashgate.

Dretske, F. (1977) Laws of Nature. *Philosophy of Science*. 44 (2). pp. 248–68.

Elder, C. (1994) Laws, Natures, and Contingent Necessities. *Philosophy and Phenomenological Research*. 54 (3). pp. 649–67.

Goodman, N. (1955) *Fact, Fiction and Forecast*. 4th edition (1983). Cambridge, MA: Harvard University Press.

Halpin, J. F. (2003). Scientific Law: A Perspectival Account. *Erkenntnis*. 58 (2). pp. 137–68.
Hempel, C. G. (1965) *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. London: Collier-Macmillan; New York: Free Press.
Howson, C. and Urbach, P. (1993) *Scientific Reasoning. A Bayesian approach*. Chicago, Illinois: Open Court.
Hume, D. (1739–40) *A Treatise of Human Nature*. Selby-Bigge, L. A. (ed.), revised by Nidditch, P. H. (1978). Oxford: Clarendon Press.
Hüttemann, A. (2007) Naturgesetze. In: Bartels, A. and Stöckler, M. (eds.) *Wissenschaftstheorie. Texte zur Einführung*. Paderborn: Mentis. pp. 135–53.
Hüttemann, A. (2013) *Ursachen*. Berlin: De Gruyter.
Jaag, S. (2015) *Explaining Laws of Nature. A metaphysical investigation into the natural principles governing the universe*. Unpublished PhD thesis.
Kant, I. (1786) *Metaphysical Foundations of Natural Science*. In: German Academy of Sciences (eds.) *Kants gesammelte Schriften* (1900–). Berlin: De Gruyter.
Kant, I (1790) *Kritik der Urteilskraft (Critique of Judgement)*. In: H.F. Klemme (ed.) (2006). Akademie Ausgabe. Hamburg: Meiner.
Kirchhoff, G. (1876) *Vorlesungen über Mechanik*. Leipzig.
Kneale, W. (1949) *Probability and Induction*. Oxford: Oxford University Press.
Ladyman, J. (2002) *Understanding Philosophy of Science*. London and New York: Routledge.
Lange, M. (2000) *Laws in Scientific Practice*. Oxford: Oxford University Press.
Lange, M. (2002) *Introduction to the Philosophy of Physics: Locality, Fields, Energy, and Mass*. Oxford: Blackwell.
Lange, M. (2005) Laws and their Stability. *Synthese*. 144 (3). pp. 415–32.
Lange, M. (2009a) *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford: Oxford University Press.
Lange, M. (2009b) Why Do the Laws Explain Why? In: Handfield, T. (ed.) *Dispositions and Causes*. Oxford: Clarendon Press. pp. 286–321.
Leibniz, G. W. (1686) On the True Method in Philosophy and Theology. In: Wiener, P. P. (ed.) (1951). *Leibniz: Selections*. New York: Charles Scribner's Sons.
Lewis, D. (1973) *Counterfactuals*. Oxford: Oxford University Press.
Lewis, D. (1983a) New Work for a Theory of Universals. In: Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press. pp. 8–55.
Lewis, D. (1983b) Extrinsic Properties. In: Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
Lewis, D. (1986) *On the Plurality of Worlds*. Oxford: Blackwell.
Lewis, D. (1994) Humean Supervenience Debugged. In: Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press. pp. 224–47.
Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
Lipton, P. (1991) *Inference to the Best Explanation*. London and New York: Routledge.
Loewer, B. (1996) Humean Supervenience. *Philosophical Topics*. 24. pp. 101–27.
Loewer, B. (2007) Laws and Natural Properties. *Philosophical Topics*. 35 (1, 2). pp. 313–28.

Loewer, B. (2012) Two Accounts of Laws and Time. *Philosophical Studies*. 160. pp. 115–37.
Maudlin, T. (2007) *The Metaphysics within Physics*. Oxford: Oxford University Press.
Mill, J. S. (1843) A System of Logic. Ratiocinative and Inductive. Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. In: Robson, J. M. (ed.) (1967) *Collected Works, Vols. VII and VIII*. Toronto: Toronto University Press.
Pargetter, R. (1984) Laws and Modal Realism. *Philosophical Studies*. 46 (3). pp. 335–47.
Popper, K. R. (1959) *The Logic of Scientific Discovery*. London: Hutchinson.
Psillos, S. (2002) *Causation and Explanation*. Chesham: Acumen.
Ramsey, F. P. (1928) Universals of Law and of Fact. In: Mellor, D. H. (ed.) *Philosophical Papers*. Cambridge: Cambridge University Press. pp. 140–4.
Reichenbach, H. (1947) *Elements of Symbolic Logic*. Macmillan Co.: New York.
Roberts, J. (1998). 'Laws of Nature' as an Indexical Term: A Reinterpretation of Lewis's Best-system Analysis. *Philosophy of Science*. 66 (3). pp. 502–11.
Roberts, J. (2008) *The Law-Governed Universe*. New York: Oxford University Press.
Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson.
Schrenk, M. (2005) The Bookkeeper and the Lumberjack: Metaphysical vs. Nomological Necessity. In: Abel, G. (ed.) *Kreativität. XX. Deutscher Kongress für Philosophie. Sektionsbeiträge Band 1*. Berlin: Universitätsverlag der Technischen Universität.
Schrenk, M. (2007) *The Metaphysics of Ceteris Paribus Laws*. Frankfurt: Ontos.
Schrenk, M. (2008) A Theory for Special Sciences Laws. In: Bohse, H., Dreimann, K. and Walter, S. (eds.) *Selected papers contributed to the sections of GAP.6, 6th international congress of the society for analytical philosophy*. Paderborn/Münster: Mentis.
Schrenk, M. (2010) The Powerlessness of Necessity. *Noûs*. 44 (4). pp. 725–39.
Schrenk, M. (2011) Interfering with Nomological Necessity. *The Philosophical Quarterly*. 61 (244). pp. 577–97.
Schrenk, M. (2014) Better Best Systems and the Issue of CP-Laws. *Erkenntnis*. 79 (10 supplement). pp. 1787–1799. doi: 10.1007/s10670-014-9642-9.
Spohn, W. (2012) *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford: Oxford University Press.
Swoyer, C. (1982) The Nature of Natural Laws. *Australasian Journal of Philosophy*. 60 (3). pp. 203–23.
Taylor, B. (1993). On Natural Properties in Metaphysics. *Mind*. 102 (405). pp. 81–100.
Tobin, E. (forthcoming 2017) *Philosophy of Science: An Introduction to Contemporary Problems*. London and New York: Continuum.
Tooley, M. (1977) The Nature of Laws. *Canadian Journal of Philosophy*. 7 (4). pp. 667–98.
Tugby, M. (2013) Platonic Dispositionalism. *Mind*. 122 (486). pp. 451–80.
Van Fraassen, B. (1989) *Laws and Symmetry*. Oxford: Clarendon Press.
Woodward, J. (1992) Realism about Laws. *Erkenntnis*. 36 (2). pp. 181–218.
Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

5 Causation

5.1 Preliminaries and agenda

The *dispositions of an object* and what that object helps to *cause* in certain circumstances are not unrelated – for example, a match is inflammable, i.e. it is disposed to light, and when struck it bursts into flames (see Section 2.1 and Chapter 6). We might want to say that *because it is so disposed*, *striking the match causes it to light*. David Lewis and Alexander Bird (see Section 3.5) argued the other way round: objects are disposed to react with r when stimulated with s because they have some substructure B that causes them to r when in s. Be that as it may, we postpone dispositions and first deal with causation per se (we return briefly to dispositions in Sections 5.5 and 5.7).

Needless to say, we concentrate on analyses of causation that suit the Empiricist spirit as we did with the analysis of counterfactual conditionals (see Chapter 3) and laws of nature (see Chapter 4). The theories most affine to Empiricism are Hume's own, but also John Mackie's (1917–81) *INUS theory*. We discuss both in Section 5.2.[1] Remember that the aim of such *Humean* theories is to say what causation amounts to without referring to *necessary connections* (MOD), *powers*, *production* (PROD) or other such *secret connections*.

After we discuss the limits of Humean analyses, also including the counterfactual one (Section 5.3), we turn to less and non-Humean ones: the *process* and *mark transfer theory* of Wesley Salmon, and Phil Dowe's and Max Kistler's successor thereof, i.e. the *conserved quantities* theories. We also turn to *mechanistic theories*, particularly Stuart Glennan's (all in Section 5.4), and manipulability theories by, for example, Georg Henrik von Wright (1916–2003), Peter Menzies and Huw Price (all in Section 5.5). We will introduce James Woodward's *interventionist theory*, where we will also briefly talk about the *causal graph theory*.

Section 5.6 will categorise the introduced theories and evaluate their extensional and intensional adequacy. In the final section (Section 5.7) we return to dispositions and ask which theories of causation help explain dispositionality. In Chapter 6 we will turn this direction of analysis on its head and get to know how *dispositional realists* define causation in terms of dispositions.

Relata for causation There are many issues surrounding causation and we cannot focus on all of them. One central question is what kind of thing the relata – cause *c* and effect *e* – of causal relations are: substances, properties, propositions, facts or events? For example, we sometimes speak of *acidic substances* to cause chemical burn, the *property of being uneducated* to cause poverty, *the fact there was a short circuit* to cause the fire, *the event of an avalanche rushing down a slope* to cause the mountain shelter to collapse.

Of course these seemingly different *relata* are not independent. A *fact* could be seen as the *instantiation of a property*: the fact that a truck had the property of so much weight caused the bridge to collapse. Similarly, we could phrase the whole affair in terms of *events*: the event that a truck with its heavy load drove onto the bridge caused its collapse.[2] Although substances are a bit out of fashion in theories of causation,[3] a novel idea for the relata of causation we have not mentioned above is that in causal graph theory *variables* and *their values* are treated as the relata of causal relations. We come to this in Section 5.5 where we also hint at one instance of substance causation. However, we will not go further into the discussion of the correct relata here. In what follows, we mean *events* when we speak of causation (unless we explicitly say otherwise).

An important point regarding the relata (of any kind) is that the relation of *c* and *e* must be of a non-logical, non-conceptual, non-mereological type otherwise it cannot count as causal – for example when you raise your left arm you raise one of your arms, and yet raising your left arm does not cause you to raise one of your arms.

Ontology, epistemology or semantics? John Mackie, whose theory we address in Section 5.2, divided the project into three kinds of analysis to explicate what causation is:

> (i) ontological (factual) – what causation is 'in the objects', as a feature of a world that is wholly objective and independent of our thoughts,
> (ii) conceptual (semantical) – what causal statements mean and/or what truth conditions they have,
> (iii) epistemic – how we know what we know about causation.
>
> (Mackie 1974: ix)

Mackie also rightly points out that the Logical Empiricists deliberately conflated these three aspects: for them, giving (ii) the meaning of a causal (or nomological or dispositional) statement is to give (iii) the verification or testing conditions for the truth of it. The ontological question (i) was deflated because most causation was supposed to be in our heads (see Section 1.2). At best, there are regularities in our perceptions (sense data) that we can take as the basis of an *objective* causal claim.

Generally, these three aspects can not be separated totally. For example, if you give the meaning of a statement (ii) in terms of truth conditions, you thereby also say what the world has to be like for the statement to be true (iii). As we will see, each theory we discuss focuses more on one aspect rather than on the other.

In Section 5.6 we will discuss further ways to classify theories of causation; however, to fully understand these categorisations it is best to know the theories first.

5.2 Regularity theories

In the seventeenth and eighteenth centuries, a widespread view on causation was that a cause (for example, a moving billiard ball bumps into another) *necessitates* its effect (for example, the second ball starts rolling) (see Section 1.2). This theory had two aspects: the first is that causes *produce* or *bring about* their effects (cf. (PROD) from Section 2.1.3); the second is that the effect must happen, given its cause (cf. (MOD) from Section 2.1.2). The distinction between these two aspects – production and necessity, (PROD) and (MOD) – wasn't made back then and even today it deserves more attention. As we know, Hume argued against both production and necessity, and he developed his own regularist theory of causation (the reader is invited to re-read the relevant passages in Section 1.2 if this reminder is too short):

> An actually occurring event *c* is a cause of an actually occurring event *e* if and only if:
>
> (i) [Contiguity] *c* is spatially in contiguity with *e*.
>
> (ii) [Succession] *e* happens temporally after *c*.
>
> (iii) [Regularity] all events of the same type as *c* are followed in spatio-temporal succession, as in (i) and (ii), by events like *e*.

As an example, consider the billiard ball *c* bumping into (i) a billiard ball *e*, whereupon *e* starts rolling (ii) and where billiard balls of the same kind have always behaved that way (iii).

This is a good start; however, counterexamples show that the theory is inadequate as it stands. Two old acquaintances make life hard for the regularity theorist: (almost) *Void Satisfactions (VC)* and *Random Coincidences (RC)* (see Sections 2.1.2 and 2.1.3).

Suppose two very peculiar events, *c** and *e**, happen to occur in spatial contiguity, (i), and in the right temporal succession, (ii). Suppose these events are so peculiar that they never repeat. Requirement (iii), regularity, is almost voidly satisfied: *all* events like those of c^* and e^* behave like c^* and e^* for there are but these two. Think, for example, of only once dialling a random hundred-digit number on your mobile. On typing the last digit, you receive a text message from the Pope by mistake. The first event was clearly not a cause of the second; however, according to Hume's analysis it would have to count as such.

More prominent as refutation are truly regular but (RC) To appreciate this example, we must first cast doubt on Hume's (i) contiguity. With Isaac Newton's (1643–1727) idea of gravitational forces that act at a distance, it has become acceptable that there is causation miles apart – for example waves are caused by the moon's gravitational forces. When (i) is dropped, we see that an example from Elliot Sober (1987) caused trouble for Hume: British bread prices and the sea level in Venice have both been positively correlated since records began. Whenever the sea levels rose, the price of bread increased as well (let's suppose that they did so in that temporal order). No one would want to claim that there is a causal relation between the two, but a simple regularity theory would have to.[4]

Remember the solutions to similar problems with dispositions: in order to be able to rule out that (VC) or (RC) count towards dispositionality, people tried to implement laws of nature in their analysis of dispositions. John Stuart Mill did just that for causation: in order for a regularity to count towards causation it has to be due to laws of nature (see Mill 1843: book III, chapters IV and V; also Davidson 1967/1993).

John Mackie's INUS condition Mill also anticipated ideas of the regularity theory, which we will discuss next. John Mackie's famous idea that an event type *C* has to be a so called *INUS condition* for an event type *E* in order for *C to count as cause* of *E*.

Insufficiency and unnecessity Mackie first points out that there is hardly ever one single event that is solely sufficient for an effect to occur. His example is a fire in a house that was caused by a short circuit. On reflection it becomes clear that other events or facts about the house are (equally) responsible for the disastrous fire: inflammable wooden panelling, the presence of oxygen, the absence of a fire sprinkling system, etc. As Mill put it, 'The real Cause, is the whole of these antecedents' (Mill 1843: 402).[5] Thus, where, for Hume, it seemed that one single factor could be *sufficient* for an effect, it becomes a crucial part of Mackie's account that there is often more than one decisive factor.

In addition to the *insufficiency* of a lonely single factor, Mackie also shows that whole causes (all factors together), although being sufficient, are *hardly ever necessary* for an effect: 'There are often several independent modes in which the same phenomenon could have originated' (Mill, quoted in Mackie 1974: 61). In other words, even without it but instead with a different complex cause the effect would still have occurred. For example instead of the short circuit (and other conditions mentioned), a burning match on an inflammable rug could have caused the very same disaster.

In short, singular causal factors (like the short circuit) are *insufficient* for the effect (there is a lot more that needs to be considered), and complex and complete causes are *unnecessary* because certain variations or different complete causes would also have led to the same outcome. These two results are the first major findings in Mackie's theory. *Insufficient*: a single factor is hardly ever sufficient alone; *unnecessary*: a whole bunch of factors that are sufficient could be replaced by a different sufficient bunch.

Necessary and sufficient We come to the second pair of conditions for causality that Mackie unearthed. Focus again on the single factor we made responsible for the house to burn down: the short circuit. We noted that it alone would not have been able to cause the fire. Other conditions had to be in place as well. Consider now a situation where these other conditions were present exactly as needed but where the short circuit did not occur. In this situation, the fire would not have happened. In other words, within this whole complex factor that was indeed present, the short circuit is *a necessary part*. Take it away (all else being equal) and the effect does not occur.

Finally, such a complex factor (take everything that preceded the house burning down together) is quite obviously sufficient for the fire. Fair enough, there could be differing overall sufficient complex conditions (exchange the short circuit for a burning match, wood for rug), but the actual one did *suffice*.

INUS Let us collect the two times two features we have found for the short circuit and the complex conditions it features in:

The short circuit is

(i) alone, just by itself, an ***I***nsufficient element for the house to burn down,
(ii) and yet it is a ***N***ecessary ingredient within a complex overall condition (wood and oxygen being present, sprinklers being absent, etc.).
(iii) This complex overall condition is ***S***ufficient for the house to burn down.
(iv) Still, this complex overall condition is ***U***nneccessary in the sense that other overall conditions (rug, match, etc.) would also lead to the house being on fire.

Swap the order of the last two entries and put the initials of the features together and you have Mackie's famous abbreviation *INUS*:

> [T]he so-called cause is [. . .] an insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result. [. . .] It will be convenient to have a short name for [the cause making condition, MS]: let us call such a condition (from the initial letters of the words italicized above) an INUS condition.
>
> (Mackie 1965: 34)

For further illustration, consider the formula $((ABC) \vee (DFG) \vee \ldots) \leftrightarrow E$. Our cause *C* might alone be insufficient for *E* but it is a necessary ingredient within complex *ABC*, which is sufficient for *E*. However, *ABC* might well be unnecessary for *E* because some complex *DFG* (and probably others: . . .) do also suffice for *E*. Thus:

> *C* is a cause of *E* iff *C* is at least an INUS condition for *E*.
>
> (Mackie 1965: 37)[6]

It is important to keep in mind that, in the acronym *INUS*, the letters do not all apply to one and the same thing: the *I* refers to a single insufficient factor within a complete cause of which this single factor is a necessary ingredient, *N*. The *S* refers to the complete sufficient cause which is also unnecessary, *U*. In short, *I* and *N* refer to a part, *U* and *S* to the whole. We should also clarify three amendments we have (tacitly) made to our concrete fire case mentioned earlier.

First, it is customary to use capital letters (*C*, *E*) when referring to *general kinds of events*, i.e. *event types*, and small letters (*c*, *e*) when referring to *concrete particular instantiations* of these kinds, i.e. *event tokens*. According to this rule, we have used capital letters to signify that Mackie's theory is for event types (see Mackie 1965: 45–50): short circuits in general (these kinds of events) are causes of such fires in general *iff* they are INUS conditions for such fires. Thus, the formula tacitly contains claims about *universal regularities*: it is never the case that any particular event *e* of type *E* occurs without a particular event *abc* (or *dfg*. . .) of type *ABC* (or *DFG*. . .) having also occurred.

Second, we said '*C* alone *might well be* insufficient' and '*ABC might well be* unnecessary for *E*' because there is no principle reason why some specific *C* should not alone suffice. Examples might come from fundamental physics where perhaps the presence of positive charge alone causes the attraction of an electron. Similarly, some subatomic reactions *E* might well be caused in one and only one way, *ABC*.

Then *ABC* would be necessary. (All this explains why Mackie wrote: *C* '*is at least an* INUS *condition*' for *E*).

Third, the dots '. . .' in '(*ABC*) ∨ (*DFG*) ∨ . . .' stand for absolutely all further complex conditions that are sufficient for *E*. Thus, whenever we find an *E* we can be sure it has been preceded by at least one of the (*ABC*) ∨ (*DFG*) ∨ . . ., which justifies the double arrow in the formula: $((ABC) \vee (DFG) \vee \ldots) \leftrightarrow E$.

Necessary factor versus necessitating factor Note the following important terminological differences. The notion of a *necessary factor* is very different from a factor that supposedly *necessitates* (anti-Humean style) some other event. A *necessary factor* is a condition without which something else could not be the case (i.e. a *condicio sine qua non*). A *necessitating factor* refers to a productive force. For example, being pregnant is one necessary precondition for giving birth to a daughter, and yet being pregnant does not necessitate birth of a *female* child.

However – and here the terminology is confusing – if there were *necessitating productive forces* (anti-Humean style) then their presence alone would be *sufficient* for the effect to occur. The effect cannot but occur if it is necessitated. Thus, necessitating factors are sufficient for something to come about.

Be that as it may, Mackie has no such necessitation in mind, but yet what does he mean by sufficiency? For Mackie, being Humean in these matters, we are justified in believing that some event is sufficient for the occurrence of a second when events of the same type have regularly followed each other. This again underlines that Mackie's theory is a sophisticated regularity theory of causation.

Causal fields and pragmatics We know that a single factor, *C*, hardly ever causes an effect *E* on its own. We might therefore be tempted to call only the whole of *ABC* a *real cause* (see Mill 1843: 402). Bearing in mind that *C* alone is (merely) an INUS condition, we could still continue, by courtesy, to speak of *C* as *a cause*.

However, there is a drawback to this liberalism: we'd be obliged to give the honorary title *cause* not only to the short circuit but also to the present wooden panelling (A) and the oxygen (B). The reader can quickly check that they are INUS conditions for *E*, but yet would we really want to say that the presence of oxygen caused the fire?

Mackie has a pragmatic strategy to avoid this. When we talk about causes for certain effects we make certain default assumptions. For example, when houses burn down we make the background assumption that oxygen was present and that firemen and -women were not constantly present. Mackie calls the facts or events that are within our default assumptions about a situation the *causal field* (Mackie 1965: 39ff). Whatever is within this field is not normally considered as an INUS condition and is thus not a cause.

Causal fields may, however, change and this underlines the pragmatic aspect of Mackie's idea. Consider this example (see Mackie 1965: 40): 'What is the cause of this man's death?' Let us assume that the answer 'It was the strong radiation he was exposed to' is correct because the radiation is indeed an INUS condition of the man's death. However, the inquirer might be dissatisfied and ask: 'Yes, sure, but of all these men that were exposed to the radiation only he died'. Now, the radiation moves into the background causal field (it's still there and important) but a different factor – for example, his overall poor health – which is also an INUS condition, moves into focus.

Do these pragmatic shifts not endanger Mackie's effort to formulate an objective theory of causation, independent of human interests? No, because, given a causal field, it is perfectly objective which factors remain as causes and which don't because it is the world's regularities (not us) that, once the field is fixed, decide which of the remaining factors are INUS conditions and which are not.

Regularity and (RC) revisited How does Mackie's regularity theory fare with regular (RC) (or derivatives thereof)? The truth is that Mackie's regularity theory is also liable to certain (RC) and Mackie acknowledges this fact (see Mackie 1974: 81 ff). We would have to toy around with logic in a way we know from (VC) and (RC) problems related to dispositions (see Section 2.1.4). This would reveal, yet again, the weakness of both the material implication and mere regularities. However, for the sake of brevity, we spare the readers the details. The omitted counterexample[7] shows that the INUS theory has to postulate causation where there is mere coincidence.

The reader might wonder at this point what progress Mackie made compared to Hume if problems of (RC) still prevail. The truth is that the advantages of Mackie's INUS theory lie more in having made explicit the differences between single insufficient but necessary factors and sufficient but unnecessary complexes plus the causal field on which the causal structures unfold.

Dissatisfied with his first account, Mackie, in *The Cement of the Universe* (1974), suggested an entirely different theory of causation: in order for there to be a causal link 'some "causal mechanism", some continuous process connecting the antecedent [. . .] with the consequent' (Mackie 1974: 82) has to unfold. What Mackie had in mind preceded the transfer and mechanistic theories which we will discuss in Section 5.4.

BOX 5.1 and 5.2 Regularity theories

- **Hume** spelled out what it means for an event *c* to cause another *e* in terms of three requirements: **contiguity, succession, regularity**. That is (i) *c* is spatially in contiguity with *e*; (ii) *e* happens temporally after *c*; (iii) all events of the same type as *c* are followed in spatio-temporal succession, as in (i) and (ii), by events such as *e*.
- **Truly regular but accidental coincidences** are problematic for this theory because it is forced to call them causal although they are not. (Think of the Venice water levels and the bread prices in Great Britain).
- **Mackie's** regularity theory is more complex and adds aspects that **Mill** already pointed out, particularly that **single factors are hardly ever sufficient** for an effect and that such an effect could most probably have been brought about by **many other different causes**. (Mackie keeps Hume's succession and regularity demand but contiguity wasn't central).
- Thus, for Mackie, *c* is a cause of *e* iff *c* is at least an **INUS condition** for *e*: 'The so-called cause is [. . .] an *I*nsufficient but *N*ecessary part of a condition which is itself *U*nnecessary but *S*ufficient for the result' (Mackie 1965: 34, emphasis added).
- We have introduced Mackie's notion of a **causal field**, which stands for the default background assumptions we make when figuring out which factors are INUS conditions and thus causes for an event.
- **Pragmatic considerations** decide which factors or circumstances are delegated to the background field and which remain explicit and are therefore candidates for being causes.
- We have also pointed out that **Mackie's theory** is not without problematic counterexamples that hinge on **random regular coincidences**.

5.3 Counterfactual theories

Hume wrote: 'We may define a cause to be an object followed by another, and where all objects, similar to the first, are followed by objects similar to the second'. This is the regularity theory we already know, and yet he immediately continued:

> Or, in other words, where, if the first had not been the second never had existed.
>
> (Hume 1748: Sect. VII, Part II, §60: 76)

Hume seems to have believed that the two ideas – to say what a cause is through regularities or through a counterfactual relation – are the same otherwise he would not have said 'in other words'. Be that as it may, 200 years later, David Lewis successfully developed the second idea into a full-blown theory that diverges considerably from pure regularity theories. In the end, Lewis's approach also crucially depends upon regularities (and definitely not on powers or production, which will be discussed later).

Unlike Mackie (see Section 5.2), Lewis confines himself to causation amongst particular events (see Lewis 1973: 161), that is, he talks about event tokens, *c* and *e*, and not, like Mackie, about event types, *C* and *E*. Also, while Lewis acknowledges that a pragmatic difference could be made between core causal factors and the more marginal background conditions in the causal field, he concerns himself, neutrally speaking, with 'what it is to be one of the causes' (Lewis 1973: 162) be it marginal or a central.

Causal dependence Lewis's initial theory comes in two steps: he defines what he calls *causal dependence* and only afterwards *causation* itself. Here is Lewis's definition of *causal dependence*:

> An event e *depends causally* on an event c *if and only if* the family of events {e, ¬e} *depends counterfactually* on the family of events {c, ¬c}, that is, both: c $\Box\!\!\rightarrow$ e *and* ¬ c $\Box\!\!\rightarrow$ ¬ e.
>
> (Lewis 1973: 165–7)

In short, 'whether e occurs or not depends on whether c occurs or not' (Lewis 1973: 166). Note that *causal* dependence is spelled out entirely in terms of *counterfactual* dependence.

Here is an example to show how the definition works: suppose a house is in fact on fire, *e*, and there has in fact been a short circuit, *c*. Does the house being on fire *causally depend* on the short circuit to have happened? According to the definition just given this amounts to whether the following two counterfactuals are true: (i) had the short circuit happened, then the house would be on fire and (ii) if the short circuit had not happened, then the house wouldn't be on fire.

You might be puzzled by the first counterfactual. It seems fairly strange to ask *whether e would have occurred had c occurred* if both actually *did* occur. Well, remember from Section 3.4 that a counterfactual with a true antecedent and a true consequent is trivially true. Thus, the first counterfactual is, at worst, an idle wheel. However, we come back to this later.

In the case of trivial truth of the first counterfactual, the second one, namely $\neg c \Box\!\!\rightarrow \neg e$, must do the crucial work: only if it is also true can we claim that there is causal dependence between *c* and *e*. Is it true that had the short circuit *not occurred* then the house *would not have been* on fire? This seems correct because in the closest scenarios, i.e. those that are very similar to ours except that there wasn't a short circuit (although everything else remains the same), the fire does not actually happen. There's no reason why it should. (If you are unfamiliar with how one evaluates such counterfactuals consult Section 3.4.)

The two counterfactuals are true: the first trivially, the second substantially. Thus, we may attribute causal dependence to the two events. A welcome result.

Let us return briefly to the first counterfactual and see what Lewis's motivation for it is. His definition of causal dependence is meant to capture what causal dependence is independent of actual occurrence. Non-actual (imaginary) short circuits and fires could still be said to causally depend on each other. For two non-actual events the first counterfactual is the substantial one and the second is trivially true!

Causal dependence is not yet causation We have succeeded in defining *causal dependence*, but we ultimately want a definition of *causation*. Is *causal dependence* not just the same as *causation*? No. Here is one difference: as we have just learned, causal dependence may remain in the abstract (i.e. may also link non-actual events) but, for Lewis, causation is something factual: we may say that *c* caused *e* only if *c* and *e* actually occurred.

Let us consider actual cases and rephrase our question: why is causal dependence of actual *e* on actual *c* not just the same as the causation of *e* by *c*? Well, causal dependence of actual events at least entails causation:

> Causal dependence among actual events implies causation. If *c* and *e* are two actual events such that *e* would not have occurred without *c*, then c is a cause of *e*.
>
> (Lewis 1973: 167)

The missing link for a full identification of causal dependence with causation is the converse implication that if *c actually causes e* then *e depends causally/counterfactually on c*. Unfortunately, this direction is not always true (see Lewis 1973). In other words, causal dependence is sufficient but not necessary for causation.[8] Why? Can we see an example where we clearly believe to have a case of causation but not of causal/counterfactual dependence? Here is one: at precisely 5pm Suzy throws a rock at a bottle (this is event *c*), the rock hits it and the bottle shatters two seconds later (event *e*). If this is not a paradigm case of causation, what is? Suzy caused the bottle to shatter: *c* caused *e*.

So far so good, and yet the story goes on. Billy, Suzy's friend, has also been present and the two had the following deal: if Suzy does not throw at precisely 5pm then Billy throws, slightly harder, at one second past 5pm. Let us assume that Billy's potentially faster rock would also have hit the bottle two seconds past 5pm (both Billy and Suzy are very good at rock throwing and they never miss). Suzy has, so to speak, *a backup* in Billy and Billy's throw was *pre-empted* by Suzy's. More accurately, it was pre-empted *early* (before it even occurred; we will discuss tricky cases of *late* pre-emption later).

Back to Suzy We agreed that her throwing definitely caused the bottle to shatter. However, does the bottle shattering, *e*, depend causally/counterfactually on her throw, *c*? Is it true that *had she not thrown*, *the bottle would not have shattered*? Unfortunately not! As the story goes Billy, Suzy's backup, would have thrown and *the bottle would still have shattered*. So, we have $\neg c \:\square\!\!\rightarrow$ *(still) e* instead of $\neg c \:\square\!\!\rightarrow \neg e$ which is required for causal/counterfactual dependence. This first Suzy/Billy story (there's one more to come) is a case where we have causation, but where we do not have causal/counterfactual dependence.

Modally fragile events Here is one way in which we could react: if Billy throws and the bottle breaks, then this is a different bottle breaking, e^*, from Suzy's, *e*. It will come a second later or earlier and the pieces of glass will fall differently because of the different strength and angle of Billy's throw. Thus, it is not the same shattering, *e*, that would occur if Billy throws, but a different one, e^*. Although $\neg c \:\square\!\!\rightarrow e^*$ is true, $\neg c \:\square\!\!\rightarrow \neg\, e$ remains untouched.

Lewis argues against this intuition. He believes that events are not so *modally fragile* as the example wants us to believe. *Modal fragility* is Lewis's term for the idea that an ever so slight change to an event would turn it into a different one (see Lewis 1973: 196–9; Lewis 2000: 86).[9] He points out that we usually are not that picky when it comes to the individuation of events. For example, we say that we postpone a seminar from September to October, not that we cancel it and schedule a new one (see Lewis 1973; 2000: 86). It is the same with Billie's e^* and Suzy's *e*: they are, after all, just the same event.

Causal chains Here is Lewis's own idea for how to deal with the early pre-emption case (see Lewis 1973: 171–2, 200). Suzy, so we intuitively said, caused the bottle to shatter; however, there was no (direct) counterfactual dependence between her throw and the shattering because if she had not thrown Billy would have and the bottle would have shattered nonetheless. However, without a direct dependence, says Lewis, there is still a chain of intermediate events leading from her throw to the bottle shattering: the stone being in her hand, the stone being in mid-air at point 1, then at point 2 . . . until it hits. For each link in this chain there are true counterfactuals: had Suzy not thrown, the stone would not have been at 1, and had the stone not been at 1 it would not have been at 2 . . . and, finally, had it not been at *n* it would not have hit the bottle. Call such a sequence of counterfactual/causal dependencies, i.e. $\neg c \,\Box\!\!\rightarrow \neg p_1 \ldots \neg p_n \,\Box\!\!\rightarrow \neg e$, *a causal chain*. (You can make this chain as tight as you like but let us focus only on a point right in the middle between Suzy and the bottle).

Lewis claims that all the conditionals in the chain are true, even in Billy's presence. Take, the first one: had she not thrown, *her* stone would not have been in mid-air. This is true even if, had she not thrown, Billy would have thrown his stone. To see why the second counterfactual is true is more complicated. For now, let us take for granted that had her stone not been in mid-air, the bottle would not have shattered. (We will return to this case later). Thus, we might want to say that Suzy caused the shattering because there is such a causal chain, a chain of true "smaller" counterfactuals (instead of just the one (false) "big" counterfactual).

Causation Here is Lewis's (almost) final definition of what causation is:

> One event is a cause of another iff there exists a causal chain leading from the first to the second.
>
> (Lewis 1973: 167)

With the *iff* we now have both directions, i.e. necessary and sufficient conditions for causation.

Before we turn to cases that caused problems for Lewis's initial theory – because our intuitions about them and what the theory says fall apart – we should look at presuppositions and consequences of Lewis's definition, which might have gone unnoticed.

Transitivity First, Lewis's definition makes causation transitive, that is, whenever c causes c^* and c^* causes e, then c causes e.[10] This is simple to see: the two causal links – c to c^* and c^* to e – connect seamlessly to a causal chain. Thus, qua chain, c also causes e. There is a debate about whether we should accept transitivity. We do not want to spend much time on this issue but we wish to give one example where transitivity is at best controversial: someone is bitten by a deadly poisonous snake. This causes him to reach for an antidote. The antidote prevents him from dying. The snakebite therefore causes him to ingest the antidote, the antidote is causally responsible for him to still be alive and, by transitivity, the snakebite causes him to be alive. This does not seem right.[11]

Truth conditions for counterfactuals and similarity of worlds We have to point out two tacit assumptions of Lewis's theory. They both concern the *truth conditions of counterfactuals*. In Section 3.5, we introduced Lewis's theory at length but here's a brief reminder: a counterfactual is true in our world iff, within the set of possible worlds where its antecedent is true, those worlds are more similar to ours where its consequence also obtains than those where it doesn't obtain. His definition depends crucially on the similarity of worlds. As we saw in Section 3.5, the parameters with which Lewis measures similarity are ranked, and this ranking is crucial to answer an otherwise deadly objection to his causal chain solution to Suzy's early pre-emption of Billy's throw. We will now explain why this is so.

Remember that we postponed the argument why the second of the "small" counterfactual dependencies – 'had her stone not been in mid-air, then the bottle would not have broken' – is true. This is tricky, particularly if someone's objection is: 'This is not right because had the stone not been in mid-air, then obviously Suzy would not have thrown it in the first place. However, if Suzy had not thrown the stone, Billy would have (according to their deal). Yet, if Billy had thrown the stone, the bottle still would have broken. Thus, it is not right to claim that if Suzy's rock had not been in mid-air, the bottle would not have broken. One of the links in the alleged causal chain does not hold its promise. This theory forces you to deny that Suzy caused the bottle to shatter'.

Backtracking This objection sounds compelling and the crucial bit of reasoning, called *backtracker*, is that had the stone not been in mid-air, Suzy would not have thrown it. It is called backtracker because the temporal order of the antecedent and the consequent is reversed: the antecedent is in the future of the consequent. Now, Lewis's way to deal with the overall objection is to argue that *all* backtrackers are wrong: 'The Proper solution [. . .] is flatly to deny the counterfactuals that cause the trouble' (Lewis 1973: 170).[12] The objection is grounded in something false. Fair enough, but what's the argument?

We hinted at the fact that the ordering of the aspects under which the similarity of worlds should be measured comes crucially into play.[13] We repeat these aspects here (see also Section 3.4):

> (i) It is of the first importance to avoid big, widespread, diverse violations of law.
> (ii) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
> (iii) It is of the third importance to avoid even small, localized, simple violations of law.
>
> (Lewis 1979, quoted in Lewis 1986: 47–8)

According to these aspects, let us evaluate both the counterfactual in the chain and the objector's opposing backtracking counterfactual.

Both counterfactuals start with 'had the stone not been in mid-air . . .'. The actual intended counterfactual continues '. . . then the bottle would not have broken'. The unwanted backtracker continues '. . . then Suzy would not have thrown'. Which one is true? Only the first, not the second. Here is Lewis's reasoning and his advice, along with his similarity standards, is:

> It will normally be best not to diverge at all from the actual course of events until just before the time of [the antecedent of the counterfactual; here: the stone not being in mid-air]. The longer we wait, the more we prolong the spatiotemporal region of perfect match between our actual world and the selected alternative [in accordance with (ii) from the list].
>
> (Lewis 1973: 171, my additions in brackets)

Now, the closest world to ours where the antecedent is true – due to a small miracle which allows the stone to disappear (small miracles are somewhat OK: (iii)) – is, according to this advice, one in which everything in the past until that miraculous point in time is just as in ours. This includes Suzy's but not Billy's throwing! Follow that world further into the future and the bottle does not break (because there is neither Suzy's nor Billy's stone).[14] Our intended counterfactual – 'had the stone not been in mid-air, then the bottle would not have broken' – is true but not the opponent's backtracker – 'had the stone not been in mid-air then Suzy would not have thrown' – and Lewis's solution to early pre-emption is secured.

Late pre-emption Why did we call Suzy's throw an *early* pre-emption of Billy's (only potential) throw? Well, there are also cases of *late* pre-emption. Here's one: Suzy and Billy play their game again. This time, however, they indeed *both* throw and they do so *simultaneously*. However, Suzy's throw was slightly harder and her rock hits the target a split second before Billy's. His stone lands amidst the pieces of glass. Suzy's throw pre-empts Billy's success very late. 'So what?', we might first say. 'What's the difference to early pre-emption?' Well, unfortunately, we cannot rely upon the causal chain story in order to save our intuition that Suzy caused the bottle to shatter because there is no such causal chain. Let us focus on what should be the chain's very last link: had Suzy's stone not been in the air (a millimetre just in front of the bottle), then the bottle would not have shattered. This is unfortunately wrong and no prohibition of backtrackers can help: the bottle would still have shattered because, as things stand now, Billy's stone was already on its way and would have destroyed the bottle. Lewis's first strategy (causal chains plus prohibition of backtrackers) is not effective here.

Trumping Here's another similar case: suppose a sergeant and a major simultaneously command 'Advance!' and their troops advance. The soldiers know that a major's command trumps a sergeant's command, and the soldiers will do what the major says, not what the sergeant says. And so here: it is because of the major's command, not because of the sergeant's, that the soldiers advance. The major caused them to move – he trumped the sergeant's command.

Trumping could be seen as an extreme of pre-emption: causal chains hit their target at the same time (unlike the stones) and they come as close as they can just before the effect. It's only there, at the very end, where one is more potent or more authoritative and, thus, alone effective.[15] We mention trumping on top of pre-emption because trumping made Lewis dismiss a first solution he had to late pre-emption – a kind of copy-and-paste idea with quasi-dependence at its heart (see Lewis 1973: 202–7). Being unsuccessful regarding trumping (see Lewis 2000: 83), we skip this intermediate amendment of his theory and immediately turn to the ultimate one which is a solution to both.

Causation as influence Lewis (2000) offers not only a solution to pre-emption and trumping, but rather a whole new theory of causation. Here is his core idea in a metaphor: suppose you come across a complicated machine and you do not know which bits are connected to which. You wiggle the first element to see what else wiggles, and then the second, and so on. Slowly, you find out what the levers and wheels are connected to in the machine. Causation between events is just the same, only that you cannot literally wiggle the first event in order to see what it causes. However, it can be wiggled counterfactually and 'seeing what else "wiggles" when you "wiggle" one or another event tells you which ones are causally connected to which' (Lewis 2000: 91).

Lewis's ultimate Definition of Causation This includes the following new definition of causal *influence* (instead of the old *causal dependence*) and also of causation itself:

> Where C and E are distinct actual events, let us say that C *influences* E if and only if there is a substantial range C_1, C_2. . . of different not-too-distant alterations of C (including the actual alteration of C) and there is a range E_1, E_2. . . of alterations of E, at least some of which differ, such that if C_1 had occurred, E_1 would have occurred, and if C_2 had occurred, E_2 would have occurred, and so on. Thus we have a pattern of dependence of whether, when, and how upon whether, when, and how.
>
> As before, causation is the ancestral: C causes E iff there is a chain of stepwise influence from C to E.
>
> (Lewis 2000: 91)[16]

We see this work best when we apply it to Suzy and Billy's late pre-emption case.[17] We wish Suzy's throw to come out as the cause of the bottle's shattering but not Billy's. So, is it true that wiggling Suzy's would wiggle the bottle's shattering? Lewis thinks that it does:

> altering Suzy's throw slightly while holding Billy's fixed would make a lot of difference to the shattering, but altering Billy's throw slightly while holding Suzy's fixed would not. Take an alteration in which Suzy's rock is heavier, or she throws a little sooner, or she aims at the neck of the bottle instead of the side. The shattering changes correspondingly. Make just the same alterations to Billy's pre-empted throw, and the shattering is (near enough) unchanged.
>
> (Lewis 2000: 92)

A similar story can be told for the trumping case (see Lewis 2000: 92). Thus, if we define causation through (wiggling) influence, we can formally attribute causation to both Suzy and the major just as our intuition does. This welcome result could almost end our introduction to Lewis's theory if there were not further problematic cases. We must at least hint at them because they are test cases for any theory of causation, not only Lewis's.

Overdetermination One is overdetermination. Reconsider Mackie's short circuit that caused the fire. Suppose there had not only been the short circuit but also that, at the same time, a burning match fell onto an inflammable rug. The fire was overdetermined: had the short circuit not occurred, the house would still have been on fire because of the match; and had the match not fallen on the rug the house would still have burnt down because of the short circuit. Wiggling any one of the two intuitive causes would not really change the resulting fire (because the other has its full impact). Thus, unintuitively, Lewis's theory would be committed to saying that neither is a cause.

Instead of giving a solution to this problem, Lewis points out that our intuitions are not as clear as we have just made it seem. Contradicting our assumption, some people might not believe that both the short circuit and the match are indeed causes. However, when intuitions are not univocal it is not fair to blame a theory for unintuitive judgements (we would disagree about what would make a satisfying answer). We might therefore be excused if we simply disregard overdetermination cases (Lewis 1973: 171, fn. 12).

Double prevention Another hard test case for theories of causation is double prevention (see Lewis 2000: 83–4): assassin$_1$ kills assassin$_2$ who would otherwise have murdered assassin$_3$. Now that assassin$_3$ is unthreatened, she successfully kills the originally intended victim. The potential prevention by assassin_2 of assassin_3 murdering the victim was itself prevented by assassin_1: a double prevention. How do we judge the causal role of assassin_1? Was he a (partial) cause of the victim's death? We leave this for the reader to decide (and to figure out what Lewis's answer should be).[18]

Towards new shores With his counterfactual wiggling idea Lewis might capture why Suzy's stone throw is the cause of the bottle shattering, despite Billy's throw in the late pre-emption cases. Yet isn't Lewis beating about the bush? Isn't the truth much more obvious: it was Suzy's not Billy's stone that hit the bottle. It was *her stone*, not Billy's that *transferred energy, mass* and/or *momentum* to the bottle. This is what makes it the cause. In light of this idea, Lewis's counterfactual wiggling seems only to mimic the actual force of that stone.

Some philosophers have given a down-to-earth physical answer to the question of what causation is rather than to utilise ephemeral counterfactuals. Their general analysis comes in terms of the transfer of conserved physical quantities, such as mass/energy or momentum. This is the next approach we will look at.[19]

BOX 5.3 Counterfactual theories

- Lewis posits the equivalence of counterfactual and causal dependence for token events c and e:
  e* depends causally on *c iff ***e* depends counterfactually on *c*, i.e.** iff **(i)** $c \mathrel{\Box\!\!\rightarrow} e$ **and (ii)** $\neg c \mathrel{\Box\!\!\rightarrow} \neg e$.
- However, causal dependence is not quite causation. Although it is sufficient, **causal dependence is not necessary for causation**. At best, causal chains are where **a *causal chain* is a tight but finite sequence of counterfactual/causal dependencies** (think of **early pre-emption** Suzy and Billy case): $\neg c \mathrel{\Box\!\!\rightarrow} \neg p_1, \ldots, \neg p_n \mathrel{\Box\!\!\rightarrow} \neg p_{n+1}, \ldots, \neg p_m \mathrel{\Box\!\!\rightarrow} \neg e$.
- However, because of **late pre-emption** (the return of Suzy and Billy) and trumping cases (major versus sergeant), Lewis had to amend his original theory. **Counterfactual wiggling** becomes the crucial test for whether there is a causal relation. Lewis calls his ultimate account ***causation as influence***:

  > Where *C* and *E* are distinct actual events, let us say that *C* **influences *E*** if and only if **there is a substantial range** $\mathbf{C_1}, \boldsymbol{C_2}$**. . . of**

different not-too-distant alterations of *C* (including the actual alteration of *C*) and there is a range E_1, E_2. . . of alterations of *E*, at least some of which differ, such that if C_1 had occurred, E_1 would have occurred, and if C_2 had occurred, E_2 would have occurred, and so on. Thus we have a pattern of dependence of whether, when, and how upon whether, when, and how. (As before, causation is the ancestral: ***C* causes *E* iff there is a chain of stepwise influence from *C* to *E*.)**

(Lewis 2000: 91)

- Critical issues remained: Lewis's stipulation of the **transitivity of causation** (which might not be true) and that (most) **backtrackers** should always be false is contentious. Also, **overdetermination** and **double prevention** are problematic.

5.4 Transfer theories

By throwing the stone, Suzy transfers energy and momentum from her hand to the stone, enough to launch it onto a trajectory with an endpoint at the bottle. On impact, the stone transfers much of its momentum to the bottle, enough to make the glass' internal bonds burst. The bottle shatters. This last part – the bottle shatters – is not true of Billy's rock in both the early and late pre-emption cases. His rock does not transfer energy to the bottle.

So goes the story of Suzy and Billy told by a physicist in semi-scientific terms. Some philosophical theories of causation, first processes and mark transfer theories of John Mackie (Mackie 1974) and Wesley Salmon (Salmon 1984), and conserved quantity theories by Phil Dowe (Dowe 1992, 2000) and Max Kistler (Kistler 2006) make use of this approach. Here is a brief outline of Kistler's conserved quantity theory:

> Two events c and e are related as cause and effect if and only if there is at least one physical quantity, P, subject to a conservation law [for example, energy, momentum, charge, etc.], exemplified in c and e, and a determinate amount of which is transferred between c and e.
>
> (Kistler 2006: 72, my addition in brackets)

Before we explore this in more detail, note that these theories are a kind of Naturalisation program. They ask whether science can discover a natural phenomenon that corresponds to the intuitions we have about causation. This seems to be what Salmon calls out for: 'Let us take Hume's challenge seriously: let us try to find a physical connection between cause and effect' (Salmon 1998: 16).[20]

An anti-Humean connection? The quote from Salmon makes it sound as if transfer theories diverge from the Humean program. This is true only to a certain extent (see Salmon 1984: 136–8, 147; Mackie 1974: 228–9). Contra Hume, transfer theories indeed point to a connection in nature that underlies causation: energy or momentum or mass that is transmitted from cause to effect.[21] However, this transfer of a physical magnitude is empirically verifiable and not postulated *a priori* like the hidden powers Hume argued against. This should be welcomed by the Empiricist. The connection by transfer also has no modal force because it is not of the necessitating kind Hume argued against. To put it crudely, it is, after all, a mere factual regularity that there is a transfer of energy in those cases. Still, it is a connection rather than none and it is not secret or inexistent as Hume thought.

We now turn to the details of the causal theories, which derive from two generations: Dowe's and Kistler's transfer theories, which we have hinted at earlier, are the successors to Mackie's and Salmon's earlier mark transfer or process theories. We will only briefly discuss the earlier historical theories, and will concentrate on the more recent theories.

Processes and mark transmission After having abandoned his INUS theory (see Section 5.2), Mackie suggested that in order for there to be a causal link amongst events 'some *continuous process* connecting the antecedent in an observed conditional regularity with the consequent' (Mackie 1974: 82, emphasis added) has to be going on. Mackie introduces a new category into our ontology: *processes* (instead of events). Salmon, on whose theory we concentrate, concurs: 'One of the fundamental changes I propose in approaching causality is to take processes rather than events as basic entities' (Salmon 1984: 139). What is a process? Salmon gives an example:

> A baseball colliding with a window would count as an event; the baseball, traveling from the bat to the window, would constitute a process.
> (Salmon 1984: 139–40)

We can infer that Suzy's stone moving through the air towards the bottle is also a process; however, discussion about processes alone does not deliver a theory of causation. In particular, not all processes are conducive to causation and Salmon goes to some effort to distinguish feasible (causal) from useless (pseudo) processes:

> The basic method for distinguishing causal processes from pseudo-processes is the criterion of *mark transmission*. A causal process is capable of transmitting a mark; a pseudo-process is not.
>
> (Salmon 1984: 142, emphasis added)

Very well, but what is a *mark transmission*? Again Salmon gives examples. If a moving car, which is a real process, scratches a wall then this single, local interaction leaves a lasting mark on the moving car. If, however, the car's shadow flits over the wall's surface and *collides* with, say, a bow window, the shadow is deformed only momentarily but resumes its shape immediately after (see Salmon 1984: 143). The shadow does not have a long-lasting mark. Moving shadows are therefore only pseudo processes.

Causation as interaction of causal processes Causation occurs where causal processes interact and they can be said to interact if they leave marks on each other. Suzy's stone hits the bottle which then shatters. The stone has a certain momentum, the bottle its internal molecular bonds. After the collision, the stone loses most of its momentum and the glass now exhibits the modified characteristic of being scattered in pieces on the ground. These two causal processes – the stone's flying and the bottle sitting there – interact and leave marks on each other. Note that Billy's rock causes no trouble for this theory, regardless of whether it's thrown early or late (see Section 5.3), because it is Suzy's stone that leaves the mark. It has therefore caused the pile of shards.

This is Salmon's early theory in very broad brushstrokes. For the sake of brevity, we have omitted many details and now turn to a more recent variant of it.

Conserved Quantities Theory Salmon accepts the variation of his theory that Dowe (1992, 2000) suggested and that Kistler (2006) later further developed. Salmon tells us what the essential change is:

> Dowe [. . .] pointed out that *causal processes transmit conserved quantities*; and by virtue of this fact, they are causal. I had come close to this point by mentioning the *applicability of conservation laws to causal interactions*, but did not take the crucial additional step.
>
> (Salmon 1994: 303, emphasis added)

In other words, causal processes are no longer identified by their abstract ability to *transmit marks* but by the fact that they possess and transmit *conserved physical quantities*:[22]

> A conserved quantity is any quantity universally conserved according to current scientific theories. Some conserved quantities are mass-energy, linear momentum, angular momentum, and charge.
>
> (Dowe 1992: 210)

If current scientific theories are correct, then, because of the (conservation) *laws of nature*, it is ascertained that whatever amount of these magnitudes you put into a system or process it will remain there unless a fraction of it is transferred to somewhere else. The sum total of the remaining amount and the transferred amount will always equal the initial amount. This makes the transfer of specifically these quantities particularly apt to figure in a theory of causation. Ephemeral quantities (*marks* according to Salmon) that come and go in an unconstrained way, would not be reliable or constant enough. Moreover, a Naturalisation of the notion of causation is achieved: what the laws are and, thus, what causation is, depends on what the natural world is like. We also see once more how intertwined the notions of causation and laws are.

Modified Process Theory of Causality Causal processes transmit these conserved quantities, but that is not a definition of either causation, causal processes or causal interactions. At the beginning of this section, we presented a definition by Kistler. Here is Dowe's 'basis of the modified process theory of causality':

> Definition 1. A *causal interaction* is an intersection of world lines which involves exchange of a conserved quantity.
>
> Definition 2. A *causal process* is a world line of an object which manifests a conserved quantity.
>
> (Dowe 1992: 210, caps removed, emphasis added; see also Salmon 1998: 251, 257)

Let us clarify the core concepts that Dowe uses here: *causal interactions* do not involve just any mark transmission but specifically only the change in the value of a conserved quantity. This exchange is, as mentioned earlier, governed by a conservation law. The essential feature of *causal processes* (distinguishing them from pseudo-processes) is the manifestation of a conserved quantity. (This catapults moving shadows and whatever features they have immediately into the realm of pseudo-processes because they do not possess energy, momentum or charge). The *world line of an object* can (we are slightly simplifying here) be indentified with its trajectory through space. An *object* is 'anything found in the ontology of science (such as particles, waves or fields), or common sense' (Dowe 1992: 210).[23]

To see the theory at work we should apply it to our example, Suzy's and Billy's stone throw. In fact, we have already done so by applying Salmon's earlier mark transmission idea to Suzy and Billy in terms of *momentum transfer.* We did so because the latter is an instance of the former. However, there is a drawback, which is discussed later.

To further understand Dowe's (and Kistler's) theory let us examine the counterexamples that the theory has been confronted with. Needless to say, most of them work against the earlier mark transfer theory, too. The counterexamples we list have become standard test cases for theories of causation.

Negative causation: absences and omissions Philosophers who are also professional gunmen report that many pistols work in the following way (see Schaffer 2000: 287–8): a striker is connected to a compression spring under tension; however, the spring does not (yet) propel the striker forward because the striker is held back by a safety pin. If let loose it would hit the cartridge which would then explode and shoot the bullet out of the barrel. All this sounds pretty easy to handle for a conserved quantity theory of causation. In fact, it sounds as if we could use it as a paradigm example: the spring's potential energy is turned into the kinetic energy of the hammer which hits the cartridge and causes it to explode. The energy of the explosion is turned into air pressure which bestows the bullet with momentum and so forth.

Unfortunately, the crux of the problem is that if we ultimately want to claim that pulling the trigger caused the bullet to hit the target we fail, at least in terms of the transfer theory. This is because by pulling the trigger with your finger you move the safety pin that kept the spring compressed. In other words, instead of there being any kind of energy transfer from your finger to the spring, striker or bullet, you are simply taking away a piece of metal. No conserved quantity reaches any of the crucial further parts of the gun from your finger. What you're achieving is an essential absence, the one of the safety pin, but you do not *cause* anything according to transfer theories.

Absence is the key word here. For transfer theories it is extremely difficult to handle what is called *negative causation*: causation by absences and omissions. Other examples include forgetting to water someone's plants while they are on holiday causing them to whither, or the double prevention case we introduced in the section on Lewis's causal theory (no transfer of a conserved quantity from you to the plant or from assassin$_1$ to the victim).

Too much causation While causation by absences causes trouble because these causal cases fall through the net of transfer theories, there are also intuitively non-causal cases which the transfer theories count as such. Take Billy's stone in the late pre-emption case again. It might have pushed forward some air molecules which reach the bottle at the moment Suzy's stone hits. These minute molecules, which got their momentum from the stone, transfer minimal energy to the bottle. Due to this transfer of energy Billy's throw has to count as a cause of the bottle's shattering. The transfer theory did not specify a minimally necessary amount.

Reductionism A further problem is that conserved quantity theories are either committed to a radical *Reductionism* for physics (see Dowe 2000: chapter 7), or they have to acknowledge that they do not capture much of what we call causation in the special sciences and everyday life. For example, suppose Suzy, instead of throwing a rock, kisses Billy and he blushes. His blushing has clearly not been brought about by the energy transfer between them (only if an esoteric reading of *energy* were allowed); rather, psychological and biological factors play the crucial role here (this example is from Collins *et al.* 2004: 14).

Interlude: dualism, pluralism, nihilism Remember that nothing in Lewis's counterfactual or Mackie's INUS theory of causation refers essentially to physical properties. For them, negative events, such as the absences of deeds, actions and happenings, might be unproblematic.[24] Other than the transfer theories, these accounts could be better suited to deal with omissions, absences and, generally, non-physical causation. The latter cases seem to be more a matter of pure counterfactual dependence rather than the transfer of a conserved quantity (see (MOD) in Section 2.1.2).

However, keep in mind that conserved quantities approaches to causation handle physical cases like Suzy's and Billie's stone throw much more straight forwardly. These are cases of causal *production* (see (PROD) in Section 2.1.3).

The fact that different causal theories score differently in different games leads some philosophers to the belief that causation is a word not for one unified phenomenon but for at least two. For example, in 'Two Concepts of Causation', Ned Hall (2004) argues that some of our intuitions about causation characterise it as production, while other intuitions characterise it more as (counterfactual) dependence.[25]

Other philosophers even go *pluralist* and claim that there are actually more than two concepts of causation (Psillos 2009).[26] Still others vote for *nihilist options*. John Norton (2003), for example, believes that our everyday intuitions about causation are so diverse that there is simply nothing in the world that corresponds to our confused idea. The concept of causation should be dropped from our philosophical vocabulary altogether (Norton 2003). This is, of course, similar to Russell's (1912) early onslaught on causation. It is also akin to the *Canberra Plan*'s method to do metaphysics, which we discuss in Section 7.6 (where we also look in more detail at Norton, Russell and Psillos).

Addendum: mechanistic theories When Mackie abandoned his INUS theory he referred to both processes and *mechanisms* in one breath: 'some "causal mechanism", some continuous process connecting the antecedent [. . .] with the consequent' (Mackie 1974: 82) Recently, some philosophers have taken this mechanistic manner of speaking literally and they indeed define causation in terms of tangible material mechanical devices, be they purposely built (like guns) or biologically evolved (like hearts or kidneys in mammals) or otherwise natural (the solar system). Energy, momentum and charge will certainly be the fuel to run these mechanisms, and yet these quantities no longer figure centrally within the relevant theory of causation. It is worth having a brief look at mechanistic theories of causation. We will largely concentrate on Stuart Glennan's work (see Glennan 1996, 2002, 2009, 2010, 2011).[27]

From the abstract to the concrete Mechanisms are real, structured objects with interconnected parts that engage into an orchestrated overall behaviour. Parts might themselves be sub-mechanisms but that is not too important for our purposes:

> A mechanism underlying a behavior is a complex system which produces that behavior by the interaction of a number of parts according to direct causal laws.
>
> (Glennan 1996: 52)[28]

Examples include combustion engines, solar systems, hearts and lungs:

> Two of the many subsystems of the human body are the cardiovascular and respiratory systems. Each of these systems has mechanisms for doing certain things (pumping blood, inhaling oxygen and exhaling carbon-dioxide).
>
> (Glennan 1996: 52)

Mechanistic causation With this characterisation and examples in mind we can already dare to look at the account of causation that is given in mechanistic terms:

> Roughly put, a mechanical theory of causation suggests that two events [*c* and *e*, say; MS] are causally connected when and only when there is a mechanism connecting them.
>
> (Glennan 1996: 64)

That is, *c* causes *e iff c* is the input or activation of a mechanism such that the mechanism's parts start to interact in such a way that the overall mechanism produces behaviour *e*.

Challenges to a mechanistic theory There is a multitude of challenges for this theory of causation. Some, like omissions and absences, are obvious and inherited from process and conserved quantity theories (change mechanism for process and conserved quantity and you have the objection). Others are more hidden but become visible when we start thinking about what a mechanism's parts do. However, note that other than the process theories, mechanistic theories might be able to handle the kiss and the gun's trigger cases: the activation of a gun turns out to be causal because releasing a safety pin is a process within the overall mechanism, be there a direct energy transfer to the next part of the mechanism or not. Also, Billy's blushing on being kissed by Suzy might be *mechanistically* explained: Mechanistas allow biological, neuronal and maybe even psychological mechanisms into their non-physicalist, non-Reductivist ontology.

Causation at the fundamental level As long as we are dealing with day-to-day macro objects like combustion engines, hearts or even the brain's neurons, mechanistic theories seem to do fine. All the involved objects have complex subsystems with interacting parts that are crucial for mechanistic explanations. However, in fundamental physics we might either hit rock bottom or fall through ephemeral clouds. At some point, there are either fundamental physical entities without a substructure (quarks or Higgs bosons or . . .) when no mechanistic story can be told about their behaviour because they do not have the requisite parts that could interact, or – and maybe worse – the fundamental physical ontology knows no discrete entities at all but only waves or fields or superstrings.[29]

Glennan (and other Mechanistas) are aware of both variants of the problem (see Glennan 2011: 809). In response, Glennan endorses dualism: 'A complete causal theory requires a [further] theory of fundamental causal interactions' (Glennan 2010: 381, my addition in brackets). According to Glennan (2002), this separate theory for causal interactions amongst fundamental phenomena could be explicated in terms of counterfactual dependence.

Preview: dispositionalism There's a different, radical way in which one could deal with the fundamental level: adopt something that is itself more akin to mechanisms, but more akin in which respect? Before we give the answer, we have to mention something about mechanisms that we have deliberately ignored so far but which is a crucial element within the mechanistic story.

From the beginning, Mechanistas use anti-Humean vocabulary to describe what a mechanism is: a complex system which *produces* behaviour (see Glennan 1996: 52) *or* entities and *activities* that are *productive* of changes (see Machamer *et al.* 2000: 3). We know this *activity* and *production* talk from our discussion of *dispositions*: dispositions (or objects that have them) are naturally *productively responsible* (PROD) for the coming about of their effects (see Section 2.1.3). Thus, in line with the acceptance of anti-Humean concepts like *production* and *activity*, the Mechanistas would be well advised to wholeheartedly embrace a Realist dispositions ontology (see Sections 2.2.3 and 6.2) where, at the fundamental level, properties are dispositional powers.

As we will see in Chapter 6, some philosophers defend the radically anti-Humean view that most or all fundamental properties are dispositional powers that produce or bring about certain reactions, *e*, if triggered by the appropriate circumstances, *c*. Note that a theory of causation in terms of (fundamental) dispositions would be structurally akin to the Mechanistas theory: *c* causes *e* iff there is a *dispositional power P* on the micro level, had by some involved fundamental entity, which is the disposition to manifest *e* when *c*-ed. Compare this to the macro level where *c* causes *e* when there is a *mechanism* connecting *c* and *e*.

Following this, the threatening causation-dualism would be overcome. For example, instead of postulating a counterfactual theory for causation on the fundamental level, say that fundamental entities have fundamental dispositions. The fundamental dispositions could then be perceived as unanalysable, basic *micro-mechanisms*.[30]

Further critique Note that apart from general Humean doubts about this move some people might worry that such a theory is running around in circles: *c causes e iff c has the power to produce e* is a bit short of informative content.[31] Yet, maybe this is just a caricature of what proponents of powers have in mind (see Section 6.4).

BOX 5.4 Transfer theories

- Differing from the other theories of causation we have introduced so far, transfer theories concentrate on **causal processes** rather than on successions of separate but causally related events.
- Transfer theories come in an older and younger variant (first **mark transmission** and then **conserved quantities transfer**) and they seek to establish the **causal connection that Hume thought was secret**.
- However, the link they do **empirically** find, namely mark or conserved quantity that is transferred in a causal process, is **not of a necessitating, modal nature** and thus **not** as **anti-Humean** as it first seems.

- Focusing on the newer theory, causation is defined roughly in this way: **a causal interaction is an intersection of two causal processes that involves an exchange of a conserved quantity**.
- If a translation into events-talk is allowed, this would read: **two events *c* and *e* are causally related iff there is at least one conserved physical quantity** (for example, energy, momentum, charge), **a determinate amount of which is transferred between *c* and *e*.**
- **Mechanistic accounts of causation** replace the transfer of a mark or the flow of energy by **mechanisms** (in more or less the literal sense of mechanism): roughly, **two events *c* and *e* are causally connected iff there is a mechanism connecting them**.
- Since the behaviour of mechanisms and their parts is described in causal vocabulary – centrally in terms of **production** or **activity** – mechanistic theories of causation are not entirely reductive and are certainly anti-Humean.
- This would be especially so if **Mechanistas** were to pair up with **Dispositionalists**.

5.5 Interventionist theories

Suppose you want to find out whether a drug is effective against a certain illness. Randomised clinical trials help you to do so. Here's roughly what they do: you give the drug to a large group of people who are suffering from the disease it is supposed to cure. You administer a placebo, not the drug, to another group of patients. If there are significantly more and/or faster cured people in the group that received the drug than in the placebo control group, you are justified to believe that the drug caused the (swift) cure.

But not so fast. There are many traps on the way. For example, if by chance you sorted the two groups such that only young and otherwise healthy people ended up in the drugs group and only elderly and those of general poor health in the placebo group then we have reason to doubt that the drug was effective – or at least as effective as it seemed to be – because the good overall physical constitution of the patients in the drugs group would have had a significant contribution to the swift healing. These, and the unhealthy in the placebo group, distorts the outcome. If the drug is as ineffective as the placebo, then this flawed experiment's outcome would seem to indicate that the drug is potent.

Thus, for the causal inferences and predictions, which we wish to make from these kinds of trials, a set of cautionary measures have to be applied for them to be reliable. Otherwise we do not gain the knowledge for effective control and manipulation we wish to obtain. This holds not only for medical trials but also for epidemiology, biomedical research, psychology, social sciences, effective policy making, econometrics, etc.

Judea Pearl (2000), a computer scientist, and philosophers Peter Spirtes, Clark Glymour and Richard Scheines (1993)[32] have devoted their work to show which statistical methods, rules and algorithms lead to successful inferences of causal relationships from the observational data (for example, statistics about drug and placebo administration, illnesses and recoveries). However, we will not discuss their rather formal approaches to the subject, but instead turn to James Woodward's (2003) work who turned the previously mentioned researchers' findings into a philosophical theory of causation.[33]

The theory in slogan form Woodward's account is: 'X is a cause of Y iff there is a possible ideal intervention on X that changes Y' (see Woodward 2003: 45). We now need to spell out the details. We do so in following Woodward back to our medical trial. Remember, we have a population of subjects that all suffer from a disease. We intervene in this group of people: some receive the drug (the treatment group) and others (the control group) receive a placebo. It is customary in the language of such experiments to

> represent the treatment received by an individual subject u_i by means of a binary variable T that takes one of two values 1 and 0, depending, respectively, on whether u_i does or does not receive the drug.
>
> (Woodward 2003: 94)

Recovery, our goal, is similarly represented as a variable R which, again, can have value 1 or 0 depending on whether the subject is cured or not. Finally, let us call our intervention to give the drug or placebo I. Now, formulated in our variables language, what we ultimately want to know is whether the intervention I which sets the value of T to 1 for an individual u_i (i.e. whether giving them the drug) would lead to the value 1 of R (i.e. to the recovery of the subject) and whether the intervention I: T=0 for u_i would lead to R=0.

We cannot both give and not give the drug to one subject and see what happens in each case, and so we give it as proxies to half the population and not to the other. We saw that some additional conditions need to be fulfilled for the outcome of such a trial to be reliable. We will now ask, with Woodward, what these conditions ideally are (see Woodward 2003: 95). Other less ideal experiments might lead to reliable results that identify causes correctly and the purpose of finding the *ideal* conditions is to make the ultimate explication of causation that Woodward is after watertight, i.e. immune to counterexamples (see Woodward 2003: 95).

Four constraints on interventions One such condition has already been suggested: certain biased divisions of our whole target group in the placebo and drug subgroups need to be excluded. In addition to T (medication or not), there are most certainly other variables that are relevant for the recovery of subjects: their age, sex, general health, etc. The distribution of such additional factors should be evenly distributed through the placebo and drugs group. Theoretically speaking, we have to demand that our intervention, i.e. giving the drug or placebo, is without bias, i.e. (statistically) independent of any other variables *Z* (age high or low, general health good or poor) that are also relevant for the value of *R* (recovery or not). This is Woodward's condition I.4. We will examine the others in reverse order (I.3, I.2, I.1) and then put everything together by number.

Suppose that the drug came in very small pills that are easily swallowed with half a glass of water but that the placebo was administered as a painful injection. These ways of intervening, *I*, (i.e. giving the medicine) would probably also have a direct effect on the well-being of our patients. That is, our way of intervening would sidetrack the causal route we wish to test, thus, the way in which we intervene should not have other direct potential effects on R other than those that the setting of T to 1 or 0 might have. This is Woodward's condition I.3.

Clearly the experiment also fails if apart from our own intervention *I*, i.e. our own giving of the drug or placebo, there are others (illegal drug dealers) that affect *T*, i.e. secretly giving the drug to someone who received the placebo from us. Our intervention *I* has to be such that other interventions on *T* are prohibited. In other words, *I* has to be the sole *switch* for *T* (see Woodward 2003: 96–70). This is condition I.2.

Finally, I.1 is a rather trivial requirement on our intervention *I*: *I* really has to be a cause of *T* being either 1 or 0 for each individual. In our experiment, *I* has to be the intervention that the patients really took the drug or placebo instead of pouring it down the drain.[34]

This concludes our search for constraints on experimental interventions that have to be fulfilled in order for our trial to be successful. Similar stories could be told about other interventions like the lowering of taxes in order to allow a certain business sector to thrive, or the sprinkling of pesticides on plants to prevent insect attacks, and so on. However, we have to turn away from concrete scientific inquiries into what causes what and move towards an abstract and philosophical explication of what causation is.

For a start, combining I.1 to I.4, we can say what an ideal intervention is:

I is an (ideal) intervention on *X* with respect to *Y* iff *I* meets the following conditions:
I.1 *I* causes *X*.
I.2 *I* acts as a switch for all other variables that might interfere with the value of *X*. That is, *I* switches *off* other interferences with *X*.
I.3 *I* has no influence on *Y* other than through *X*, i.e. *I* is not a cause of any causes of *Y* that are distinct from *X*.
I.4 *I* is (statistically) independent of any other variables Z that also cause *Y*.
(Woodward 2003: 98)

Note two things: (i) in our example *X* was *T* and *Y* was *R*; (ii) an intervention on some *X* is always an intervention on *X* relative to some *Y*. If, for example, a healthy tan instead of recovery from illness is your aim, the original *I* might not be a relevant intervention (see Woodward 2003: 103). We will now explore these four conditions further (see Woodward 2003: section 2.2, 38ff).

Wiggling Woodward First, we can express them in Lewis's 'wiggling' vocabulary (see Section 5.3). I.1 says simply that with your intervention *I*, you wiggle *X*. I.2 demands that nothing else wiggles *X*. I.3 requires that your wiggling *X* with *I* does not also wiggle *Y*, or only so via the route of *X* and in no other way. I.4 says that you should not wiggle only those *X*s that are, for other reasons, biased towards or against *Y*, i.e. there should be a proper balance within the wiggled group.

Causal graphs Within interventionist theories of causation like Woodward's, conditions such as those mentioned earlier and also causal relations in general are often accompanied and illustrated by graphical representations, which can be causal graphs or path diagrams. This explains the name *causal graph theory* that is often associated with Pearl's, Spirtes *et al.*'s and Woodward's theories.

The second way we can rephrase I.1 to I.4 is therefore by graphical representations. The overall situation is this: $\mathrm{I} \Rightarrow \mathrm{X} \Rightarrow \mathrm{Y}$, i.e. *I* tempers with *X* which, in turn, changes *Y*. This already includes what we would draw for I.1: $\mathrm{I} \Rightarrow \mathrm{X}$. For I.2, we could draw arrows from *I* to other Z_i that could, next to *I*, also influence *X*, $\mathrm{I} \Rightarrow \mathrm{Z_i}$, and specify that (in the following equations) *I* switches off the Z_i. For I.3, we add $\mathrm{I} \not\Rightarrow \mathrm{Y}$. The statistical independence, I.4, is not easily represented in such equations. We could, somewhat artificially, write: for any Z_i, if $Z \Rightarrow Y$ then $I_{(\mathrm{Zi}/\neg\mathrm{Zi})} \Rightarrow X$ to indicate that *I* should be statistically independent of other *Y*-relevant factors like Z_i.

Causal models and structural equations Clearly our use of variables and arrows is a simplification or, less pejoratively, an *abstraction* of what really goes on at the sickbed. Our variables only stand for one factor T (drugs/no drugs) or R (recovery/no recovery). For a more realistic picture, other variables would also have to be introduced (for example, S: male/female; A: age low/age high) and considered in a larger causal graph. In this way we would achieve more complex models of what goes on in reality.

In the literature, people often speak of appropriate models of variable X causing variable Y or that a variable X taking a certain value causes variable Y to take a specific value. Of course, variables in an abstract model do not cause anything. What is meant in this manner of speaking is that the event, which the change in value of a variable represents, causes the event behind the other variable's change.

Structural equations Turning from facts or events back to the abstract model, we wish to mention, while not going into depth, that causal relations can be cast into so called *structural equations*. If the medication is indeed effective, we could write in its simplest form the following equation: $R = f(T)$, meaning that recovery is a function of drug intake. Both T and R can, in our simple model, only take the values 1 and 0. The outcome hoped for is that $R = f(T) = T$, i.e. that recovery and drug intake are positively correlated (if T is 1, R is; if T is 0, R is). Taking further variables on board can make these equations fairly complex and medical trials will most certainly only give statistical answers that say something about probabilities rather than strict one–one correlations. Equations might therefore look more like: $p(R) = f(T, A, O, S)$, where the probability p of recovery R is a complicated function of drug use T, age A, overall health O and sex S.

From type-level to token-level actual causation Equipped with our knowledge about ideal interventions (which meet I.1 to I.4) we can revisit Woodward's definition of causation (here in a slightly lengthier version than earlier:

> X causes Y iff (i) there is a possible ideal intervention that changes the value of X such that (ii) if this intervention were carried out and all other variables besides X and Y are held fixed at some value (maybe by other interventions), then the value of Y (or the probability of some value of Y) would change.[35]
>
> (Woodward 2003: 45, 55)

If the outcome of our medical trial is positive, it shows the researchers that *generally* the administering of that tested drug causes recovery in patients. In other words, we gain knowledge about a *type-level causal claim*: Xs generally cause Ys.

We might, however, also be interested in what the relation is

> between such claims and claims involving *actual*-, *singular*-, or *token-causal* claims, which we may understand as claims to the effect that X's assuming some actual value on some particular occasion (for some particular individual) caused Y to assume some actual value on that occasion[.]
> (Woodward 2003: 45, 55)

In other words, we return to our concrete question whether Suzy with her particular, actual stone throw on that afternoon caused that very bottle to shatter. Here is Woodward's definition for *actual causation* (AC):

> (AC1) The actual value of X equals x and the actual value of Y equals y.
>
> (AC2) There is at least one route R from X to Y for which an intervention on X will change the value of Y, given that other direct causes Z_i of Y that are not on this route have been fixed at their actual values. [. . .]
>
> Then X equals x is an actual cause of Y equals y if and only if both conditions (AC1) and (AC2) are satisfied.
> (Woodward 2003: 77, I replaced '=' with 'equals')

For clarification, we will now apply this to the Suzy and Billy case, but let us first define some variables: S shall represent Suzy's throw and it might take the values 1 if she does throw and 0 if she doesn't (S corresponds to X in the definition). B, standing for the bottle breaking or staying undamaged, is also a variable that takes the values 1 or 0 (B corresponds to Y). Finally, we represent Billy and his throwing or not throwing by the initial W of his full name William. Again, $W = 1$ or $W = 0$. Billy, i.e. W, is the only other potential cause Z_i of Y that the definition mentions.

We can now write down some structural equations: B takes the value of S or W, whichever is higher or if they are equal then that equal value. In other words, when either of the two or both throw ($S = 1$ or $B = 1$ or $S = B = 1$) then the bottle breaks ($B = 1$), but if neither throws, the bottle remains intact. The equation $B = \max(S, W)$, which stands for 'B takes the maximal value of S and W' expresses these dependences.[36]

But did Suzy and Billy not have a deal? If Suzy throws ($S = 1$), Billy refrains from throwing ($W = 0$), but if she doesn't throw, he does, i.e. if $S = 0$ then $W = 1$. This agreement has to be captured by an additional equation, namely $W = 1 – S$. Figure 5.1 shows this situation in a causal graph.

So, suppose Suzy does throw, $S = 1$. Then, according to the deal and the equations, $W = 0$ because $W = (1–S) = (1–1) = 0$, i.e. Billy does not throw. Still, according to $B = \max(S, W) = \max(1, 0) = 1$ the bottle shatters.

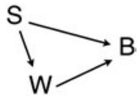

Figure 5.1

All this is preparation work. Let us see whether Suzy is the actual cause *according to Woodward's definition*: $S = 1$ is an actual cause of $B = 1$ if and only if both conditions AC1 and AC2 are satisfied. AC1 is the actual value of $S = 1$ and the actual value of $B = 1$. Both really take place: Suzy throws and the bottle breaks. How about AC2? The actual value of the only Z_i around, namely *W*, is 0. Let us hold it fixed as AC2 asks us to do. In other words, we imagine that we handcuff Billy (or similar) so that he cannot throw. Then, so AC2 further, we intervene in *S*, i.e. change its value. We set *S* to 0 (i.e. do not allow Suzy to throw). Would this intervention in *S* change the value of *B*? Yes! According to our scenario and equation $B = \max(S, \mathrm{W})$ which it represents, we get $B = \max(0, 0) = 0$. Thus, AC2 is fulfilled.

With this example we have killed two birds with one stone: we have seen *how* and, most important, that Woodward's definition *does* work in accordance with our intuition in an actual causation case. Note, moreover, that the scene depicted is the *early pre-emption* Suzy and Billy scenario. In other words, Woodward's theory can handle early pre-emptions.

At this point we have reached our first goal: to acquire a basic understanding of Woodward's theory of causation, especially of actual causation. We now have to turn to some details, doubts and consequences.

Circularity Maybe the reader felt uneasy throughout our definition attempts: the characterisation of the conditions for an ideal intervention employ *causal language* at a number of points. For example, it says that the intervention variable *I* must *cause X* (see I.1). Thereby, the notion of an intervention becomes itself a causal notion (see Woodward 2003: 104). In the definition of actual causation, reference is also made to *interventions*, to *direct causes* and to *fixing* actual values. All this is *causal vocabulary*.

Here is a serious reservation about the adequacy of Woodward's definition: it's hopelessly circular! We now aim to dispel this worry in order that the 'hopeless' in 'hopelessly circular' can be dropped. For a start, note that it is not surprising that we end up with an explication that contains causal terms if we start with actual, practical, scientific methods that are applied to find out *what causes what* and not what *causation*, metaphysically speaking, is: scientists are simply not interested in a reductive analysis of the concept of causation. On the philosophical side of things, Woodward comments:

> There is a very widespread tendency in philosophical discussions of causation and explanation to assume that any interesting account of these notions must be 'reductive.' [. . .] The account that I present is not reductive, and I am skeptical that any reductive account will turn out to be adequate. [. . .] An account of causation and explanation can be worthwhile and illuminating without being reductive.
>
> (Woodward 2003: 20–1)

In the end, the reader has to judge whether getting to know Woodward's theory has been illuminating (as we think it is). Here is another observation.

Permissible possibilities We find many locutions of the kind '*X* causes *Y* iff there is a *possible* intervention', 'if an intervention *were carried out*', etc. in Woodward's theory. One reason for these modalities (possibility/counterfactuals) to enter is straightforward: there are lots of causal events in the world where no *actual intervention* has taken place or will ever take place, and yet in order for there to be causation, interventions only have to be *possible* not *actual*, otherwise many intuitive paradigm cases of causation would not count as such.

Of course, the possibility of interventions is not meant to refer to 'the present technological powers of human beings' (Woodward 2003: 46). If this were so, the definition of causation would have 'the unacceptable consequence that X cannot cause *Y* when human beings lack the power to manipulate *X*' (Woodward 2003: 46). Thus, Woodward has to extend the class of interventions to those that transcend the human factor:

> Although there will be realistic cases in which manipulations carried out by human beings will qualify as interventions [. . .] [there's] no reference to human activities or to what human beings can or can't do. Notions such as 'human agency' and 'freely chosen action' do not occur as primitives.
>
> (Woodward 2003: 103–4, my addition in brackets)

However, what kind of possibility is meant if not possibility within human reach? The truth is that it has to be a metaphysical possibility and transgress what the laws of nature allow. For example, the moon orbiting the earth causes the tides. What kind of intervention would show that this causal relation is achieved? Well, we would have to take the moon away or, at least, change it's orbit to see whether the tides also disappear (or change). However, any such intervention, if it is in accordance with the laws of nature, would mean, for example, that some massive collision with a gigantic comet has to happen. But this would not be the surgical intervention that our definitions demand: this massive intervention would *itself* alter the tides and thus I.3 would be violated (see Woodward 2003: 128–32).

> We thus arrive at the following conclusion: commitment to a manipulability theory leads unavoidably to the use of counterfactuals concerning what would happen under conditions *that may involve violations of physical law*.
>
> (Woodward 2003: 132, emphasis added)

In other words, Woodward's theory has to allow even for those (logically) *possible interventions* that are *counter-nomic* against the laws. The reader is reminded of Lewis's counterfactual theory of causation where (small) miracles in nearby worlds are required, and Woodward agrees that his theory has some aspects in common with Lewis's. However, there are differences. The most obvious is that Lewis's but not Woodward's analysis is meant to be a reduction of causation to entirely non-causal and, at best, ultimately non-modal vocabulary.[37]

Absences, double prevention, late pre-emption, absences/omissions We could now check how Woodward's theory deals with absences, double prevention, late pre-emption, and absences/omissions. We will focus on the latter. Are omissions a problem for the interventionist? They are not.

Remember representations by variables: variables may assume the value 0, which stands for the event not having occurred. As long as there is a potential intervention on that variable, setting it to 1, for example, which would have a change in the target variable as a result, there is also the potential to establish a causal relation. You did not water the plants, $W = 0$, and so they died $P = 0$. If had I intervened (reminded you of your duty), you would have watered them, $W = 1$, and they would not have died $P = 1$. All is well. (This is, in fact, very close to why counterfactual theories can handle the case: see Lewis's counterfactual dependence).

Briefly, the *double prevention* cases work in the same way, except that we also need the notion of an *indirect cause* which we skipped for the sake of brevity (see Woodward 2003: 52–7). The trickiest case for Woodward is that of *late pre-emptions*. A step towards a solution is that the variables must be finer grained: Suzy, instead of only throwing or no throwing, can also take a bigger or smaller stone, spin it or not, etc. In other words, S can take a whole range of other values than 1 and 0. Of course, our target variable, B, representing the bottle and its behaviour must take a fuller range of values too (the pieces of glass being more or less scattered). As the reader will guess, the rest is akin to Lewis's wiggling idea.

Agency theories We do not want to end our section on Woodward without mentioning that there is a precursor of interventionist theories, namely *agency theories of causation*, which already had the idea of manipulation at their core. The agency theorists – for example, Georg Henrik von Wright, Peter Menzies and Huw Price – have some anti-Humean sentiments:

> Empiricists need to keep in mind that human subjects have access to the world in two ways: as observers, certainly, but also as *agents*, capable of intervening in the processes of the world at will.
>
> (Menzies and Price 1993: 194–5)[38]

On the basis of this thought, these theorists formulated their central thesis:

> An event *A* is a cause of a distinct event *B* just in case bringing about the occurrence of *A* would be an effective means by which a free agent could bring about the occurrence of *B*.
>
> (Menzies and Price 1993: 187)

This is fairly similar to interventionist theories and it shares many threats and promises. The potential problem of circularity, for example, also occurs in that both *bringing about* and *effective means* are causal terms within the explanans (also PROD comes to mind again). Menzies and Price react to the allegation of circularity:

> We all have direct personal experience of doing one thing and thence achieving another. [. . .] These cases provide direct non-linguistic acquaintance with the concept of bringing about an event.
>
> (Menzies and Price 1993: 194–5)

Causation can then be legitimately defined with the help of this basic concept without conceptual circularity; however, the explicit reference to human agency might smell of an unacceptable anthropocentricity. In their defence, the proponents of that theory reply that the dependence of causation on human action is of an epistemological or conceptual rather than an ontological kind. Quite credibly they even claim that it is unclear whether mere passive observers, who are not also agents, would have a concept like causation at all.[39]

With these few words we must leave interventionist and agency theories behind. In fact, we have to wrap up our findings in this chapter on causation. As always, here is a summary:

BOX 5.5 Interventionist theories

James Woodward's interventionist theory of causation has two roots:

- Earlier **manipulation theories** like the **agency theories and causal graph theory** and **causal modelling**
  We have described both these ancestors briefly.
- Briefly, Woodward's account of causation is that ***X* is a cause of *Y* iff there is a possible ideal intervention on *X* that changes *Y***.
- For this definition, the notion of an **intervention** is essential. Taken from the methodologies of actual science (medical trials, for example) we have collected, with Woodward, adequacy conditions for when a **manipulation** counts as an **ideal** (worthwhile) intervention. We have explicated what interventions are and how to use them for the characterisation of causation.
- For that purpose, **modality** and **counterfactual conditionals** have to be taken on board as well: a **possible** intervention is one that **would lead to such-and-such if it were performed**.
- Woodward's account is **non-reductive** in two senses: causal vocabulary like 'intervention' (spelled out in terms of causation itself) appears in the explanans, and no elimination of causation to non-modal Humean vocabulary is attempted. Woodward doubts that an adequate reductive account will ever be found and he believes that an **account of causation can be illuminating without being reductive**. Seen from a Humean perspective, this might be the biggest disadvantage of the theory.

5.6 Classifications

We offered a tripartite categorisation of causal theories in Section 5.1, where we asked, with Mackie, whether they are more on the ontological, epistemological or semantical side. We wish to offer a further three dichotomies along which theories of causation can be classified and also to see where and how the theories fit in. Finally, we introduce two criteria causal theories should meet in order to be adequate.[40]

Humean versus non-Humean One crucial dichotomy is whether a theory of causation is *Humean or non-Humean*, that is, whether its aim is to reduce causal locutions into a language that is free of causal vocabulary and lacks reference to powers, capacities or dispositions in particular.

For the Empiricists, this would also mean rephrasing in observational terms; for the post-Empiricist Humeans, like Lewis, it is enough if the analysans make no reference to modal terms.

The latter is the case for the early Mackie (note that the *necessary* and the *sufficient* in his definition is the logical relation that the material implication discloses, not the anti-Humean necessitating relation; see Section 5.2). Yet, is this the case for Lewis? At first, it might not seem so because modal counterfactual conditionals appear in his analysis, but because we are already acquainted with his theory of counterfactuals we know that he also offers a Humean analysis of those in order that his theory of causation in the end is Humean.

Apart from its special incarnation as a mechanistic theory of causation, which is decidedly non-Humean, other process, mark transfer or conserved quantities theories are formulated by and large in a Humean spirit.[41] It is true that they postulate a connection (energy flow, for example) but this connection is empirically testable and non-modal in nature.

Agency theories are quite anti-Humean in that they explicitly say that regularities are just not enough for causation. They also believe that the arguments Hume himself offered against agency as a basis for causation to be invalid (see Hume 1748: Sect. VII, Part I: 67; especially endnote §52). Whether Woodward's interventionist theory is Humean depends on how the counterfactuals he makes use of are spelled out – if in Lewisean style then maybe. However, Woodward is more comfortable spelling them out in terms of dispositional capacities (see Section 5.7). A further fact should definitely catapult him out of Humean territory: his analysis is not reductive, which brings us to the next dichotomy.

Reductive versus non-reductive Humean analyses tend to be reductive, anti-Humean don't. Reductive are those theories that do not allow for causal vocabulary within the *definiens* of their definitions. Non-reductive theories do not mind causal vocabulary to appear again as long as no vicious definition circle results and the characterisation remains enlightening.

There are non-reductive, non-Humean theories that we have not yet touched upon: Dispositional Essentialism, of course, which is to come in Sections 6.2 and 6.4), but also theories by, for example, Curt J. Ducasse (1881–1969; Ducasse (1926)) or Elizabeth Anscombe (1919–2001; Anscombe (1971)) to which we cannot do justice here.

Intrinsic versus extrinsic We can finally ask whether a theory describes (actual) causation as something wholly intrinsic, internal to the particular causal process or the succession of events. Hume, Mackie and Lewis would deny this: whether a particular and actual *c*-*e*-sequence is causal depends on whether it is an instance of a regularity. For Hume this is quite explicit; in the INUS conditions or the counterfactuals it is more implicit.

In any case, such regularities over and above the particular event are extrinsic to it and cover other events elsewhere.

Whether the process and conserved quantities theories are intrinsic or extrinsic depends upon our focus: the very mark or quantum of energy that is transferred is intrinsic to the relevant process; however, in the case of conserved quantities, it is crucial that it is a regularity or, better, a law that these quantities are conserved. In this sense these theories depend on generalities that are extrinsic to the particular event.

Mechanistic theories often claim to be intrinsic: the causal oomph is said to be internal to the mechanism. (The earlier theories by Ducasse and Anscombe are also intrinsic).

For Woodward's theory, it depends on how the central counterfactuals he makes use of are evaluated. If they are true because of certain intrinsic dispositions that the objects have that are involved in the causal process, then maybe his is an intrinsic theory. If they are Lewisean style, then they are extrinsic.

Note that those who believe that causation is something intrinsic to each specific causal event are often called *Singularists* (as opposed to *Generalists*) because intrinsicality seems to preclude reference to external *regularities*. If you hold an extrinsic theory of causation you may take these regularities on board (and thus become a *Generalist*).[42]

Extensional and intensional adequacy These are 3 × 2 categories that theories of causation can fall into.[43] We can classify causal theories along two further axes: we can ask whether they are *extensionally* and *intensionally adequate*. The former is the case when all and only those event sequences or processes which we intuitively consider to be causal count as such according to the relevant theory. Early and late pre-emption, double prevention, absences and omissions make it difficult for theories to be fully extensionally adequate, even if they cover most other cases.

A theory of causation is *intensional adequate* if the platitudes that we (or most of us) associate pre-theoretically with causation are incorporated into the theory. Remember, some of the features that have been intuitively associated with causation include causes precede effects; an effect does not also cause its cause, i.e. causation is asymmetric; causes produce effects; effects depend counterfactually on causes; causation is something objective in the world; generally, when a certain type of cause is instantiated, a certain type of effect follows (see Hüttemann 2013: 57–64).

Not everyone has all of these intuitions and, most probably, there is no such thing as a totally intensionally adequate philosophical theory. Some subsets of all intuitions might even contradict other subsets. This is not too worrying as long as we allow revisions of some intuitive entries or we are willing to drop some in light of philosophical, theoretical reflection (as Hume, for example, dropped necessitation from his concept of causation after careful consideration).[44]

Total intensional adequacy, if it existed, might have extensional adequacy as a consequence. The other direction is doubtful. Remember our remarks about Lewis's or Mackie's theories (see Section 5.2 and 5.3): some anti-Humeans might nag that even if some Humean theory gets things extensionally right they still do not get at the real core of causation, namely that there is some productive power or force that drives the whole machine. Scratching the surface only, they miss intensional adequacy, or so the anti-Humeans might insist.

Finally, note that there can be a tension between the two adequacies: those theories that see some kind of productive link between cause and effect as intensionally constitutive will have difficulties in capturing, extensionally, cases of omission and absence. Those theories that see counterfactual dependence as the core of intensional adequacy can handle those cases well, but they have their difficulties in including, extensionally, pre-emptions and overdeterminations, and these are cases that the first kind of theories have little trouble with. Maybe this calls for splitting our concept of causation in two (see Section 5.4) – but we leave this for future research to decide.

BOX 5.6 Classifications

We can distinguish:

- **Humean and non-Humean**,
- **Reductive and non-reductive**,
- **Intrinsic and extrinsic** theories of causation.

The appropriateness of any theory of causation can be measured along the two axes of **extensional and intensional adequacy.**

(This section has been so brief that these issues can be easily retrieved from the main text).

5.7 Causation and dispositions

In this book, we have taken the history of how philosophers have approached *dispositions* or *dispositional predicates* as a recurrent theme. The sideways we stepped on, next to the main route, included *counterfactuals*, *laws* and, as in this chapter, *causation*. How do the theories we have encountered here relate to the issue of dispositions?

An explicit analysis of dispositions that depends partially but also crucially on causation came from David Lewis. We introduced his account in detail already in Section 3.5 on counterfactuals because counterfactuals are also a central element in his theory. A short version of his theory is: x is disposed to show reaction r on stimulus s iff if s came about and x's intrinsic character was unchanged, then the stimulus and x's intrinsic setup would *cause* the reaction r. By now we know what Lewis means when he talks about causation and one could, in principle, replace reference to causation in his definition of dispositions with what has been said in Section 5.3 (although this would make the definition impracticably long). In theory, we could also insert into Lewis's definition any of the other approaches to causation we learnt about here.

We look instead at other possibilities and also offer an idea that might be worth pursuing in future research. An option for the *conserved quantities theories of causation* to tackle dispositions could be: if causation is essentially the transfer of energy then the disposition of a system to react upon being stimulated might be redescribed in terms of potential energy. *Potential energy* is a technical term from physics: it is the energy stored in a system due to its position in a gravitational, electromagnetic or other field. Derivatively, energy can also be stored due to chemical bonds or mechanically in, for example, springs. In order for this energy to be released it is sometimes necessary to trigger it with some activation energy. Think, as an example, of a rock that holds steady because of friction. If you push it slightly it roles all the way down the hill and its potential energy is transformed into kinetic energy. Now consider dispositions (for example, inflammability) as stored potential energy (the chemical bonds in the match's tip) that is released as heat (the inflammation of the match) upon receiving activation energy (striking it on a rough surface).

A theory which is akin to the idea just offered is Andreas Hüttemann's, which we will discuss further in Section 6.4.3. Also Maudlin's, which was outlined in Section 4.6, bears similarities. Hüttemann's theory shares something with Woodward's account: both believe that dispositions or capacities are more fundamental than causation or even than laws of nature. Woodward writes:

> We should think of laws of nature as abstract and idealized descriptions of powers and capacities possessed by particular objects and systems [. . .].
> (Woodward 1992: 211–12)[45]

Dispositions enter the theory of causation in that Woodward believes the dispositions to be the truthmakers of counterfactuals and, remember, he needs the truth of counterfactuals for merely possible, non-actual interventions.

Note, that we have again reached a turning point. Causation, Woodward and Hüttemann claim, is not there to analyse dispositions but we can analyse causation with the help of dispositions. Remember that we have already encountered such 180-degree turns – for example, Lange's idea to analyse what a law of nature is in terms of counterfactuals rather than the other way round (see Section 4.5). The following chapter is devoted entirely to such a Copernican turnaround: Dispositional Essentialism takes dispositions to be basic, and everything else, causation, laws and counterfactuals, have to be spelled out in their terms. Remember that we have hinted at such a possibility when we saw how entangled mechanistic theories are with dispositions (see Section 5.4).

Notes

1 Also in an Empiricist spirit, Arthur Burks (1951, 1955), Wilfried Sellars (1958) and Arthur Pap (1958) tried to explicate causal relations by introducing syntactical rules for causal conditionals in a formal language. These attempts were not very influential and we therefore leave them aside here.

2 We could characterise, very roughly, an event as a particular change of an object's properties at a specific time (e.g. the heavy truck's location changed at t_0 and the carried weight of the bridge changed accordingly). For two classic papers on what *events* are, see Davidson (1969) and Kim (1976).

3 What was meant by substance in, for example, Leibniz's substance causation – 'Only a substance can be a real cause' (Leibniz 1704: 65) – is anyway not something profane like the acid we referred to provisionally above but rather some fundamental, foundational entities – in Leibniz: monads – that ground the existence of everything else (see Section 1.1).

4 Historically, it was Thomas Reid's (1710–96) example of night following day that was meant to be trouble for Hume because although night and day have been following each other contiguously and in temporal succession ever since the beginning of time, they do not cause each other (Reid 1788: 342).

5 Remember Lewis's idea of an x-complete cause in our counterfactuals chapter (see Section 3.5). This passage in Mill about the 'whole cause' is one of his reference points.

6 Also, a cause 'is an event which precedes [i.e. temporally succeeds] the event of which it is a cause' (Mackie 1965: 33). Hart and Honoré (1985) had a similar idea (see also Wright 1985): in order for it to be a cause, an event has to be a *N*ecessary *E*lement in a *S*et of conditions which are together *S*ufficient for the outcome (they speak of a *NESS* condition).

7 For those who are familiar with Mackie's work, we are alluding to the Manchester and London workers example. Michael Baumgartner (2008: 339ff) offered a convincing strategy for how Mackie could deal with the (RC) case within Mackie's INUS paradigm.

8 It is a bit unfortunate to have causal dependence (which is nothing but counterfactual dependence) and causation because causal dependence and causation are so easily confused. It's best to think of counterfactual dependence immediately when we read causal dependence.

9 For temporal fragility, see Paul 1998.

10 Lewis connoisseurs will have noticed that we have reversed the standard order of representing things here. Lewis, who definitely wants causation to be transitive, first makes sure that it is transitive through his definition of causal chains and only then he uses the chain to cope with early pre-emption cases. We have done this the other way round.

11 For more examples see Collins *et al.*, 2004: 40–1.

12 More precisely, Lewis says that they are false in standard contexts but let us leave this aside.

13 See Lewis 1973: 163–4, 170–1; 1979: 47–8. In fact, Lewis partially arranges the criteria in the way he did in order to make backtrackers wrong. This has been criticised by some philosophers (for example, Horwich 1987: 171ff and Woodward 2003: 137): arranging the list such that backtrackers

are false without further reasons for that ordering seems to be an ad hoc move. (Note, however, that we have given some independent intuitions for why the list should be as it is in Section 3.4). Moreover, in the case of some fairly ordinary backtrackers it is doubtful whether they should count as false. For example, someone alone in their kitchen in rage takes a plate out of the cupboard and smashes it onto the floor. The plate shatters. 'Had the plate (in that situation) not shattered, the person would not have thrown it'.

14 Note that the counterfactual 'Had Suzy not thrown, Billy would have' which is based on their deal 'If I don't throw, you do' is not a backtracker. The real throw (antecedent) and Billy's possible throw (consequent) are here located at the same time.

15 The term *trumping* is Jonathan Schaffer's invention. The example of the soldiers is van Fraassen's (see Schaffer 2000).

16 In the original, the last sentence is in brackets immediately after the previous sentence. I have altered it for better legibility.

17 It solves the early pre-emption case, as the readers will easily convince themselves.

18 There are further critical issues, including *delayers* and *hasteners*. Eating earlier that morning delayed hunger in the afternoon. Wiggle the meal (i.e. have it a little sooner or later) and you wiggle the hunger (it comes slower or faster, more or less strong). However, eating in the morning did not cause hunger in the afternoon. Or so says our intuition contra what Lewis's wiggling theory would have to say. *Omissions*, *absences* and *kisses* (see Section 5.4) are further general problem cases of theories of causation, yet which Lewis can handle.

19 Lewis is aware of the fact that the transmission of energy is a good answer to the Suzy–Billy case; however, he believes that not all cases of causation can be captured that way (see Lewis 2000: 91). Indeed, we will see the limits of that theory later.

20 Later we will get to know the so called *Canberra Plan*, which is a research program that has these kinds of Naturalisation at its heart (see Section 7.6).

21 For reasons of brevity, henceforth *energy* stands for all other quantities in question as well.

22 Mackie, in his ancestral process theory, anticipated the essentiality of conserved quantities (Mackie 1974: 217) and Fair (1979) published his 'Causation and the Flow of Energy'.

23 We wish to mention that Dowe faces some difficulties when it comes to spelling out identity criteria for objects which he 'solves' by claiming that *identity* is a primitive, unanalysable notion but we leave the details aside (see Dowe 2000: 98–107).

24 This move to accept negative events into one's metaphysics is also not without its problems because once you allow my omission to water the plants as a cause of their dying, why is then the Pope's omission to water the plants not also a cause? We cannot argue with the legal or moral responsibilities that I have but not His Holiness because this would lead us to a different philosophical realm: ethics or legal philosophy instead of metaphysics.

25 Hall's (2004) notion of production is not the same ours (see Section 2.1.3), but it is near enough to make this point here.

26 On pluralism, see De Vreese's (2006) article 'Pluralism in the philosophy of causation' in the special issue of *Philosophica*.

27 Other seminal papers include Machamer *et al.* (2000), Bechtel and Abrahamsen (2005), Craver (2007) and Bogen (2008). Nancy Cartwright's (2009) *nomological machines* are not so unlike mechanisms either. For a concise overview, see Jon Williamson's (2011) 'Mechanistic Theories of Causation Parts I and II'.

28 A concise overview of definitions for mechanisms can be found in (Hedström and Ylikoski 2010: 51).

29 See, for example, Ladyman and Ross's infamous 2007 *Every Thing Must Go*; note their play on words with *everything*.

30 Nomological necessitation (see Armstrong's (1997: 227) theory of lawhood, including causation) is similar in this respect. It could also be conceived as a micro-mechanism. In fact, Glennan discusses the Armstrongian possibility (Glennan 2011: 811–13) for the fundamental level.

31 Bert Leuridan (2010) and Laura Franklin-Hall (2011) provide harsh but valuable critique of mechanisms along these lines, including mechanistic theories of explanation and causation. For a general evaluation of mechanisms, see Woodward (2013). There are also philosophers, including Illari and

Williamson (2013) and Bogen (2008), who see value in turning to unanalysed powers or activities within mechanisms.

32 The epilogue in Pearl's (2000) work, the article by Scheines (1997) and Hitchcock's contribution (Hitchcock 2009) are more accessible than the rather formal treatments within their books. Noteworthy is the idea to treat causality, with the help of the theory of causal nets, as a theoretical concept (Schurz and Gebharter 2015).

33 We should mention that Woodward had already presented interventionist ideas as early as the early 1990s (see, for example, Woodward 1992).

34 There is a condition, I.5, that Woodward mentions but he suggests that we think of it as being already covered by the others, I.3 in particular. We will therefore leave it aside.

35 We will discuss Woodward's '*possible*' intervention later. That all *other variables*, i.e. other causal factors besides X and Y should be *held fixed* at some value, makes the intervention on X '*surgical*' (see Woodward 2003: 130, 133). The conditions for interventions in I.2 to I.4 are thereby sharpened.

36 I have borrowed these structural equations from Hüttemann (2014).

37 In Section 4.6 of Woodward (2003: 133ff), he compares Lewis's to his own theory. Some authors – for example, Alexander Reutlinger (2013: chapters 3 and 4) – see no real difference between them.

38 See also von Wright 1973: 114. It would not be surprising if these authors knew Schopenhauer's (1844: 195) castle metaphor on the same issue and with similar formulations.

39 See also Reichenbach (1938: 117), von Wright (1973: 117) and Dummett (1978: 333).

40 I have adopted the three dichotomies from Stathis Psillos (2002: 127–33). The two adequacy conditions come from Andreas Hüttemann (2013: 63–4). Psillos's and Hüttemann's textbooks are highly recommended for further introductory reading on the subject of causation (the latter only if you read German).

41 To be fair to Kistler, we must mention that other than his conserved quantity theory he sympathises with anti-Humean dispositional Realism (Kistler 2006: especially chapter 6).

42 Psillos (2002: 129) offers reasons why the agency account could count as *Singularist* but yet extrinsic.

43 There are, in principle, eight possible combinations. If you take the earlier introduced three-fold distinction of ontological, conceptual and epistemic theories of causation (see Section 5.1), you have 24 possibilities to explore. To figure out whether all combinations find their proponents (or make sense at all) is a valuable exercise for the reader.

44 Compare what has been said here about extensional and particularly intensional adequacy to the 'Canberra Plan', which we will discuss in Chapter 7.

45 For more on laws, see Woodward (1992: 202–3; 2003: 236 ff). Also compare this to Cartwright (1989).

Bibliography

Anscombe, E. (1971) Causality and Determination. In: Sosa, E. and Tooley, M. (eds.) *Causation*. Oxford: Oxford University Press. pp. 88–104.

Armstrong, D. M. (1997) *A World of States of Affairs*. Cambridge: Cambridge University Press.

Baumgartner, M. (2008) Regularity Theories Reassessed. *Philosophia*. 36 (37). pp. 327–54.

Bechtel, W. and Abrahamsen, A. (2005) Explanation: A Mechanist Alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*. 36 (2). pp. 421–41.

Bogen, J. (2008) Causally Productive Activities. *Studies in the History and Philosophy of Science Part A*. 39 (1). pp. 112–23.

Burks, A. (1951) The Logic of Causal Propositions. *Mind*. 60 (239). pp. 363–82.

Burks, A. (1955) Dispositional Statements. *Philosophy of Science*. 22 (3). pp. 175–93.

Cartwright, N. (1989) *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Collins, J., Hall, N. and Paul, L. A. (eds.) (2004) *Causation and Counterfactuals*. Cambridge, MA: MIT Press.
Craver, C. (2007) *Explaining the Brain*. Oxford: Oxford University Press.
Davidson, D. (1967) Causal Relations. *Journal of Philosophy*. 64 (21). pp. 691–703.
Davidson, D. (1969) The Individuation of Events. In: Rescher, N. (ed.) *Essays in Honor of Carl G. Hempel*. Dordrecht: Reidel. pp. 216–34.
De Vreese, L. (2006) Pluralism in the Philosophy of Causation: Desideratum or Not? *Philosophica*. 77 (1). pp. 5–13.
Dowe, P. (1992) Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science*. 59 (2). pp. 195–216.
Dowe, P. (2000) *Physical Causation*. Cambridge: Cambridge University Press.
Ducasse, C. J. (1926) On the Nature and the Observability of the Causal Relation. *Journal of Philosophy*. 23 (3). pp. 57–68.
Dummett, M. (1978) Bringing About the Past. In: Dummett, M. (ed.) *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press. pp. 333–50.
Fair, D. (1979) Causation and the Flow of Energy. *Erkenntnis* 14 (3). pp. 219–50.
Franklin-Hall, L. (2011) The Emperor's New Mechanism. Unpublished.
Glennan, S. (1996) Mechanisms and the Nature of Causation. *Erkenntnis*. 44 (1). pp. 49–71.
Glennan, S. (2002) Rethinking Mechanistic Explanation. *Philosophy of Science*. 69 (53). pp. 342–53.
Glennan, S. (2009) Mechanisms. In: Beebee, H., Hitchcock, C. and Menzies, P. (eds.) *The Oxford Handbook of Causation*. Oxford: Oxford University Press. pp. 315–325.
Glennan, S. (2010). Mechanisms. In: Psillos, S. and Curd, M. (eds.) *Companion to the Philosophy of Science*. London and New York: Routledge. pp. 367–84.
Glennan, S. (2011) Singular and General Causal Relations: A Mechanist Perspective. In: McKay Illari, P., Russo, F. and Williamson, J. (eds.) *Causality in the Sciences*. Oxford: Oxford University Press. pp. 789–817.
Hall, N. (2004) Two Concepts of Causation. In: Collins, J., Hall, N. and Paul, L. A. (eds.) *Causation and Counterfactuals*. Cambridge, MA: MIT Press. pp. 225–76.
Hart, H. L. A. and Honoré, A. M. (1985) *Causation in the Law*. 2nd edition. Oxford: Oxford University Press.
Hedström, P. and Ylikoski, P. (2010) Causal Mechanisms in the Social Sciences. *The Annual Review of Sociology* 36. pp. 49–67. doi: 10.1146/annurev.soc.012809.102632.
Hitchcock, C. (2009) Causal Modelling. In: Beebee, H., Hitchcock, C. and Menzies, P. (eds.) *The Oxford Handbook of Causation*. Oxford: Oxford University Press. pp. 299–314.
Horwich, P. (1987) *Asymmetries in Time: Problems in the Philosophy of Science*. Cambridge, MA: MIT Press.
Hume, D. (1748) *Enquiries concerning Human Understanding and concerning the Principles of Morals*. 17th impression of the 3rd OUP edition. Selby-Bigge, L. A. (ed.). Oxford: Clarendon Press.
Hüttemann, A. (2013) *Ursachen*. Berlin: De Gruyter.

Hüttemann, A. (2014) Scientific Practice and Necessary Connections. *Theoria*. 29 (1). pp. 29–39.
Illari, P. M. and Williamson, J. (2013) In Defense of Activities. *General Philosophy of Science*. 44 (1). pp. 69–83.
Kim, J. (1976) Events as Property Exemplifications. In: Brand, M. and Walton, D. (eds.) *Action Theory*. Dordrecht: Reidel. pp. 159–77.
Kistler, M. (2006) *Causation and Laws of Nature*. Routledge Studies in Contemporary Philosophy. London and New York: Routledge.
Ladyman, J. and Ross, D. (2007) *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
Leibniz, G. W. (1704) *New Essays on Human Understanding*. Remnant, P. and Bennett, J. (eds., trans.) (1982). Cambridge: Cambridge University Press.
Leuridan, B. (2010) Can Mechanisms Really Replace Laws of Nature? *Philosophy of Science*. 77 (3). pp. 317–40.
Lewis, D. (1973) Causation. *Journal of Philosophy*. 70 (17). pp. 556–67.
Lewis, D. (1979) Counterfactual Dependence and Time's Arrow. *Noûs*. 13. pp. 455–76. Quoted in Lewis 1986: 32–51.
Lewis, D. (1986) *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.
Lewis, D. (2000) Causation as Influence. *Journal of Philosophy*. Special Issue: Causation. 97 (4). pp. 182–97. Quoted in Collins, Hall and Paul 2004: 75–106.
Machamer, P., Darden, L. and Craver, C. F. (2000) Thinking About Mechanisms. *Philosophy of Science*. 67 (1). pp. 1–25.
Mackie, J. L. (1965) Causes and Conditions. *American Philosophical Quarterly*. 2(4). pp. 245–64. Quoted in Sosa and Tooley 1993.
Mackie, J. L. (1974) *The Cement of the Universe*. Oxford: Clarendon Press.
Menzies, P. and Price, H. (1993) Causation as a Secondary Quality. *The British Journal for the Philosophy of Science*. 44 (2). pp. 187–203.
Mill, J. S. (1843) A System of Logic. Ratiocinative and Inductive. Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. In: Robson, J. M. (ed.) (1967) *Collected Works, Vols. VII and VIII*. Toronto: Toronto University Press.
Norton, J. D. (2003) Causation as Folk Science. *Philosophers' Imprint*. 3 (4). pp. 1–22.
Pap, A. (1958) Disposition Concepts and Extensional Logic. In: Feigl, H., Scriven, M. and Maxwell, G. (eds.) *Minnesota Studies in the Philosophy of Science, Vol. II*. Minneapolis, MN: University of Minnesota Press. pp. 196–224.
Paul, L. A. (1998) Keeping Track of the Time: Emending the Counterfactual Analysis of Causation'. *Analysis*. 58 (3). pp. 191–8.
Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Psillos, S. (2002) *Causation and Explanation*. Chesham: Acumen.
Psillos, S. (2009) Causal Pluralism. In: Vanderbeeken, R. and D'Hooghe, B. (eds.) *Worldviews, Science and Us: Studies of Analytical Metaphysics: A Selection of Topics From a Methodological Perspective*. Singapore: World Scientific Publishers. pp. 131–51.

Reichenbach, H. (1938) *Experience and Prediction – an Analysis of the Foundations and Structure of Knowledge*. Chicago, IL: University of Chicago Press.

Reid, T. (1788) *Essays on the Active Powers of Man*. Edinburgh: John Bell. (Cambridge: Cambridge University Press, 2011).

Reutlinger, A. (2013) *A Theory of Causation in the Social and Biological Sciences*. New York: Palgrave Macmillan.

Russell, B. (1912) On the Notion of Cause. In: Mumford, S. (ed.) (2003) *Russell on Metaphysics. Selection from the Writings of Bertrand Russell*. London and New York: Routledge. pp. 163–82.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Salmon, W. (1994) Causality without Counterfactuals. *Philosophy of Science*. 61 (2). pp. 297–312.

Salmon, W. (1998) *Causality and Explanation*. Oxford: Oxford University Press.

Schaffer, J. (2000) Causation by Disconnection. *Philosophy of Science*. 67 (2). pp. 285–300.

Schaffer, J. (2004) Trumping Perception. In: Collins, J., Hall, N. and Paul, L. A. (eds.) *Causation and Counterfactuals*. Cambridge, MA: MIT Press. pp. 59–73.

Scheines, R. (1997) An Introduction to Causal Inference. In: McKim, V. and Turner, S. (eds.) *Causality in Crisis?* Notre Dame, IN: University of Notre Dame Press. pp. 185–99.

Schopenhauer, A. (1844) *Die Welt als Wille und Vorstellung*. 2nd, enlarged edition. Payne, E. F. J. (trans.) (1958). New York: Dover.

Schurz, G. and Gebharter, A. (2015) Causality as a Theoretical Concept: Explanatory Warrant and Empirical Content of the Theory of Causal Nets. *Synthese* (advance online publication). doi: 10.1007/s11229-014-0630-z.

Sellars, W. (1958) Counterfactuals, Dispositions, and the Causal Modalities. In: Feigl, H., Scriven, M. and Maxwell, G. (eds.) *Minnesota Studies in the Philosophy of Science, Vol. II*. Minneapolis, MN: University of Minnesota Press. pp. 225–308.

Sober, E. (1987) The Principle of the Common Cause. In: Fetzer, J. (ed.) *Probability and Causation: Essays in Honor of Wesley Salmon*. Dordrecht: Reidel. pp. 211–29.

Spirtes, P., Glymour, C. and Scheines, R. (1993) *Causation, Prediction and Search*. New York: Springer.

Von Wright, G. H. (1973) On the Logic and Epistemology of the Causal Relation. In: Sosa, E. and Tooley, M. (eds.) (1993) *Causation*. Oxford: Oxford University Press. pp. 105–24.

Williamson, J. (2011) Mechanistic Theories of Causation Parts I and II. *Philosophy Compass*. 6 (6). pp. 433–44. doi: 10.1111/j.1747-9991.2011.00400.x

Woodward, J. (1992) Realism about Laws. *Erkenntnis*. 36 (2). pp. 181–218.

Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Woodward, J. (2013) Mechanistic Explanation: Its Scope and Limits. *Proceedings of the Aristotelian Society*. Supplementary Vol. LXXXVII. pp. 39–65.

Wright, R. W. (1985) Causation in Tort Law. *California Law Review*. 73 (6). pp. 1735–1828.

6 Dispositional Essentialism

6.1 Back to pre-Humean metaphysics

The way up and the way down are one and the same.
(Heraclitus 535–475 BC)

There is something striking about the historical development of the metaphysics of science in the past hundred years, which is reflected in the chapter structure of this book: the Empiricists wanted to eliminate dispositional predicates or dispositions in favour of observational language (see Section 2.1). Later Humean-minded philosophers attempted to get rid of dispositions in favour of categorical properties (see Section 2.2). In order to succeed with this programme, these philosophers also had to shed light on further concepts or issues from the metaphysics of science, like counterfactual conditionals, laws of nature and causation. We took the opportunity to introduce their and other philosophers' theories thereof (Chapters 3, 4 and 5). Particularly in Sections 3.5, 4.7 and 5.7, we became acquainted with the ways in which theorists made use of counterfactuals, laws and causation in order to define dispositionality.

However, we also observed that the original radical Empiricist/Humean programme had to be weakened considerably. Bit by bit, some anti-Humean, anti-Empiricist assumptions had to be taken on board (for example, that natural properties exist), otherwise the prospect for successful theories of counterfactuals, laws or causation seemed bleak. This is one reason why philosophers became more and more happy to leave Humean metaphysics behind.

Now, with Dispositional Essentialism – the theory discussed in this chapter – everything is turned on its head. Not only is the existence of dispositions, of essences and metaphysical necessity taken for granted, moreover, counterfactuals, laws and causation are explained on the basis of dispositions, essences and necessity. This chapter is therefore the complement to our starting point in Chapter 1. We have arrived back – hopefully in a spiral and not in a plain circle – to anti- or pre-Humean metaphysics from Aristotle, the Scholastics or, in some respects, the Rationalists.

In Section 6.2, we introduce *Dispositionalism* and reasons to believe in it. We then outline another anti-Humean development, *Essentialism*, in Section 6.3. We enter semantic territory when we reason for Essentialism, whereas the defence of Dispositionalism happens on ontological and epistemic grounds. Section 6.4 combines the two views in *Dispositional Essentialism*. We assess this by looking at what it can contribute to the metaphysics of science as a whole and particularly to theories of counterfactuals, laws and causation.

6.2 Dispositionalism

6.2.1 What is Dispositionalism?

Dispositionalism is the view that some or all properties are of a dispositional nature. When an object has such a dispositional property (like being soluble or being inflammable) it reacts with a certain manifestation (dissolving or being in flames) when triggered, i.e. when being in certain stimulus conditions (being in water or striking a matchbox). Dispositionalism is a Realist stance regarding such properties. It says that dispositions – with their intrinsic trigger-manifestation profile – belong to nature's fundamental building blocks. They are not just a shorthand form to pick out categorical properties plus the nomological roles they happen to play (see Sections 2.2.1 to 2.2.4).

Dispositional properties, if they are real, bring two features with them which would equip the world with anti-Humean elements. First, *modality* (MOD): dispositions are related to what could or must happen (the reaction) if something else were the case (the trigger). Such possibilities or necessities are modal facts. They transcend that which is merely the case. Second, *responsibility for production* (PROD): if a thing has a disposition then it itself together with the appropriate trigger can be responsible or can produce the respective reaction. Dispositionalism wholeheartedly endorses these anti-Humean consequences. (For more on (MOD) and (PROD), see Sections 2.1.2 and 2.1.3.)

Before we continue to give reasons to hold this view, note again that there are many synonyms for *disposition* or *dispositional property*: *capacity*, *tendency*, *power*, etc. The authors use all these terms in the quotes used in this chapter. Occasionally, they make subtle distinctions between them; however, for our purposes we can ignore these differences and use all terms synonymously.

6.2.2 Indirect reasons for Dispositionalism

So, why should we believe in Dispositionalism? 'Why should we not?' could be the response of the Dispositionalists who might continue: 'Why should we instead adopt the unintuitive opposing view of *Categoricalism* or *Quidditism*?'. This could be a pressing question indeed because, first, this alternative view has some hard-to-bite consequences, such as the unknowability of these properties' natures, the possibility of them swapping their roles unnoticed, and that of them being 'idlers' (see Sections 2.2.3 and 2.2.4). Second, as we shall see in detail in Section 7.8, there may be no convincing reason to believe in *Hume's Dictum*, which is the underlying Humean credo that motivates Categoricalism/Quidditism.

If compelling the doubtfulness of Quidditism and of Hume's Dictum are (only) *indirect* reasons to hold Dispositionalism that capitalise on the assumed weakness of the opponent's view, another (third) indirect argument is to point to the long trajectory from the Empiricists' early attempts to analyse dispositional predicates (see Section 2.1) to more recent endeavours that remove dispositional properties (see Section 3.5). As we know, these analyses were confronted with severe difficulties in the guise of counterexamples ((VC), (RC), finks and antidotes) and whether they ultimately succeeded remains a matter of debate. Thus, like Martin, we could come to the Realist, Dispositionalist conclusion that if 'counterfactuality or strong conditionality cannot explain dispositions, then there is no place to turn but to actual first-order dispositions or powers' (Martin 1994: 7).

Reversing the Humean agenda A fourth argument for Dispositionalism is more constructive. It started to be developed once the frustrations with the Humean programme became more widespread. Some philosophers in favour of Dispositionalism started to apply the kind of argument that David Lewis once used in order to argue for his realist assumption regarding natural properties. He claimed that to believe in the existence of natural properties is 'so commonsensical and so serviceable – indeed, was so often indispensable – that it was foolish to try to get on without it' (Lewis 1999: 1–2). Dispositionalists now make the very same move, but for dispositions. They claim equally that it is so commonsensical and so serviceable to assume that dispositions are real (for novel theories of laws or causation, for example) that it is foolish to try to get on without them. (These kinds of indispensability arguments will be discussed in Section 7.4).

We have mentioned three indirect reasons for Dispositionalism: (i) the opposing view has hard-to-bite consequences, (ii) there are no convincing reasons for the underlying principle, namely *Hume's Dictum*, of the opponent's view, (iii) the Reductionist, conditional analysis of dispositions fails. We have also offered a more direct, fourth reason: (iv) Dispositionalism could be serviceable for other areas within metaphysics of science and thereby constitute a worthy competitor to the Humean programme.

In the next section we will introduce two direct, positive reasons from scientific practice for Dispositionalism that were given by Nancy Cartwright, who has been one of the courageous spokespersons for Dispositionalism at times when most philosophers were still hostile towards this position.

6.2.3 Direct reasons for Dispositionalism

With her seminal paper collection *How the Laws of Physics Lie* (1983) Nancy Cartwright pioneered the Dispositionalism that started to thrive in the last decades of the twentieth century. Both of her arguments in favour of dispositions came from scientific practice. They proceed, so to speak, from methodology to ontology:

> If you want to find out how a scientific discipline pictures the world, you can study its laws, its theories, its models, and its claims [. . .]. But you can also consider not just what is said but what is done. How we choose to look [practically, experimentally, MS] at the world is just as sure a clue to what we think the world is like as what we say about it.
>
> (Cartwright 1992: 69)[1]

Cartwright's first rationale behind adopting Dispositionalism has to do with the fact that the regularities the laws of nature allegedly describe or entail and that have been the core element of at least the regularity theories of lawhood[2] are surprisingly hard to find in our world. The title of Cartwright's book indicates the point that where law statements are universally quantified or seem to imply universal truths (for example, all copper conducts electricity, haemoglobin binds oxygen), the world surprises us with many exceptions to these claims and, therefore, law statements literally 'lie'.

***Ceteris paribus* laws** Before we start with Cartwright's ideas about how to resolve this tension, we must turn briefly to the subject of *ceteris paribus* laws (which we omitted in Chapter 4 on laws of nature in order not to complicate the issues introduced).

An alleged law of nature is said to be a '*ceteris paribus* law' (shortened to '*cp*-law') if it does not hold under any circumstances but only in *normal* or *ideal* circumstances, *provided unfortunate events do not happen* or *nothing interferes*. The literal meaning of '*ceteris paribus*' is *all else being equal*, which is, of course, a misnomer, but it is often used in the literature to stand for any of the previously mentioned clauses. We will also use it in this broad sense.

Typical examples for laws that have such proviso clauses include 'provided the supply remains constant and other things are as usual, the price of a product increases with growing demand', 'all bodies fall with the same speed, *ceteris paribus*' and also 'normally, haemoglobin binds O_2'. Take the first 'Law' example: it fails when there is a war, when a meteor crashes onto earth, when people have taken drugs and/or do not act rationally, etc. The second 'law' holds when there is no air resistance or when there is no additional electromagnetic field, and similarly for the third 'Law'. An attached proviso-clause like '*ceteris paribus*' wishes to indicate these possible exceptions. It also aims to save the law from being falsified or from 'lying' (in Cartwright's terms) in these unfortunate cases.

However, simply attaching *cp* clauses to law statements that express imperfect regularities is an all too easy solution because such a proviso clause, if it's meaning is left imprecise, could be misused as an immunisation strategy. On being confronted with an alleged counterexample, a defender of the *cp*-law could always refer to the proviso: 'Well, the law does not hold in these cases. . . but, as I said, it normally does'. However, if this were allowed, the empirical sciences would be in danger of resembling pseudo-sciences, such as astrology or homeopathy which are champions in shielding themselves from being falsifiable.

Making explicit what *cp* means, however, is no simple task. An alleged law statement 'All *F*s are *G*s, *ceteris paribus*' is in danger of being tautologous or incomplete: tautologous if we specify or define the *ceteris paribus* clause by saying 'All *F*s are *G*s, except in those cases where *F*s are not *G*s'; incomplete if the *ceteris paribus* clause is thought to stand for an exclusion clause (in the antecedent of the law) of possible interferences *A*, *B*, *C*, etc. The problem with the latter variant is that we most likely have to leave a gap in our statement 'All *F*s are *G*s, unless *A* interferes, or *B* interferes, or, . . .' for we most probably do not know all the interferers, not least because there might be an infinite number.

Capacities to the rescue A lot has been written about the *cp* issue;[3] however, we have to restrict ourselves to a specific attempt to deal with *cp* clauses: Cartwright's. *Dispositions*, according to her, are key to dissolving the problem of *cp*-laws or 'lying laws' cause.

Cartwright's central claim is that the world's 'regularities are in no way ontologically fundamental. They are the consequence of the operation of capacities' (Cartwright 1989: 140) Needless to say, capacities are then taken to belong to the fundamental ontological building blocks of the world.

What lies behind this declaration? The idea is to change the subject matter of what law statements are supposed to mean: where it has been assumed so far that they are or capture or entail regularities of the form 'whenever there's an *F* there occurs a *G*', Cartwright tells us, in light of the fact that these alleged regularities can often be frustrated, to reinterpret their content.

According to Cartwright, the subject matter of law statements are ascriptions of dispositions, powers, capacities or 'natures' as she sometimes calls them ('Natures are something like powers' (Cartwright 1992: 48)). That is, instead of 'where there's an *F* there occurs a *G*, *cp*' we should read them as '*F*s have the capacity to be or produce *G*s'. Where '*F*s are *G*s' might well need a proviso clause because, in unfortunate circumstances, some *F*s might not exhibit the feature *G*, Cartwright's reinterpretation is immune to such unwanted exceptions: *F*s will still have the capacity to be or produce *G*s even if the circumstances are so unfortunate that the so disposed *F* does, on this occasion, not yield a *G*.

Look at some of our previous examples and a Cartwrightean reinterpretation thereof: 'all bodies fall with the same speed, *ceteris paribus*' is now 'all bodies have *the tendency* to fall with the same speed'; and 'normally, haemoglobin binds O_2' becomes 'haemoglobin has *the capacity* to bind O_2'.

What makes these reinterpretations possible is that dispositions can be instantiated without manifesting their reaction. In other words, the newly interpreted laws that ascribe dispositions are strict – the disposition is there regardless, even when it does not display its manifestation – and, thus, it looks as if Cartwright-laws do not face the problems of *ceteris paribus* laws because they do not need that proviso clause. Or, to put it more favourable to the original *cp*-laws, we could say that the Cartwrightean move to dispositions provides an analysis of what is meant by *cp*-law statements: 'All *F*s are *G*s, *ceteris paribus*' is a law of nature *iff* all *F*s have *the capacity* to be *G*s.

Remember that we have delved into *cp-laws* and the capacities solution to the intricacies they bring with them because we were looking for an argument for Dispositionalism. Here it is in summary: if you subscribe to Dispositionalism the problem of *cp*-laws finds an elegant solution.[4]

Critique of Cartwright's move But hang on, all this is just a cardsharp's trick, isn't it? Did we not have a very similar problem for analyses of dispositional predicates? Did we not have the trouble to exclude the workings of finks and antidotes in the antecedent of the dispositional counterfactuals (see Section 3.5)? And did we not see that this was as deep a problem as the *cp* clause is for laws? If we now 'strictify' the law in saying that such and such always and without exception has the capacity to do something we merely veil the problematic *ceteris paribus* clause by the dispositional term.

Cartwright's reaction to this allegation is straightforward and it fits perfectly to the 180-degree turn we announced at the beginning of this chapter:

> They [the *ceteris paribus* conditions, MS] must be introduced if one is stuck with the project of reducing [. . .] capacities [. . .]. But I advocate giving up that programme entirely – especially since it seems not to work – and accepting that capacities [. . .] are real things in nature. There is, I think, no other view of nature that can give an adequate image of science.
>
> (Cartwright 1989: 169–70)

In other words, if we accept that capacities are real and exist in their own right then we no longer need to provide a reductive analysis of them. Thus, regarding dispositions, we do not face the problem of *cp* clauses anymore.

Still, can this be all there is to say? Remember the earlier example of the immunisation strategy that astrologers use: faced with the apparent failure of a horoscope (you did not become a millionaire) the astrologer might say in her defence that this was due to some unforeseeable celestial movements and that she still was quite right to say that you had a *disposition* and a *strong tendency* to become rich quick. Now, aren't the proponents of capacities not in danger of letting capacities degenerate into horoscope-like entities if they do not specify triggers, circumstances and manifestations precisely? Reduction or not: aren't criteria for when we are justified in ascribing capacities essential?

The measurement of capacities Cartwright is aware of this pressing desideratum and the title of her second book – *Nature's Capacities and their Measurement* (1989) – indicates her answer: the workings of capacities can be empirically researched. We can set up experiments in which they display their manifestation on being triggered fully and undisturbed. We can also inquire, in the laboratory, which factors would lead to altered manifestations, to no manifestation at all, or even to the loss of the capacity. In the words of Cartwright's book title: capacities can be measured and it is because of this possibility that capacities are, unlike horoscopes, scientifically kosher and not speculative, esoteric inventions:

> The most stringent kind of Empiricism that seems to me to make sense is the Empiricism of practice that I advocated throughout; the Empiricism that demands that each quantity be measured and each claim be tested. And the principle argument of this book is that causes and capacities are as empirical in that sense as it is possible to be.
>
> (Cartwright 1989: 167)

The transportability argument This point leads us straight onto the second, more direct argument in favour of the assumption that (some of) nature's properties are capacities. (Remember, the first is that capacities help deal with the issue of *cp*-laws). As already mentioned, in scientific practice ideal experimental laboratory setups allow us to empirically research how, for example, quarks, protons, atoms, molecules and more complex entities behave *in the lab environment*. However, Cartwright asks how it is that we can project what we measure in highly specific lab conditions to messy real life circumstance[5] where disturbances of all kinds might influence what happens (Cartwright has a name for the human-made artefacts that produce regularities in the lab: 'nomological machines').

Cartwright's answer, which best explains this practice, is the assumption that what we measure in the lab are *a system's capacities*, its dispositions to behave. What makes lab experiments so useful is, first, that

> these are the circumstances where the feature under study operates [. . .] without hindrance or impediment, so that its nature [capacity, MS] is revealed in its behaviour.
>
> (Cartwright 1992: 50)

Second, as long as the system itself is fairly stable and the outer-lab circumstances, although messy, are not destructive, the system's capacities remain what they are, i.e. they are transported from the lab to the outer world. Thus, within certain boundaries, the system still *tends* to do what it did in shielded circumstances:

> The capacity is something they [systems or objects that are experimentally tested, MS] can be expected to carry with them from situation to situation.
>
> (Cartwright 1989: 145)

This is what enables us to predict a system's behaviour in less than ideal conditions:

> One must suppose that the causal possibilities [of a capacity, MS] that are established in that situation [the lab, MS] continue to obtain in various different kinds of situation.
>
> (Cartwright 1989: 147)

> You always assume that they will *try to behave* in new arrangements as they have tried to behave in others. They will, in each case, act in accordance with their nature [power, MS].
>
> (Cartwright 1992: 49, emphasis added)

Although in ordinary circumstances the capacities might manifest themselves less cleanly or there will be manifestations of other capacities which interact with the first ones, the core of what has been found out in the lab carries over: the dispositions' productive powers (PROD): 'they [. . .] try to behave' and their modal profile (MOD): 'the causal *possibilities* [. . .] continue to obtain'.[6]

This story of the *transportability of capacities* from one situation to another plays a very central role in Cartwright's argumentation because it also justifies why we gain scientific insight by what we are doing in the lab.[7] On a meta-level Cartwright uses this story as an argument for a capacities ontology:

> One cannot make sense of modern experimental method unless one assumes that laws are basically about natures [powers, MS].
>
> (Cartwright 1992: 47–8)

In other words, Cartwright argues that the existence of capacities is a necessary precondition for the possibility of our sciences. We will evaluate these types of indispensability argument or inferences to the best explanation in Section 7.3.

Three critical issues regarding Cartwright's theory We must acknowledge that, occasionally, a system's capacity might be lost entirely due to unfortunate circumstances – for example, haemoglobin molecules might lose their capacity to bind O_2 completely when they are sickle celled. The transportability argument lives on the assumption that systems' capacities are stably carried over from situation to situation; however, this is not always the case and it has the following consequences:

(i) In order to make proper use of our knowledge of the capacities of things, we also need information about the range of their proper workings or their stability, i.e. we need knowledge of abnormal or less than optimal circumstances. However, this leads straight back to the problem of *ceteris paribus* conditions.[8]

(ii) For a second, fairly general critique of Cartwright's argument, remember that it has the structure of an inference to the best explanation: nothing explains our scientific practice of lab research better than the assumption that capacities exist and are transportable. It might even be meant as an indispensability argument: one cannot make sense of modern experimental methods unless one assumes that laws are about capacities (see Cartwright 1992: 47). Is this true, though? Can the denier of capacities really not make sense of lab practice? Or at least in as good a way as capacities-talk can?

In answer to this, two things must be considered: first, although Cartwright's newly interpreted law statements are disposition ascriptions, these statements themselves are still good old regularity statements. Cartwright says this explicitly: which entities have which capacities 'is a mere "brute-fact connection" (Cartwright 1992: 48). Second, in out-of-lab situations, objects are not isolated and various dispositions will interact (for example, objects will have charge and mass). These interactions will most certainly be regular and law-governed rather than randomly happening. These interaction laws are not, or at least not obviously, about (further) dispositions of systems or objects. Thus, we have a second place where laws play a role.

This double reference to laws could encourage us to suspect that we can also, if not equally well or even better, make sense of our scientific practice without reference to capacities. We can say that what is transported from the lab are the entities' properties – neutrally phrased, without special mentioning of dispositions – and because the laws of nature are also true outside the lab it is possible to explain in terms of (categorical) properties and laws (interaction and other) why our lab results can be projected onto the world: all entities and their properties engage in a multitude of laws and the interaction laws amongst them specify how law superposition leads to the overall behaviour.[9] If this draft of a theory can be made more precise, Cartwright can certainly not claim that *only* dispositions talk can explain transportability. Whether she can insist that dispositions are the *better explanation* remains to be argued.

(iii) A final, third critique against Cartwright could be that she does not offer a systematic metaphysics of what the capacities she invokes exactly are (for a detailed discussion, see Psillos 2002: 192–6).

For example, in Cartwright's capacities ontology it is not clear whether capacities are dispositions of 'particles' like electrons (Cartwright 1999: 65), substances like aspirin (see Cartwright 1989: 3), 'arrangements of components' or 'factors' (Cartwright 1999: 50) or whether they are second-order properties of properties (see Cartwright 1999: 70) or all of these together.

All these critical points do not diminish Cartwright's merits as a pioneer of Dispositionalism. Other early founders of Dispositionalism have done a great job to define what dispositions are, metaphysically speaking, especially in contrast to categorical properties (see Sections 2.2.3 and 2.2.4) and we can consult their work[10] as complementation of Cartwright's oeuvre if we wish a tighter, more precise ontology.

This concludes our introduction of Dispositionalism and the ways it can be defended. We move now from the philosophy of science, where these arguments were taken, to semantics. Here (see Section 6.3) we will get to know the thesis of *Essentialism*, which we later combine with Dispositionalism to *Dispositional Essentialism*.

BOX 6.1 and 6.2 Dispositionalism

- **Dispositionalism is the view that dispositions, capacities, causal powers, tendencies, etc. are ontologically basic**: there need not be – in fact, there cannot even be – a reduction to other ontological categories like categorical properties and laws or causation. No reductive counterfactual analysis need or can be given.
- Dispositionalists believe that there are the following reasons in favour of their view:

  (i) The **hard-to-bite consequences of the opposing view, Quidditism** (namely the unknowability of these properties' real natures, the possibility of them swapping their roles, there possibly being 'idlers'),

  (ii) the **groundlessness of Hume's Dictum** (which underlies Quidditism),

  (iii) the **failure of the Humeans' attempt to conditionally analyse dispositions away**,

  (iv) **an indispensability argument for dispositions**.

- Two further reasons of the indispensability kind have been given by Nancy Cartwright, who is a pioneer of Dispositionalism:

  (i) **Dispositionalism is the only or at least best solution to the problem of *cp*-laws**;

  (ii) the **scientific practice** to do **research under unnatural, ideal set-ups in the lab** can best be explained by the assumption that **capacities are measured** there and then **transported** into natural conditions, where they **still tend to produce what they yielded in the lab**.

- Two **counterarguments** against her views are that anti-Dispositionalist defenders of **laws could equally well explain scientific practice** (so that Cartwright's inference to the best explanation is doubtful) and that Cartwright's capacities ontology remains imprecise, including unanswered questions of dispositions' stability.

6.3 Essentialism

Dispositionalism is a view from the *metaphysics* or *philosophy of science* that postulates modal (perhaps necessary) connections in nature which Hume and his followers have denied. Within the *philosophy of language*, there is a parallel and, at first, unrelated development that also concluded that certain facts hold necessarily or that certain propositions are true in all possible worlds.

The two philosophers who are most associated with this view are Saul Kripke and Hilary Putnam (1926–2016). They argued (1980 and 1975 respectively) for *metaphysically necessary connections in nature* on the grounds of *Externalist semantics*, including considerations about *direct reference*, a *causal theory of reference*, *rigid designation* and also on the basis of *a posteriori scientific discovery*. Since these deliberations come first and foremost from the philosophy of language, we need to look indirectly in order to appreciate their impact.

As you will see, a lot will relate back to things we have said about predicates and properties in Section 2.1. Philosophy of language and metaphysics are interconnected at this junction. Remember also that considerations regarding language and meaning were our and the Logical Empiricists' starting point when discussing dispositional predicates (see Section 2.1) and the Verificationist theory of meaning (see Section 1.4.1). Thus, in a way, we return to where we have started.

6.3.1 Semantic Internalism

A proper name like 'Kurt Gödel' refers to a specific person, namely to Kurt Gödel. A predicate like 'gold' applies correctly to all the gold in the world. A word philosophers of language use for the entity a name refers to, or the set of things a predicate is correctly attributed to, is 'extension'. Kurt Gödel is the *extension* of the name 'Kurt Gödel' and the set of all golden things is the *extension* of 'gold'.

Extensions are not meanings It is tempting to say that these extensions, i.e. the thing(s) the name refers and the predicate applies to, are what you know when you understand them, i.e. when you know what the predicate and the name mean. That is, it seems prima facie right to say that what we grasp when we understand the *meaning of the name* 'Kurt Gödel' or the *meaning of the predicate* 'gold' is that they have these respective extensions. This, however, can't be right. Meanings are finer grained than extensions. It is easy to see this: we only have to realise that there are *co-extensive* terms (terms with the same extension) that differ in *meaning*. (In what follows, we will focus not on names but on predicates because they are the linguistic pendants to what we are primarily interested in: properties).

Take the two predicates 'renate' and 'cordate'. They are the technical terms for 'creature with kidneys' and 'creature with a heart'. As it happens, all renates are cordates (and vice versa). That is, the set of animals with hearts and the set of animals with kidneys coincide. They are the same set. That is, both predicates have the same extension.[11] Still, you might have learned something new when you were informed that all renates are cordates. That you have learned something new would, however, be dubious if all there is to the meaning of 'renate' and 'cordate' is their extension, because knowing their meaning (their extension) would be knowing the truth of 'all cordates are renates' on the basis of their meaning alone. Because of this odd result we must postulate that meanings are something terms have over and above their extensions (so that you can know their meaning and still be surprised when learning that they are co-extensional, if they are).

Carnapian intensions For that reason, Carnap postulates that predicates, in addition to their *extension* (i.e. the set of objects the predicate is correctly attributed to) also have an *intension* (see Carnap 1956: 16–24, 118–32). What are intensions? Well, abstractly speaking, they are functions from sets of possible worlds into extensions, but this is hard to digest without further information.

Let's go back to our example. As we know, both 'renate' and 'cordate' have the same extension, and yet it is easy to imagine that there could have been creatures that developed hearts without the need of a kidney blood cleansing system and there could be animals with kidneys that do not let circulate their blood via hearts but instead by, say, continuous vein contractions in their bodies. Rephrased in possible worlds talk (see Section 3.3), we could say that there are possible worlds where there are renate creatures without hearts, others worlds where there are cordate creatures without kidneys, and yet other worlds where both is the case. That is to say that in these other possible worlds the extensions of 'renate' and 'cordate' differ.

According to Carnap, the intension of a predicate is a mechanism or function that 'looks' at each possible world (ours, others) and determines individually for these worlds what the extension is for that predicate in each world. In short, intensions are extensions per world. With this notion at hand we can say that 'renate' and 'cordate' have different intensions because in some worlds their extensions vary.

Finally, we can identify a predicate's *meaning*, that what predicates must have over and above their extensions, with these *intensions* and thereby take care of the phenomenon that there can be co-extensional predicates that differ in meaning.[12]

Intensions in speakers' heads Intensions, as functions from possible worlds into extensions, are fairly abstract entities. They seem not to be the kinds of things we ordinary people have in mind when we grasp the meaning of predicates and formulate our sentences. Carnap was aware of that fact and in an addendum to his *Meaning and Necessity* (Carnap 1956: 233–47), he brings intensions, these theoretical constructs, home, i.e. into the heads of natural language users. Here's what Carnap thinks the intension of a predicate is, not in the abstract, but for a speaker:

> [t]he intension of a predicate '*Q*' for a speaker *X* is the general condition which an object *y* must fulfil in order for *X* to be willing to ascribe the predicate 'Q' to *y*.
>
> (Carnap 1956: 242)

As a speaker you will, consciously or tacitly, have some criteria in your mind for when a predicate fits and when not. For 'horse', being four-legged, having hair, a tail and a long face might belong to your criteria. These you check when confronted with sample objects. If they fit the bill you count them in, if not, don't.

That there should be such criteria available to us, i.e. general conditions which an object *y* must fulfil (to count as a horse, say), is very much in accordance with the Logical Empiricists' Verificationist doctrine (see Section 1.4.1): there must be clear – and *observable* in the most radical formulation – criteria for its application, otherwise the term under concern is dubious. If, for example, for the predicate 'ectoplasmic' there are no such criteria, then that predicate is, according to the Empiricist, nonsensical; in other words, it has no meaning, i.e. intension.

The list of criteria an object has to meet for a predicate to be correctly ascribed to it, fulfils precisely the role that was abstractly asked of intensions, namely to decide for any situation (or possible world) which object belongs to the predicate's extension and which does not. In other words, criteria lists in speakers' heads are *a concrete realiser of the abstract function* which was formulated theoretically for intensions.[13]

We now can solve the renate/cordate riddle from the beginning. You have grasped the meaning (i.e. intension) of these predicates, that is, you know some general condition which an object must fulfil (in this case, having kidneys) in order for it to count as renate. Likewise, with 'cordate', you know some general condition which an object must fulfil (in this case, having a heart) in order for it to count as cordate. Having grasped these meanings, you still might not know the fact that all renates are cordates. Therefore, (knowing the) meaning is not (knowing the) extension.

The conventional nature of intensions Here's another link back to the Empiricist/Verificationist thoughts regarding language we discussed earlier in the book: the Empiricists' affinity to *Nominalism*. This was the view that there is not much more to a *property* than the set of objects some *predicate* refers to because any arbitrary set of objects can be named an abundance of property results (see Section 2.1.4).

This translates into our extension/intension dichotomy in the following way. Within a language community, the intensions, i.e. criteria for when an object belongs to the extension of a predicate, are convention guided. Although these conventions are handed over in the language acquisition process from parent to children, they are still, according to the Empiricist, in a sense arbitrary. There are no natural, divine or otherwise decried meanings/intensions to our predicates. Relating this to Nominalism: intensions of predicates are arbitrary and therefore their extensions are as well. If there is no more to a property than being the set of objects predicates refer to, then which properties there are is likewise arbitrary. We will come back to the relation of predicates (especially their intensions) and properties again, with some surprising twists. Here, we want to focus more on their conventional nature.

Conventionalism holds true, so Carnap believes, especially in the sciences where we can artificially define new technical terms and fix their meanings as we please. Meanings are, in a sense, exchangeable human tools and there is no other guideline for meaning fixations than that they should be practical for our scientific or technological goals, that they are suitable for prediction and explanations. In relation to the course of sciences (here the example is *being water*) Carnap observes:

> In the oldest books on chemistry, for example, there were a great number of statements describing the properties of a given substance, say water or sulphuric acid, including its reactions with other substances. *There was no clear indication as to which of these numerous properties were to be taken as essential or definitory for the substance.*
>
> (Carnap 1956: 241, emphasis added)

Later, as the sciences develop,

> there was a *greater intensional precision*. On the basis of the theory of chemical elements, slowly with increasing explicitness certain properties *were selected as* essential. For a compound, the molecular formula (e.g. 'H_2O') *was taken as definitory* [. . .] for the elementary substances, first certain experimental properties were more and more clearly *selected as definitory*, for example the atomic weight, later the position in Mendeleev's system. Still later [. . .] nuclear composition *was regarded as definitory*, say characterized by the number of protons (atomic number).
>
> (Carnap 1956: 241–2, emphasis added)

We highlighted here the conventional character of the business to fix intensions: 'were selected as', 'was taken as', 'was regarded as'. In the view of the Empiricists, those conventions were chosen merely for *practical* benefits: ‘We can at the present time observe *the advantages already obtained* by the explicit conventions which have been made’ (Carnap 1956: 242, emphasis added).

Intensions fix extensions Returning to the basic idea to identify meanings with intensions and intensions with criteria lists we must note something else that is striking about Carnap’s semantics: the meaning/intension of a predicate determines the set of things the predicate applies to. That is, *intensions determine extensions*. In the extension of a predicate is whatever fulfils the intension’s criteria. As a corollary, therefore, terms with the same intension also possess the same extension.

This highlighted changeable, conventional character of meanings/intensions together with the fact that intensions fix extensions have a striking and problematic consequence, especially for scientific discourse. Before we start with these problems, we let Hilary Putnam (1975), a student of Carnap and later a critic of his theory of meaning, summarise Carnap’s view:

> In his early years Carnap thought of understanding a term as possessing the *ability to verify* whether or not any given entity falls in the extension of the term. In terms of intensions: ‘grasping’ an intension would amount, then, to possessing the ability to verify if an entity *e* in any possible world *x* belongs to *f(x)* or not [f(x) being the intension-function which determines for world *x* the extension of the predicate under concern; MS].
>
> (Putnam 1975: 185)

The problem of incommensurability For the ancient Greeks, gold was as valuable as it is today. As folklore has it, Archimedes was asked by King Hiero II to determine whether a certain crown was made entirely of gold or whether the goldsmith had secretly substituted some of the gold with a less valuable metal. (Archimedes and Hiero will have said ‘χρυσός’ (chrysos) not ‘gold’ but let us leave translation issues aside).

Although Archimedes was allegedly successful in detecting the forgery in the case of the crown, we can easily imagine – and will do so with Putnam (see 1975: 153ff) – that there was another piece of metal, *X*, which he, with his ancient methods, would not have been able to identify as forgery: ‘Archimedes would have said that our hypothetical piece of metal *X* was gold, *but he would have been wrong*’ (Putnam 1975: 153).

This seems possible. However, there is an alternative *relativistic* way to describe the situation. We could instead say, ‘Archimedes had his theory and his meaning for the word "gold". So, for him, that piece of metal would indeed have been gold. His *gold* is simply not our *gold* and which theory you subscribe to – his, ours, a future one – is arbitrary'.

In fact, this latter view is one Carnap and others might well be committed to: Archimedes intension for 'gold' might have been 'bright yellow and soft, malleable metal with a certain density' and, thus, a certain piece of metal that fulfils this intensional criteria might well fall within Archimedes extension for gold. However, as today's intension for 'gold' has essentially 'chemical element with atomic number 79', that piece might today be considered as fool's gold (or something like it).

Consequently, on Carnap's view, there is no trans-historical fact of the matter whether 'This is gold!' is objectively true or false. Worse, it is hard to defend within this view that science has made any progress: if Archimedes talked about something else when using the word 'gold', i.e. when what he meant by 'gold' is not what we mean by 'gold', then our theories about allegedly golden things cannot be compared and it is not possible to say that the one theory is truer than the other or that one of them should succeed the other. They simply speak about different things.

This odd consequence of theory change – when Carnapian (and similar) semantics are assumed – was first highlighted and also partially defended by Thomas Kuhn (1962), Paul Feyerabend (1924–94; see Feyerabend 1975) and others. The phenomenon has become known under the name '*incommensurability of theories*'.[14]

Putnam believes that this picture can't be right. Scientists and laypeople assume that there is real scientific progress. The visible technological progress, which is based on the science behind, attests to it. Therefore, Putnam believes that the semantic theory behind Kuhn's and Feyerabend's incommensurability problem (i.e. Carnap's or similar ones) must be wrong. A theory that has incommensurability as a consequence can't be right because we do and can compare scientific theories and assess their progress. However, if Carnap's theory of meaning for (scientific, natural kind) terms is wrong, which semantics is correct?

6.3.2 Semantic Externalism

Some metaphysical preparation work needs to be done to appreciate the next semantic move. In Sections 2.1.4 and 4.3, we saw the need for the assumption that nature comes equipped with her own properties. If we are lucky our sciences discover them. The most fundamental properties of physics (being negatively charged, having a certain mass, etc.) are legitimate candidates. Other than non-natural properties which reflect human interests (like 'vegan approved', 'product made in China' or 'impressionist painting'), these natural properties do not depend at all on humans and their ways to refer to and describe the world.

Carnap and natural properties We know that the assumption of the existence of natural properties is against the Empiricists' everything-goes attitude regarding properties, and yet Carnap acknowledges that this attitude is hard to maintain. By the time *Meaning and Necessity* (from which we took the above intensions theory) was published in 1956, he no longer had reservations to adopt, tentatively, a Realist/ Physicalist language and to speak as if (natural) *properties* existed:

> [*Property*] is not to be understood in a mental sense, that is, as referring to a process of imagining, thinking, conceiving, or the like, but rather to *something objective that is found in nature* and that is expressed in language by a designator.
>
> (Carnap 1956: 20)[15]

Now, note a grand possibility that opens up here: if you accept that there are physical properties or features out there in the world and that objects either have or do not have these properties then it is a natural step to identify the meaning/intension of (some) *predicates* with these *properties* that exist in reality:

> By the *intension* of the predicator 'P' we mean the property P; by its *extension* we mean the corresponding class. [. . .] The term 'property' is to be understood in an objective, physical sense, not in a subjective, mental sense.
>
> (Carnap 1956: 16)

In other words, once we have (natural) properties at hand, Carnap can start to identify intensions /meanings with properties, instead of identifying intensions/meanings with criteria lists in the heads of speakers (as he did earlier). In fact, such properties (as much as the original criteria lists) would fulfil precisely the role or function that was abstractly asked of intensions for predicates, namely to specify for any world or situation, what belongs to the extension of the predicate – those objects that have those properties. In other words, criteria lists or properties can be concrete role-fulfillers of the intension's abstract function. Properties are worldly realisers of it; lists in speakers' heads are mental realisers.

All this points to the path we will go down with Hilary Putnam and Saul Kripke,[16] but note that Carnap himself did not develop his idea further in this semantic direction. (In fact, he was quick to add anti-Realist caveats to this properties Realism).[17] One of the welcome consequences of the new semantics will be that the Kuhn-Feyerabend threat of incommensurability can be avoided.

Natural properties and natural kinds Let us, as Kripke and Putnam do, take the full realist option: natural properties do exist (see Section 7.5). Actually, we add a further but similar ontological category: next to natural properties, Realists often also speak of natural kinds. Typical examples include being a tiger, being water, being gold, etc. What are natural kinds and how do natural kinds differ from simple natural properties?

For our purposes, we give an answer that comes from Katherine Hawley and Alexander Bird (2011). Natural kinds encapsulate many natural properties in one: natural kinds are complex natural properties (see Hawley and Bird 2011: 208).[18] Hawley and Bird give two central examples (and distinguish them further from each other but we leave this aside). The first is that methane molecules form a natural kind:

> For something to be a methane molecule, it must have four parts which each instantiate being a hydrogen atom, and a fifth part which instantiates being a carbon atom. Moreover, those five parts must stand in the appropriate geometrical and physical relations to one another, i.e. they must instantiate the relevant relational universals in the relevant order.
>
> (Hawley and Bird 2011: 209)

The second is that:

> Electrons form a natural kind of a different sort to methane molecules. To be an electron is to instantiate each of several specific properties (having charge $-1.602176 \times 10^{-19}$ C, having mass 9.109382×10^{-31} kg and having spin 1/2). We can thus identify the kind with a conjunctive universal, i.e. a complex universal with electron-mass, electron-spin, electron-charge and identity amongst its proper parts. Each electron instantiates this conjunctive complex universal.
>
> (Hawley and Bird 2011: 211)

In short, natural properties are the very basic, fundamental features of reality, natural kinds are compositions of those, either by the same thing (electrons, protons, quarks) instantiating a couple of them together (having charge, having mass, etc.) or by a structured compound of entities (methane or water or nitro-hydrochloric acid) that possess natural properties. It is important to note that the composition is also natural – just like nature is equipped with fundamental natural properties, so has nature a natural way to compose them.

We introduced natural kinds on top of natural properties because Kripke and Putnam mostly talk about kinds rather than properties, and so will we from now on.

Direct reference We turn back to semantics. Two parts of Kripke's and Putnam's semantic idea about how natural kind predicates work[19] will become extremely important for us: *direct reference* and *rigid designation* (the latter we will discuss later). *Direct reference* is almost self-explanatory when we reconsider Carnap's opposing picture. Carnap believed that whether or not an object belongs to the extension of a predicate, or in this case a natural kind term, is determined by the intension of that predicate or term (i.e. through the list of criteria speakers have in mind).[20]

Kripke's and Putnam's theory works without this detour through intensions, i.e. through descriptions that sort out the correct objects. The theory of *direct reference* says, on the contrary, that a natural kind term, 'gold', say, *refers directly*, i.e. unmediated by some description, to the right objects. But how does it do so? Metaphorically speaking, how does a natural kind term find its referents when it cannot search for them with a description at hand which they should fulfil?

Here's the idea (see Putnam 1975: 141ff). It is clearly a bit idealised but we can imagine the following scenario: some medieval alchemists were the first to discover, or synthesise in their cauldrons, nitro-hydrochloric acid. They might well have proclaimed in the presence of their fellows: 'Let us call *this liquid* [pointing to a sample of it], and everything that is like it "aqua regia"!' (this being the old, Latin name, i.e. royal or king's water, for nitro-hydrochloric acid). That kind of stuff has, so to speak, been baptised at that moment. Thus, the extension of a natural kind term like *aqua regia* is fixed not through the detour of intensions but first and foremost by directly pointing and referring to a paradigmatic sample of these kinds. *Aqua regia*, from that moment on, refers to that kind of stuff.[21]

Two assumptions have to be made in the background for this to work. First, in order to have all the nitro-hydrochloric acid in the world in the extension of *aqua regia*, and not just the sample of that acid, a similarity assumption has to be made: *this* (pointing to the sample) and *everything like it* is *aqua regia*. That there are other occurrences (of the same kind of stuff) that are exactly the same as the one pointed to is somewhat part of the assumption that nature itself comes equipped with natural kinds. It seems legitimate to believe that there is something about these occurrences of the kind that makes them alike.

Causal chains of reference Second, future language users (alchemists and scientists), who have not been present at the baptism situation, have to cooperate with the former language users insofar as they have to tacitly intend to also use the term *aqua regia* to refer to the kind of stuff to which their fellows originally pointed to (and to similar stuff). In other words, when the term *aqua regia* is handed down from speaker to speaker, a causal chain is established from speaker to speaker that is ideally traceable back to the original baptism. Reference to the same kind of stuff, when using the term 'aqua regia', is secured: the extension of *aqua regia* is partially fixed by direct reference to a sample and this sample is a natural kind of stuff the world comes equipped with.

The reader might have spotted an incongruency in our story. We said that natural kind terms refer directly to the respective natural kind without the detour of intensional descriptions. However, in our example we made the alchemists say 'this *liquid*' and we can easily imagine that he might have added '*which can dissolve the noble metals*', in order to distinguish it from the bottle of ordinary water which happened to stand on the table as well. In other words, descriptions did play a role in the baptism situations when picking out the intended sample.

Stereotypes Putnam agrees; however, he makes plausible that these initial descriptions later become irrelevant and are not a constitutive meaning of the term. Indeed, the first reference to fixing descriptions might even turn out to be false for the named natural kind. For example, it turned out that *aqua regia*, i.e. nitro-hydrochloric acid, cannot dissolve all noble metals (gold and platinum do dissolve; however, titanium, iridium and tungsten withstand its acidic power). In other words, the alchemists' description 'which can dissolve the noble metals' was merely an initial disposable tool to fix the reference. Once the reference is fixed, natural kind terms refer directly to the natural kind stuff without the detour of the disposable and maybe wrong descriptions.

Putnam believes that descriptions (correct and incorrect ones) merely belong to what he calls the '*stereotype*' associated with the respective stuff by most or many people. However, these stereotypical descriptions are, and this is very important to note, very different from Carnap's intensions, which would be the essential means to fix the reference of the term (remember: intensions fix extensions). Stereotypes are not essential – even disposable (see Putnam 1975: 169–70).

Scientific progress and alleged incommensurability again Remember the dire consequences of Carnap's semantics for natural kind terms: if it were right (i.e. if intensions were to fix extensions), then succeeding scientific theories would talk past each other because where one would speak about stuff that fulfils the descriptions a, b and c, the successor theory might talk about stuff that fulfils c^*, d and f. When Newton, for example, talked about mass stuff he had in mind that it has the feature to remain constant even when accelerated to a velocity v ($m_v = m_0$).

When Einstein talked about stuff that fulfils $m_v = m_0/\sqrt{(1-v^2/c^2)}$, Einstein's theory is not succeeding Newton's theory and it is even less so a better theory. Rather, Einstein talks about something else because his word 'mass' means something difference to Newton's 'mass', and thus their theories are incommensurable.

This seems an odd thing to say. Our intuition is that Einstein's theory succeeds Newton's and that it makes the latter's theory better. In Putnam's (and also Kripke's) theory of natural kind terms this intuition can be met. 'Mass', for Kripke and Putnam, is a natural kind term that refers first and foremost to stuff like *this* (pointing to some massive objects like an apple or a dumb-bell). In other words, the reference of 'mass' has been fixed (long before Newton) and without the detour of some description. Newton did think something different about that kind of stuff than Einstein, and yet Einstein, in discovering something new about mass, corrected the stereotype we had of mass before the formulation of the relativity theory. Our stereotype for 'mass' was updated but 'mass' still refers directly to the same kind of stuff, mass.

Kuhn's and Feyerabend's incommensurability suggest that later, intuitively better theories cannot be compared in their quality to older theories because, according to these two philosophers, the theory's terms have different meanings, i.e. speak about different entities. This no longer applies because in Putnam's semantics,

> later theories in a mature science were, in general, better descriptions of *the same entities that earlier theories referred to*. In my opinion the hypothesis that this is right is the only hypothesis that can account for the communicability of scientific results.
>
> (Putnam 1975: 155, emphasis added)

Internalism versus Externalism At this point, we can fully understand why Kripke's and Putnam's semantics is called *Externalist* and why Carnap's is rightly named *Internalist*. The *Internalist* thinks that speakers have, *internal in their heads*, criteria for what gold, water, molybdenum and other natural kind terms refer to. These inner states of the mind of the speaker play a crucial role for the meaning of natural kind terms. Their extension is picked out through these descriptions. The *Externalist* – Putnam offers the infamous slogan '"meanings" just ain't in the head!' (Putnam 1975: 144)[22] – believes that 'water' (etc.) refers directly to water, i.e. the stuff *external to our heads*, out there in the world. Later, we might find out a lot about that stuff and add it to our stereotype. However, again, the extension comes first and the world has a lot to say about what we should add to the stereotype because descriptions are apt if and only if they trace reality correctly.

Rigid designation In addition to *direct reference*, we also mentioned *rigid designation* as one of the two important features of Externalist Semantics: natural kind terms (like 'water') not only directly refer to the respective stuff (here water) but they also *designate* their target kind *rigidly*. What does this mean?

Even if we discover that what we once believed about water is wrong, the term 'water' continues to stick to the originally baptised kind of stuff (here H_2O). A kind term rigidly and irreversibly designates the original.

The thesis of rigid designation can be and has been pushed even further. Even if we fabulate and imagine the world in alternative ways, 'water' still refers (in that alternative world) only to stuff that is like our water. Putnam offers his readers his famous twin earth thought experiment to endorse this view:

> [w]e shall suppose that somewhere in the galaxy there is a planet we shall call Twin Earth. Twin Earth is very much like Earth; in fact, people on Twin Earth even speak English. In fact, apart from the differences we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is exactly like Earth. [. . .]
>
> One of the peculiarities of Twin Earth is that the liquid called 'water' is not H_2O but a different liquid whose chemical formula is very long and complicated. I shall abbreviate this chemical formula simply as XYZ. I shall suppose that XYZ is indistinguishable from water at normal temperatures and pressures. In particular, it tastes like water and it quenches thirst like water. Also, I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water, that it rains XYZ on Twin Earth and not water, etc.
>
> If a spaceship from Earth ever visits Twin Earth, then the supposition at first will be that 'water' has the same meaning on Earth and on Twin Earth.
>
> (Putnam 1975: 139–40)

However, on further inspections, the visitors will find out that the extension of 'water' in the twin earthlings' sense of the word ('water$_{TW}$', so to speak) is the set of all samples that consist of *XYZ* molecules and not, as in our own sense of 'water' (better: 'water$_{E}$'), the set of all samples that consist of H_2O molecules (see Putnam 1975: 141).

After the discovery, we would clearly say that their stuff just isn't water (even if they might call it 'water') because water is that kind of stuff (pointing to a sample here on normal earth, which is H_2O) and not *XYZ*. 'Water' as a term used by ordinary earthlings always rigidly refers to H_2O and nothing else. Other stuff *might be called similarly* by other language communities – twin earthlings, for example – but *it just isn't water*. (To use our indices again: water$_{E}$ is not water$_{TW}$!)

Note that this was true even at times, say in 1750, when neither we nor the twin earthlings had an inkling about the chemical structure of the respective liquids, that is, when their and our beliefs (stereotypes!) about the respective chemically different liquids, H_2O and *XYZ*, were identical: tasteless, thirst-quenching, transparent liquid, occurring in oceans, lakes and seas. Even then, according to the consequence of direct reference and rigid designation, 'water' as used by earthlings referred strictly to H_2O, and 'water' as used by twin earthlings referred strictly to *XYZ*.

In his Twin Earth example, Putnam imagined another planet or possible world where some stuff, *XYZ*, had identical surface features to those of water, although clearly that stuff, *XYZ*, isn't water despite superficial appearance. If, on the other hand, we *really* wish to imagine water on a different planet or in another possible world we must think of it as H_2O because if we only imagine some stuff that is similar in appearance we have not imagined water but just something that looks like it. If we imagine water on a different planet or in another possible world we imagine H_2O.[23]

6.3.3 Essences

In all possible worlds, it is true that water is (or has as its substructure) H_2O. Truth in all possible worlds, however, equals necessity. Hence, according to Putnam's theory, we can say that *necessarily water is* H_2O. This brings us to a very important topic at the core of this section.

Natural kind terms refer to all and only those samples that are *of the same kind of stuff* as the first, baptised sample. For this to work, we made a background assumption: there is something that unites these samples, that makes them one kind of stuff. In other words, the sample of liquid we or someone referred to in the baptism situation as a paradigm example for water stands in some natural relation to all other of its occurrences (and thus all samples stand amongst each other in this relation). In the case of water, Putnam explicitly speaks of the relation as '*being of the same kind of liquid as*'.

> Suppose I point to a glass of water and say 'this liquid is called water' [. . .]. My 'ostensive definition' of water has the following empirical presupposition: that the body of liquid I am pointing to bears a certain sameness relation (say, x is the same liquid as y, or x is the same$_L$ as y) to most of the stuff I and other speakers in my linguistic community have on other occasions called 'water'.
>
> (Putnam 1975: 141–2)

Three things are immensely important here: first, there are features that unite natural kinds and they are, according to Putnam, not some ephemeral surface features but underlying deeper, more fundamental chemical substructures. Natural kinds,

> do have a hidden structure, and it is good scientific methodology to use the name (i.e. the kind term, MS] to refer rigidly to the things that possess that hidden structure, and not to whatever happens to satisfy some [superficial, MS] description.
>
> (Putnam 1975: 163)

Second, of these uniting features/underlying structures, we can say that they belong *necessarily* to the respective natural kind. Water is necessarily, essentially, H_2O.

Third, and also important, what that essence is can be empirically discovered by science – for water, for example, *x* is the same liquid as *y* (namely water) if and only if *x* and *y* share the underlying feature of being H_2O. In other words, the underlying essence of natural kinds is or can be discovered aposteriorily.[24]

Aposteriority Taking these three aspects together, the 'same general hidden structure (the same "essence", so to speak)' (Putnam 1975: 153) is empirically revealed, i.e. what water necessarily is, namely H_2O, is an *a posteriori* discovery. To underline its aposteriority further, we can contrast this necessity to logical or conceptual necessities. The fact that water is H_2O is not obvious from an analytic, conceptual viewpoint. People prior to the eighteenth century were competent users of the term 'water' without having the slightest idea about its essential substructure. This is unlike the case where you master the word 'bachelor' and, because of your language competence, you know *a priori* that all bachelors are unmarried (see Section 1.3). You can't analytically make such a link from water to H_2O purely on your language competence.

Metaphysical Necessity Putnam and Kripke were therefore convinced that the *a posteriori* discovered necessity of water being H_2O – or gold having the atomic number 79, or electrons having unit charge $e^- = 1.602 \times 10^{-19}$ Coulombs, to give other examples of natural kinds essences – is a different kind of necessity than conceptual necessity. It's a *de re* necessity, discoverable in things, in the world, and they called it a '*metaphysical necessity*'.

This is a striking result, especially when we think back to Humean or Logical Empiricist themes. Remember that Hume had two arguments against necessary connections (see Section 1.2). First, that we have *no a priori means* to discover necessary connections. This is still widely accepted amongst Hume's supporters and his opponents. However, many philosophers no longer accept Hume's second claim that we have no *a posteriori* epistemic access to necessity because they have been convinced by Kripke's and Putnam's arguments for such connections.[25]

This possibility of the existence of *a posteriori* discoverable necessities or essences of natural kinds (for example, that water is necessarily/essentially H_2O) is an important revelation, and one that many philosophers have accepted and utilised in their metaphysical theories. We discuss a crucial application in the next section (see Section 6.4) but we can let one of its later protagonists summarise the stance of *natural kinds Essentialism*:

> The distinctions between natural kinds are based on facts about their essential natures or structures, not on how we find it useful, convenient, or natural to classify them. Thus, membership of a natural kind is decided by nature, not by us; and the question of whether something is or is not a member of a given natural kind can never be settled just by fiat or arbitration. This question can only be settled by discovering whether what is to be classified has the essential properties or structure of the kind in question.
> (Ellis 2001: 19)

Before we turn to more details of Brian Ellis's metaphysics, we wish to hint at how Putnam's and Kripke's view can be criticised.

A critique of Externalist semantics Going through Kripke's and Putnam's arguments from the philosophy of language it might seem as if the metaphysical result of necessary connections in nature or the essences of natural kinds can be gained from considerations about *direct reference* and *rigid designation*, i.e. prima facie, it might look as if semantics yields this ontological outcome. However, this is not so.

Nathan Salmon, Alan Sidelle and Scott Soames could convincingly show that there is some heavyweight metaphysics going on in the background and it is widely acknowledged that if we do not stick Essentialist assumptions into the argument from the start we won't get necessity or essences out either.[26] We cannot go through the details of the arguments here but remember that Putnam had to assume that there is 'a hidden structure' in natural kinds and that,

> it determines what it is to be a member of the natural kind, *not only in the actual world, but in all possible worlds*.
> (Putnam 1975: 160, emphasis added to indicate the modal impact)

To highlight the Essentialist input that is smuggled in, consider these alternatives: there could have been no underlying structure or, indeed, we could stipulate that the substructure is not essential but rather the surface appearance. Why not say '*x* is the same liquid (natural kind) as *y* if and only if *x* and *y* share their superficial features (being a transparent, thirst-quenching liquid)'?

Of course, the hypothesis that it is the chemical substructure (if anything) is more intuitive and probably more useful: scientifically discovered features are fairly reliable and figure in our best theories about the world. It is not unreasonable at all to assume that they (if anything) are the essential ones. However, as reasonable as the Essentialist assumption is, it is a prior assumption put into the semantic argument from the outset, and one that needs separate support.[27]

Where could this support come from? Here's a possible route: with Armstrong, Lewis and other philosophers we introduced the belief in natural properties. They thought they were *indispensable* for the best theories of laws of nature or, slightly weaker, that *nothing can explain better* what a law of nature is than a theory that presupposes the existence of natural kinds (see Sections 2.1.4 and 4.3). If we accept this indispensability argument (or inference to the best explanation, IBE) for the existence of natural properties, then the step to the belief in natural kinds and their essences is small indeed. If we already accept that the external world comes equipped with her own properties, isn't it therefore natural to also accept that nature provides the world with her own kinds that have their naturally given essential substructures, features and properties? Indeed, this is the route many Essentialist take (see Ellis 2001: 145–50; Lowe 1998, 2006). We will discuss and weigh *IBE*s and *indispensability arguments* in Sections 7.3 and 7.4.

Aristotelean Natures The Essentialism presented here is the view that natural kinds and properties have certain of their features necessarily. Phrased in terms of possible worlds, an Essentialist might express their view thus: to say that water is essentially H_2O is to say that, necessarily, anything that is a sample of water is composed of H_2O molecules, i.e. that there is no possible world where a sample of water does not consist of H_2O (but, say, *XYZ* instead).

Some philosophers, notably Kit Fine, turn this view on its head. They argue that essences stem from the identity or nature of the respective kind and that they are not to be expressed by or reduced to necessities. Rather, according to Fine,

> it seems [. . .] that far from viewing essence as a special case of metaphysical necessity, we should view metaphysical necessity as a case of essence.
> (Fine 1994: 9, also Fine 2002).

We do not have the space to discuss this twist further, but note that it brings Essentialism closer to the theses of one of its very early founding fathers: Aristotle. In his *Metaphysics, Book VII: Zeta,* Aristotle speaks of an essence of a thing as '*to ti ên einai*' ('the what it was to be') or '*to ti esti*' ('the what it is'). We find claims like 'the essence of a thing is what it is said to be in respect of itself' (Aristotle 1029b14) and also claims targeting kinds more directly: 'species of a genus have an essence' (Aristotle 1030a11–12). With this historical embedding of the view, we end our section on essence.

Back to the core issues in Metaphysics of Science We said at the beginning of Section 6.3 that semantic Externalism seems at first unrelated to the philosophy or metaphysics of science. We have seen that this is not so when we presented Putnam's reasoning against the relativity and incommensurability of scientific theories which would be the result of Internalism, according to Kuhn and Feyerabend.

In any case, in the wake of Kripke's and Putnam's arguments, many philosophers saw the chance of a revival for anti-Humean metaphysics of a broader kind. Talk of essences and metaphysical necessity became admissible again, either because intuitions we hold were unearthed and made attractive again or because the essence assumptions were judged to be indispensable or immensely serviceable to metaphysics. Here, speaking of laws of nature, Psillos comments on this historical development:

> It was Kripke's liberating views in the early 1970s that changed the scene radically. By defending the case of necessary statements, which are known a posteriori, Kripke [1972] made it possible to think of the existence of necessity in nature which is weaker than logical necessity, and yet strong enough to warrant the label necessity. [. . .] As a result of this, the then dominant view of laws as mere regularities started to be seriously challenged.
>
> (Psillos 2002: 161)

Kripke, in the final addendum (g) to his *Naming and Necessity*, cautiously suggested such an extension of his findings:

> A good deal of what contemporary philosophy regards as mere physical necessity is actually necessary tout court. The question how far this can be pushed is one I leave for further work.
>
> (Kripke 1980: 164)

Despite this, a tight and deliberate connection between *Dispositionalism* and Kripke–Putnamian *Essentialism* only started to be made apparent in the early 2000s. It bears the telling name of '*Dispositional Essentialism*'. Our next section is devoted to it and its applications to counterfactuals, laws and causation.

BOX 6.3 Essentialism

- The modern revival of Essentialism is due to the **Externalist semantics** of **Hilary Putnam** and **Saul Kripke**: the **extension** of a natural kind term is **fixed by direct reference** to a paradigmatic sample of the kind out there in the **external world** (‘*this* is water’, accompanied by a pointing gesture to a sample of water) together with three **background assumptions**:
  - (i) **nature indeed has the kind *water* as one of her building blocks**,
  - (ii) the term ‘**water**’ **refers rigidly** to (samples of) water,
  - (iii) **all samples of it, across possible worlds, share the same underlying chemical structure**.
- This underlying chemical structure is **discovered empirically** due to scientific enquiry (in the case of water, it happens to be the chemical structure H_2O) and so the Essentialist conclusion is an ***a posteriori* necessity** claim: it is *a posteriori* necessary that every sample of water has the chemical structure H_2O i.e. water is essentially H_2O.
- Carnapian **Internalist Semantics** – so called in opposition to Externalism because, here, the reference of natural kind terms is fixed through intensions, i.e. **criteria lists of features, available to the speaker in their heads**/minds, which entities need to fulfil in order to count as a member of the kind – has no commitment to Essentialism. In good old Empiricist fashion, the criteria in the list are freely exchangeable and, if defended, then only for mere pragmatic reasons but not because they belong to the natural essence of the kind.
- The postulations of **natures or essences would mean accepting necessary connections in nature**. Water is *necessarily H_2O*; gold has *necessarily* 79 neutrons, etc. This is why modern Humeans are disinclined to believe in such essence claims.

6.4 Dispositional Essentialism

In Section 6.2, we introduced *Dispositionalism*, the view that some or all properties are of a dispositional nature, and we discussed arguments in its favour, notably those by Nancy Cartwright. In Section 6.3, we introduced *Essentialism*, the view that natural kinds have some of their features (their underlying substructures, for example) necessarily, and we discussed how Essentialism became rehabilitated, particularly due to semantic arguments given by Saul Kripke and Hilary Putnam.

Dispositionalism, Essentialism and Dispositional Essentialism Consider Dispositionalism and Essentialism together and you get *Dispositional Essentialism*, the view that natural kinds essentially possess some dispositional properties. Note that Dispositionalist Essentialism combines two conceptually distinct modalities: the modality of essence, i.e. the necessary possession of certain features, and the (counterfactual) modality dispositions bring with them which binds stimuli and manifestations together. An example for a Dispositional Essentialist claim is: ‘Necessarily, salt is soluble’, where NaCl is the natural kind and solubility the respective disposition it *essentially* has and which makes it true that if salt *were* put in water it *would* dissolve.

Dispositional Essentialists often extend their view to *properties in general* (not only kinds): all properties (at least the fundamental ones) have dispositional features essentially. An example might be *being electrically charged*: necessarily, something has that property just in case it would feel an attractive/repellant force if there were other charges around. In other words, charge essentially brings with it (some would even say ‘can be identified with’) the disposition to be attracted/repelled if in the vicinity of another charge.

A prominent early defender[28] of the stance that all properties are dispositional is Sydney Shoemaker:

> What makes a property the property it is, what determines its identity, is its potential for contributing to the causal powers of the things that have it. This means, among other things, that if under all possible circumstances properties X and Y make the same contribution to the causal powers of the things that have them, X and Y are the same property.
>
> (Shoemaker 1979/1984: 256)

Note that one does not have to combine Dispositionalism with Essentialism: Cartwright believes in dispositions but denies essences: ‘We no longer identify natures [i.e. dispositional powers] with essences’ (Cartwright 1992: 47, my addition in brackets) and Kripke and Putam never explicitly mention dispositions as essences of natural kinds. However, since the mid-1990s the combined has found its defenders, notably Alexander Bird (2007) and Brian Ellis (2001).[29]

The Bookkeeper and the Lumberjack It is important to realise that modal connections in nature enter the stage with both Dispositionalism and Essentialism. Essences account for necessary connections (between being water and being H_2O, for example), and dispositions (see Section 2.1.2) comprise a counterfactual modality: if something has a disposition then it could or would behave in a certain way if something else were the case.

For the moment, never mind the difference that essences are attached to necessity and dispositions to counterfactuals. We shall later see how Dispositional Essentialists smoothly move from the one modality to the other.[30] Here, we wish to underline another striking difference between essences and dispositions. For this purpose, we need to consider the following dichotomies: synchronic or atemporal versus diachronic relations; cataloguing versus producing (see Mumford 2004: 168; Schrenk 2010: 727).

In Dispositional Essentialism, the Essentialist part is responsible for the cataloguing, the '*bookkeeping*': for each natural property or kind there are the respective necessarily associated dispositions – for example, for salt it is solubility, for electrons it is being negatively charged, etc. A table of all natural properties and kinds in one column and their dispositional powers in the next would comprise the worlds' inventory list that would tell us in which natural kind or property we would find which disposition (or, metaphorically speaking, in which toolbox we would find which instrument). There is no messing about: the inventory list is engraved in stone – in other words, what it says is metaphysically necessary. The essences are eternal or atemporal, for all places, in every world.[31]

If this passive bookkeeping is a synchronic or atemporal affair – bookkeepers are no men of action – how does change come about in the world? Clearly, someone else has to do the hard work of such diachronic affairs – the cutting, pushing, chopping, tearing. Dispositional Essentialism knows who fulfils this role – *the lumberjack's* role: dispositions! Their job is to provide the diachronic productive force so that salt dissolves at $t+\Delta t$ when put in water at t, and electrons attract and make positrons move when in their vicinity (see Schrenk 2005).

This dovetails perfectly with the observation we made earlier in the book (see Section 2.1.3): dispositions are not only associated with (MOD), the counterfactual modality, but also with (PROD), responsibility for production. If something has a disposition, then it, together with its dispositional property and the appropriate trigger, is *productively responsible* for the respective reaction to come about.

As we know, these two features, (MOD) and (PROD), are not independent and, historically, modality and production have been lumped together: 'efficacy, agency, power, force, energy, necessity, connexion, and productive quality, are all nearly synonymous' (Hume 1739–40: Book I, Part III, Sect. XIV: 157). But, why should this be so? Why should 'it is necessary that when c then e' equal 'c necessitates e' equal 'c causes e'

equal '*c* produces *e*'? (Metaphysical) necessity and production might well be two separate issues (see Section 1.2).

Although there have been attempts to separate (causal) production and (metaphysical) necessity (in fact we have encountered three such possibilities even if their proponents have not conceptualised their moves under these labels), their disentanglement still awaits a satisfactory success. This will become clear in the rest of this chapter where we have plenty of opportunity to see how the Essentialist's synchronic or atemporal necessity and the Dispositionalist's diachronic production sometimes harmonise and sometimes collide.

The three places where production and necessity were separated include Salmon's account of causation on the basis of transfer of energy or momentum, which unearths the connection Hume thought was secret, although it had no modal impact (see Section 5.4); Armstrong's nomological necessitation, which is much closer to production than to metaphysical necessity (see Section 4.4); and Cartwright's non-Essentialist Dispositionalism (see Section 6.2).[32] With Mumford and Anjum's account of causation we will introduce a fourth attempt (see Section 6.4.3).

The Plan of Battle for the rest of Section 6.4 Here is a short, elliptical way to defend Dispositional Essentialism: the idea that (natural) kinds and properties have dispositional essences is so serviceable that 'it would be foolish to try to get on without it' and, moreover, if a 'metaphysical hypothesis is serviceable then this is a reason to think it is true'.[33] Thus, Dispositional Essentialism is the correct metaphysics of science or, to say it in the words of Brian Ellis, one of its defenders:

> If your system strikes your readers as being simpler, more coherent, or more promising than any alternative for dealing with the recalcitrant difficulties of other systems, then this may be a good enough reason to buy it.
>
> (Ellis 2001: 262)

All this of course stands and falls with the theory *really being serviceable* and really being 'simpler, more coherent, or more promising'. We want the readers to be able to judge for themselves at the end of this chapter. In Chapter 7, we will discuss generally the validity of such kinds of serviceability arguments.

To provide them with the necessary material for judgement, we will look at how Dispositional Essentialists develop truth conditions for counterfactual conditionals (see Section 6.4.1), theories of laws of nature (see Section 6.4.2) and of causation (see Section 6.4.3) on the basis of their view. If they fare better than all the (Humean) competitors we discussed in the earlier chapters (see Chapters 3, 4 and 5), then, according to Ellis's argument, Dispositional Essentialism is the right metaphysics.[34]

We will also realise that different philosophers lean more towards essence and (MOD) and others more towards (PROD). In fact, we will see that Brian Ellis (on counterfactuals in Section 6.4.1) and Alexander Bird (on laws in Section 6.4.2) are more in the 'bookkeeping' business; Andreas Hüttemann (on causation in Section 6.4.3) softens that variant; and Stephen Mumford and Rani Lill Anjum, through and through 'lumberjacks', will overpower us with (PROD).

6.4.1 Dispositional Essentialism and counterfactuals

How about this: the fact that object *o* is fragile, i.e. that it possesses the disposition of fragility, *is* the truthmaker of the conditional claim '*o* would shatter if struck in circumstances *C*'. With Brian Ellis and his variant of Dispositional Essentialism we can generalise this idea so that we do not only have truth conditions for counterfactuals that are directly linked to a disposition but also to any counterfactual you like. Ellis writes:

> The proposition 'If A were the case, then B would be the case' will be true on such a theory [Dispositional Essentialism, MS] if and only if in any world of the same natural kind[s] as ours in which 'A' is true, in circumstances as near as possible to those that actually obtain, 'B' must also be true.
>
> (Ellis 2001: 278)[35]

Let us turn to some details: Ellis's answer is a derivative of standard possible worlds semantics for counterfactual conditionals (see Section 3.4). The core remains the same: look at a possible world most similar to ours where *A* is true. If, there, *B* is also true, the counterfactual is true; if not, it is false.

However, one aspect changes: Ellis spells out the crucial notion of *similarity of worlds* in terms of 'being of the same kind'. Worlds qualify as similar (enough) only if they are of the same kind, i.e. are inhabited by the same natural kinds and natural properties as the actual world. Thus, different to Lewis's semantics, in this case it is possessing the same natural kinds and not similarity of laws (let alone Humean conceived laws) that is most important for closeness of worlds.

There are difficulties for Ellis's view that result from this twist with Dispositional Essentialism's core. For a start, Ellis is faced with the problem of spelling out what exactly it means to keep things as near as possible to those that actually obtain except for the variation of *A* (see Ellis 2001: 279).

Remember, for *A* to obtain, but the world to be otherwise just like ours, Lewis had to introduce the idea of a small miracle, i.e. a violated law (see Section 3.4). However, although Lewis's metaphysics allows miracles (the world's parts are loose and separate and freely variable), Ellis's doesn't: things happen according to the essences of kinds, with necessity, in all possible worlds, no miracles allowed. Thus, for Ellis, to change the actual world such that *A* is the case might well mean to change a hell of a lot more in the past so that we'd hardly recognise that world anymore as similar to ours.

Worse is to come. Some of the counterfactuals scientists consider have antecedents that are counter-nomic or, here, counter-dispositional essences (for some examples including Millikan's oil drop experiment, see Jenkins and Nolan 2012). Consider, for example, a scientific experiment where scientists want to find out whether a natural kind has essence *A* or *A**. They might reason in the following way: if the right hypothesis were *A* then, in our experiment, we would measure *B*; however, if the right hypothesis were *A** we'd observe B*. Whatever they actually will measure, *B* or *B**, either *A* or *A** will turn out to be in conflict with the essence of the involved kind. Thus, one of the initially envisaged counterfactuals cannot be properly accounted for because, according to Dispositional Essentialism, there is no possible world where its antecedent can be made true and counterfactuals with necessarily false antecedents are, on standard accounts, trivially true. However, we feel that both counterfactuals are worth considering or even essential for scientific practice.

Ellis is aware of these difficulties and wrote: 'I am persuaded that this objection is indeed a serious one' (Ellis 2001: 279; Ellis attributes the argument to Bigelow). He aims to resolve the problem by first pointing out that there is a difference between epistemic possibility (what we can imagine) and real or metaphysical possibility (what nature's kinds and their dispositional profile allow to happen).

Next, Ellis proposes not to focus on the *truth conditions* for the problematic counterfactuals (where metaphysically possible worlds are the ones to be considered) but to turn to the *assertability conditions* of these counterfactuals (see Ellis 2001: 282; Ellis 1979). For assertability, imaginations (i.e. epistemic possibilities) are just fine and they might well take metaphysical impossibilities on board. Counterfactuals with counter-nomical counter-essential antecedents can therefore be assessed through assertability conditions.[36]

Other modal operators Besides counterfactual conditionals, there are other important modalities, most importantly *necessity* and *possibility*. Dispositionalism and Essentialism have the resources to account for both, if they make the 180-degree turn properly, in order that dispositionality and essence come first and are the primitive notions whereas possibility and necessity are derivatives thereof. Ellis (2001: section 8.6) gives a theory of necessity in terms of Dispositional Essentialism in his book, *Scientific Essentialism*. We have already considered the relation between necessity and essences at the end of Section 6.3.3 (Aristotelean natures) and do not want to go into detail here.[37]

Instead, we wish to take the opportunity to introduce the work of other Dispositionalists and focus not on how necessity but how *possibility* can be derived. Andrea Borghini and Neil Williams (2008), Jon Jacobs (2010) and Barbara Vetter (2015)[38] all rely on dispositionality in order to define this modality (they need not and do not consider Essentialism). Except for small differences, all four philosophers point to the central idea that something is possible iff it is the (actual or not) manifestation of a disposition:

> State of affairs *S* is possible iff there is some actual disposition *d*, the manifestation of which is (or includes) *S*.
>
> (Borghini and Williams 2008: 26)

> [S]ome proposition or truth T is possible just in case there is some actually instantiated property (or property complex) that is a power for some other property (or property complex) that would be a truthmaker for T.
>
> (Jacobs 2010: 236)

> It is possible that p just in case something has (or had, or will have) an iterated potentiality for it to be the case that p.
>
> (Vetter 2015: 246; where Vetter writes 'potentiality' we might, for our purposes, read 'dispositionality')

We wish to point out some difficulties for these theories (of which we have only introduced the rudimentary core). The first we take from David Yates (2014) and Jennifer Wang (2014).[39]

Yates writes:

> It is possible that none of the actual contingent existents exists, but there is no way to account for this possibility in terms of the powers of any subset of those very things.
>
> (Yates 2014: 412; attributing this difficulty to Cameron (2008))

In other words, there could have been nothing instead of something, so we might believe; however, none of the existing objects' dispositions can account for that possibility. Jennifer Wang subsumes the very same critique under the label 'global versus local' possibilities and adds:

> [W]hile Dispositionalism seems well-suited to account for possibilities that are local, diachronic, and gradable, it cannot account for certain possibilities that are global, synchronic, or absolute.
>
> (Wang 2014: 13)

Here is a further example that hints at what she is getting at – Dispositionality is a local matter (here and now, of one object), and yet certain negative and universal possibilities are global:

> Under normal conditions, a glass won't break when placed stably on a table – so it's possible that the glass doesn't break. Second, it's possible that JFK was not shot.
>
> (Wang 2014: 7)

Our short version of dispositional possibility makes the problem obvious: something is possible iff it is the manifestation of a disposition/potentiality. But of which disposition is *JFK was not shot* the manifestation? Of which is *the glass does not break*?

A clash of essential necessities and dispositional possibilities A further difficulty emerges only when Dispositionalism is combined with Essentialism. Suppose we do get a tenable definition of possibility through dispositionality. Now, because possibility and necessity are interdefinable – necessarily p iff it is not possible that not p – we get a definition for necessity for free. But haven't we already seen that essences also generate necessity? If water is essentially H_2O, it is necessary that any sample of water is also a sample of H_2O, and if essences can vouch for necessity then they can also for possibility: possibly p iff it is not necessary that not p.

Here is the pressing question: are those possibilities/necessities generated by essences and those by dispositions of the same kind? And if so, is their scope in harmony? Or are they different in kind and therefore not potentially competing. It seems that the answer could lie along the axis of the dichotomy *synchronic or atemporal* necessity versus *diachronic or dynamic* necessity introduced at the beginning of Section 6.4: first, the atemporal bookkeeping necessity is the one essences deliver; second, the lumberjack's (incorporating (PROD)) is the one dispositions bring with them. Actually, Wang's earlier quote indicates: 'Dispositionalism seems well-suited to account for possibilities that are [. . .] *diachronic* [. . .]' and yet dispositions 'cannot account for [. . .] possibilities that are [. . .] *synchronic*' (Wang 2014: 13). If essences cannot account for synchronic possibilities, then the earlier postulated division of labour – the bookkeeper versus the lumberjack – is ideally achieved by essences and dispositions.

All this is fairly speculative and we have entered territory for future research. We therefore leave these intriguing questions behind and quickly turn to what Dispositional Essentialism has to say about laws of nature. There, we will again encounter different necessities (and (MOD) and (PROD)).

6.4.2 Dispositional Essentialism and laws

Consider the following propositions: (i) what a law, *L*, says is bound (necessitated) to happen; (ii) the law *L* holds necessarily. The difference between (i) and (ii) can be easily seen when considering Armstrong's theory of lawhood (see Section 4.4). If the law, in this case *N*(*F*, *G*), holds then any occurrence of *F* necessitates the occurrence of *G*. *G* must happen when *F* happens. This is what (i) says. However, for Armstrong, *N*(*F*, *G*) itself holds only contingently, i.e. this law could not have held. This is the denial of (ii).

In Humean accounts of lawhood, the distinction between (i) and (ii) is somewhat blurred. In the Humean world everything happens only contingently – there aren't any necessitating/necessary connections (contra (i)) – and because laws are mere descriptions of what contingently happens, they themselves inherit this contingence (contra (ii)).

The theory of laws that Dispositional Essentialists put forward is stronger than Armstrong's view. In such a theory, it turns out that the laws could not have been different; they are not contingent but necessary[40] (affirmation of (ii)) and, moreover, what they say happens with absolute necessity (affirmation of (i)). This view is endorsed by Brian Ellis and Alexander Bird.[41]

The derivation of laws from essences and dispositions Bird, on whose ideas we will concentrate here, derives his theory of fundamental laws in his *Nature's Metaphysics* (2007).[42] Here's a reconstruction of the argument. First, there is the core Dispositional Essentialist thesis that properties and kinds, at least perfectly natural ones, have dispositional essences:

(Essence) $\qquad \Box(Px \rightarrow D(S, M)x)$

where $\Box$ stands for metaphysical necessity, *P* for a natural property, and *D(S, M)* for the disposition *D* to react with manifestation *M* when stimulus *S* occurs. Second, there is a premise about dispositions:

(Dispositions $\quad \Box(D(S, M)x \leftrightarrow Sx \mathbin{\Box\!\!\rightarrow} Mx)$

In other words, that *x* has disposition *D(S, M)* is equivalent to (i.e. can be identified with) the counterfactual claim that if *x* were in stimulus conditions *S*, then it would manifest *M*. Bird assumes this equivalence to be analytically true, hence the necessity operator in front.

From *(Essence)* and *(Dispositions)* we can derive:[43]

(I) $\qquad \Box(Px \rightarrow (Sx \mathbin{\Box\!\!\rightarrow} Mx))$

So far, this is only the stage setting for Bird's argument, with the core claims of Dispositional Essentialism at centre stage. Here come the details. Suppose, continues Bird, that some object x in some possible world w has the property P and that the stimulus S occurs, i.e. suppose that

(II) in w: $Px \wedge Sx$

Now, because of (I), $Px \rightarrow (Sx \mathbin{\Box\!\!\rightarrow} Mx)$ is also true in w and so, via *modus ponens*, with Px from (II) we get $(Sx \mathbin{\Box\!\!\rightarrow} Mx)$. Second, we take Sx from (II) to get, again via *modus ponens* (which also holds for this counterfactual conditional): Mx. We summarise our finding:

(III) in w: Mx

Summarising (I) – (III), we can say that we get Mx under the assumption $Px \wedge Sx$. So, we may write:

(IV) at w: $(Px \wedge Sx) \rightarrow Mx$

Also, because x was a freely chosen variable, we can generalise:

(V) at w: $\forall x((Px \wedge Sx) \rightarrow Mx)$

Moreover, because also world w was chosen arbitrarily, we can universally quantify over worlds, i.e. we can put a necessity operator in front of the formula:

(Conclusion) $\Box\forall x((Px \wedge Sx) \rightarrow Mx)$

The conclusion is a necessarily true, universally quantified formula, i.e. a formula that fulfils the desiderata for natural laws such as truth, objectivity, generality/universality, support of counterfactual conditionals and (relevant for anti-Humeans to which Dispositional Essentialist clearly belong to) necessity. Bird's conclusion, therefore, is that starting with *(Essence)* and *(Dispositions)* we can derive (the) laws of nature.

Evaluation of Bird's derivation We owe our readers a word on Bird's second premise, *(Dispositions)*. Three things seem strange: First, isn't the Dispositionalist precisely someone who does not believe in the counterfactual elimination of dispositions? Second, didn't Bird argue vehemently against a simple counterfactual analysis because of antidotes which, on top of finks, falsify the analysis (Bird 1998; Section 3.5)? There is a third internal friction which we will come to next.

A Dispositionalist can resolve the first apparent inconsistency by insisting that although the biconditional of *(Dispositions)* holds, dispositions are indeed irreducible real properties and that they are what makes the counterfactual true, not vice versa. We do not find an explicit remark to that effect in Bird; however, it is fairly clear that some such interpretation is right, especially when we consider his arguments against non-dispositional Quiddities (Bird 2007: 70–80) and the overall *Dispositionalist* credo of the book.

Bird argues against the second seeming contradiction in that he first underlines that his derivation of laws is meant only for the fundamental level (the laws of physics, say). Then he argues that, on that level, finks are impossible (Bird 2007: 60–2) and antidotes unlikely (Bird 2007: 62–3). Finks are impossible, according to a summarised version of his argument, because fundamental entities have no substructure which could be finked. Moreover, for a fink to operate there has to be a point in time in between the stimulus time and the manifestation time. Bird argues that on the fundamental level there isn't such a point in time. So, no fink can operate.[44]

Antidotes, i.e. interferers that do not take away the disposition but fiddle with its manifestation, might, however, be possible but Bird states the hope that in fact there aren't any:

> The direction of the development of physics with ever fewer fundamental properties and corresponding forces indicates that the prospects for antidote-free fundamental properties and thus strict laws only at the fundamental level are promising.
>
> (Bird 2007: 63; also see 24–42)

This hope – how the empirical sciences may develop – is questionable.[45] We will return to this issue after we have introduced the third aspect that makes Bird's use of a counterfactual doubtful.

Bird states elsewhere in *Nature's Metaphysics* that 'necessarily if the potency is instantiated and receives its stimulus, then the manifestation will occur' (Bird 2007: 64) and we find statements to the same effect in Ellis (2001: 286): 'Therefore, [. . .] for all x, necessarily, if x has p, and x is in circumstances of the kind C, then x will display an effect of the kind E'. However, this suggests that it is not the *counterfactual conditional* he uses above – $(Px \rightarrow (Sx \mathbin{\Box\!\!\rightarrow} Mx))$ – but a *necessary material conditional* – $\Box\, ((Px \wedge Sx) \rightarrow Mx)$ – that binds the stimulus to a manifestation. And, think about it, the *necessary material conditional* actually fits better to Bird's argument: if there are neither antidotes nor finks on the fundamental level, then in his derivation of laws of nature, he would be justified to use the stronger *necessary material conditional*.

No work for lumberjacks Suppose if this were right, then there would be a striking consequence: metaphysical necessity (of essences) would also be the modality (MOD) in dispositions and, presumably, necessity would also account for dispositions' productive power (PROD). This idea seems to be guided by a parsimony principle: there is but one necessitating force, namely necessity, and no other necessitation in addition to it.[46] It would also be in line with Hume's equation of necessity with productive power, only that Hume denied their existence while Bird and Ellis affirmed them.

If Bird and Ellis are right, then there is no separate modality in dispositions. Essentialism's metaphysical necessity does all the work. Thus, the question we asked at the beginning of Section 6.4, namely whether dispositions create a different modality than essences, is answered negatively. There would not be a distinction in kind between diachronic, productive relations (remember the lumberjack) and atemporal correlations (the bookkeeper).

However, whether this is a viable move has been contested by some philosophers see, for example, Eagle (2009), Schrenk (2010). Their argument is twofold. First, they doubt, with due respect to Bird, that there can be no antidotes on the fundamental level. Next, they point out that Bird's early argument against the counterfactual analysis on the basis of antidotes (see Section 3.5) also works against Bird's and Ellis's later assumed necessary conditional because the latter entails the respective counterfactual conditional: $\forall x \, \Box \, (Fx \rightarrow Gx) \models \Box \, \forall x \, (Fx \, \Box\!\!\rightarrow Gx)$. If the counterfactual analysis fails, the analysis with a necessary material conditional fails a *forteriori*.

If Eagle's and Schrenk's arguments are correct, then metaphysical necessity can, after all, not be the (MOD) in dispositions and, consequently, something else would have to vouch for dispositions' (PROD).[47] In some sense, Bird (and Ellis) fell for Hume's line (in that they bought his equation of production with necessity) where they could (maybe should) have separated the two and just have defended one of them.

Some Dispositionalists, notably Stephen Mumford and Rani Lill Anjum, capitalise on this finding and even see an advantage in the dispositions' lack of necessity (see Mumford and Anjum 2011a, 2011b). We'll come to their approach later when we discuss accounts of causation in terms of Dispositionalism.

Necessity *tout court* Never mind the three doubts we have just mentioned regarding Bird's derivation of laws from Dispositional Essentialism. If his theory can be defended, it brings with it both that laws necessitate what happens *and* that laws themselves are necessary. These were items (i) and (ii) with which we started Section 6.4.2. Now that the laws of nature could not have been different (ii), turns out to be a hard bullet to bite. Not only do philosophers of science engage in counterfactual reasoning like 'if the laws of nature had been different then such and such could be true', scientists also occasionally hypothesise that the law of gravitation, say, could evolve over time, namely in that the gravitational constant G^* changes in magnitude throughout the history of the universe. Thus, against Bird, it seems, the laws could have been different. Also, compared to Section 6.4.1, if laws are metaphysically necessary then all counterfactuals with counter-nomic antecedents come out as trivially true.

Bird saw the need to dedicate a whole chapter of his book (Bird 2007: 169–88) to why the Dispositional Essentialist can nonetheless stick to the necessity of laws. We cannot go into the details; however, briefly Bird distinguishes, just as Ellis (see Sections 6.4.1), between epistemic readings of counter-nomic facts and metaphysical ones. It is the latter that are strictly impossible; however, *for all we know* (this is the epistemic reading) the laws could have been different.[48]

We wish to offer a possibility for quasi-counter-nomic facts that remains open to Dispositional Essentialism, even on metaphysical grounds: nothing within Dispositional Essentialism forbids that there are other possible worlds that have different kinds and properties with different essential dispositional profiles: '*schmellectrons*' with triple positive charge, for example. In such a world, the respective laws are therefore also different, and so although electrons could not possibly be differently charged, there could be *schmellectrons,* which are very much alike electrons in all other respects. The kind of Coulomb's law that these schmellectrons' dispositional essence generates is, therefore, also slightly different (it's 'Schmoulomb's law'). If all this is possible, then one way of saying that Coulomb's law could be different is to say that there could be other worlds that are not so very different to ours except that there are *schmellectrons* which are very much alike electrons but yet they obey *schmoulomb*'s law.[49]

Have all laws been captured? There are, next to the trouble with necessity, other issues with Bird's derivation of laws of nature from Dispositional Essentialism. We might wonder, for example, whether all candidates that scientists believe to be laws are captured, but here is a problem: science postulates the so called conservation laws of energy, momentum, charge, etc. They hold, roughly, that the total energy, momentum, charge, etc. of all closed systems or a certain physical/chemical reaction remains the same. It is hard to see how the conservation laws for systems can be deduced from the fundamental properties of their parts and their dispositional powers (see Schurz 2011: 140).

Remember the difficulties we encountered when Dispositionalists tried to define possibility in terms of dispositions (see Section 6.4.1)? In that case, one of the problems was that it was hard for dispositions to account for possibilities that are global. It is similar here: conservation laws seem to be too global a matter to be captured by local dispositions of individual particles.[50]

We have seen some pros and cons of a derivation of the fundamental laws of nature from dispositional essences of fundamental properties.

We have to leave these issues behind now because we must turn to theories of causation that are based on dispositions.[51]

6.4.3 Dispositional Essentialism and causation

The following could be the start of a rudimentary dispositional theory of causation: *C causes E*, *iff* some things within the *C* event have the disposition to manifest with event *E* and the *C* event includes the trigger for those dispositions. For example, Suzy's stone broke the bottle, i.e. caused the bottle to break because a stone has the disposition to break fragile objects like that bottle (the manifestation) if thrown with a certain momentum (the triggered stimulus).

Dispositional theories of causation are more fine grained than this.[52] We will discuss Hüttemann's and parts of Mumford's and Anjum's theory in this section. Our goal is twofold: not only do we wish to disclose their theories of causation but we also aim to locate their points of view on the axes between the Essentialist's necessity and the dispositional (MOD) and (PROD).

The interference theory of causation Possible disturbances of a disposition's manifestation or a causal process have repeatedly been a stumbling block for analyses of dispositions and analyses of causation: finks and antidotes (see Sections 2.1.3 and 3.5), pre-emptions and preventions (see Chapter 5), to name but a few. We have also seen how such disturbances make it hard to believe that cause and effect or stimulus and manifestation are necessarily connected (see Section 6.4.2).

Hüttemann's strategy to meet the disturbances challenge to causation is twofold.[53] First, he issues a kind of Copernican revolution: causation has to be understood as interference into processes which, were they not so disturbed, would not change. In other words, interferences are not a nuisance to causation – *they are causation*. Second, other than Bird, he takes on board some of the critique against necessity in dispositions that came from Eagle and Schrenk and weakens necessity to *conditional necessity*.

The basis of Hüttemann's thought is that objects or complex systems *are disposed to display* a certain default behaviour (here lies the core for why his is a dispositional theory). As a paradigm example, Hüttemann quotes Newton's first law:

> Every body continues in its state of rest or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed upon it.
> (Hüttemann 2013a: 108)

According to Hüttemann, many law-statements should be interpreted that way, i.e. as statements about dispositions to display some default behaviour or process unless something intervenes. There are examples from other sciences: ‘that economic systems in which inflation rises yield higher unemployment rates (if nothing interferes)’ (Hüttemann 2013a: 119) or that the population of predators follows, if undisturbed, that of prey (see Hüttemann 2013b: 182).

Such ‘quasi-inertial laws’ describe what happens if systems are uninterrupted, i.e. if nothing intervenes (see Hüttemann 2013a: 108). In other words, these laws ascribe a special kind of disposition to objects or compound systems. These dispositions are special in the following sense: first, their manifestations are temporally extended processes or behaviours and, second:

> the triggering conditions for the dispositions appealed to are purely negative. The quasi-inertial behaviour becomes manifest in the absence of disturbing factors (antidotes).
>
> (Hüttemann 2013a: 109)

Thus, formulated explicitly in dispositional language, Newton’s first law becomes: every mass has the disposition to continue in its state of rest or of uniform motion in a straight line (= manifestation) if nothing interferes, i.e. no forces are impressed upon it (= trigger).

Complex systems When we are concerned with a complex system’s inertial disposition to behave in some default way (if undisturbed), then the overall disposition of that system as a whole is, according to Hüttemann, the resultant of the joint contributions of the compound’s subsystems’ dispositions. That is, the dispositions of the parts of a complex system ‘*contribute* to the behaviour of the compound’ (Hüttemann 2013a: 109). Hüttemann takes for granted that such a composition of subdispositions of subsystems to create the (inertial) disposition of the whole is law governed:

> the presence of a subsystem’s disposition makes a difference to the behaviour of the compound and [. . .] the difference it makes depends on the law of composition.
>
> (Hüttemann 2013a: 109)

Let us accept this idea of composition as unproblematic (but see later for a critical remark on adding laws to the ontology).

What, therefore, is a cause and what is an effect? Hüttemann quotes Mach and also Hart and Honoré:

> In general we only feel the need to ask for a cause, if a (unexpected) change has occurred.
>
> (Mach 1896: 432, my translation)

> The notion, that a cause is essentially something which interferes with or intervenes in the course of events which would normally take place, is central to the common-sense concept of cause.
>
> (Hart and Honoré 1985: 27)

Hüttemann implements this 'common-sense' idea into his theory of dispositions that exhibit default behaviour: *E*, an effect, is the deviant behaviour diverging from the disposition's default manifestation process (what normally takes place) due to an (unexpected) interfering or intervening factor, *C*, the cause. Thus, according to Hüttemann:

> A cause is an actual disturbing factor (antidote) to the default behaviour that a system is disposed to display.
>
> (Hüttemann 2013a: 113)

In other words,

> if the default behaviour does not occur, the law tells us that there must be some factor that interacted with B. The event of this factor interacting with the default process is the cause of the later non-occurrence of the default behaviour.
>
> (Hüttemann 2013a: 113)

Thus, interferences are causes and causes are interferences. Here is an example: billiard ball *B* is at rest. According to Newton's first law, its inertial default behaviour is to stay in this state. It manifests its inertial disposition. However, billiard ball *A* bumps into it and therefore interferes with *B*'s inertial behaviour. *B* is deflected and starts rolling. It's rolling is the effect of *A*'s interference, the cause. Note aside that *A*'s default behaviour to continue rolling is also disturbed by *B*: the collision with *B* is also a cause for *A*'s change of direction.

In support of Hüttemann's theory, in addition to its prima facie plausibility, we can point out that it can handle pre-emption, double prevention and negative causation cases well (see Chapter 5). We pick out pre-emption for illustration: regardless of whether there is a potential backup cause (a further disturbing factor) that would have interfered had the present one not disturbed the inertial process (in the earlier example, another slower billiard ball *C* approaching *A* as well), the actual interferer (*A*) qualifies as the cause (see Hüttemann 2013a: 115).

Conditional necessity Is Hüttemann's theory more on the side of productive powers or necessities? Other than Bird, Hüttemann focuses on the dispositional aspect and not as much on the Essentialist part. Still, he sees the dispositions' productive force (PROD) fuelled by necessity or, more accurately, by a 'conditional metaphysical necessity'.[54] In fact, there seems to be a parsimony principle at work again: Hüttemann aims to give a unified account of dispositional, nomological and causal modalities (see Hüttemann 2013a: 122). (Conditional) metaphysical necessity is regarded as the binding modal glue in all of these cases.

Let us reconstruct Hüttemann's idea. He proposes that the laws of the compositions of dispositions (remember subsystems and their subdispositions) are a metaphysically necessary affair: it is the nature of such systems to possess their dispositions to behave in their specific default ways should nothing interfere. Second, he believes that if something interferes, then the according behaviour is also guided by (conditional) necessity: if nothing else (other than the first interference) disturbs the system, then the system's deviant behaviour unfolds with necessity:

> Now assume that it is part of the essence of systems that they behave according to the laws of composition. Then the following holds: It is necessarily the case that, *if nothing interferes,* the compound system manifests its behaviour. It is furthermore necessarily the case that, if something interferes, the system displays a different behaviour of this or that kind. In other words, if we assume that the laws of composition hold with metaphysical necessity, then the manifestation of the dispositions of the systems happens with conditional metaphysical necessity.
>
> (Hüttemann 2013a: 122, emphasis added)

Hüttemann's theory is new and still awaits critical assessment. One critique would be to point out that his ontology is less parsimonious than his fellow Dispositionalists'. He needs laws as well as dispositions and dispositions as well as categorical properties (this results from his ideas regarding compositions of parts to wholes where structural features of parts and wholes are categorical).

Second, the phrase which is so crucial for his theory of a conditional necessity – *if nothing interferes* – has been a stumbling block for the past century (remember, most strikingly, Martin's finks (see Section 3.5)). Hence, it is questionable whether this proviso which is at the heart of Hüttemann's conditional necessity is as innocent as it looks.

Finally, Hüttemann's necessity, even if only conditional, is still a bookkeeper's business. In fact, this conditional necessity *just is* metaphysical necessity, only the circumstances under which it is supposed to hold differ from those Bird had in mind. Thus, as with Bird's theory, dispositions' modality (MOD) and their productive responsibility (PROD) made way for (conditional) metaphysical necessity. In this respect, Hüttemann's account of causation is akin to Bird's account of laws. Not so Mumford's and Anjum's, to which we come next.

The vector account of dispositions and causation The core idea of Mumford's and Anjum's non-reductive theory of causation (see Sections 5.6 and 5.7) is briefly introduced as:

> Causation occurs when powers exercise themselves. (6)
> A cause should be understood as something that disposes towards an effect. (19)
> Effects are brought about by powers manifesting themselves. (7)[55]

However, rather than focusing on their overall theory, developed in much more detail in *Getting Causes from Powers* (Mumford and Anjum 2011a), our focus will be on their stance regarding metaphysical necessity, (MOD) and (PROD) because it is here that their theory refreshingly differs most crucially from the other Dispositionalists' accounts.[56] In fact, maybe together with Cartwright (see Section 6.2.2), Mumford and Anjum are the only pure Dispositionalists amongst those we have discussed, in the sense that they do not resort to Essentialist metaphysical necessity in the way that Ellis and Bird do.

Rather, they wish to postulate a '*sui generis* modality' (MOD) for dispositions[57] which shall also and centrally function as a productive power (PROD): 'disposing towards an effect' is meant in a *bringing about* sense.

This desired productive aspect (PROD) is most prominently visible when Mumford and Anjum bring their vector metaphor into play (Mumford and Anjum 2011a: 70–4). We are asked to think of an operating disposition (i.e. a causal factor) as if it was something like a (Newtonian) force which pushes an object into a certain direction: similar to forces, causes push towards certain effects.[58]

However, just as an object might not move (or deform) even though a force operates on it (namely because other forces are active as well and pull in opposite directions), so might dispositions/causal factors be operative and yet be only partially successful (or not at all) when other interfering powers are present as well.

The overall causation of an actually occurring effect *E* is, thus, a matter of dispositional powers working in concert: similar to the vector-addition of a multitude of Newtonian forces, so do individual causal factors (which are seen as operating dispositional powers) combine, enforce or hinder each other to result in the overall cause *C* (akin to the resultant force of force vector addition) which brings about the overall outcome *E*.

Mumford's and Anjum's thoughts are very similar to what Mill had to say on the presence and combination of a multitude of causal factors:

> The [total] cause, then, philosophically speaking, is the sum total of the conditions positive and negative taken together; the whole of the contingencies of every description, which being realized, the consequence *invariably follows*.
>
> (Mill 1843: 217, emphasis added, my addition in brackets)

However, there is also a difference. Unlike Mill (and Bird, Ellis and Hüttemann for that matter), Mumford and Anjum deny that the actual consequence, E, 'follows invariably', i.e. '*necessarily*', even when C is the (actual) *sum total of the conditions positive and negative taken together* (see Mumford and Anjum 2011a: 53–70).

The powerlessness of necessity Their argument for this conclusion is that there could always be a further additional factor, A, to that initial sum total, C, that could disturb the outcome:

> No matter how big C becomes, it still cannot exclude the possibility of prevention. There remains some possible A that could prevent E even if all that is included within C occurs. […] All that is needed is some possible interferer, and no finite list – even if it includes negative factors – can exclude every possible interferer. No matter how big an antecedent is, it can always be strengthened.
>
> (Mumford and Anjum 2011a: 66–7; we exchanged their variables for our C, A and E)

In possible worlds talk, this reads: where, in the actual world, C (the actual sum total of the conditions positive and negative taken together) *does* indeed produce E, there is still a possible world where both $C + A$ come about (a new sum total consisting of the old plus some more) and E^*, not E, is produced.[59]

Next, they borrow an argument from (Schrenk 2010: 731–2): necessity is monotonic, i.e. it holds:

> Monotonicity of necessity:
>
> If $\Box(C \rightarrow E)$ then, come what may (an additional A, for example), still: $\Box(C \wedge A \rightarrow E)$.

Now, because there is a world where $C \wedge A$ and $\neg E$ (according to their previous argument), we can write $\Diamond\neg(C \wedge A \rightarrow E)$. This is equivalent to: $\neg\Box\ (C \wedge A \rightarrow E)$. Thus, by *modus tollens* applied to monotonicity, we get $\neg\Box\ (C \rightarrow E)$.

Therefore, with Schrenk, Mumford and Anjum conclude that Mill's '*invariably follows*' cannot be interpreted as: *E* follows *with necessity* from *C*. To put it the other way round, necessity is powerless: it can't be what puts productive, causal force into dispositions.

However, the intention behind Mumford's and Anjum's vector account of dispositional causation is that there has to be something that is binding causes and effects together instead of being Humean loose and separate events. They therefore postulate that dispositions have a modality in their own right – the earlier mentioned *sui generis* dispositional modality. They develop this idea further in *Getting Causes from Powers* (Mumford and Anjum 2011a: 175–94).

This *sui generis* modality (MOD) which is not necessity is also somehow responsible for the productive power (PROD) Mumford and Anjum crucially associate through the forces metaphor with dispositions. At the beginning of Section 6.4 we were tempted to say that their approach is very much on the lumberjack's side of the spectrum and less on the bookkeeper's. This was because (MOD) is degraded to less than necessity (their *sui generis* modality) and (PROD) is so centrally important within the forces metaphor.

Revisiting the anti-necessity stance However, when we look closer, we find that there is *a sense of necessity* that Mumford and Anjum acknowledge (see Mumford and Anjum 2011a: 76): for any two worlds w_1 and w_2, if w_1 and w_2 have exactly the same total distribution of powers and interferers in the past, *C*, then these worlds show exactly the same resultant manifestation, *E*, in their future.

Since this claim starts with a quantification over all possible worlds ('for any two worlds…') we felt compelled to speak of *a sense of necessity*. Thus, let us define necessity* in the following way:

> Necessarily*, when the sum total of all powers and interferers in the past are taken together, *C*, then the effect, *E*, comes about in the future = *def.* For any two worlds w_1 and w_2, if w_1 and w_2 have exactly the same total distribution of powers and interferers in the past, *C*, then they show exactly the same resultant manifestation, *E*, in the future.

In this sense, the sum total of all powers and interferers *does*, after all, *necessitate** the outcome.[60]

In fact, their vector theory seems to presuppose this kind of necessity because their idea of vector summation relies on the fact that the same vector additions have always the same vector sums. (All this holds at least for deterministic powers, and, at this point, Mumford and Anjum do not assume indeterminism (2011a: 74–6)).

Manifestation partners and simultaneity The dispositional theory of causation that Mumford and Anjum develop in their *Getting Causes from Powers* (2011a) is much richer than the few aspects – causation occurs when powers exercise themselves in concert, powers are vector-like, there is an anti-necessitarian *sui generis* modality – we have outlined here. They also argue, for example, that dispositions need manifestation partners – sugar is disposed to dissolve when in water and water is disposed to dissolve sugar (Mumford and Anjum 2011a: 34–8) – and that causes do not precede their effects, but rather that causes and effects happen simultaneously (Mumford and Anjum 2011a: 106–29). This latter thought is also incorporated into a process theory.

Of course, their views have found their critics. Luke Fenton-Glynn, for example, criticises the simultaneity view on the basis that it is inconsistent with current physical theory (Glynn 2012: 1102) and the process part of the theory is criticised by Anjan Chakravartty (2013: 898). Mumford's and Anjum's claim that causes can be fruitfully represented as vectors also came under fire. One could argue that, generally, too much is left at a metaphorical or gestural level. More specific arguments from Glynn state that the vectorial representation of powers is misleading because, unlike with forces, there is no metric to measure causal powers (Glynn 2012: 1103), that although 'forces compose in a linear, additive fashion, the linear, additive composition of causal factors seems to be the exception rather than the rule' (Glynn 2012: 1103) and also that when Mumford and Anjum do finally endorse a principle of compositional pluralism (Mumford and Anjum 2011a: 96–101) they thereby concede that 'causal powers often are not very much like classical forces after all' (Glynn 2012: 1105). Mumford and Anjum have also received critique for their anti-necessitarian stance and their belief in a *sui generis* modality (see Smart and Talibard (2012), Cross (2012, 2014) and Bird (2016)).

Dispositions ground necessities of diachronic affairs Here is a somewhat speculative suggestion differing from and exceeding Mumford's and Anjum's work. For the suggestion's impact, we must underline an aspect we have, for simplicity, neglected a little: time! Remember that a dispositional power's job is a diachronic one: we are concerned with an antecedent stimulus/cause at t and a consequent manifestation/effect at $t + \Delta t$. This should, in fact, also hold for necessity*, where we were concerned not with single dispositions but total distribution of powers and interferers in concert. Necessity* also links, diachronically, the past to the future (compare this to (Maudlin 2007: 174) in Section 4.6).

Remember something else: at the beginning of Section 6.4, we underlined a difference between dispositions and essences. Essentialist claims are atemporal affairs: if the essence of electrons is to be negatively charged, then this is eternally or atemporarily so. There, we also had Kit Fine say about the relation between (atemporal) essence and (metaphysical) necessity:

> It seems […] that far from viewing essence as a special case of metaphysical necessity, we should view metaphysical necessity as a case of essence.
> (Fine 1994: 9; also 2002)

Inspired by this quote, we might want to start thinking similarly about (diachronic) dispositions and (diachronic) necessity*: it is the latter that sum totals of the former achieve, i.e.

> it seems that far from viewing dispositionality as a special case of (diachronic) necessity*, we should view the (diachronic) necessity* of $(C\ (t) \rightarrow E\ (t+\Delta t))$, for sum totals C to E, as a case of the individual dispositions within C working in concert.

It's powers that drive (with necessity*) the goings on in the world. It's essences that maintain (with necessity) the order. Taking our old acquaintances (PROD) and (MOD) on board, we could add: the concert (i.e. sum total) of the individual dispositions' productive powers (PROD) add up to (MOD), here necessity*.[61]

This gesture towards a novel account brings the section on dispositions and causation to an end. After a brief summary, we will add some polemical remarks in Section 6.5.

BOX 6.4 Dispositional theories of counterfactuals, laws and causation

- In Section 6.4 we first **wedded Essentialism to Dispositionalism,** which we individually introduced in Sections 6.3 and 6.2, respectively. In order to get acquainted with the couple – **Dispositional Essentialism** – we suggested learning what its features are, i.e. to see how Dispositional Essentialists can **give accounts of counterfactuals, laws and causation**. This way, we turn the book's goal in Chapters 2–5 (and most of metaphysics of science's history) on its head, namely to analyse dispositions in terms of these other categories.
- Dispositional accounts of **counterfactuals (and other modalities)** have to reverse the direction of counterfactual analyses of dispositions. Instead of interpreting the attribution of fragility to some object *o* as the assertion that the counterfactual

conditional '*o* would shatter if struck in circumstances *C*' is true, **Dispositionalists say that the fact that *o* is fragile makes the counterfactual true**. This core theory can be extended to counterfactuals that are not closely tied to a single object in that a **multitude of dispositions of more complex situation and their potential joint manifestations** can be taken into consideration.

- **Possibility and necessity. Possibility** might be analysed in dispositional terms in the following way: **it is possible that *p* iff something has (or some things have) the disposition to *p***. Since **necessity is definable in terms of possibility,** the modality of **necessity** would also be captured. However, because **essences could ground necessity** too (and thereby possibility, see Section 6.3) there could also be friction between the necessity/possibility emerging from essence to that emerging from dispositionality.
- **Laws.** Laws of nature can be **derived from the core claims of Dispositional Essentialism, namely from (Essence) $\Box$ (*Px* $\rightarrow$ *D*(*S*, *M*)*x*) and from (Dispositions) $\Box$ (*D*(*S*, *M*)*x* $\leftrightarrow$ (*Sx* $\Box\rightarrow$ *Mx*))**. Alhtough the laws that result from the derivation, $\Box\forall x((Px \wedge Sx) \rightarrow Mx)$, fulfil the anti-Humeans' desideratum that **laws govern** what happens in the world, the **laws themselves also come out as metaphysically necessary**. This might be seen, even by anti-Humeans, as being over the top.
- **Causation.** We introduced two different dispositional accounts of causation. Hüttemann's says that, '**A cause is an actual disturbing factor (antidote) to the default behaviour that a system is disposed to display**' (Hüttemann 2013a: 113). Mumford and Anjum claim that **a cause is 'something that disposes towards an effect**' (Mumford and Anjum 2011a: 19).

6.5 Back to monads?

As we have repeatedly said, everything comes full circle in this chapter. Remember Hume's onslaught on causation and necessary connections as conceived by the *Rationalists* and how the *Logical Empiricists* and later neo-Humeans continued to defend Hume's credo. The unobservability of modalities like necessity was one of the original reasons to dismiss them: 'External objects as they appear to our senses, give us no idea of power or necessary connection' (Hume 1748: Sect. IV, Part I: 63–4).

In the chapters between our first encounter with Hume (see Section 1.2) and this present chapter, we have become witnesses to the arguments that led many people to turn their back on Hume. For example, Dispositionalists, Essentialists and Dispositional Essentialists of our century are no longer convinced of his arguments. As we saw earlier, some of the strongest anti-Humean claims were issued by Alexander Bird and Brian Ellis: 'necessarily if the potency is instantiated and receives its stimulus, then the manifestation will occur' (Bird 2007: 64).

This might remind some readers of statements like Hobbes's whom we quoted as an example for the *Rationalists*: 'All the effects that have been, or shall be produced, have their necessity in things antecedent' (Hobbes 1655: 9.5, emphasis added). Or compare Mumford's and Anjum's powers to Leibniz's monads: 'the internal principle of a substance that brings about change' and which 'can be called appetition' (*Monadology* §15).

Remember how Russell first reacted to Leibniz: 'I felt – as many others have felt – that the *Monadology* was a kind of fantastic fairy tale [. . .] wholly arbitrary' (Russell 1900/1992: xxi). Would he also have written this about Dispositional Essentialism? And would he also have continued with enthusiasm that, on further inspection, 'suddenly a flood of light was thrown on all the inmost recesses of [this] philosophical edifice. I saw how its foundations were laid, and how its superstructure rose out of them'?

We leave it to the readers to judge for themselves; however, we offer some tools to assess modern metaphysics in our next chapter on meta-metaphysics.

Notes

1 With due respect to Quine (1953), to be is not only to be the value of a bound variable, to be is also, according to Cartwright (1992), to be presupposed by experimental practice.
2 Including Lewis's, but see also Braddon-Mitchell (2001) and Schrenk (2014).
3 See the articles by Earman *et al.* (2002) and Reutlinger and Unterhuber (2014) in the two special issues of *Erkenntnis*; and also the Stanford Encyclopedia entry (Reutlinger *et al.* 2011).
4 For why things are not so simple, see Schrenk (2007).
5 See *The Dappled World,* to use another title of Cartwright's books (1999). However, the most concise formulation of her transportability argument can be found in Cartwright (1992: 49ff).
6 Here, Cartwright relies heavily on John Stuart Mill's philosophy and she devotes section 5.5 of her book to Mill (Cartwright 1989: 170–9).
7 The transportability argument from scientific practice has been further developed in Hüttemann (1998, 2007) and Bartels (2000).
8 Some philosophers – Brian Ellis and Alexander Bird who are discussed later, but not Cartwright (see Cartwright 1992: 48) – say at this point that some dispositions belong to the essence of some fundamental entities – for example, that electrons are necessarily negatively charged. As a consequence, at least for those fundamental kinds of entities, their capacities cannot be lost in any circumstance and, thus, the *cp* issue is, for them, irrelevant. We cover Essentialism and especially Dispositional Essentialism in Sections 6.3 and 6.4.
9 For a reply to this critique, see Cartwright (1992: 56ff).
10 See, for example, Rom Harré and Edward. H. Madden (1975), Hugh Mellor (1974, 2000), George Molnar (1999; 2003: section 1.3 and chapter 9), Alexander Bird (2007: section 2.2), Brian Ellis (2001: chapter 3) and Stephen Mumford (1998: chapter 4). It is probably Stephen Mumford's (1998) book *Dispositions* that sparked metaphysics of science's renewed interest in Dispositional

Realism and contributed most to its rise since the late nineties. It is recommended as a starting point if you are interested in more details on dispositions.

11 For the sake of the argument, we ignore possible mutations or accidents. This example is Quine's (1953: 21).

12 There are residing troubles. Although the meaning difference of 'renate' and 'cordate' can be captured with the notion of intension, it cannot be grasped for predicates like 'triangular' and 'trilateral': every triangle is, with mathematical necessity, both 'triangular' and 'trilateral'. That is, there is no possible world where some triangles are one but not the other. Thus, even the intensions of these two predicates coincide and the meaning difference between the two has to be sought elsewhere. In these cases, meanings transcend intensions and philosophers of language speak of this phenomenon as hyper-intensional. By the way, not only in light of the possibility of hyper-intensionality, the original *intensions* should have been called '*hyper-extensions*' (because that is what they are: sets of extensions of possible worlds) and the new hyper-intensions merely 'intensions'. Unfortunately, my term 'hyper-extensions' has never been established. If you want to learn more about extensions, intensions and other things from the philosophy of language, a good starting point is David Braun's entry in the *Routledge Companion to Philosophy of Language* (Russell and Graff Fara 2012: 9–17).

13 The relation between intensions as abstract entities and the criteria in people's heads are slightly more complicated than here depicted. Carnap's way of speaking is here that we *grasp* ('*erfassen*') these abstract entities when we engage in the relevant thought processes (see Putnam 1975: 138–9; Carnap borrows this term from Frege). This is unimportant for our purposes.

14 We should not omit that Kuhn and Feyerabend had other ways to argue for the incommensurability of sciences, like the sociological, historical fact that there are often abrupt paradigm shifts within scientific enterprise which are not reason and argument guided but which are rather ad hoc or dependent on who has the power and rhetoric in a lab or within a funding body to direct the path of research.

15 Hand-in-hand with this change of mind comes the move from a pure observational language to a physical language (see Section 2.2.1). 'If someone decides to accept the thing language, there is no objection against saying that he has accepted the world of things [and its properties, MS]' (Carnap 1956: 207–8).

16 In what follows on natural kind terms (rather than names) we focus on Putnam's arguments. However, as mentioned earlier, Kripke put forward very similar ideas (in Lecture III of his *Naming and Necessity* (1980) so that we will often speak of Putnam's and Kripke's ideas even if we quote only from Putnam.

17 Carnap made clear that *realism*, as he understands it, is merely a *serviceable* way of speaking, and yet, the service that this realist language provides is by no means an argument for a through and through Metaphysical Realism: 'The efficiency, fruitfulness, and simplicity of the use of the thing [and properties] language may be amongst the decisive factors [for its choice]. [. . .] However, it would be wrong to describe the situation by saying: "The fact of the efficiency of the thing [and properties] language is confirming evidence for the reality of the thing [and properties] world"; we should rather say instead: "This fact makes it advisable to accept the thing [and properties] language"' (Carnap 1956: 208, my additions in brackets).

Note something striking here: later metaphysicians such as David Lewis or David Armstrong consider such serviceability arguments completely differently. For them, these arguments *do* amount to a justification of the belief in the reality of natural properties. Contrast Carnap's quote to what Lewis said regarding the assumption of the existence of natural properties: it is 'so commonsensical and so serviceable – indeed, was so often indispensable – that it was foolish to try to get on without it' (Lewis 1999: 1–2; Section 4.3). It is interesting how philosophical intuitions have changed. Gillian Russell has put today's metaphysicians' and semanticists' attitude nicely: 'Sometimes it is not a refutation of a skeptical view that tempts philosophers away from it, but rather the anticipated fruitfulness of approaches that begin from less skeptical assumptions' (Russell and Graff Fara 2012: 9–17). In Section 7.4, we will assess such serviceability arguments in general.

18 They give this answer after having rejected two alternatives, namely that natural kinds are particulars or that natural kinds are *sui generis* entities, neither universals nor particulars. This latter view is defended by E. J. Lowe and, independently, by Brian Ellis. Ellis postulates not only natural kinds of objects but also properties, processes and events (Lowe 1998, 2006; Ellis 2001). Hawley and Bird's

view is akin to Armstrong's, despite differences they point out in their paper: 'Kinds mark true joints in nature. But it is not clear that we require an independent and irreducible category of universal to accommodate the kinds' (Armstrong 1997: 67–8).

19 Actually, the same holds for *proper names*, not only for *natural kind terms*. In fact, both Kripke and Putnam believe that natural kind terms ('gold', 'water', 'electron' etc.) behave just like names. Indeed, natural kind terms are, according to Putnam and Kripke, names for natural kinds (or samples thereof).

20 This is the first of Carnap's theories we have introduced. Let us skip the later-mentioned predicates–properties idea he had.

21 Even if we believe the core of this idea, it remains a contested issue about what kind terms really refer to: is it the extension, comprising all individual samples, or the natural kind, conceived of as a universal? We leave this question to the philosophers of language.

22 We should probably add, at this point, that Putnam changed his mind frequently on these matters. In his book *Reason, Truth and History* (1981) he sympathises again with Internalism.

23 Remember in Chapter 1 that we, Hume and almost everyone else before Kripke and Putnam equated the conceivable with the possible. It turns out now that, at least superficially, we can conceive something that turns out to be impossible: here, water being *XYZ*. Only if we add 'Well, you did not *really imagine* water to be *XYZ* but you imagined something else being *XYZ*' might we be able to restore the entailment from (really) conceivable to possible.

24 We can link what has just been said to what we learnt from Hawley and Bird about natural kinds: for any natural kind (i.e. complex universal) there is a substructure (i.e. a certain composition or conjunction of universals) which makes the natural kind what it is (see Hawley and Bird 2011: 219).

25 Note, that neo-Humeans need not reject all the essence claims we stumbled across. There is the possibility to argue that cases like water being H_2O might not fall under the scope of Hume's Dictum because Hume's Dictum is restricted to wholly distinct existences. However, if water is identical to H_2O then 'they' are not distinct. In other words, constitutional essence claims could be acceptable (although causal necessities are still rejected). See Wilson (2010) for an evaluation of this possibility.

26 See Salmon (1981/1982), Sidelle (1989), Soames (2006) and also Mellor (1977: 309). For concise reconstruction of Kripke's and Putnam's argument, see Salmon (1981/1982: 161–75).

27 According to its critics, Externalist semantics show at best that speakers implicitly assume Essentialism for natural kinds, but it provides no argument for that assumption. Authors like Sidelle (1989: 37) go so far as to say that this assumption results merely from conventions, from familiar appeals to what we can imagine or are willing to say. This would clearly catapult metaphysical necessity out of the world again and we are much closer to the conventionalisms of Carnap and the Empiricists.

28 Other proponents of this and similar views include Achinstein (1974: 274): 'I believe that saying that if two properties are identical then anything is such that necessarily if it has one it has the other and is such that necessarily if its having the one causes the other or is caused by something then its having the other causes or is caused by the same thing conforms with, or at least is not violated by, intuitions we have about "same property"'. Popper (1959: 424–5), here giving an epistemic argument, sympathizes with the view: 'In general, the dispositional character of any universal property will become clear if we consider what tests we should undertake if we are in doubt whether or not the property is present in some particular case'. See Whittle (2008) for a suggestion to revise Shoemaker's theory.

29 Note that Achinstein and Shoemaker, amongst others (see previous endnote), can claim to have defended similar views much earlier, and yet they have not applied their view to the subjects we discuss next in the extended way that the Dispositional Essentialists did.

30 Both are also inter-definable on pure logical grounds: $\Box p$ iff $(\neg p \mathrel{\Box\!\!\rightarrow} \bot)$, i.e. necessary is that whose negation counterfactually implies a contradiction (see Williamson 2007). However, this is not the connection we will focus on.

31 The belief in natural properties or kinds and their essences is very much akin to the creation myth of the monotheistic Abrahamic faiths (Judaism, Christianity and Islam). Remember: in the beginning God created the heavens and the earth. . . . water, land, sky, animals . . . vegetation: 'plants bearing

seed according to their kinds and trees bearing fruit with seed in it according to their kinds' (Genesis 1:12). The Genesis book speaks not exactly about what we, today, think to be the natural kinds, including their essences, but the general belief in a world that is well ordered and categorised into natural properties, kinds and their essences is the same. Combine all this with some of the dispositional Essentialists' claim that there is but one modality, metaphysical necessity that lords over the world – and Dispositional Essentialism seems to be a biblical theory. Note that Humean metaphysics, especially classical and Logical Empiricism, is much more akin to Buddhism (see, for example, Gopnik 2009).

32 However, in her 'No God, No Laws' Cartwright (2007: 7) makes explicit that she does equate causal or dispositional connections with necessity: 'And a regularity is just a collection of paired events: *B* follows *A* once, *B* follows *A* again, it does so again, and again and … It doesn't matter in what mode the regularity occurs, whether for instance this kind of pattern would continue […] in different possible worlds. A regularity is just a collection of paired events and a collection does not make any of its members happen'.

33 These two statements have been adopted for Dispositional Essentialism from what David Lewis said about the assumption that natural properties exist (see Lewis 1999: 1–2; 1986: 3).

34 We will not need both Essentialism and Dispositionalism for all our purposes. Sometimes Dispositionalism alone suffices.

35 Ellis writes 'kind' (singular) because he believes our world as a whole, with all its subkinds, to be a natural kind. This does not matter for our concerns.

36 Whether this solution is acceptable or not we cannot discuss it here any further; however, there is a related problem with laws of nature to which we turn in the section on dispositions and laws (see Section 6.4.2). Note, however, that Humeans can capitalise on this difficulty for Essentialism to handle counterfactuals (see Schaffer 2005: section 7.8).

37 An in-depth assessment of Dispositional Essentialism and the grounding of modalities can be found in Jaag (2014). More recent critique can be found in Vance (2014).

38 Others who have put forward dispositional theories of possibility include Contessa (2010), Molnar (2003) and Mumford (2004).

39 Both authors are in principle sympathetic to the general Dispositionalist idea so that they, in addition to pointing out the difficulties, offer improvements. Yates focuses largely on the formal (logical) adequacy of the Dispositionalists' idea to spell out possibility and necessity. He brings into play the axioms of the different modal logical systems *T*, *K*, *S*4 and *S*5.

40 For a hybrid view, where some laws turn out to be necessary and others contingent, see Tahko (2015).

41 To summarise, we have (i) and (ii) with Bird and Ellis; the combination (i) but not (ii) with Armstrong; and not (i) and not (ii) with the Humeans. The reader might wonder whether the fourth logically possible combination is defensible: not (i), i.e. the laws do not necessitate what goes on, but (ii) they are nonetheless necessary themselves. This is tricky because if the laws are necessary, (ii), then it is hard to see how what they say could not happen necessarily (not (i)). In other words, (ii) seems to imply (i). Such a position – not (i) and (ii) – has hardly ever been explored, but Alastair Wilson recently suggested something very close to it in *The Nature of Contingency* (forthcoming). Here is the idea: similar to logical or mathematical truths, necessary laws could *determine,* in the sense of (logical) entailment, what happens; however, *determine* (in this sense) is different from *govern* or *necessitate* as produce (PROD) as in (i). Thus, (i) could be denied on the ground that the determination of what happens by law is not its necessitation.

42 See Bird (2007: 43–50) and also Bird (2005). We ignore certain details regarding the scope of the necessity operator and the issue of *cp*-laws. For Ellis's theory, see Ellis (2001: chapter 6) and for more on the relation between dispositions and laws see McKitrick (2005) and (Dorato and Esfeld 2015).

43 This and what is to follow are some moves which, if you know your beginners' logic and some modal logic, you will follow easily. If not believe us that they are logically kosher.

44 See Bird (2007: 61) for why there aren't such points in time. The reader might also wonder why the most straightforward argument against finks isn't simply *Essentialism*: if dispositions are essentially possessed no fink can take them away. The thought seems to be that sometimes it is the fact

that some disposition of a certain kind or value is possessed essentially; however, which one of that kind or value is a contingent matter. Bird gives the example of *spin up* or *spin down*, which are the two values the disposition *spin* of an electron can essentially have.

45 Note that for the less than fundamental laws, for which antidotes and finks are possible, Bird aims to turn a vice into a virtue. His idea is that in such cases the dispositional counterfactual bears a *cp* clause which is inherited by the derived laws. Thus, he hopes to give a theory of *cp*-laws in passing (see Bird 2007: 59–60; this is similar to Cartwright, in Section 6.2.3).

46 'Thou shalt have no other gods before me': see endnote 31.

47 It should be noted that Eagle and Schrenk differ in what the consequence of their argument should be, but the details are not relevant here. Eagle opts for Humeanism whereas Schrenk argues for an anti-Humean connection in nature (PROD) that is not necessity (MOD).

48 Bird relies on what Kripke and Putnam have put forward to account for the seeming contingency of water being H_2O (while this, according to them, is necessary).

49 This move is also open to Shoemaker, who uses laws to express powers had by properties essentially (see Section 6.4 and (1998), where he alludes to this solution).

50 For a solution that seems a little ad hoc, see Ellis (2001: 249ff): the world as a whole is a natural kind and it has the dispositional essence to the effect that closed systems conserve energy.

51 For other critical assessments of Dispositional Essentialism and laws in general, see Smart and Barker (2012) who argue that Dispositional Essentialism is not too much different to Armstrongian necessitarianism modified by a thesis about property identity; Corry (2011) who challenges the Dispositional Essentialist to respond to Cartwright's argument that there are no strict regularities; and Drewery (2005) who doubts that essences are ontologically more basic than laws. In a review of Bird's book, Barbara Vetter (2009) offers many valuable critiques. She also makes the striking observation that, in some respects, Bird's metaphysics exhibits the Humean feature of 'just one little thing, and then another' which Dispositional Essentialism aims so much to avoid (Vetter 2009: 321).

52 There is a paper collection on dispositions and causation, see Handfield (2009). Also, the mechanistic theory of causation can, in certain respects, be regarded as a dispositional theory of causation (see Section 5.4). For a critical assessment of a theory of causation that is both dispositional *and* essentialist see Chakravartty (2008) who argues (like Cartwright) against essence, but for Dispositionalism.

53 We write this as if Hüttemann developed his theory in direct response to that problem. This is not the case. The genesis of his theory is independent of that hurdle. Still, because it does mention and offer a potential solution to it we thought it is legitimate to unfold his theory as if it was a direct response. Note also that (as Hüttemann himself points out), Tim Maudlin (2007) has a similar approach (see Section 4.6). Three points of deviation are: (i) Hüttemann's ontology incorporates dispositions whereas Maudlin is (only) a realist regarding laws; (ii) Hüttemann but not Maudlin defends metaphysical necessity; and (iii) Maudlin is, in the end, a reductionist (of everything to physics) while Hüttemann is not committed to reducing the other sciences to physics.

54 Hüttemann (2013b: 27) mentions that we have already found the idea of a conditional necessity in Francisco Suárez's (1548–1617) *Disputationes Metaphysicae* (1597).

55 We say *non-reductive* because the crucial notion of dispositionality is a causal notion itself. Mumford and Anjum welcome that their theory is non-reductive (Mumford and Anjum 2011a: 7–11). Note, however, that they prefer the term 'powers'. For our purposes we can ignore the subtle differences between powers and dispositions and use the two terms synonymously.

56 See also Mumford and Anjum (2010, 2011c) and also Jacobs (2007: 183ff) for a similar but different dispositional theory of causation.

57 The idea for a *sui generis* 'necessary connection that dispositions or causal powers bring to the world' is already in Mumford (2004: 168).

58 For similar if not entirely matching thoughts on forces as a metaphor for causes, see Creary (1981), Corry (2006) and Schrenk (2011).

59 We have some doubts here that this argument works. It is the force vector addition model that tempts us to think that $C + A$ is a further possible world because for any sum of forces C, there does seem to be a further force A such that $C + A$ is the new resultant. However, this idea, projected onto

possible worlds, might conflict with orthodox models of possible worlds as maximal-consistent states of affairs, for example, as something like Carnap's state descriptions (see Section 2.3), where worlds (or total *C*s) are rows (of *T*s and *F*s) in truth tables. If so, it is doubtful whether the addition of some *A* is really possible to any *C* according to Mumford and Anjum. Here is the direction in which the counterargument might go: We may assume that *C*, the sum total of all conditions *positive and negative taken together*, is the specification of a whole possible world ('sum total of all conditions positive and negative' – including powers and forces – suggests such a reading), i.e. a row in a truth table. Then 'positive conditions' correspond to *T*s, 'negative' or absent conditions to *F*s in rows of the truth table. Yet, in this picture, no *A* can be added to *C*. It would not be clear in this view of sum totals/possible worlds what that would mean: entries in rows can 'change' from *T* to *F* or *F* to *T* but this means, in effect, to move to a different row in the truth table which is not the original *C* (in fact, which is inconsistent with *C*). In other words, there is no $C + A$ if *C* is a complete row (= a sum total). $C + A$ is a conceptual impossibility if this view, or some similar one, of possible worlds and sum totals is assumed and it is incorrect or confusing to say that 'no matter how big an antecedent is, it can always be strengthened'. However, let us assume for the sake of the argument, that *A* can be added to any total *C*, as force vector addition seems to suggest.

60 It would be interesting to investigate what the logical relation is between orthodox necessity and this necessity* (if they really are distinct). The necessity of ($C \rightarrow E$) surely implies the necessity* of ($C \rightarrow E$). The converse should not hold otherwise Mumford's and Anjum's argument that not even sum totals of causes necessitate fails. In this case, their overall message would shrink to '*single causal factors* (smaller than sum totals) do not necessitate' but that would be fairly uncontroversial.

61 Should orthodox necessity and necessity* turn out to be equivalent (see endnote 60), then we could even claim that the diachronic necessity of ($C(t) \rightarrow E(t+\Delta t)$), for sum totals C, is constituted by the individual dispositions within C working in concert. This would be the opposite to the view we attributed to Bird and Ellis in Section 6.4.2., specifically in 'No work for lumberjacks'.

Bibliography

Achinstein, P. (1974) The Identity of Properties. *American Philosophical Quarterly*. 11 (4). pp. 257–75.

Aristotle (n.d.) *Metaphysics*, *Book VII: Zeta*.

Armstrong, D. M. (1997) *A World of States of Affairs*. Cambridge: Cambridge University Press.

Bartels, A. (2000) The Idea Which We Call Power. Naturgesetze und Dispositionen. *Philosophia Naturalis*. 37. pp. 255–68.

Bird, A. (1998) Dispositions and Antidotes. *Philosophical Quarterly*. 48 (191). pp. 227–34.

Bird, A. (2005) The Dispositionalist Conception of Laws. *Foundations of Science*. 10 (4). pp. 353–70.

Bird, A. (2007) *Nature's Metaphysics*. Oxford: Clarendon Press.

Bird, A. (2016) Overpowering: How the Powers Ontology Has Over-reached Itself. *Mind*.

Borghini, A. and Williams, N. (2008) A Dispositional Theory of Possibility. *Dialectica*. 62 (1). pp. 21–41.

Braddon-Mitchell, D. (2001) Lossy Laws. *Noûs*. 35 (2). pp. 260–77.

Cameron, R. P. (2008) Truthmakers and Modality. *Synthese*. 164 (27). pp. 261–80.

Carnap, R. (1956) *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd, enlarged edition. Chicago, IL: University of Chicago Press.

Cartwright, N. (1983) *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Cartwright, N. (1989) *Nature's Capacities and their Measurement*. Oxford: Oxford University Press.

Cartwright, N. (1992) Aristotelian Natures and the Modern Experimental Method. In: Earman, J. (ed.) *Inference, Explanation, and other Frustrations. Essays in the Philosophy of Science*. Berkeley, CA: University of California Press. pp. 44–71.

Cartwright, N. (1999) *The Dappled World*. Cambridge: Cambridge University Press.

Cartwright, N. (2007) No God, No Laws. PhD thesis (online). Available from: www.isnature.org/Files/Cartwright_No_God_No_Laws_draft.pdf.

Chakravartty, A. (2008) Inessential Aristotle: Powers without Essences. In: Groff, R. (ed.) *Revitalizing Causality: Realism about Causality in Philosophy and Social Science*. New York: Routledge. pp. 152–62.

Chakravartty, A. (2013) Realism in the Desert and in the Jungle: Reply to French, Ghins, and Psillos. *Erkenntnis*. 78. pp. 39–58.

Contessa, G. (2010) Modal Truthmakers and Two Varieties of Actualism. *Synthese*. 174 (3). pp. 341–53.

Corry, R. (2006) Causal Realism and the Laws of Nature. *Philosophy of Science*. 73 (3). pp. 261–76.

Corry, R. (2011) Can Dispositional Essences Ground the Laws of Nature? *Australasian Journal of Philosophy*. 89 (2). pp. 263–75.

Creary, L. (1981) Causal Explanation and the Reality of Natural Component Forces. *Pacific Philosophical Quarterly*. 62 (2). pp. 148–57.

Cross, T. (2012) Review of Anna Marmodoro (ed.): *The Metaphysics of Powers: Their Grounding and Their Manifestations*. Notre Dame Philosophical Reviews (online). Available from: http://ndpr.nd.edu/news/31889-the-metaphysics-of-powers-their-grounding-and-their-manifestations-2/ (accessed 10 December 2015).

Cross, T. (2014) Review of Stephen Mumford's and Rani Lill Anjum's *Getting Causes from Powers*. *Dialectica* 68 (4). pp. 614–19.

Dorato, M. and Esfeld, M. (2015) The Metaphysics of Laws: Dispositionalism vs. Primitivism. In: Bigaj, T. and Wutrich, C. (eds.) *Metaphysics in Contemporary Science*. Poznan Studies. Brill/Rodopi. pp. 403–24.

Drewery, A. (2005) Essentialism and the Necessity of the Laws of Nature. *Synthese*. 144 (3). pp. 381–96.

Eagle, A. (2009) Causal Structuralism, Dispositional Actualism, and Counterfactual Conditionals. In: Handfield, T. (ed.) *Dispositions and Causes*. Oxford: Oxford University Press. pp. 65–99.

Earman, J., Glymour, C. and Mitchell, S. (eds.) (2002) Ceteris Paribus Laws. *Erkenntnis*. Special Issue. 57 (3).

Ellis, B. (1979) *Rational Belief Systems.* Oxford: Blackwell.

Ellis, B. (2001) *Scientific Essentialism*. Cambridge: Cambridge University Press.

Feyerabend, P. (1975) *Against Method*. London: New Left Books.

Fine, K. (1994) Essence and Modality: The Second Philosophical Perspectives Lecture. *Philosophical Perspectives*. 8. pp. 1–16.

Fine, K. (2002) The Varieties of Necessity. In: Gendler, T. S. and Hawthorne, J. (eds.) *Conveivability and Possibility*. Oxford: Oxford University Press. pp. 253–81.

Glynn, L. (2012) Review of Stephen Mumford's and Rani Lill Anjum's *Getting Causes from Powers*. *Mind*. 121 (484). pp. 1099–1106.

Gopnik, A. (2009) Could Hume Have Known about Buddhism? Charles Francois Dolu, the Royal College of La Flèche, and the Global Jesuit Intellectual Network. *Hume Studies* 35 (283). pp. 5–28.

Handfield, T. (ed.) (2009) *Dispositions and Causes*. Oxford: Oxford University Press.

Harré, R. and Madden, E. H. (1975) *Causal Powers: A Theory of Natural Necessity*. Oxford: Blackwell.

Hart, H. L. A. and Honoré, A. M. (1985) *Causation in the Law*. 2nd edition. Oxford: Oxford University Press.

Hawley, K. and Bird, A. (2011) What are Natural Kinds? *Philosophical Perspectives*. 25 (1). pp. 205–21.

Hobbes, T. (1655) De Corpore (Concerning Body). Molesworth, W. (ed.) (1839) *The English Works, Vol. I*. London: Scientia Aalen.

Hume, D. (1739–40) *A Treatise of Human Nature*. Selby-Bigge, L. A. (ed.), revised by Nidditch, P. H. (1978). Oxford: Clarendon Press.

Hume, D. (1748) *Inquiries Concerning Human Understanding and Concerning the Principles of Morals*. 17th impression of the 3rd OUP edition. Selby-Bigge, L. A. (ed.). Oxford: Clarendon Press.

Hüttemann, A. (1998) Laws and Dispositions. *Philosophy of Science*. 65 (1). pp. 121–35.

Hüttemann, A. (2007) Causation, Laws and Dispositions. In Kistler, M. and Gnassounou, B. (eds.) *Dispositions and Causal Powers*. New York: Routledge. pp. 207–19.

Hüttemann, A. (2013a) A Disposition-based Process-theory of Causation. In: Mumford, S. and Tugby, M. (eds.) *Metaphysics of Science*. Oxford: Oxford University Press. pp. 101–22.

Hüttemann, A. (2013b) *Ursachen*. Berlin: DeGruyter.

Jaag, S. (2014) Dispositional Essentialism and the Grounding of Natural Modality. *Philosophers' Imprint*. 14 (34). pp. 1–21.

Jacobs, J. D. (2007) Causal Powers: a Neo-Aristotelian metaphysics (unpublished PhD thesis). Indiana University. Available from: http://www.jonathandjacobs.com/resources/papers/Jacobs-Thesis.pdf.

Jacobs, J. D. (2010) A Powers Theory of Modality: or, How I Learned to Stop Worrying and Reject Possible Worlds. *Philosophical Studies*. 151 (2). pp. 227–48.

Jenkins, C. and Nolan, D. (2012) Disposition Impossible. *Noûs*. 46 (4). pp. 732–53.

Kistler, M. and Gnassounou, B. (eds.) (2007) *Dispositions and Causal Powers*. New York: Routledge.

Kripke, S. (1980) *Naming and Necessity*. Oxford: Blackwell.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Lewis, D. (1986) *Philosophical Papers*, *Vol. II*. Oxford: Oxford University Press.

Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lowe, E. J. (1998) *The Possibility of Metaphysics: Substance, Identity, and Time*. New York: Oxford University Press.

Lowe, E. J. (2006) *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Clarendon Press.
Mach, E. (1896) *Die Prinicipien der Wärmelehre*. Leipzig: Barth.
Martin, C. B. (1994) Dispositions and Conditionals. *Philosophical* Quarterly. 44 (174). pp. 1–8.
Maudlin, T. (2007) *The Metaphysics within Physics*. Oxford: Oxford University Press.
McKitrick, J. (2005) Dispositions and Laws of Nature. *Synthese*. 144 (3). Introduction.
Mellor, D. H. (1974) In Defense of Dispositions. *The Philosophical Review*. 83 (2). pp. 157–81.
Mellor, D. H. (1977) Natural Kinds. *British Journal for the Philosophy of Science*. 28. pp. 299–312.
Mellor, D. H. (2000) The Semantics and Ontology of Dispositions. *Mind*. 109 (436). pp. 757–80.
Mill, J. S. (1843) A System of Logic. Ratiocinative and Inductive. Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. In: Robson, J. M. (ed.) (1967) *Collected Works, Vols. VII and VIII*. Toronto: Toronto University Press..
Molnar, G. (1999) Are Dispositions Reducible? *Philosophical Quarterly*. 49. pp. 1–17.
Molnar, G. (2003) *Powers: A Study in Metaphysics*. Mumford, S. (ed.). Oxford: Oxford University Press.
Mumford, S. (1998) *Dispositions*. Oxford: Oxford University Press.
Mumford, S. (2004) *Laws in Nature*. London and New York: Routledge.
Mumford, S. and Anjum, R. L. (2010) A Powerful Theory of Causation. In: Marmodoro, A. (ed.) *The Metaphysics of Powers: Their Grounding and Their Manifestations*. New York and London: Routledge. pp. 143–59.
Mumford, S. and Anjum, R. L. (2011a) *Getting Causes from Powers*. Oxford: Oxford University Press.
Mumford, S. and Anjum, R. L. (2011b) Dispositional Modality. In: Gethmann, C. F. (ed.) *Lebenswelt und Wissenschaft*. *Deutsches Jahrbuch Philosophie 2*. Hamburg: Meiner. pp. 380–94.
Mumford, S. and Anjum, R. L. (2011c) Spoils to the Vector – How to Model Causes if You Are a Realist about Powers. *The Monist*. 94 (1). pp. 54–80.
Popper, K. R. (1959) *The Logic of Scientific Discovery*. London: Hutchinson.
Psillos, S. (2002) *Causation and Explanation*. Chesham, UK: Acumen.
Putnam, H. (1975) The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science*. 7. pp. 131–93.
Putnam, H. (1981) *Reason, Truth and History*. Cambridge: Cambridge University Press.
Quine, W. V. O. (1953) Two Dogmas of Empiricism. In: Quine, W. V. O. *From a Logical Point of View*. Cambridge, MA: Harvard University Press. pp. 20–46.
Reutlinger, A., Schurz, G. and Hüttemann, A. (2011) Ceteris Paribus Laws. In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Winter 2014

edition (online). Available from: http://plato.stanford.edu/archives/win2014/entries/ceteris-paribus/ (accessed 10 March 2015).
Reutlinger, A. and Unterhuber, M. (eds.) (2014) Ceteris Paribus Laws Revisited. *Erkenntnis*. Special Issue. 79 (10). pp. 1703–1847.
Russell, B. (1900/1992) *A Critical Exposition of the Philosophy of Leibniz*. London: Routledge.
Russell, G. and Graff Fara, D. (2012) *The Routledge Companion to Philosophy of Language*. Abingdon, UK: Routledge.
Salmon, N. (1981/1982). *Reference and Essence*. New York: Prometheus Books.
Schaffer, J. (2005) Quiddistic Knowledge. *Philosophical Studies*. 123. pp. 1–32.
Schrenk, M. (2005) The Bookkeeper and the Lumberjack: Metaphysical vs. Nomological Necessity. In: Abel, G. (ed.) *Kreativität. XX. Deutscher Kongress für Philosophie. Sektionsbeiträge Band 1*. Berlin: Universitätsverlag der Technischen Universität. pp. 849–56.
Schrenk, M. (2007) *The Metaphysics of Ceteris Paribus Laws*. Frankfurt: Ontos.
Schrenk, M. (2010) The Powerlessness of Necessity. *Noûs*. 44 (4). pp. 725–39.
Schrenk, M. (2011) Interfering with Nomological Necessity. *The Philosophical Quarterly*. 61 (244). pp. 577–97.
Schrenk, M. (2014) Better Best Systems and the Issue of Cp-laws. *Erkenntnis*. 79 (10 supplement). pp. 1787–99. doi: 10.1007/s10670-014-9642-9.
Schurz, G. (2011) Review of Alexander Bird: Nature's Metaphysics. *Erkenntnis*. 74 (1). pp. 137–42.
Shoemaker, S. (1979/1984) Identity, Properties and Causality. In: Shoemaker, S. *Identity, Cause and Mind*. Cambridge: Cambridge University Press. pp. 312–42.
Shoemaker, S. (1998) Causal and Metaphysical Necessity. *Pacific Philosophical Quarterly*. 79. pp. 59–77.
Sidelle, A. (1989) *Necessity, Essence and Individuation: A Defense of Conventionalism*. Ithaca, NY: Cornell University Press.
Smart, B. and Barker, S. (2012) The Ultimate Argument against Dispositional Monist Accounts of Laws. *Analysis*. 72(4). pp. 714–22.
Smart, B. and Talibard, M. J. (2012) Review of *Getting Causes from Powers*. *Philosophy in Review*. XXXII (5). pp. 407–9.
Soames, S. (2006) The Philosophical Significance of the Kripkean Necessary Aposteriori. *Philosophical Issues*. 16: Philosophy of Language. pp. 288–309.
Tahko, T. (2015) The Modal Status of Laws: in Defence of a Hybrid View. *The Philosophical Quarterly*. Unpublished.
Vance, C. (2014) Dispositional Modal Truthmakers and the Necessary Origin. *Philosophia*. 42 (4). pp. 1111–1127.
Vetter, B. (2009) Review of Alexander Bird: Nature's Metaphysics. *History of Philosophy and Logical Analysis*. 12. pp. 320–8.
Vetter, B. (2015) *Potentiality. From Dispositions to Modality*. Oxford: Oxford University Press.
Wang, J. (2014) The Modal Limits of Dispositionalism. *Noûs* (online). doi: 10.1111/nous.12080. pp. 1–16.

Whittle, A. (2008) A Functionalist Theory of Properties. *Philosophy and Phenomenological Research*. 77. pp. 59–82.
Williamson, T. (2007) Philosophical Knowledge and Knowledge of Counterfactuals. *Grazer Philosophische Studien*. 74 (1). pp. 89–123.
Wilson, A. (manuscript) The Nature of Contingency.
Wilson, J. (2010) What is Hume's Dictum, and Why Believe It? *Philosophy and Phenomenological Research*. 80 (3). pp. 595–637.
Yates, D. (2014) Dispositionalism and the Modal Operators. *Philosophy and Phenomenological Research*. 89 (1). pp. 411–24.

7 Epilogue

Meta-metaphysics

7.1 Metaphysics, anti-metaphysics and metaphysics again

All things recur eternally.
(Nietzsche 2003: 237)

We have walked a long path from Rationalist metaphysics and the classical Empiricists' critique thereof, through the Kantian reconciliation of the two, to the Logical Empiricists' radicalisation of their classical ancestors' views All this was covered in Chapter 1. In the subsequent chapters, which systematically targeted dispositions (Chapter 2), counterfactuals (Chapter 3), laws of nature (Chapter 4), causation (Chapter 5) and dispositional essentialism (Chapter 6), we followed the development of metaphysics of science since the Logical Empiricists' hostilities.

As we have seen, during the past century philosophers became more and more happy to loosen the strictures of Empiricism's radical anti-metaphysical stance. They saw the necessity to gradually add to their theories those entities and concepts that were considered dubious before: natural properties, dispositions, necessity and essence to name the most striking ones. According to one of their arguments, if these notions are not taken on board again, many concepts that are crucial to science, such as causation or law of nature, will not find proper explications. This development culminates in the contemporary resurrection of metaphysics of science. The shift of perspective can be observed in each of our chapters. We give some examples:

Dispositions: whereas the Empiricists want to eliminate dispositional predicates in favour of observational language, the dispositional essentialists believe that nature's properties themselves are essentially dispositional (Chapter 2 versus Chapter 6).

Counterfactuals: in the wake of Logical Empiricism, where counterfactuals were analysed in terms of laws (taken to be regularities) and logical entailment, Lewis helps himself to his possible worlds realism (Section 3.2 versus Section 3.4).

> **Laws:** Where laws were first seen as mere epistemic tools to make predictions (inference tickets), Maudlin later assumes that they belong to the fundamental building blocks of the universe (Section 4.2 versus Section 4.6).
>
> **Causation:** Where Mackie analysed causation in terms of necessary and sufficient conditions, Armstrong sees causation as the instantiation of nomological necessitation (Section 5.2 versus Section 4.4).

However, in the last decade analytic metaphysics has come under fire again. One of the strongest statements comes from James Ladyman and Don Ross:

> Contemporary analytic metaphysics, a professional activity engaged in by some extremely intelligent and morally serious people, fails to qualify as part of the enlightened pursuit of truth, and should be discontinued [. . .]. We care a great deal about philosophy, and are therefore distressed when we see its reputation harmed by its engagement with projects and styles of reasoning we believe bring it into disrepute, especially among scientists.
>
> (Ladyman and Ross 2007: vii)

Most philosophers would agree that a critical re-evaluation of metaphysics' blossoming within the past decades is required. However, not all of them are as sceptical as Ladyman and Ross.

In Box 7.1 we list some publications on this thriving topic, which has now come to be known under the name 'meta-metaphysics'. As you can see from this list and also from the quotes in this chapter, the literature on meta-metaphysics is fairly new. We will discuss this literature in close relation to the earlier chapters of the book. This will give us plenty of opportunity to consider which meta-metaphysical justifications were in fact given or should have been given for the vindication of the respective first order metaphysical claim.

BOX 7.1 Recent publications on meta-metaphysics and the relation between metaphysics and science

Braddon-Mitchell and Nola (2009) *Conceptual Analysis and Philosophical Naturalism.*

Brandom (2012) Metaphilosophical reflections on the idea of metaphysics.

Callender (2011) Philosophy of science and metaphysics.

Chakravartty (2010) *A Metaphysics for Scientific Realism: Knowing the Unobservable*.

Chakravartty (2013) *Realism in the Desert and in the Jungle: Reply to French, Ghins, and Psillos.*
Chalmers, Manley and Wasserman (2009) *Metametaphysics: New Essays on the Foundations of Ontology.*
Godfrey-Smith (2012) *Metaphysics and the Philosophical Imagination.*
Hawley (2006) Science as a guide to metaphysics?
Ladyman (2007) *Does Physics Answer Metaphysical Questions?*
Ladyman (2012) *Science, Metaphysics and Method.*
Ladyman and Ross (2007) *Every Thing Must Go: Metaphysics Naturalized.*
Maudlin (2007) *The Metaphysics within Physics* (esp. the epilog).
Ney (2012) *Neo-Positivist Metaphysics*
Paul (2012) Metaphysics as modeling: the handmaiden's tale.
Price (2009) *Metaphysics after Carnap: The Ghost Who Walks?*
Ross, Ladyman and Kincaid (2013) *Scientific Metaphysics.*
Sider (2009) Ontological realism.
Takho, T.E. (2015) *An Introduction to Metametaphysics*.
Van Fraassen (2004) *The Empirical Stance* (esp. chapter 1)
Wilson, M. (2010) What Can Contemporary Philosophy Learn from Our 'Scientific Philosophy' Heritage?
Wilson, J. (2013) Three dogmas of metaphysical methodology.

7.2 Simplicity, strength and their balance as a tool to judge metaphysics

With their Verificationist theory of meaning the Logical Empiricists had a clear criterion for what counts as bad or unjustifiable metaphysics: unless some observation could prove a claim or theory to be true or false there's no cognitive value to that claim or theory and, consequently, because metaphysics is very likely to transcend empirical investigations, simply all metaphysical statements lack meaningful content. In other words, the phrasing '*bad* or *unjustifiable* metaphysics' we used earlier is, in light of Logical Empiricism, a pleonasm (see Section 1.4.1).

Nowadays, hardly anyone believes in Verificationism of meaning. One might think that metaphysics is better off because even if all metaphysical claims transcend the empirical they can still be meaningful. However, if there is indeed no observable evidence (i.e. there are no metaphysical predictions or retrodictions or experiments for empirical tests to check whether metaphysics corresponds to the facts), then it is hard to see how to rationally justify and defend metaphysical claims, be they meaningful or not. For example, whether there is nomological necessitation, *N*, according to Armstrong, can never be settled by observation even if the question about whether *N* exists is acceptable as meaningful. How can we therefore justify our belief (or disbelief) in *N*?

Here's an idea: for justification, let us look at internal coherence, simplicity and comprehensiveness of the theory rather than for its correspondence with empirical observations. Quite a few philosophers have proposed something to that effect:

> If your system strikes your readers as being simpler, more coherent, or more promising than any alternative for dealing with the recalcitrant difficulties of other systems, then this may be a good enough reason to buy it.
> (Ellis 2001: 262)

> Scientific control of metaphysical theories is given theory internal through the criteria of consistency, logical coherence, clarity, comprehensiveness, depth.
> (Meixner 1999: 128, my translation)

> [Metaphysicians'] methodology is rather quasi-scientific. [. . .] Theoretical insight, considerations of simplicity, integration with other domains (for instance science, logic, and philosophy of language), and so on, play important roles.
> (Sider 2009: 385)

> Metaphysicians use standards for choosing theories that are like the standards used by scientists (simplicity, comprehensiveness, elegance, and so on).
> (Sider *et al.* 2007: 6)

This univocal assent sounds promising. Unfortunately, we have good reasons to doubt that this method will work. For one thing, these criteria are minimal requirements to any theory, be that metaphysics, science or literary theory. Let us put it this way: who would want their theories to be complicated, inconsistent, incoherent, unclear, shallow and incomprehensive? We might suspect that, in their fundamentality, the standards of simplicity, strength and coherence are possibly too indiscriminatory and insufficient to settle metaphysical disputes.

Moreover, the question arises how we could make each of these demands precise enough in order to be able to measure the overall value of metaphysical theories for comparison. What exactly is *insight*? What is *depth*? And how do we quantify how insightful and how deep one metaphysical theory is compared to another?[1]

We accept this challenge for one of the listed criteria and investigate what could be done to make this sample case more precise. In order to see if we can spell out what is meant by *simplicity* (or at least one aspect of simplicity) of a metaphysical theory, we introduce an age old principle, *Ockham's Razor*.

Ockham's Razor Educated people love to formulate this principle in Latin: *Entia non sunt multiplicanda praeter necessitatem* (entities must not be multiplied beyond necessity). This formulation, however, is not Ockham's, who never explicitly mentions but only indicates his belief in something like it in his theological writings. The mentioned Latin phrase apparently comes from John Ponce of Cork (1599–1661). Ockham also never spoke of a razor. This metaphor – razors were used to scrape ink off paper to correct errors (think of *tabula rasa*, Latin for shaved/erased tablet, i.e. a blank slate) – comes from the nineteenth century mathematician William Rowan Hamilton (1805–65, see his *Discussions* 1852: 590).[2]

What, now, is Ockham's parsimony principle? It says in essence that theories which need to postulate fewer entities for their explanatory power than their rivals should be preferred. Modern biology, for example, sees no need to postulate any vital forces that keep organisms alive and, therefore, no biologist believes in theories that postulate the existence of these forces. Causal modelling theory suggests that one should prefer a theory that posits a single cause for multiple, regularly co-occurring events to a theory that invokes multiple unrelated causes for each of these events (see Maudlin 2007: 76), in other words, the simpler the ontology of a *scientific theory*, the better that theory.

One might also want to say that a *metaphysical theory* with simpler, more modest ontological postulates is to be preferred to extravagant, complex ontologies. For example, if metaphysicians can explain what laws of nature are or what causation is without the postulation of powers or necessary connections in nature, then such a theory is to be preferred to anti-Humean rivals. Here is another metaphysical example for such ontologically parsimonious tendencies: reduction or elimination attempts which try to get rid of certain kinds of entities (or properties) in favour of a smaller set of fundamental building blocks can count as an Ockhamian barber job (remember the example that modern day Humeans try to reduce dispositions to categorical properties plus laws and/or causation (see Sections 2.2.4, 2.2.5 and 2.5)). This (amongst other things) also arrives at a simpler ontology.

Can we trust Ockham's Razor, though? Is it a good principle to decide between theories? Ockham's Razor is a principle that is quite successful when it comes to empirical scientific theories. There, we are able to argue in favour of it on inductive grounds: parsimonious scientific theories have, more often than not, been more successful than opulent rivals. For metaphysics, however, no such inductive testing by additional empirical observations is possible. It is uncertain whether, in metaphysics, simplicity tracks truth. Therefore, it is less clear whether Ockham is successfully applicable within metaphysics as well.

Lacking such reason and in the absence of justification, the urge to apply the Razor in metaphysics might come down to mere aesthetic preferences. As a matter of taste, but for no other objective rational reason, we might prefer early twentieth century Bauhaus metaphysics to seventeenth century Baroque ontologies. The problem is that tastes vary considerably and not everyone agrees with Quine's famous statement that an 'overpopulated universe is in many ways unlovely. It offends the aesthetic sense of us who have a taste for desert landscapes' (Quine 1948/1953: 4).[3]

Note one further critical point: the background assumption has so far been that ontologies can be easily compared in relation to their sizes, as if we were always only confronted with ontologies that are the sub- or superset of the other (one containing kinds of entities a, b, c; the other only a and b). However, we are confronted with a trickier case when we wish to judge which of the following three ontologies is richest: $\{a, b, c\}$, $\{a, b, d\}$, $\{e, f, g, h\}$. Needless to say, this is problematic for both a scientific and a metaphysical application of Ockham.

Strength, coherence, clarity, etc. Now, simplicity was only one example from our criteria list for good metaphysics and we might well suspect that we also await difficulties for the other entries like 'clarity' (Meixner) or 'being promising' (Ellis), etc. Not only might the praised standards of strength, coherence, clarity, etc. be too indiscriminatory and insufficient to settle metaphysical disputes, they might also be too vague and imprecise.

Ellis writes:

> A metaphysics has to be argued for, and defended, on many different fronts at the same time. The only way of doing this is to set it out carefully, displaying its range and overall coherence, contrast it with other positions, show how it deals with philosophical issues in various fields, and answer specific objections. The form of the argument is therefore [. . .] an argue by display. You show your wares and invite people to buy them.
>
> (Ellis 2001: 262)

However, like it or not, all you might be able to offer are sham packages because there is no well-calibrated scale with which you could objectively display the weight of your wares. In this regard, Lewis self-consciously raises doubts that such an Ellis-style bazaar evaluation will help arrive at watertight conclusions:

> Maybe the prize is higher than it seems [. . .]. Maybe the price is not right; even if I am right about what theoretical benefits can be had for what ontological costs, maybe those benefits are just not worth those costs. Maybe

> the very idea of accepting controversial ontology for the sake of theoretical benefits is misguided.
>
> (Lewis 1986: 4–5)[4]

It seems, therefore, that we will have to look for more and different possibilities to arrive at, justify and evaluate metaphysical claims.

7.3 Inferences to the best explanation

One of the key arguments that Nancy Cartwright has offered in support of Dispositional Realism was the *transportability argument* (see Section 6.2.3). How come, she asked, that we can project what we measure in ideal laboratory conditions to messy real life circumstances where disturbances of all kinds might influence what happens? Her answer was that what best explains this is the following assumption: what we measure in the lab is a system's capacities, that is, its dispositions to behave. Now, as long as the system itself is fairly stable and the outer-lab circumstances, although messy, are not destructive, its capacities remain. They are *transported* from the lab to the outer world. Within certain boundaries, the system's dispositional behaviour is also reduplicated outside. This is what enables us to make predictions about the system's doings in less than ideal conditions.

Note the hidden dialectic here: the assumption that what is measured in the lab are stable *dispositions* of a system is supposed *to explain in a better way* than alternative theories (specifically Humean anti-Dispositionalist accounts) why lab findings are transportable to the outside world. In other words, Cartwright uses a so called 'inference to the best explanation' (IBE).[5]

Now remember Armstrong's *nomological necessitation* (see Section 4.4). There, the assumption of the existence of a necessitation relation was justified by an inference to the best explanation too, unless you accept the second-order universal of nomological necessitation, *N*, so the story goes, you cannot really grasp what laws of nature are and, for that matter, what causation is. Therefore, even if we can't empirically verify the existence of *N*, it is still reasonable to assume its existence in nature because of this IBE: 'The postulation of that extra thing is a case of *inference to the best explanation*' (Armstrong 1983: 55).

The IBE argumentation that what the 'best' (available) explanation says exists *does actually exist* is very attractive for metaphysicians because it seems to allow the stipulation of entities that could otherwise not be justified and especially not empirically by observation. It is, however, not unproblematic. First note that IBEs pull in the opposite direction of the Razor principle: where Ockham aims to shave away excessive growth, IBEs are there to enrich our ontology (see Maudlin 2007: 179–83). The two methods therefore need to be negotiated when weighing metaphysical theories.

Similar to our discussion of Ockham, we will focus first on the use of IBE within the empirical sciences to evaluate their tenability. Remember our first example of a successful IBE application in astronomy (see Section 4.4): in 1920, Adams and Leverrier explained why Uranus's orbit diverged from what Newtonian theory predicted by the postulation of the existence of a further planet, later to be called 'Neptune'. Unfortunately, there was yet another adventure in space, this time unsuccessful: Leverrier later also postulated the existence of a planet, 'Vulcan', which was this time used to explain Venus's divergence from its predicted Newtonian orbit. However, for all we know, this planet does not exist. The difference between the predicted and the actual result can be better explained by Einstein's relativity theory.[6]

This is the reason why we wrote '*best*' in quotation marks earlier and secretly added the word 'available': quite obviously Leverrier's alleged 'best' explanation was merely the best at hand to him at that time. Later, Einstein's theory proved to be better. Although IBEs deliver prima facie justification, they are not watertight, deductively valid arguments. For the sciences, this is not a major disaster because, luckily, IBEs await later *empirical verification* or *falsification*: the planet is ultimately discovered or its nonexistence proven because all possible places have been observed as empty space. Moreover, in the sciences, the general success rate of IBEs can be inductively tested (just as with Ockham's Razor): if IBEs prove to be largely successful, they are a useful, reliable tool even if they are not infallible and occasionally go wrong.

However, all this should ring alarm bells for metaphysicians because of the following discontinuity between scientific and metaphysical IBEs: metaphysical IBEs target entities which cannot be found in the realm of the empirically testable. Cartwright's capacities and Armstrong's necessitation transcend the observable. Thus, we cannot gather corroborating data through any empirical test. We can neither confirm nor empirically contradict a specific metaphysical claim and, therefore, we also cannot accumulate a positive track record for a positive induction about the general success of IBEs. In other words, the justification of IBEs within metaphysics will probably always only be a matter of how convincing, coherent, strong and maybe aesthetic the inferences are.[7]

It will come as no surprise that there are plenty of examples where philosophers disagree about their respective IBEs. Here are two striking examples: although Lewis agrees with Armstrong's IBE to natural properties (in fact, Lewis has been convinced by Armstrong thereof), Lewis disagrees with Armstrong's IBE to nomological necessitation (see Section 4.3). Or think of the step that some make from the existence of natural kinds and properties to their essences which others will not find plausible (see Section 6.3). IBEs, just as Ockham's Razor, therefore prove to be a controversial tool to arrive at metaphysical (existence) claims.

To end this section, we highlight statements from Laurie Paul as proponent and Helen Beebee as opponent of IBEs. Despite the gap between science and metaphysics, Laurie Paul claims that 'if we accept inference to the best explanation in ordinary reasoning and in scientific theorizing, we should accept it in metaphysical theorizing' (Paul 2012: 22).[8] Helen Beebee, who also compares IBEs in the sciences to those in metaphysics, questions the legitimacy of IBEs in metaphysics. She starts with a positive assessment within the sciences:

> [b]ased on the success of chemistry and particle physics and perhaps the sciences more generally [we] think that inference to the best explanation is a good mode of inference to deploy in those particular contexts.
>
> (Beebee 2009: 456)

However, she has doubts when it comes to metaphysics and writes ironically:

> It doesn't follow that it is a good mode of inference to deploy, completely generally, in metaphysics. Indeed, if we regard inference to the best explanation as a mode of inference that trumps all other considerations [. . .]. [Then it] isn't very surprising that what comes at the end of the chain of inference is God.
>
> (Beebee 2009; this quote is from Beebee's review of John Foster's *The Divine Lawmaker*)[9]

7.4 Transcendental arguments, serviceability and indispensability

We have discussed that Lewis's argument for the assumption that the world comes equipped with perfectly natural properties was an inference to the best explanation. Remember what he said regarding this assumption: it is 'so commonsensical and so serviceable – indeed, was so often *indispensable* – that it was foolish to try to get on without it' (Lewis 1999: 1–2, emphasis added). Focusing on the italicised notion, instead of calling this an inference to the best explanation, we could label it as an 'indispensability argument'. These arguments have the following structure: if, in order to arrive at *X* the postulation of *Y* is necessary, and you do indeed want *X*, you cannot but postulate *Y*. *Y* is *indispensable* for *X*. Thus: *Y*!

Before we continue to unearth the general structure of this kind of argument, note that it is a recurrent theme: we have already encountered the need to believe in the existence of natural properties for a specific theory to work in our dispositions chapter, where Kaila's analysis of dispositional predicates won't work without the assumption of natural classes or some such (see Section 2.1.4).[10]

Transcendental arguments With *indispensability arguments*, we move closer to an argument that we encountered at the beginning of our book: *Kantian transcendental arguments* (see Section 1.3). Remember that transcendental arguments claim to unearth the necessary conditions for the possibility of something. Couched in this structure, Lewis's argument reads that unless you postulate natural properties there is simply no possibility to define what a law of nature is (see Section 4.3). So, if you do not want to give up your project (in this case, to explicate what laws of nature are) you must believe in the existence of natural properties.[11]

How close Lewis's and other modern day indispensability arguments really are to transcendental arguments is a complex issue. Certainly, Lewis's and other recent arguments are not, like the Kantian original, meant to be *synthetic* judgements arrived at in an *a priori* way (see Section 1.3); however, just what status these modern arguments have if they are not synthetic *a priori* is not so easy to say and definitely worth future research. Clearly, the type of target phenomenon in Lewis's argument is also different to Kant's. For Kant, it is sense experiences of the outer world, which are said to be possible only if our mind structures phenomenal input in a causal and spatio-temporal order; for Lewis, however, it is a theory of laws that is said to be in need of the postulation of perfectly natural properties.

Serviceability – indispensability's little sister A detailed comparison of transcendental with indispensability arguments within metaphysics of science would not only go beyond the scope of this book, but research on this topic is also still missing. For this reason, we shift to a different adjacent form of arguments: when praising natural properties Lewis (see earlier) does not initially claim their *indispensability*, but rather he starts with an appraisal of their *serviceability*.[12] In Section 6.4, we also saw Dispositionalists use serviceability arguments for their thesis, and we should accept Dispositionalism, according to Ellis, because of its serviceability for the project to give accounts of counterfactual conditionals, laws and causation.

Whatever we may think of *indispensability* (or *transcendental*) arguments, it is clear that if they succeed they are much more compelling than arguments that only show the *serviceability* of some metaphysical postulation. This is because serviceability (but not indispensability) may have competitors: many other postulations instead of *Y* might be equally *serviceable* for *X*, and we need further reasons to decide whether *Y* or one of the competing *Y**, *Y*** are to be preferred (or exclusively true). If, however, the indispensability of *Y* for *X* can be proven, then regardless of what else can or has to be postulated, the need for *Y* is ascertained, and *Y*'s postulation is on safe grounds. If they are obtainable, then indispensability arguments are to be preferred to serviceability arguments.

In addition to this degrading of serviceability, remember how differently serviceability (or 'efficiency' and 'fruitfulness' as Carnap says) was judged as a means to come to metaphysical conclusions. Whereas Lewis claims that the serviceability of a hypothesis 'is a reason to think it is true' (Lewis 1986: 3), Carnap denies this and concludes with a pragmatical advice:

> The efficiency, fruitfulness, and simplicity of the use of the thing language may be amongst the decisive factors [for its choice]. [. . .] However, it would be wrong to describe the situation by saying: 'The fact of the efficiency of the thing language is confirming evidence for the reality of the thing world'; we should rather say instead: 'This fact makes it advisable to accept the thing language'.
>
> (Carnap 1956: 208, my addition in brackets)

Indispensability arguments within IBEs Note that we used Lewis's argument for natural properties as both an example for an indispensability/serviceability argument and also as an example for an inference to the best explanation. How do IBEs relate to serviceability, indispensability and transcendental arguments?

For IBEs, we will select the best explanation from the pool of explanations that suffice to explain a certain phenomenon. Certain metaphysical (existence) postulates might be essential parts of such an explanation without which it would fail. In other words, these metaphysical postulations are an indispensable, necessary part of this best sufficient explanation. Thus, we have here an application of indispensability within IBEs: first, establish what the *best explanation* is for some phenomenon and then extract what the *indispensable* elements of that explanation are, for example existence postulations.[13]

In our example, natural properties are indispensable for the (alleged) best theory of laws. The existence of natural properties, according to the claim, is a necessary precondition for that theory of laws to work. We should also believe in that latter theory because, according to Lewiseans at least, it is considered to be the best we have.[14]

It is this latter caveat that might often spoil the game for heavyweight metaphysics: which theory is considered best depends on so many, probably subjective, aesthetic factors (see earlier) that we end up arriving at best at conditionals – for example, if you consider *this* to be the best theory, then *that* will be a necessary ingredient; if you prefer another theory it will have its own indispensable assumptions. Establishing conditionals of these forms is certainly interesting, but they, by themselves, do not carry the truth of their antecedents with them and we therefore do not arrive at unconditionally indispensable metaphysical assumptions.

Maybe conditional truth is all metaphysics should aim for. Some consider this not to be a disaster: if metaphysics' 'principal aim is not belief, or even knowledge, but understanding' (Brandom 2012: 22), then that goal can be considered achieved when we comprehend these conditionals. On that note, we turn away from possible argument forms within the metaphysics of science and instead look at a belief that is widely shared by contemporary metaphysicians. We will return to the issue of metaphysical methodologies in Section 7.6.

7.5 Science, scientific Realism and the metaphysics of science

Philosophers who engage with the metaphysics of science tend to sympathise in one way or another with science itself. Just like the Logical Empiricists, they see science, in addition to logical reasoning, as the single most important, most reliable path to truth. For that reason, they not only want the contents of their metaphysics to be close to the findings of the sciences (for example, if science says there are no vital forces active in living beings, metaphysics should not postulate them), but they also want, if possible, some scientific methods to be methods of metaphysics. This we have seen when discussing *Ockham's Razor* and *inferences to the best explanation* in both the scientific and the metaphysical realm.

Metaphysics' affinity to the sciences Laurie Paul once made fun of a too-uxorious pro-attitude of metaphysics to science in saying that, with such an attitude, metaphysics is at best a 'handmaiden to science' (Paul 2012: 2).[15] Others, however, emphasise and endorse the role model function of science to metaphysics. Tim Maudlin, for example, writes in the introduction to his *Metaphysics within Physics* book:

> Physical theories provide us with the best handle we have on what there is, and the philosopher's proper task is the interpretation and elucidation of those theories. In particular, when choosing the fundamental posits of one's ontology, one must look to scientific practice rather than to philosophical prejudice.
>
> (Maudlin 2007: 1)[16]

Alexander Bird and Craig Callender (although ultimately arriving at very different metaphysical theories) concur:

> If there is a contradiction between metaphysics and physics, metaphysics must give way.
>
> (Bird 2007: 8)

> Laying bare the metaphysical assumptions of our best [scientific] theories of the world is a crucial and important part of understanding the world. And metaphysical speculations, when anchored in systematic theorizing connected to epistemically worthy pursuits, can aid our search for new and better theories of the world, and hence, better science.
>
> (Callender 2011: 48)

The understandable fear behind these remarks is certainly that without any guidance of science metaphysics would degenerate into the wild speculations that were shunned by the logical Positivists.

Scientific Realism A particular (as yet unmentioned) philosophical stance about science and its findings is very appealing to metaphysicians of science: *Scientific Realism*. Stathis Psillos, one of its main defenders, divides this stance into three sub-theses:

> **The Metaphysical Thesis:** The world has a definite and mind independent structure. (This distinguishes Scientific Realism from Idealism or Empiricist Phenomenalism.)[17]
>
> **The Semantic Thesis:** Scientific theories should be taken at face value. They are truth-conditioned descriptions of their domain, both the observable and unobservable. (This distinguishes Scientific Realism from Empiricist Instrumentalism.)
>
> **The Epistemic Thesis:** Mature and predictively successful scientific theories are well-confirmed and approximately true of the world. (This distinguishes Scientific Realism from pessimistic attitudes which doubt scientific progress.)[18]
>
> (All taken from Psillos 2007: 226–7; see also Psillos 1999)

A Scientific Realist mainly believes that nature really is how our best sciences say it is, i.e. that the world really is made out of those kinds of entities that science postulates. This holds especially true for unobservable entities, the existence of which is postulated by well-confirmed scientific theories. For example, Scientific Realism tells us that we should believe in the existence of quarks as fundamental physics describes them to be rather than adopt a mere Instrumentalist attitude towards them. The latter, Instrumentalism, would roughly claim that quark-theory is a useful tool to make predictions and to explain why certain phenomena occurred; however, what the world is really like, i.e. whether entities like quarks really exist, should at least be left open, if not dismissed, as a nonsensical question.

Metaphysics of science and Scientific Realism Metaphysics of science is a very broad topic, and yet it certainly includes the question of what exists. *Scientific Realism* is a specific answer to this question: the ontology of the best sciences is true, i.e. those kinds of entities science postulates exist.

What would the alternative be? Well, should you be a scientific *anti-Realist* about the entities science postulates (you could be an Empiricist Instrumentalist, for example), then you are unlikely to hold a strong position when it comes to the other topics that metaphysics of science is concerned with (for example, laws, causation, natural kinds). In other words, you will not say that whether electrons and positrons really exist is an empty question, but claim that being an electron, *E*, nomologically necessitates bearing unit charge, *C*. As a scientific anti-Realist, you are more or less forced to hold Instrumentalist views also about causation, laws and natural kinds.

This confirms our hunch that Scientific Realism is at least an ally to metaphysics of science, but it is not to say that Scientific Realism predetermines any specific answers to the other metaphysical questions. Although many metaphysicians of science do in fact sympathise with Scientific Realism, their opinions about laws and causation diverge. Armstrong, Bird, Cartwright, Dowe, Ellis, Glennan, Kistler, Lange, Lewis, Mackie, Martin, Maudlin, Mumford, Salmon, Woodward (to name but a few who have figured in our book) are considered to be Scientific Realists, and yet, as we have seen, hardly any of them agree on theories of laws, causation or counterfactuals.

The reason why there is no unambiguous route from science plus Scientific Realism to metaphysics (put another way, the reason why metaphysical theories can hardly ever be extracted from physics in a unique and uncontroversial way) is obvious. First, as we have seen, there is already a controversy when it comes to the validity of (metaphysical) argument figures: not every Scientific Realist believes that Ockham's Razor, IBEs or indispensability arguments are applicable in metaphysics and the potential means to arrive at metaphysical truths, and where there is disagreement about the methods, there is likely to be disagreement about the results.[19]

Second, even when two philosophers agree on all this – on Scientific Realism and on the metaphysical methods – they need not arrive at the same conclusions. Lewis and Armstrong on laws are a key example: both philosophers are Realists and both use IBEs/indispensability arguments, and yet their theories diverge.

In short, although metaphysicians of science typically agree that Scientific Realism is correct and that metaphysics must bear some affinity to science, these shared presuppositions do by no means guide them towards unanimous answers to metaphysical questions.

7.6 The Canberra Plan

We lack a universally agreed manual about how to get, in an unambiguous way, from science to metaphysics. This was the conclusion of Section 7.4), and this is what we have just highlighted again in Section 7.5 on scientific Realism. However, (some) Scientific Realists have one final, as yet unmentioned recipe up their sleeve: they suggest that we follow what has been called the *Canberra Plan*. This plan to *tame* our metaphysical intuitions with the help of the sciences is even seen by some as the legitimate heir to Logical Empiricist programmes like Carnap's *Aufbau* (see Leitgeb 2011: 266, fn. 1). Why Canberra? David Lewis, Frank Jackson and other philosophers who were associated with the Research School of Social Sciences in Canberra, Australia, are the main proponents of the Plan, thus, it took its name from that city.

Platitudes The Plan advises us first to analyse those philosophically interesting concepts like causation, law of nature, natural kind, etc. which we use, without having precise definitions in mind, in our everyday language, in our folk theories, and also in the sciences. 'Analysis' means to collect and systematise all platitudes that we ordinarily associate with these concepts. For example, we might have the intuition that causes always precede their effects, that nothing causes itself, that there is no causation at large distances, etc.

The second step is to look for entities (broadly conceived and including processes, forces, fields, etc.) that our scientific theories postulate and that would best fit the list of intuitions we collected for the previously mentioned concepts. In other words, we investigate whether there is something in the world, as science describes it, that fits the roles and features our pre-theoretical intuitions demand for causation, laws, etc. Since Canberra planners take Scientific Realism for granted, i.e. that the world, as science *describes* it, *is* the world as it really is (at least approximately), we allegedly latch onto reality with this method.

Referents If the search is successful (also partial fulfilment of the entries might count as success), then a scientific counterpart for the pre-theoretical concepts is found. If, however, there is no realiser for the cluster of roles we have compiled, then the concept has to be discarded as untenable, at least for a proper metaphysical theory (in everyday life it might still be somewhat useful). Craig Callender comments on the Canberra Plan:

> Conceptual analysis determines the Ramsey sentence[20] that best describes the role we want some X to play, e.g., causation, but then science tells us what the world is like and whether there is anything that actually realizes that role. The enterprise of metaphysics is then very modest, for Canberra Plan metaphysics assumes that we know what the world is like. But that was what metaphysics originally was supposed to tell us!
>
> (Callender 2011: 51, endnote 3)[21]

Causation as sample case We can reconstruct theories of causation (see Chapter 5) as the result of Canberra Planning.[22] If we are very generous with our interpretation, we may start with Hume and regard him as a precursor of the Canberra Planners. Hume first collects platitudes from what the folk or metaphysicians think causation is: spatial contiguity, temporal succession, regularity, necessary connection. Hume then shows that one entry on that list, namely the idea that causation is necessitation or some such connection, finds no counterpart in the world.[23] Despite the failure of the last entry, Hume endorsed the other intuitions on his platitudes list – that causally related events happen in spatial vicinity, that causation is forward in time and that all this happens regularly – and formed his regularity theory of causation from these elements.

A second theory of causation can be more plausibly reconstructed in terms of the Canberra Plan: the transfer theory. Remember that some philosophers – Wesley Salmon, for example – did not buy Hume's claim that a (causal) connection cannot be found in the world: 'Let us take Hume's challenge seriously: let us try to find a physical connection between cause and effect' (Salmon 1998: 16). Indeed, Salmon and later Dowe, Kistler and others claimed to have found a referent (or referents) within our best sciences that can fulfil that role: it is the transfer between processes of energy or momentum or charge or other conserved physical quantities that can serve as that phenomenon in the world. The (folk) platitude that some kind of connection between cause and effect has to exist is therefore vindicated: something that can play the role of causal connections is found and described by science.

How can a Canberra Planner reconcile Hume and Salmon? Well, the discordance between them can easily be resolved by pointing out that science, in Hume's time, was not able to find the world's counterpart of the connection platitude. Metaphysical progress, and this is a corollary that the Canberra Plan wholeheartedly welcomes, is connected to or following scientific progress. Some call this kind of metaphysics, which progresses together with scientific advances, 'Inductive Metaphysics'.

Critique of the Canberra Plan Have we therefore found a sound method for doing metaphysics of science? This would be too good to be true. As with IBEs or the definition of simplicity or other metaphysical principles we have introduced and critically evaluated, so it is with the Canberra Plan: where there is some vagueness there is room for interpretational freedom. We said that the search for platitude-realisers might be considered successful where something in the world is found that fulfils *some subset* of the entries on the platitudes list. However, how many items are enough? And what if two different subsets are fulfilled by different referents? What if the platitudes are from the start too diverse to promise successful satisfaction by one thing?

We stick to the concept of causation as our sample case and, with John Norton, turn to Russell's famous critique thereof. Again, we reconstruct their arguments in terms of the Canberra Plan. Note, first, that Norton believes Salmon's process view to be 'the most promising of all present views of causation', which 'seeks to answer most responsibly to the content of our mature sciences' (Norton 2003: 6), and yet he argues in his article 'Causation as folk science' (Norton 2003) against Salmon. Why? Well, he comes to the conclusion that, at least in the fundamental sciences, nothing can be found that answers to sufficiently many folk intuitions of causation and, thus, that there is ultimately no such thing as causation in the world.[24]

For his argumentation, Norton refers to the *locus classicus* for such causal nihilism: Russell's 'On the notion of cause' (Russell 1912).[25] There, Russell lists the following platitudes which he believes are commonly associated with causation:

1. **Sufficiency:** Causes are sufficient for their effects; that is, it is always the case that if the cause occurs, then the effect occurs (see Russell 1912: 7–12). [. . .]
2. **Time-asymmetry:** Causes precede their effects in time, but not vice versa (see Russell 1912: 13–16).
3. **Causal asymmetry:** The causal relation is asymmetric; that is, if A causes B, then B does not cause A (see Russell 1912: 10).
4. **Locality:** Cause and effect are local and distinct events in space-time region r, where r is 'something short of the whole state of the universe' (see Russell 1912: 7)

(I borrowed this concise list and references to Russell from Reutlinger 2013: 152)

However, Russell claims that nothing in physics satisfies these pre-theoretic folk ideas. We cannot go into the details of his argumentation, but here is a rough idea: the crucial point against (2) and (3) is that fundamental physical equations are (time) symmetric, and also that physical laws make no difference between the past and the future (see Russell 1912). In other words, nothing in the world of fundamental physics can account for the required asymmetries (2) and (3).

Russell's reasons against sufficiency (1), is that there is always the possibility of interfering factors so that even if the alleged cause is there the effect could be prevented from happening. Thus, an alleged cause (alone) is never sufficient. Russell's reason against locality (4) is related. You might say in reaction to the doubts about (1) that surely if you take the whole universe's stage at one point in time to be the cause of some future event then, because you have fixed everything there is to be fixed, there is nothing left that could interfere and spoil the effect. If that is the case, you might rescue sufficiency; however, with the whole universe on board, you then contradict locality as spelled out in (4) (compare Russell's to the argument against necessity in 6.4.3). Russell therefore came to the conclusion that:

> the word 'cause' is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable.
>
> (Russell 1912: 164)

Norton concurs:

> [A]t a fundamental level, there are no causes and effects in science and no overarching principle of causality.
>
> (Norton 2003: 22)

What is the upshot of this detour? Well, as you can easily verify, there is a significant overlap of Russell's list and Hume's (temporal asymmetry, contiguity) where both have left out connectedness. Remember that at the beginning of our critical evaluation of the Canberra Plan we asked how many platitudes need to be fulfilled and what there is to do if two different subsets are fulfilled by different candidates.

Here, we are in a situation where these questions become pressing: Salmon (Section 5.4) claims that one crucial platitude finds its relatum in the world – the 'secret' connection is to be identified with the transfer of conserved quantities – and yet Norton and Russell claim that none of the other platitudes are fulfilled. What shall we conclude? Is causation real? Is it the transfer of energy? Or should the concept of causation be dropped because of the emptiness of the other intuitions we might have? It seems hard to say which weighing of goods is the right one.

Dualism and pluralism Similarly (but avoiding either of the two options – nihilism or opting to put a lot of weight on one platitude), some philosophers came to believe that causation is a word not for one unified phenomenon but for at least two. In our chapter on causation we mentioned Ned Hall's 'Two Concepts of Causation', which argues that one set of our intuitions about causation characterises it as production;[26] the other set as (counterfactual) dependence (Hall 2004).

Others, like Stathis Psillos, opt for a pluralism and claim that there are more than just two concepts of causation (Psillos 2009).[27] Psillos argues that different kinds of causation at best family resemble each other, but there are no sufficient and necessary conditions for *the one* concept of causation (see Wittgenstein's (1921) notion of family resemblance concepts like *game* or *art*). Of course, dualism and pluralism presuppose that there are more platitudes in addition to those Russell saw unfulfilled that can be somehow satisfied (counterfactual dependence being a candidate).

We conclude that, just as with other metaphysical methods, principles or guidelines, the Canberra Plan is not a recipe that leads to unique results. Depending on which or how many platitudes you wish to save, you might succeed – or not – when trying to find a (partial) satisfier of the folk's intuitions. Causation has, of course, only been one example for metaphysical theories. Much the same holds true for theories of laws, disposition, natural kinds, etc.

Lewis on Causation and the Plan Before we turn to a kind of complementary approach to the Canberra Plan in Section 7.7, we wish to inform the reader of a curiosity: David Lewis, although being convinced that the Canberra Plan is a worthwhile and successful recipe when it comes to the metaphysics of the mental (see Lewis 1980), of colour (see Lewis 1997) or of value (see Lewis 1989), rejects it as a method of arriving at a theory of causation.[28]

Here's why Lewis believes that the Plan does not work for causation: what the folk calls causal is too gerrymandered and too widely varied to yield a single natural referent. In other words, the Canberra Plan is bound to fail from the start – before we even watch out for candidate role fulfillers – because the folk's intuitive platitudes are too wild a bunch.[29]

Now, the first reaction could be similar to Russell's and Norton's: the concept of causation has to be dropped from serious discourse; however, as we know, Lewis does not go down this route. On the contrary, because causation is such an entrenched concept in our everyday and scientific discourse, Lewis aims to arrive at a theory of causation that sidetracks the Canberra Plan. We learnt about his theory in Section 5.3 and do not want to reiterate it here, but instead we want to introduce a test criterion for its tenability (and that of other metaphysical theories) which is somewhat complementary to the Canberra Plan. This leads us to Section 7.7.

7.7 Extensional adequacy

In whatever way Lewis or some other metaphysician arrives at their theory of causation (or some other metaphysical notion), one crucial criterion for the theory's success is its *extensional adequacy* (see Section 5.6). A theory of causation is *completely extensionally adequate* when all and only those event sequences or processes that we intuitively consider to be causal count as such, according to the respective theory. Weaker, and more realistically, a theory is *mostly extensionally adequate* if there is a great overlap between intuitions and theory when it comes to what they judge to be causal.

Remember Suzy and Billy and other similar cases (double prevention, overdetermination, pre-emption, omission, etc.) that Lewis and Lewiseans discussed and the trouble they caused. Those were actually cases of mismatch: what the folk believe are clear cases of causation the theory says are not, or vice versa. In other words, these cases made it difficult for the theory to be considered extensionally adequate even if it covers many other cases.

Such a mismatch counts, prima facie, against a metaphysical theory because if the mismatch is too big the theory loses touch with what ordinary people and also scientists take causation to be. What such a theory defines, therefore, is at best a new technical concept, one that might be useful in its context, but one that has little to do with the concept we wished to elucidate in the first place.

Even if there are only negligible divergences between theory and intuition, ideally some additional reasons would be given for why we should accept the mismatch. However, how big the difference can be and precisely which kinds of cases one should allow is, as so often, a matter of metaphysical taste.

The complement to Canberra The extensional adequacy test for the tenability of metaphysical theories is somewhat complementary to the Canberra Plan. Here's why. The extensional adequacy test asks which *individual cases* the folk count, *unreflectedly*, as causal, i.e. it asks for the *extension* of the pre-theoretical folk concept. It then checks whether the philosophically reflected metaphysical theory under consideration – which aims to explicate a rational, tenable concept of causation – captures those cases.

The Canberra Plan is the other way round – it asks the folk for their intuitions for what causation is, i.e. it asks for what the folk believe the *intension* of their word 'causation' is. It then checks whether there is a possible *extension* in the world, as science describes it, which fulfils the intension's platitudes.

In other words, the extensional adequacy test says: here's the extension of the (folk) concept under concern, now find a metaphysical theory, i.e. the philosophically reflected intension for the concept, that captures these and only these cases. The Canberra Plan says that here is the intension of the (folk) concept under concern, now find a scientifically certified counterpart in the world that meets most of (or the most important of) the intuitions that are captured in the folk's intension.

For the case of causation, Lewis has chosen the extensional adequacy test to justify his theory instead of arriving at it through the Canberra Plan. With this remark, we end this section on extensional adequacy and also, almost, our overall meta-metaphysical considerations of potential methods and principles that help to arrive at metaphysical theories.

With Ockham's Razor, inferences to the best explanation, indispensability, serviceability, and transcendental arguments, Scientific Realism, the Canberra Plan and the extensional adequacy test, we still cannot claim to have given a complete list – in fact, we will add *intuitions* and *the argument from unknowability* to our metaphysical toolkit in the next section. We also confess to having only scratched the surface of those principles discussed. Ideally, a follow-up book *The Meta-Metaphysics of Science* would deal in depth with those issues.

However, two things might have become clear: first, there are many and often competing methods or principles for how metaphysics of science should be conducted and second, even where philosophers agree on their methods they still hardly ever lead them to the same conclusions. In fact, the whole book and the many first order metaphysical theories discussed (of dispositions, counterfactuals, laws, causation, etc.) attest to this.

There is one issue left that we need to have a close look at: *Hume's Dictum*. Apparently, an unexpected minority of only about 30 per cent of a representative group of professional philosophers believe in this metaphysical credo (at least with regards to laws of nature: see Bourget and Chalmers 2013: 28). Still, as we have seen throughout this book, it has fuelled debates within the metaphysics of science for nearly three hundred years.

BOX 7.2–7.7 Methods and principles of metaphysics of science

In this chapter on *Meta-metaphysics* we made explicit a handful of methods, principles and guidelines for how to conduct metaphysical research. Many of them we encountered throughout the book. Here's a list:

- **Ockham's Razor** and **inference to the best explanation** (IBE) are two principles that pull in different directions: **Ockham asks us to prefer the theory of two or more candidate theories that are required to postulate the existence of the least number of (kinds of) entities**, whereas **IBEs can be used to enrich our ontologies beyond what is in front of our eyes**: those entities (even if unobservable) that the best explanation of some phenomenon postulates shall be taken for granted.
- In the sciences, both procedural instructions have proven to be successful. Evidence for their empirical successes can be given for both Ockham and IBE. However, lacking this possibility of empirical confirmation **it is unclear whether, in metaphysics, these principles are acceptable.** Nonetheless, we saw that many metaphysicians are happy to invoke them.
- **Indispensability** or **transcendental arguments lay bare the necessary existence postulations of or other preconditions for some theory.** If the existence of perfectly natural properties is a precondition for a specific theory of laws to work, then if you want to endorse that theory, you need to believe in natural properties. However, why you should want to believe in *that* theory in the first place remains to be justified – maybe this can be done with a preceding IBE.
- Many metaphysicians of science see **metaphysics as a handmaiden of the sciences**. In particular, **Scientific Realism** is often taken for granted and one's theories of laws, causation, etc. must at least cohere with scientific findings or, better, be somehow extracted from science.
- It is difficult to say how this can be done in unambiguous ways. One programme that suggests a concrete recipe is the **Canberra Plan**: take **metaphysically interesting concepts of both our everyday life and pre-philosophical scientific thinking** (dispositions, causation, laws, for example), extract our intuitions about them and **compile a list of features they ought to fulfil**. Next, consult our best scientific theories and **check whether there are kinds of entities to be found in reality that meet these desiderata**. If so, let them be the referents of our pre-theoretic concepts, if not drop the concept from philosophical/scientific discourse (and maybe explain how we previously formed the respective error-laden ideas and made use of them). We took the conserved quantity

theory of causation as an example that can be seen as the result of Canberra Planning.

- **Adequacy regarding the intuitive extension**, a kind of complement to the Canberra Plan, is a reasonable demand unless there are intelligible reasons for (some) divergence.
- We will add **intuitions** and **the argument from unknowability** to our metaphysical toolkit in the next section on Hume's Dictum.

7.8 Why defend Hume's Dictum?

In Section 2.2 on the current metaphysics of dispositional properties we said that neo-Humean Categoricalists stand against anti-Humean Dispositionalists (see Section 2.2.4). The former believe that all dispositional properties are reducible to categorical properties, the latter think that a reduction is not possible. Categoricalists defend their anti-dispositions view because they believe in *Hume's Dictum*. Here's a version of it from Ludwig Wittgenstein, who, in his early *Tractatus* years, subscribed to views that were close to those of Logical Empiricism: 'A necessity for one thing to happen because another has happened does not exist' (Wittgenstein 1921: 6.37).

As we saw in Section 2.1, dispositions also bring necessity to the world because, if some properties were dispositional, some combined instantiations of properties would not be possible: *soluble* sugar cannot be both in water (in otherwise perfect trigger circumstances) and yet not dissolve, i.e. *it is impossible* for soluble sugar not to dissolve in those circumstances. So, if you do believe in Hume's Dictum – that nothing necessitates something else's occurring (or its absence) – then you cannot also believe in dispositional properties. Hume's Dictum obliges you to be a Categoricalist.

However, why trust Hume's Dictum in the first place? Why disbelieve in necessary connections?[30] This is the meta-metaphysical issue we are concerned with in this section.

***Non-observability* cannot be a reason** Remember from Section 2.2 that having made the transition from sense data to natural properties, from sentences and their verification to direct talk about the world, we have long been liberated from the Positivists' limitations of observability and the Verificationist principle of meaning. Jessica Wilson confirms:

> Contemporary proponents of HD [Hume's Dictum] don't accept Hume's strict Empiricist constraints; rather, they are typically happy to allow that we can justifiably believe in the existence of entities lying beyond the reach of experience (e.g., as a matter of inference to the best explanation).
> (Wilson 2013: 151)[31]

Why do so many philosophers still share Hume's and the Empiricists' scruples about necessity, causal necessitation, powers, etc. although they neither subscribe to Verificationism nor observability issues any longer?

Psychological assessments We will start with three reasons that are more of a psychological than of a philosophical nature. Potential real reasons of a rational, philosophical kind will be discussed later.

(i) Our first quasi-psychological explanation is that neo-Humeans project what Empiricists used to believe about sense perceptions onto nature: the world with its own properties behaves just as the flux of sense data presumably does. As the latter is a flow of mutually independent (if maybe regular) perceptions – one little sense datum then another – so is the world nothing but a mosaic of unconnected categorical quiddistic properties (see Section 2.2).

(ii) Digging slightly deeper, we might find that neo-Humeans simply associate necessary connections with bad metaphysics. Remember that Hume opposed the Rationalists like Descartes and Leibniz because he felt that their metaphysical speculations is idle talk, neither provable nor falsifiable: 'there are no ideas, which occur in metaphysics, *more obscure and uncertain*, than those of power, force, energy or necessary connexion' (Hume 1748: Sect. VII, Part I, §49: 60–1, emphasis added). Hume wished to discipline philosophy and liberate it from fabulating. For similar reasons, the Logical Empiricists targeted Hegel and Heidegger. Today, and justifiably so, philosophers still do not want to relapse into mere speculation; therefore, necessities, which were central to what seemed to be esoteric metaphysical storytelling back then, is still mistrusted today (for neo-Humeans, dispositions might have the taste of something like Leibnizian monads with their desire or will or appetite for particular ends). Necessary connections, dispositions, etc. are found to be 'weird or spooky or strange' (Maudlin 2007: 71).

(iii) The third psychological interpretation for why people cling to neo-Humeanism is given by both Maudlin and Wilson:

> [Hume's Dictum, MS] serves as a foundational assumption and methodological guide in a wide range of metaphysical debates, constitutive of a broadly Humean approach to metaphysical theorizing.
>
> (Wilson 2013: 151)

And

> the Humean project is very seductive: one is given a delimited set of resources and set the task of expressing truth conditions for some class of

> propositions in those terms. To win the game is to get the truth conditions to come out in a way that is, largely, intuitively correct. Proposed solutions can be counter-exampled, counter-examples can be reinterpreted, intuitions can be bartered off against each other. If a proposed analysis fails, there is always the hope that one more widget, one extra subordinate clause, can set things right again. No end of cleverness can be deployed both on offense and defense.
>
> (Maudlin 2007: 2. Think of Lewis on causation and the back and forth between analyses and counterexamples in Section 4.3.)

In other words, Humeanism, and especially Lewis's grand project of proving Humean Supervenience to be right, is a great normal science research programme[32] with a given set of assumptions and tools. Everyone can objectively check whether a move has been conducted correctly within that set of rules. All this gives neo-Humeanism the veneer of scientificity the Logical Empiricists also claimed for themselves.

Regardless of its truth or falsity, it is undeniably a very aesthetic philosophical edifice that is formulated in an admirably clear and concise way (as the readers of the Humean sections on counterfactual conditionals (see Section 3.4), laws (see Section 4.3) and causation (see Section 5.3) will agree). For those who enjoy good philosophical puzzles merely for the sake of puzzle solving (and maybe not so much because of their love of wisdom), Humean Supervenience offers a lot of material.

Philosophical reasons These were (not much to be trusted) explanations of a psychological layman. We must turn to proper rational, philosophical reasons for (or against) Hume's Dictum. However, note that:

> it is a curious fact that the proponents of the contemporary Humean programme – Lewis included – having abandoned the Empiricist theory of thought that underwrites Hume's rejection of necessary connections provide precious little by way of motivation for the view.
>
> (MacBride 2005: 127)

Intuitions One first possibility – an option that is only marginally better than the *prejudicial reasons* earlier – is to claim that Hume's Dictum is *true by intuitions*.[33] However, anti-Humeans will simply counter with their competing intuitions in favour of necessities. A stalemate results where it is not clear how to resolve it. Both Craig Callender and Barry Loewer warn that our:

> intuitions are historically conditioned and possibly unreliable and inconsistent.
>
> (Callender 2011: 44)[34]

> [W]e shouldn't take such intuitions all that seriously. [. . .] it is not plausible to think that our intuitions should be reliable guide to the fundamental nature of reality.
>
> (Loewer 2012: 132)

Loewer also gives a telling example where our pre-scientific intuitions go wrong:

> Most people have the intuitions that continued application of force is required to keep a body in motion and that the heavier the object, the faster it falls. Obviously these intuitions are misguided. Why should intuitions concerning laws [as example for a metaphysical theory, MS] be more reliable?
>
> (Loewer 1996: 116)

It seems we have to watch out for more reliable reasons for (or against) Hume's Dictum.

An epistemic argument: the threat of unknowability The *threat of unknowability* argument (which is our second philosophical argument, after *intuitions*) goes like this: what we can *in principle* know about the world through our senses is only some subset of all occurrent facts, and even if we knew them all we would still not know which ones are necessary. The necessary transcends the actual. Being truth in all possible worlds, not only in ours, it far exceeds the empirically knowable in principle. Humeans just have to add 'the further premise that what isn't knowable in principle *isn't* in principle' (Maudlin 2007: 71, emphasis added) and move swiftly on from epistemology to ontology – that what cannot be known in principle does not exist. Thus, we come to reject the idea that there is any necessity in the world.[35]

However, this Empiricism affine epistemic view is in danger of being committed to anti-Realisms about too many things. It cannot be easily restricted to necessity claims. In fact, the guideline 'what isn't knowable in principle *isn't* in principle' either (i) has to reconstruct the existence of other unobservable entities in hard to believe ways or (ii) it leads to a far reaching Scepticism about many other facts as well (those from the past, for example). In this vein, Tim Maudlin (2007: 74ff) launches the following two-step argument against this quasi-Empiricist anti-necessity stance:

(i) Socrates had a specific blood type but because medical science around 400 BCE was not advanced enough to test it, his blood type did not belong to the known facts. Therefore, there is no fact of the matter regarding his blood type or, more radically, his blood type does not exist. Because of this untenable consequence, the 'what isn't knowable in principle isn't in principle' guideline has to be dropped.

However, the defenders of the guideline will surely counter that, in order to avoid this bizarre consequence, there is the '*in principle*' proviso. Other than necessities, Socrates's blood type, even when *not in fact known*, is still *knowable in principle* and therefore we can rely on its existence.

However, according to Maudlin, this response is 'baroque' and 'self-defeating'. Here's his reasoning: *knowability in principle* refers not only to actual but also to the merely possible evidence, i.e. Socrates's blood type *could have been known, had it been tested.* Since the latter counterfactual is true, according to the proponent of the *knowability in principle principle*, his blood type exists. However,

> this has everything exactly backwards: we think that there is a determinate (but unknown) fact about how such tests would have come out exactly because we think there is a determinate (but unknown) fact about what the blood type was.
>
> (Maudlin 2007: 74)

In other words, we believe that we could have known the blood type because it does exits, but not that, because we could in principle know it, we are justified to say it exists. This rescue attempt of the existence of some unobservables is hard to believe.

(ii) If the defenders of the guideline choose the radical way instead and do not try to rely on the 'in principle' proviso, then this raises immediate awkward, anti-Realist consequences in our ontology. In addition to past events (Socrates's blood type), other non-observable entities would also be in danger of non-existence: 'Quarks and their flavors, as well as wavefunctions, are neither the data of experience nor constructs from them' (Maudlin 2007: 76, see also 75). In this case, the 'what isn't knowable (in principle) isn't in principle' argument against necessity turns out to be a slippery slope. It might, in fact, make us slide into an external world Scepticism or, at least, to an Instrumentalism.

Neither of (i) and (ii) is attractive. The *what isn't knowable in principle isn't in principle* guideline, which was meant to be used as an epistemic reason against the belief in necessities, might have too many untenable consequences. Maudlin therefore concludes that no anti-necessity view should be based on such sceptical epistemological foundations (see Maudlin 2007: 76).[36]

Methodological justifications I: Ockham's Razor Maybe there are *methodological justifications* (instead of epistemic ones) in favour of Humeanism. One such method would be to listen to Ockham's advice (see Section 7.2): fundamental necessities in nature would be an additional metaphysical ingredient over and above the factual and, should it be possible to base all our metaphysics on the Humean mosaic alone (see Sections 2.2.4 and 2.2.5) i.e. should it be possible to cater for causation, laws, etc. without necessities, then we had better adopt this sparser metaphysics without supplementary modal facts.

This reasoning presupposes, first, that Ockham's Razor is indeed a good methodological advice in metaphysical circumstances (see Section 7.2) and, second, that we can indeed dispense with necessities. The whole history of the metaphysics of science is a testament that the latter is an unresolved issue, and whether one follows the methodological advice of Ockham at all might well be a matter of taste, namely whether one prefers desert landscapes or baroque ornamentation, and whether one or the other aesthetic preference traces what nature is really like is hard to tell.

Methodological justifications II: IBEs Jessica Wilson discusses a further possible methodological argument to those we targeted in Sections 7.2 to 7.6: maybe a version of *inferences to the best explanation* (IBEs) works in favour of Hume's Dictum. Hume's Dictum, according to Wilson, 'serves as the cornerstone for a fairly comprehensive framework of philosophical theses and positions' (Wilson 2010: 634–5) and if one could establish 'that such a cornerstone is *the best way* of systematising philosophical theory' (Wilson 2010: 634–5, emphasis added) one would have reason to believe it.

Such a kind of *inference to the best explanation* for Hume's Dictum is given by Jonathan Schaffer (2005). He claims that the best account of counterfactuals requires that there are no necessary connections, i.e. the (allegedly) best theory of counterfactual makes the truth of Hume's Dictum indispensable.[37] Here's why he thinks so:

Suppose you support the counterfactual 'had the electron not been charged then it would have moved according to the law of gravitation alone'. Now, remember from (Section 3.4) how one of the dominant theories of counterfactuals suggests we evaluate this counterfactual: we look at the nearest possible world where the antecedent is true and check whether the electron has indeed moved by gravitation alone. However, this presupposes that it is possible for electrons not to be charged. We need the possibility of free recombinations of the two properties (being an electron/being charged), otherwise we cannot check possible worlds where they come apart because there would not be such worlds.[38] In other words, Schaffer is right in at least this respect: for *that* theory of how we evaluate counterfactuals, there mustn't be necessary connections between properties.[39]

So far, this is only a conditional claim: *if* you subscribe to that theory, then you need Hume's Dictum. But why should you subscribe to it?

Schaffer additionally claims that *that theory* is *the best theory* for counterfactuals. Actually, prima facie, this is not such an unbelievable assumption: remember from Sections 6.4.1 and 6.4.2 how hard it was for Ellis or Bird to account for counterfactuals. It looks then as if Schaffer is right and his argument is an inference to the best explanation for Hume's Dictum.

Despite this, at least two things are unclear: first, counterfactuals are not everything metaphysicians think about. A metaphysicist must also cater for laws, causation and properties and so on. Is the anti-necessities view the best for all these areas? Is it, as Wilson put it, a cornerstone for *a number of philosophical theses* and positions? This is important because the case of counterfactuals clearly cannot be decisive in isolation, even if, on its own, it turns out to be better off without necessities. The whole metaphysical package has to be considered – but things start to become fishy. In fact, the unavailability of an overall evaluation is what fuels the ongoing debates between Humeans and anti-Humeans which we have followed throughout this book.

Second, even the isolated case of counterfactuals might not be so clear after all. We have seen that Ellis suggests to turn from truth conditions to assertability conditions in order to be able to simulate free recombinations of properties (see Section 6.4.1) and how we could resolve the issue with a surrogate possibility: the possible existence of alien properties like *being a schmellectron* (see Section 6.4.2). These and other 'alternative frameworks based in the denial of HD [Hume's Dictum] would have to be shown not to have similar advantages of fruitfulness and systematicity' (Wilson 2010: 634–5).

Although it has to be acknowledged that Schaffer's argument does score points for Hume's Dictum, the overall balance still has to be calculated.

Methodological justifications III: Physicalism We turn to a final potential methodological justification for Hume's Dictum. When paired with Humean Supervenience claims (see Section 2.2.5), it is often perceived of as a promising tool or research programme to formulate and defend the thesis of *Physicalism*.[40] Humean Supervenience is, as Lewis writes, a 'metaphysics built to endorse the truth and descriptive completeness of physics more or less as we know it' (Lewis 1986: x). We might therefore be able to launch a serviceability argument (see Section 7.4) for Hume's Dictum when combined with Humean Supervenience: both together are serviceable to defend Physicalism.

Remember, first, that the supervenience claim here says that the complete (macro) state of the world – including both macro properties and macro objects, as well as the truth of counterfactual conditionals and other modal facts, facts about laws, causation, objects, minds, etc. – supervenes on the distribution of the *local*, intrinsic, fundamental (micro) properties of each space-time point. When supervenience is combined with Hume's Dictum, these properties are thought of as being categorical: one property instantiated here, unconnected to another property instantiated there. If these fundamental properties are physical, the thesis that ultimately everything is physical, and thus the theses of Physicalism, is indeed supported.

However, as we have pointed out already in Sections 2.2.3 to 2.2.5, there is no obstacle for the anti-Humean to also be a Physicalist or materialist. The anti-Humeans will only disagree in so far as they believe that some (or all) of the fundamental physical properties in the mosaic are dispositional or otherwise non-categorical. Other than that, they can endorse this supervenience claim as easily as the Humeans, and therefore vote for the reducibility of everything else to that physical basis.

For that reason, we can see that Hume's Dictum combined with Humean Supervenience is at best *serviceable* for Physicalism but not *indispensable*. The third potential methodological reason for Hume's Dictum would therefore not be a strong argument, if it were a tenable argument *at all* (see Section 7.4). We will now discuss why it probably isn't.

When we 'look to scientific practice rather than to philosophical prejudice' (Maudlin 2007: 1), Humean Supervenience, in its original form, probably does not accord with present day physics and therefore, for strong reasons, it can't be a metaphysics which is serviceable to 'the truth and descriptive completeness of physics' in the first place.

The findings in fundamental physics which contradict elements of Humean Supervenience come from quantum mechanics, particularly those regarding 'non-separability' and 'entanglement'. Barry Loewer gives a concise outline of why there is a major friction:

> It seems pretty clear that contemporary physics does dream of non-Humean properties. I have in mind so called 'entangled states' that are responsible for quantum nonlocality [. . .]. The entangled state of a pair of particles fails to supervene on the intrinsic properties of the separate particles. That is, the local properties of each particle separately do not determine the full quantum state and, specifically, do not determine how the evolutions of the particles are linked. Since we have reason to believe that quantum theory is true, we have reason to think that HS [Humean Supervenience, MS] is false.
>
> (Loewer 1996: 103–4)[41]

The proponent of Humean Supervenience must therefore give up the idea of local, separated, intrinsic categorical properties being the sole supervenience base that constitutes the fundament; non-local, unseparated, entangled affairs also have to be taken on board. In light of current physics, Hume's Dictum and Humean Supervenience fail in their original form.[42]

We won't go into further details of quantum mechanics here; instead we wish to mention quickly that in addition to these lessons from scientific *theory*, scientific *practice* might also speak against neo-Humeanism. Maudlin claims that in scientific practice, researchers

> seek laws, announce laws, and use laws but they do not even attempt to analyze them in terms of the total physical state of the universe or anything else.
>
> (Maudlin 2007: 67)

In other words, in scientific practice, from the moment that laws have been (epistemically) discovered, they are treated (ontologically) as basic, i.e. they are immediately given a place within the fundamental physical state of the universe (see Section 4.6) where they are treated as parts of the supervenience base and not something that itself supervenes thereupon.[43] Both current scientific theory and its practice might therefore speak against Humean Supervenience.

In summary, it is unclear whether Hume's Dictum is a metaphysics to defend Physicalism and, therefore, the third potential methodological justification for Hume's Dictum – its serviceability for Physicalism – also proves to be questionable. This ends our pros and cons of methodological and other philosophical arguments regarding Hume's Dictum (plus Humean Supervenience). In fact, this ends our meta-metaphysics chapter.

BOX 7.8 Why still believe in Hume's Dictum?

- Once the step from **Empiricist Phenomenalism** to a **Realist properties ontology** has been made, there is, prima facie, little reason to stick to an anti-connections metaphysics. The neo-Humeans' **no necessary connections between distinct entities** demand is questionable.
- In this section we gave three **psychological reasons** why some philosophers might uphold Hume's Dictum. We also discussed five **potential philosophical arguments** in favour of it and gave **theoretical and practical scientific reasons** to be sceptical of Humean Supervenience.

- The **first psychological reason** was that neo-Humeans might **project what Empiricists used to believe about perceptions onto nature**: the world with its properties behaves just as, presumably, the flux of sense data. The **second** was that **necessary connections might be associated with bad metaphysics**, like spooky or strange Leibnizian monads. **Third**, **Humean Supervenience is indeed an attractive normal science research programme** with transparent, fixed rules and clear standards for success.
- In relation to **philosophical arguments**, we mentioned, first, that Hume's Dictum could be **true by intuitions** but dismissed this argument as having little weight because the anti-Humeans could evoke counter intuitions.
- Second, we discussed **the argument from unknowability** which moved from epistemology (that necessities can in principle not be known) to ontology (that they do not exist). We pointed out that this argument might prove too much because it is hard not to slip into more radical Scepticisms in addition to the rejection of necessity.
- Third, we discussed **methodological justifications**: **Ockham's Razor**, an **indirect inference to the best explanation**, and the claim that **Humean Supervenience is the best research programme to defend Physicalism**. We presented **two scientific reasons** against the latter: a theoretical one coming from quantum mechanics, and a second one coming from scientific practice.

7.9 Polemical remarks

We cannot help but finish with some polemical remarks about both the Humean and the anti-Humean camps.

The neo-Humeans' biggest problem is that their demands may start to look unconvincing: once the step from Empiricist Phenomenalism to a Realist properties ontology has been made (and Humeans do make it, as we have seen in Lewis's philosophy) there seems to be little reason (indeed, it seems to be contrived) to stick to an anti-connections in nature metaphysics.

However, the neo-Humeans biggest advantage is that they have an ethos: their philosophy comes with clearly set out presuppositions and fixed constraints in order that their research program can be conducted following these parameters. Everyone can objectively check whether a move within that game has been made correctly or is against the rules.

The latter is a virtue some anti-Humean programmes occasionally lack and, without any constraints on their epistemology, they might fall victim to speculative metaphysical system building. However, their advantage might be an intuitively more credible, more commonsensical metaphysics.

Here's a notable meta-observation that Chris Swoyer and Francesco Orilia call the 'fundamental ontological tradeoff'. It describes, spot on, the development of the past century's metaphysics (in their terms 'ontology') of science:

> The fundamental ontological tradeoff reflects the perennial tension between explanatory power and epistemic risk, between a rich, lavish ontology that promises to explain a great deal and a more modest ontology that promises epistemological security. The more machinery we postulate, the more we might hope to explain – but the harder it is to believe in the existence of all the machinery.
>
> (Swoyer and Orilia 2011)

Notes

1 Ted Sider arrives in similar ways at similar questions: 'Worries about contemporary ontology begin as worries about its epistemology [...] The main ontological positions seem internally consistent and empirically adequate, so all the weight of theory-choice falls on the criteria; but are the criteria up to the task? What justifies the alleged theoretical insights? Are criteria that are commonly used in scientific theory (for example, simplicity and theoretical integration) applicable in metaphysics? How can these criteria be articulated clearly? And what hope is there that the criteria will yield a determinate verdict, given the paucity of empirical evidence?' (Sider 2009: 385).

Note that our title for this section, 'Simplicity, strength and their balance as a tool to judge metaphysics', is a homage to Lewis's best system idea for laws of nature. Here we use the idea on a meta-level for a metaphysical best system, so to speak.

2 Hamilton is known to physicists for his reformulation of Newtonian mechanics, namely the 'Hamiltonian' reformulation. For a historical discussion of Ockham's Razor see Thorburn (1918: 347). We met Ockham in Section 4.3 as a proponent of Nominalism.

3 Tim Maudlin, for example, discusses Ockham's razor in detail (see Maudlin 2007: 3–4, 179–81) and points out that there is no rational reason to believe in it (Maudlin 2007: 179). He prefers his ontology '*mit Schlag*', meaning with cream (Maudlin 2007: 4).

4 Here, Lewis is considering the pros and cons of his possible worlds modal realism.

5 IBEs are also called 'Abduction' in reference to Charles Sanders Peirce, who had already discussed species of this kind of argument (Peirce 1878).

6 For more on IBEs, see Peter Lipton (1991), especially pages 70–4 on philosophical IBEs. A shorter assessment can be found in Schurz (2008).

7 There might also be IBEs in the sciences for empirically undetectable entities, superstrings or Everettian possible world. Or shall we say that, at this point, science turns into metaphysics? Some philosophers see such a methodological continuity from science to metaphysics and think that metaphysics is just a little further into the unobservable abstract realm than theoretical physics.

8 See also Hawley (2006) and endnote 19 where we discuss Scientific Realism.

9 For a criticism of IBEs in metaphysics and the alleged continuity between metaphysics and science, see Saatsi (forthcoming) who also nicely reconstructs the arguments of the proponents of metaphysical IBEs. See also Stephen Biggs's (2011) paper 'Abduction and modality' in which he argues that IBEs provide a basis for modal metaphysical deliberation.

10 There are famous indispensability arguments not only in metaphysics but also in the philosophy of mathematics. Both Quine (1948/53, 1976) and Putnam (1979a, 1979b, 2012) argue that the indispensability of mathematical entities like numbers or sets for the empirical sciences entitles us to the belief in their existence.

11 A caveat aside, we are only interested in unearthing the general structure of the argument but not in filling in the details of why the postulations are necessary nor in evaluating whether the argument is watertight.

12 Lewis not only calls natural properties 'serviceable' but also his assumption of the reality of possible worlds: this 'hypothesis is *serviceable*, and that is a reason to think it is true' (Lewis 1986: 3, emphasis added).

13 Indispensability, especially of existence postulations, does not always have to be a part of IBEs. It can be freestanding, without reference to explanations.

14 We struggled earlier to categorise what indispensability arguments are if not *synthetic a priori*. Hosted within IBEs, indispensability claims are indeed arrived at by pure thought, in an *a priori* fashion. Are they synthetic or analytic? They say of a certain theory (of laws, for example) that it cannot deliver what we want it to deliver unless we make certain postulations. So, as long as they claim something about a theory (and not the world), we remain on analytic grounds.

15 Paul (2012) argues for more autonomy of metaphysics to science. She is, of course, playing here with the medieval '*philosophia ancilla theologiae*' (Latin for 'philosophy is the handmaiden to theology') claim.

16 Below, (Section 7.7), we will see an example of how Maudlin arrives at metaphysical conclusions from the theory and practice of science, and in Section 4.7 we already encountered some of his meta-metaphysical considerations.

17 Here, the move from austere Nominalism (as the view that any set of things forms a property) to the belief in natural properties (see Section 2.1, 3) can count as a conversion from Empiricist Phenomenalism to scientific Realism.

18 Psillos has in mind Kuhn's and Feyerabend's incommensurability thesis (see Kuhn 1962; Feyerabend 1975) which, in its most radical form, says that subsequent scientific theories cannot be compared to each other if their technical terms differ in such a way that a translation from one to the other is not possible, and when they cannot be compared then it is a stronger reason why it is not possible to say that the latter theory has made progress compared to the former. For more on this matter, see Sections 6.3.1 and 6.3.2, which also reflects the opposing views mentioned in the *Semantic Thesis*.

19 In an interesting way, Katherine Hawley links the belief in Scientific Realism to the use of IBEs within metaphysics (see endnote 8). She first points out that we arrive at scientific Realism through an IBE: 'Scientific Realism, as I shall understand it here, is the view that there are cases where the involvement of a claim about an unobservable entity in an empirically successful scientific theory provides reason to think that the claim is true' (Hawley 2006: 458). Second, she argues that someone who accepts this IBE, the conclusion of which is the affirmation of Scientific Realism, should also accept further IBEs within metaphysics of science: 'Scepticism [about metaphysics, MS] is typically based upon rejection of the inferential methods of metaphysicians, and rejection of inference to the best explanation in particular. As such, [such a Scepticism, MS] is incompatible with standard versions of Scientific Realism' (Hawley 2006: 454).

20 Never mind the details of Ramsey sentences, just consider them to be the conjunctions of all the collected platitudes and intuitions we alluded to earlier.

21 For more precise statements of the Plan and more detailed analyses see Nolan (1996), and all the authors in Braddon-Mitchell and Nola (eds.) (2009; especially page 384), as well as a review of the latter by Haukioja (2009: 1). More recently, Douglas Kutach has, under the unspecific name of 'empirical analysis', developed the Canberra Plan further: 'The *empirical analysis* of *X* is the engineering of a conceptual framework optimized in the service of the scientific explanation of whatever empirical phenomena motivate our possession of a concept of *X*, especially insofar as they are characterized in terms of experiments.' (Kutach 2013: 10). A novelty is that Kutach focuses on how and why creatures like us came to possess those concepts that we now wish to give a precise philosophical analysis of. Kutach is especially interested in the business of explicating what causation is but his overall account is also a valuable recommendation for how science can be used to guide metaphysics.

22 In fact, we have already mentioned in Section 5.4 that this could be done. It has to be said, though, that none of the theories we now discuss were developed with the Canberra Plan in mind. This, here, is our reconstruction.

23 Of course, Canberra Planning and Hume's argument vary in that Hume said there is *no sense impression* that would correspond to a causal connection where the Canberra Plan might claim that *science does not postulate it*.

24 To be fair to Norton, we must confess that what he says is more cautious: '[T]he concepts of cause and effect are not the fundamental concepts of our science and [. . .] science is not governed by a law or principle of causality. This is not to say that causal talk is meaningless or useless – far from it. Such talk remains a most helpful way of conceiving the world [. . .]. What I do deny is that the task of science is to find the particular expressions of some fundamental causal principle in the domain of each of the sciences' (Norton 2003: 1).

25 For detailed discussions thereof, revisit Russell's Republic in Price and Corry (2007).

26 Hall's notion of production is not exactly ours, (PROD), in Section 2.1.3, but it is near enough to make this point here. Counterfactual dependence clearly resembles (MOD).

27 On pluralism, see the special issue of *Philosophica* (De Vreese 2006).

28 I rely in this section on David Liebesman's 'Causation and the Canberra Plan' (2011).

29 See Liebesman (2011: section 2). There, see also Section 3 for another reason it fails according to Lewis.

30 Neo-Humeans extend their denial of necessary connections to entities of any ontological category: events, properties, kinds, etc. and also to any 'sort of connection (of the appropriate modal strength [. . .]) between such entities' (Wilson 2010: 597; in her paper: 'What is Hume's Dictum, and Why Believe It?'). Neo-Humeans would therefore, for example, not only deny that causes necessitate their effects (something dynamic, synchronic through time (see Sections 5.2 and 5.3)) but probably also that gold necessarily has atomic number 79 or that water consists necessarily of H_2O (a static, atemporal (constitutive) relation, see Section 6.3). There is a caveat which we mentioned in Section 6.3: if Hume's Dictum is restricted to *distinct* existences, then cases like water being necessarily identical to H_2O do not fall under its scope.

31 Also remember Tim Maudlin's statement to the same effect (Section 2.2.4). Maudlin dedicates a whole chapter, "'Why be Humean?" (Chapter 2: 50–77), of his book *The Metaphysics within Physics* to the question of neo-Humeanism.

32 This is a term used in the epistemology of science regarding issues of theory and theory change. It was established by Thomas Kuhn and stands for periods of science where the basic assumptions are untouched and research is conducted within a fixed set of paradigms (Kuhn 1962: especially chapter IV: 'Normal Science as Puzzle Solving; Section 6.3.2).

33 The label 'prejudicial' is borrowed from Maudlin. He divides the space of reasons pro and contra Humeanism into four areas: prejudicial, semantic, epistemological and methodological (see Maudlin 2007: 71). We'll turn to all types of reason in what follows. For a more detailed discussion of *intuitions* and relevant quotes see Wilson (2010: 619ff). Wilson also discusses and rejects the possibility that Hume's Dictum could be *analytically true*. A. J. Ayer has also put forward such an argument (see Wilson 2010: 607ff).

34 Callender speaks here specifically of our 'murky' modal intuitions.

35 Note that this argument bypasses the Logical Empiricist route from epistemology through Verificationst semantics to the meaninglessness of necessity claims. More precisely, this direct route from epistemology to ontology (without semantics) does not arrive at a meaninglessness claim but at a non-existence claim. (For neo-Humeans, necessity claims do make sense! It is only the existence of necessary connections that is in doubt).

36 Maudlin's original argument is targeted at the Neo-Empiricist line of argument in Earman (1984: 195). Later, Earman together with Roberts, refined their argument (Earman and Roberts 2005) in that they try to put a wedge between the no-necessities conclusion and the more radical Scepticisms about all sorts of objects. They wish to avoid the threat of a slippery slope (Earman and Roberts 2005: 279–82).

37 For the relation between indispensability and IBE arguments, see Section 7.3.

38 We have changed Schaffer's own example and his argumentation slightly (Schaffer 2005). For a critique of his argument, see Wilson (2015).

39 As we know from Section 3.4, the counterfactual would then be trivially true, which would be acceptable for the case we consider. However, the counterfactual 'had the electron not been charged it would have turned into a pink elephant' would also be true. This is unacceptable.

40 Maudlin, whose arguments we will discuss here, distinguishes a 'separability' thesis (more or less our *Hume's Dictum*) and 'physical statism' (more or less our *supervenience*). The slight differences are not important here (see Maudlin 2007: 51).

41 Maudlin (2007: 53–64) offers a detailed analysis of these findings..

42 To be fair to Lewis, we must add that he was well aware of this fact; however, he remained optimistic: 'Therefore if I defend the *philosophical* tenability of Humean Supervenience, that defence can doubtlessly be adapted to whatever better supervenience thesis may emerge from better physics' (Lewis 1999: 226).

43 Note that this argument from scientific practice might not only be directed against supervenience (as done here), but also directly against Hume's Dictum. Andreas Hüttemann (2014), who supports Cartwright's reasoning (see Sections 6.2.3 and 7.3), adds further convincing reasons from scientific and/or technological practice for necessary connections and, thus, against Hume's Dictum.

Bibliography

Anderson, F. and Becker Arenhart, J. R. (2016) Metaphysics within Science: against Radical Naturalism. *Metaphilosophy*. 47. pp. 159–80.

Armstrong, D. M. (1983) *What is a Law of Nature?* Cambridge: Cambridge University Press.

Beebee, H. (2009) Review of John Foster's *The Divine Lawmaker*. *British Journal for the Philosophy of Science*. 60 (2). pp. 453–57.

Bennett, K. (2015) There is no special problem with metaphysics. Philos Stud. 173. pp. 21–37.

Biggs, S. (2011) Abduction and Modality. *Philosophy and Phenomenological Research*. 83 (2). pp. 283–326.

Bird, A. (2007) *Nature's Metaphysics*. Oxford: Clarendon Press.

Bourget, D. and Chalmers, D. (2013) What Do Philosophers Believe? *Philosophical Studies*. 3. pp. 1–36.

Braddon-Mitchell, D. and Nola, R. (eds.) (2009) *Conceptual Analysis and Philosophical Naturalism*. Cambridge, MA: MIT Press.

Brandom, R. (2012) Metaphilosophical Reflections on the Idea of Metaphysics. *Philosophia*. 40 (1). pp. 13–26.

Callender, C. (2011) Philosophy of Science and Metaphysics. In: French, S. and Saatsi, J. (eds.) *Continuum Companion to the Philosophy of Science*. London: Continuum. pp. 33–54.

Carnap, R. (1956) *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd, enlarged edition. Chicago, IL: University of Chicago Press.

Chakravartty, A. (2010) *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge: Cambridge University Press.

Chakravartty, A. (2013) Realism in the Desert and in the Jungle: Reply to French Ghins, and Psillos. *Erkenntnis*. 78. pp. 39–58.

Chalmers, D., Wasserman, R. and Manley, D. (eds.) (2009) *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

De Vreese, L. (2006) Pluralism in the Philosophy of Causation: Desideratum or Not? *Philosophica*. 77 (1). pp. 5–13.
Earman, J. (1984) Laws of Nature: the Empiricist challenge. In: Bogdan, R. J. (ed.) *D. M. Armstrong*. Dordrecht: Reidel. pp. 191–223.
Earman, J. and Roberts, J. (2005) Contact with the Nomic. A Challenge for Deniers of Humean Supervenience about Laws of Nature. Part II: The Epistemological Argument for Humean Supervenience. *Philosophy and Phenomenological Research*. 71 (2). pp. 253–86.
Ellis, B. (2001) *Scientific Essentialism*. Cambridge: Cambridge University Press.
Feyerabend, P. (1975) *Against Method*. London: New Left Books.
Godfrey-Smith, P. (2012) Metaphysics and the Philosophical Imagination. *Philosophical Studies*. 160. pp. 97–113.
Hall, N. (2004) Two Concepts of Causation. In: Collins, J., Hall, N. and Paul, L. A. (eds.) *Causation and Counterfactuals*. Cambridge, MA: MIT Press. pp. 225–76.
Hamilton, W. (1853) *Discussions on Philosophy and Literature, Education and University Reform*. MacLachlan and Stewart: Edinburgh.
Haukioja, J. (2009) Review of David Braddon-Mitchell and Robert Nola (eds.): *Conceptual Analysis and Philosophical Naturalism*. Notre Dame Philosophical Reviews (online). Available from: http://ndpr.nd.edu/news/24124-conceptual-analysis-and-philosophical-naturalism/ (accessed 6 March 2015).
Hawley, K. (2006) Science as a Guide to Metaphysics? *Synthese*. 149 (37). pp. 451–70.
Hume, D. (1748) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. 17th impression of the 3rd OUP edition. Selby-Bigge, L. A. (ed.). Oxford: Clarendon Press.
Hüttemann, A. (2014) Scientific Practice and Necessary Connections. *Theoria*. 29 (1). pp. 29–39.
Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
Kutach, D. (2013) *Causation and Its Basis in Fundamental Physics*. Oxford University Press.
Ladyman, J. (2007) Does Physics Answer Metaphyscial Questions? *Royal Institute of Philosophy Supplements*. 61. pp. 179–201.
Ladyman, J. and Ross, D. (2007) *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
Ladyman, J. (2012) Science, Metaphysics and Method. *Philosophical Studies*. 160. pp. 31–51.
Leitgeb, H. (2011) New Life for Carnap's Aufbau? *Synthese*. 180. pp. 265–99.
Leitgeb, H. (2013) Scientific Philosophy, Mathematical Philosophy, and All That. *Metaphilosophy*. 44. pp. 267–75.
Lewis, D. (1980) Mad Pain and Martian Pain. In: Block, N. (ed.) *Readings in Philosophy of Psychology*, *Vol. 1*. Cambridge, MA: Harvard University Press. pp. 216–22.
Lewis, D. (1986) *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.

Lewis, D. (1989). Dispositional Theories of Value. *Proceedings of the Aristotelian Society Supplement*. 63. pp. 113–37.
Lewis, D. (1997) Finkish Dispositions. In: Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press. pp. 133–51.
Lewis, D. (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
Liebesman, D. (2011) Causation and the Canberra Plan. *Pacific Philosophical Quarterly*. 92 (2). pp. 232–42.
Lipton, P. (1991) *Inference to the Best Explanation*. London and New York: Routledge.
Loewer, B. (1996) Humean Supervenience. *Philosophical Topics*. 24 (1). pp. 101–27.
Loewer, B. (2012) Two Accounts of Laws and Time. *Philosophical Studies*. 160 (1). pp. 115–37.
MacBride, F. (2005) Lewis's Animadversions on the Truthmaker Principle. In: Beebee, H. and Dodd, J. *Truthmakers: The Contemporary Debate*. Oxford: Oxford University Press. pp. 117–40.
MacLaurin, J. and Dyke, H. (2012) What Is Analytic Metaphysics for? *Australasian Journal of Philosophy*. 90. pp. 291–306.
Maudlin, T. (2007) *The Metaphysics within Physics*. Oxford: Oxford University Press.
Meixner, U. (1999) Die Metaphysik von Ereignis und Substanz. Eine Skizze der Grundgedanken. In: Kanzian, C. and Siebenrock, R. (eds.) *Gottesentdeckungen*. Wien: Thaur. pp. 127–43.
Ney, A. (2012) Neo-Positivist Metaphysics. *Philosophical Studies*. 160. pp. 53–78.
Nietzsche, F. (2003) *Thus Spoke Zarathustra*. London: Penguin Books.
Nolan, D. (1996) *The Credo of the Canberra Planners* (online). Available from: http://sites.google.com/site/professordanielnolan/credo (accessed 10 March 2015).
Nolan, D. (forthcoming) *Method in Analytic Metaphysics*. In: Cappele, H., Gendler, T. S. and Hawthorne, J. (eds.) The Oxford Handbook of Philosophical Methodology. Oxford: Oxford University Press.
Norton, J. D. (2003) Causation as Folk Science. *Philosophers' Imprint*. 3 (4). pp. 1–22.
Paul, L. 2012 Metaphysics as Modeling: the Handmaiden's Tale. *Philosophical Studies* 160 (1). 1–29.
Peirce, C. S. (1878) Deduction, Induction, and Hypothesis. In: Hartshorne, C. and Weiss, P. (eds.) (1931–1935) *Collected Papers*. Cambridge, MA: Harvard University Press. pp. 2619–44.
Price, H. (2009) *Metaphysics after Carnap: The Ghost Who Walks?* In: Chalmers, D., Wasserman, R. and Manley, D. (eds.) Metametaphysics. New Essays on the Foundations of Ontology. Oxford: Oxford University Press. pp. 320–46.
Price, H. and Corry, R. (eds.) (2007) *Causation, Physics, and the Constitution of Reality. Russell's Republic Revisited*. Oxford: Clarendon Press.
Psillos, S. (1999) *Scientific Realism. How Science Tracks Truth*. London: Routledge.

Psillos, S. (2007) *Philosophy of Science A–Z*. Edinburgh: Edinburgh University Press.

Psillos, S. (2009) Causal Pluralism. In: Vanderbeeken, R. and D'Hooghe, B. (eds.) *Worldviews, Science and Us: Studies of Analytical Metaphysics: A Selection of Topics From a Methodological Perspective*. Singapore: World Scientific Publishers. pp. 131–51.

Putnam, H. (1979a) What is Mathematical Truth. In *Mathematics Matter and Method: Philosophical Papers, Vol. 1*. 2nd edition. Cambridge: Cambridge University Press, pp. 60–78.

Putnam, H. (1979b) Philosophy of Logic. In: *Mathematics Matter and Method: Philosophical Papers, Vol. 1*. 2nd edition. Cambridge: Cambridge University Press, pp. 323–27.

Putnam, H. (2012) Indispensability Arguments in the Philosophy of Mathematics. In: H. Putnam, *Philosophy in an Age of Science: Physics, Mathematics and Skepticism*. Cambridge, MA: Harvard University press, Ch. 9.

Quine, W. V. O. (1948/1953) On What There Is. In: Quine, W. V. O. (ed.) *From a Logical Point of View*. Cambridge, MA: Harvard University Press. pp. 1–19.

Quine, W. V. O., (1976) Carnap and Logical Truth, reprinted in *The Ways of Paradox and Other Essays*. Revised edition. Cambridge MA: Harvard University Press, pp. 107–32.

Reutlinger, A. (2013) *A Theory of Causation in the Social and Biological Sciences*. New York: Palgrave Macmillan.

Robus, O. M. (2015) Does Science License Metaphysics? *Philosophy of Science*. 82. pp. 845–55.

Ross, D., Ladyman, J. and Kincaid, H. (eds.) (2013) *Scientific Metaphysics*. Oxford: Oxford University Press.

Russell, B. (1912) On the Notion of Cause. In: Mumford, S. (ed.) (2003) *Russell on Metaphysics. Selections from the Writings of Bertrand Russell*. London: Routledge. pp. 163–82.

Saatsi, J. (forthcoming) Explanation and Explanationism in Science and Metaphysics. In: Slater, M. and Yudell, Z. (eds.) *Metaphysics and the Philosophy of Science*. Oxford: Oxford University Press.

Salmon, W. (1998) *Causality and Explanation*. Oxford: Oxford University Press.

Schaffer, J. (2005) Quiddistic Knowledge. *Philosophical Studies*. 123. pp. 1–32.

Schurz, G. (2008) Patterns of Abduction. *Synthese*. 164 (2). pp. 201–34.

Sider, T. (2009) Ontological Realism. In: Chalmers, D., Manley, D. and Wasserman, R. (eds.) *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press. pp. 384–423.

Sider, T., Hawthorne, J. and Zimmermann, D. (2007) *Contemporary Debates in Metaphysics*. Oxford: Blackwell.

Stanford, P. K. (forthcoming) So Long And Thanks For All The Fish: Scientistic Metaphysics, the Metaphysics of Science, and the Benefits of Integrative Naturalism.

Swoyer, C. and Orilia, F. (2011) Properties. In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Winter 2011 edition (online). Available from:

http://plato.stanford.edu/archives/win2011/entries/properties/ (accessed 10 March 2015).

Takho, T.E. (2015) *An Introduction to Metametaphysics*. Cambridge: Cambridge University Press.

Tallant, J. (2015) Metaphysics, Intuitions and Physics. *Ratio*. 28. pp. 286–301.

Thorburn, W. M. (1918) The Myth of Occam's Razor. *Mind*. 27 (107). pp. 345–53.

Van Fraassen, B. C. (2004) *The Empirical Stance*. Yale: Yale University Press.

Wilson, J. (2010) What is Hume's Dictum, and Why Believe It? *Philosophy and Phenomenological Research*. LXXX (3). pp. 595–637.

Wilson, J. (2013) Three Dogmas of Metaphysical Methodology. In: Haug, M. C. *Philosophical Methodology: The Armchair or the Laboratory?* London: Routledge. pp. 145–66.

Wilson, J. (2015) Hume's Dictum and Metaphysical Modality: Lewis's Combi-Natorialism. In: Loewer, B. and Schaffer, J. (eds.) *The Blackwell Companion to David Lewis*. Oxford: Wiley-Blackwell. pp. 138–58.

Wilson, M. (2010) What Can Contemporary Philosophy Learn from Our 'Scientific Philosophy' Heritage? *Noûs*. 44. pp. 545–70.

Wittgenstein, L. (1921) *Tractatus logico-philosophicus*. In: Ostwald, W. (ed.) *Annalen der Naturphilosophie*, *Vol. 14*. McGuiness, B. and Pears, D. (trans.) (2001). London: Routledge Classics. pp. 185–262.

Afterword

This book has traced one possible path through the metaphysical jungle and we have not seen all the creatures that live in it. Alternative routes would have shown different animals. Of those we have not encountered we must mention a few.

Suppose you throw a die many, many times. If it is a fair die the frequency with which it shows a six will be roughly one in six throws. We say that the probability for the die to land on six is 1/6th. If you are a Humean you will want to analyse this probability in terms of the aforementioned frequency. However, if you are an anti-Humean you might want to refer to the die's propensity, where propensities are meant to be dispositions, to show a certain manifestation with a certain probability. Also, many laws of nature, especially those of quantum mechanics, are of a probabilistic nature. They say how probable a certain event is, given that certain circumstances obtain. Now, this whole area of *probabilistic laws*, of *propensities* and of *probability* in general we have left aside.

Causation has, it seems, a direction in time: causes happen prior to their effects. However, this temporal asymmetry seems lost on the fundamental level of reality: the most fundamental laws happen to be temporarily symmetrical. Starting with this friction, we could have asked questions about the true *nature of time*. Relatedly, for example, via Einstein's relativity theory, we could also have said a lot about the *nature of space*.

Much more information on properties and universals could be given. The *theory of tropes*, for example, which lies in some respects in between Nominalistic properties and universals, has been neglected.

Also, next to (neo-Humean) Supervenience, we could have spent time on the notions of *reduction*, *emergence* and *grounding*. Next to Dispositional Essentialism we could have introduced *ontic structural realism*, an interesting competitor which also opposes Humeanism.

We had to be selective: it would have been impossible to do justice to all these issues. Hopefully, the book will have stimulated you to pursue further studies and equipped you with the necessary knowledge and tools to be able to do so.

Further reading on these topics

- **Probabilities and propensities**

  Eagle, A. (ed.) (2010) *Philosophy of Probability: Contemporary Readings*. London: Routledge.

  Handfield, T. (2012) *A Philosophical Guide to Chance: Physical Probability*. Cambridge: Cambridge University Press.

- **Space and time**

  Dainton, B. (2010) *Time and Space.* 2nd edition. Durham: Acumen.

- **Properties**

  Edwards, D. (2014) *Properties*. Polity.

  Mellor, D. H. and Oliver, A. (eds.) (1997) *Properties*. Oxford: Oxford University Press.

- **Reduction, emergence, grounding**

  Beckermann, A. (1992) Supervenience, emergence, and reduction. In: Beckermann, A., Flohr, H. and Kim, J. (eds.) *Emergence or Reduction? Prospects for Nonreductive Physicalism.* Berlin/New York: De Gruyter.

  Correia, F. and Schnieder, B. (2012) Grounding: an opinionated introduction. In: Correia, F. and Schnieder, B. (eds.) *Metaphysical Grounding: Understanding the Structure of Reality.* Cambridge: Cambridge University Press.

  Clark, M. J. and Liggins, D. (2012) Recent work on grounding. *Analysis*. 72 (4). pp. 812–23.

- **Structural realism**

  Ainsworth, P. M. (2010) What is ontic structural realism? *Studies in History and Philosophy of Science*. 41, B (1). pp. 50–7.

  French, S. and Ladyman, J. (2011) In defence of ontic structural realism. In Bokulich, A. and Bokulich, P. (eds.) *Scientific Structuralism* (Boston Studies in the Philosophy of Science: Volume 281). Dordrecht: Springer. pp. 25–42.
  Ladyman, J. (2002) Science, metaphysics and structural realism. *Philosophica.* 67. pp. 57–76.

- **All *Stanford Encyclopedia of Philosophy* Articles on these topics are very useful starting points too.**

Index