A POWERFUL PARTICULARS VIEW OF CAUSATION

R. D. Ingthorsson

A Powerful Particulars View of Causation

This book critically examines the recent discussions of powers and powers-based accounts of causation. The author then develops an original view of powers-based causation that aims to be compatible with the theories and findings of natural science.

Recently, there has been a dramatic revival of realist approaches to properties and causation, which focus on the relevance of Aristotelian metaphysics and the notion of powers for a scientifically informed view of causation. In this book, R.D. Ingthorsson argues that one central feature of powers-based accounts of causation is arguably incompatible with what is today recognised as fact in the sciences, notably that all interactions are thoroughly reciprocal. Ingthorsson's powerful particulars view of causation accommodates for the reciprocity of interactions. It also draws out the consequences of that view for the issue of causal necessity and offers a way to understand the constitution and persistence of compound objects as causal phenomena. Furthermore, Ingthorsson argues that compound entities, so understood, are just as much processes as they are substances.

A Powerful Particulars View of Causation will be of great interest to scholars and advanced students working in metaphysics, philosophy of science, and neo-Aristotelian philosophy, while also being accessible for a general audience.

R.D. Ingthorsson is an Associate Professor in Philosophy at the University of Helsinki, Finland. He is the author of *McTaggart's Paradox* (Routledge, 2016) and co-editor of *Mental Causation and Ontology* (2013).

Routledge Studies in Metaphysics

Philosophy of Time and Perceptual Experience
Sean Enda Power

Fundamental Causation
Physics, Metaphysics, and the Deep Structure of the World
Christopher Gregory Weaver

Eliminativism, Objects, and Persons
The Virtues of Non-Existence
Jiri Benovsky

Thisness Presentism
An Essay on Time, Truth, and Ontology
David Ingram

Metametaphysics and the Sciences
Historical and Philosophical Perspectives
Edited by Frode Kjosavik and Camilla Serck-Hanssen

Authority and the Metaphysics of Political Communities
Gabriele De Anna

Language and Word
A Defence of Linguistic Idealism
Richard Gaskin

Neo-Aristotelian Perspectives on Formal Causation
Edited by Ludger Jansen and Petter Sandstad

A Powerful Particulars View of Causation
R. D. Ingthorsson

A Powerful Particulars View of Causation

R. D. Ingthorsson

Contents

Preface

There are two things I want to get across to the reader before they begin reading this book: (i) whom I owe thanks to, and (ii) what kind of book this is. Gratitude is first priority, but gratitude to friends and family, i.e. those I owe the most, will be dealt with in privacy.

I also owe thanks to Anna Marmodoro and all the wonderful people frequenting the seminars in her *Power Structuralism in Ancient Ontologies* and *Metaphysics of Entanglement* projects. I happened to live in Oxford in the period 2010–15 (enjoying the status of Academic Visitor for six months) and could take full advantage of the incredible line-up of invited speakers and reading seminars. As an invited speaker I could also put some of the ideas in this book to the test.

Material from this book has also been presented to audiences for critical scrutiny in Innsbrück, Geneva, Ligerz, Warsaw, Aarhus, Helsinki, Tampere, Gothenburg, Linköping, Lund, Stockholm, Södertörn, Umeå, Kristiansand, Oslo, Pittsburgh, Cologne, Pistoia, Macerata, and Milan.
Now, I am afraid that many of those I owe the most will find themselves criticised the most in this book. I hope it may soften the blow if I explain that this is because I engage the most with the philosophers that stimulate me the most. I have said many times before that the best a philosopher can hope for is to write something that someone finds important enough to refute. They can hope to also be adored, but that is frivolous and vain. Don't do it.

To make up for the lack of adoration expressed in the book, let it be known, in contrast to the criticism I offer, that if I were to make a list of philosophers (in no particular order) from which I have learnt

the most about causation and powers (including where I think they are wrong) it would look like this: Mario Bunge, Ingvar Johansson, Dorothy Emmet, Michael Esfeld, Jonathan Lowe, Stephen Mumford, Rani Anjum, Hobbes, William Huemer, Ben Kovitz, Johanna Seibt, Wesley Salmon, Anna Marmodoro, David Hume, David Armstrong, John Heil, Aristotle, Alexander Bird, Anjan Chakravartty, Craig Dilworth, W.E. Johnson, C.B. Martin, J.S. Mill, Tim Maudlin, Phil Dowe, and Peter Simons. If I appear critical of the views of these thinkers, it is because they have written something worth refuting. If I don't seem to express clearly enough how they have inspired me, it is because it is difficult to keep track of such things. Unfortunately, the good things you agree with tend to meld into a background against which the flaws and mistakes appear clearly.

People who have offered very useful commentaries of some part of the text herein, or papers/presentations/grant proposals on which it builds, are Ingvar Johansson, Jonathan Lowe, Sophie Gibb, William Huemer, Johanna Seibt, Anna Marmodoro, John Heil, Robin Hendry, Anjan Chakravartty, Peter Simons, Phillip Blum, Mark Steen, Lorenzo Azzano, Andrea Raimondi, Simone Gozzano, Paul Needham, Rani Anjum, Olivier Massin, Henry Laycock, Holly Andersson, Akiko Frischhut, Jani Hakkarainen, Markku Keinänen, Tuomas Tahko, Tobias Hansson Wahlberg, Robin Stenwall, Fredrik Stjernberg, Martin Berzell, András Szigeti, Harald Wiltsche, James Lennox, Toby Friend, John Pemberton, Brian Prince, Tamer Nawar, David Yates, Avril Styrman, Florian Fischer, Cord Friebe, Anne Sophie Meinck, Cristian Kanzian, Eric Olson, Ludger Jansen, Max Kistler, Petter Sandstad, Alex Broadbent, Uwe Meixner, and many, many more. Thank you all.

Now, what kind of book is this? It is a book that deals with ideas and arguments rather than philosophers. And it is an attempt to develop what I think is a new and promising theory about the nature of causation which has the potential to unify our understanding of causation, constitution, persistence, properties, and change. The approach is therefore all at once broad in terms of the issues it addresses, and yet narrow since I try to avoid getting entangled in the nuances and quirks of individual philosophers. I do worry that many readers will be uneasy about my approach, especially the lack of detailed engagement with particular thinkers. Even when you cannot really tell if anything relevant is missing, the mere fact that it isn't explicitly addressed can set alarm bells off. My impression is that the philosophical community does have a tendency to tie ideas and arguments to individuals and is uncomfortable with the use of paraphrase to merge ideas expressed with redundant dissimilarity by different people, into a view and/or argument that we cannot tie explicitly to an individual. I think it tends to be perceived as bad scholarship, unless done really well.

In my defence, I appeal to the fact that one has to adapt the means to the ends. This book takes a very big-picture approach, in bringing

together ideas about causation, constitution, persistence, process, properties, and more. To deal with the ideas involved in these issues, while paying attention to the details and nuances of everyone involved, is too big an enterprise, at least for me, and would provide a very difficult read for you. Simplification and paraphrase, hopefully without too much loss of precision, seems the only way. I hope the effect is that the book is perceived to get quickly to the point and moves clearly in a direction that appears promising, without getting entangled in details; it may also at every turn generate the feeling that more can be said, and things remain to be explored. Take it as inspiration to explore the loose ends and say what more there is to say.

It is important to keep in mind that I am not trying to prove that my view is the only option, but rather setting a stage on which my view can plausibly be perceived as a promising option. I don't even dare say I am presenting you with an inference to *The Best* explanation, but possibly an inference to a hitherto overlooked explanation that should be seen as *among the best*. I don't believe in arguments that definitively prove a point, at least not in issues as complex as this, and I even doubt it is possible to construe an unambiguous inference to the best explanation. All I can hope for is to have provided a new context that provides the philosophical community with an opportunity and reason to re-evaluate their previous choices and conclusions. However, many of you will probably detect omissions and flaws that you think tip the balance in another direction. If you do, you have the aims of your future research clear in front of you. For my part, this is by no means intended to be the final say in the matter, and I reserve full right to change my mind about everything in light of future research.

R.D. Ingthorsson

Acknowledgements

1   Introduction

The 21st century has seen a dramatic revival of realist approaches in philosophy to the issue of properties and causation (Ellis 2001; Ingthorsson 2002; Molnar 2003; Heil 2003, 2012; Lowe 2006*a*; Bird 2007*a*; Martin 2008; Marmodoro 2010; Mumford & Anjum 2011; Jacobs 2011; Tugby 2012; Cartwright & Pemberton 2013; Vetter 2015; Livanios 2017; Williams 2019). Central to the realist revival is the notion of *powers*, i.e. the idea that natural properties are best understood as determinate ways an object can affect other objects and be affected by them. Indeed, philosophers not only investigate what role this notion might play in the elucidation of causation, but a host of interrelated notions such as natural laws, natural kinds, and agency.

As I will argue, we are really seeing a re-examination of what used to be the *standard view* of causation from Aristotle until Hume (henceforth, 'the standard view'), notably that causation involves the production of changes through the interaction of powerful particulars. This standard view is perhaps most frequently associated with Aristotelianism but it is also—with slight variations—at the core of atomism, Scholasticism, and the natural philosophy of the early Enlightenment, and is accurately presented to modern readers by W.E. Johnson (1924), and recently advocated by Dorothy Emmet (1985) and Ingvar Johansson (1989/2004). Mainstream philosophy has for some time shunned the standard view in favour of neo-Humean accounts, for, I will argue, meta-philosophical reasons that really don't have much to do with causation. It has to do with scepticism about the kind of rational theorising required for the standard view to make sense. And yet, it seems to me, people in general continue to live their lives as if the standard view is true (including neo-Humeans), and the natural sciences implicitly endorse it.

1.1   The Aims of This Book

This book is a constructively critical examination of the causal realist revival. It is critical in the sense that it aims to show (i) that contemporary realists often inadvertently incorporate into their characterisation

of powers and causation certain ideas that really belong to the dominant neo-Humean framework that they are in opposition to, (ii) that the incorporation of these neo-Humean ideas are an obstacle to the further development of powers ontology, and (iii) that some of the Aristotelian components of the standard view on which contemporary powers-based accounts are built are arguably incompatible with the theories and findings of modern science; in other words, they are *falsified* by science. The last point is particularly relevant for those causal realists who think that metaphysical theories should be compatible with the theories and findings of the empirical sciences, or, rather, that seek a reflective equilibrium between philosophy and the empirical sciences. The book is constructive in that it suggests a way to avoid the fallacies involved in (i)–(iii) to arrive at a modified powerful particulars view that is compatible with modern science. Here I build on what I have previously argued in 'Causal Production as Interaction' (Ingthorsson 2002).

To briefly elaborate on the first critical concern, my worry is that although contemporary causal realists are in a certain sense revolting against the prevailing neo-Humean framework—which has dominated the debate for so long—and have done a lot to challenge it, they are still in some ways implicitly conforming to it. First, many enough still operate under the assumption that the task of a powers-based metaphysics is to provide a semantic analysis of dispositional ascriptions of the kind we find in ordinary language (for instance, McKitrick 2018). Second, many still operate with an understanding of powers that is almost indistinguishable from the *simple conditional analysis* of dispositions (SCA), championed by reductionists like Carnap (1936), Ryle (1949), and Goodman (1954). Third, on a related note, many enough also implicitly accept—if not for anything else then by making use of SCA (or a derivate)—that the nature of powers must be accounted for in terms of some observable correlate, i.e. a so-called 'manifestation', which is the core idea of empiricism. Fourth, and final, many either fully accept the neo-Humean conception of causation as a two-place relation between Kim-style events—albeit events that are meant to be understood as involving the instantiation of powers and not categorical properties—or simply make use of the two-place conception for the sake of convenience because they think it is of no particular consequence to conform to established ways of expressing the issue.

All four ways in which contemporary causal realists conform to the neo-Humean framework are implicit in the widespread acceptance of the idea, clearly derived from SCA, that a property P is a power if it is instantiated by a particular, a, of which it is true to say that 'a will φ in circumstances S', where 'φ' is to be understood in terms of an observable change, either in a itself (if we are talking about a passive power) or in some object b on which a exerts an influence (when we are dealing with an active power). I will call this the *dispositional analysis of powers*,

because we are expected to understand the nature and identity of a power solely in terms of some particular observable behaviour exhibited in given circumstances by the objects that possess them, or the objects affected by them. These behaviours are roughly what the term 'manifestation' refers to. Very few seem to find it strange that we are expected to understand the nature of a power, P, in terms of something that is not identical to P, but a consequence of P.

As I have argued elsewhere, and will return to in Chapter 8, the dispositional analysis of powers goes hand in hand with the idea that powers are essentially relational properties and thus cannot be fundamental; it characterises powers as being for their nature and identity dependent on their manifestations. In the extreme, they are described as *pure powers*, i.e. as properties whose nature is *exhausted* by whatever manifestation they are able to bring about, and hence properties that do not determine what an object currently *is*, but only what it can *do*, provided some or other circumstances obtain in future. This idea in turn is at the heart of the idea that powers/dispositions—rather than being themselves occurrent properties in their own right—are mere potentialities inhering in objects until the moment they are actually instantiated in the realisation of the manifestation. The problem with this view comes out clearly in the most popular objection to *pandispositionalism*—the view that all properties are powers—notably that if all powers are mere relations to manifestations it leads to a vicious infinite regress (Campbell 1976: 93; Swinburne 1980: 316–19; Foster 1982: 68; Robinson 1982: 114; Martin 1997; Armstrong 1997: 80; Lowe 2006*a*: 138; Engelhard 2010; Ingthorsson 2015). In Chapter 8, I side with Martin (2008), Heil (2003), Jacobs (2011), and Esfeld & Sachse (2011: Ch. 2.1), to defend the view that powers are both qualities and powers at once; they are *powerful qualities*.

It is a further consequence of combining a dispositional analysis of powers with powers-based accounts of causation that causation ends up being characterised in a way that bears close affinity to the neo-Humean account of causation. Notably, as a two-place relation between temporally distinct entities; between the exertion of powers and the realisation of their manifestation. Friends of powers sometimes even characterise the *relata* of the causal relation in terms of Kim-style events, just as neo-Humeans are wont to do, i.e. as an event e_1 in virtue of instantiating a power F, bringing about an event e_2 instantiating a property G. There are some proponents of powers that explicitly reject the two-place relation characterisation and insist that causation is really a continuous process (Ingthorsson 2002; Chakravartty 2005; Mumford & Anjum 2011). I think this is the way to go, but more needs to be said about the nature of that process.

The second critical concern is that the standard view is arguably falsified by modern physics. This was first argued by Mario Bunge (1979), and

later by me (Ingthorsson 2002, 2007, 2019). Roughly, the standard view depicts interactions between powerful particulars as involving a *unidirectional* exertion of influence of an active particular (the 'Agent') upon a passive particular (the 'Patient'). The notion of power is accordingly understood to involve a distinction between active and passive powers, which account for the ability of an Agent to exert influence, and the ability of a Patient to be influenced, respectively. However, according to modern physics, unidirectional actions *do not exist*; all interactions are perfectly reciprocal in a way that excludes the characterisation of the particulars involved as either 'active' or 'passive' (Bunge 1979; Resnick, Halliday, & Krane 2002). If this is right, all notions deriving from or influenced by the idea of unidirectional actions risk being false by the same measure.

This flaw, although serious, is not fatal, and in my previous work I have sketched a way to accommodate the reciprocity of interactions within a powerful particulars view (2002). Indeed, I think the resulting view offers a range of very attractive possibilities. First, it allows us to understand the constitution of compound entities as a causal phenomenon, in a way that has affinity to what Karen Bennett calls *diagonal building* (2017: 85). Second, it offers a way to understand the persistence of compound entities as a causal phenomenon. Third, it allows us to understand compound objects as causal processes while still retaining their status as substances; it may serve to reconcile the presumed incompatibility of process and substance ontologies, or at least clarify what kind of process/substance ontologies are incompatible, and which are compatible.

1.2   The Outline of the Book

The outline of this book, very roughly, is as follows. In Chapter 2, I discuss the relation between powers-based and neo-Humean accounts. I argue that they do not need to be understood as contrary to each other, if we understand the neo-Humean accounts as attempts to *describe* causation without any deeper ontological commitments, while powers-based accounts attempt to *explain* why causation occurs in the way it is described. Of course, once neo-Humean accounts of causation are linked to the ontological commitments of neo-Humean metaphysics, i.e. the view that there are no substantial connections between anything in this world, they are incompatible with powers-based accounts.

In Chapter 3, I introduce in greater detail the view I call the standard view, and discuss how to best understand causal realism today. I argue that contemporary realists are neglecting certain foundational issues that were part and parcel of causal realism before the reductive neo-Humean views rose to power in the philosophical community, and that this explains why contemporary realists tend to accept the above mentioned neo-Humean elements in their characterisation of powers and of causal relations.

In Chapter 4, I present a hitherto largely overlooked problem for powers-based accounts, which is that they are implicitly assuming certain features of the standard view that are incompatible with modern physics. While outlining the problem, I develop a modified powerful particulars view, which aims to correct the flaw of the standard view.

In Chapter 5, I reassess the arguments against causal necessity and argue that they do not address any view that has actually been put forward to defend causal necessity, as well as showing that both the standard view and my modified view is immune to them.

In Chapter 6, I argue that my account of causation in terms of interaction allows us to understand the *constitution* of compound objects, and their *persistence* over time, as thoroughly causal phenomena. Indeed, in Chapter 7, I continue to argue that my causal account of constitution and persistence also offers a way to reconcile *process* and *substance ontology*; it presents compound objects as causal processes without robbing them of their status as substances.

In Chapter 8, I address the question of how to understand powers and argue, in opposition to the categorical/dispositional distinction, that they can be conceived of as *powerful qualities*. Finally, in Chapter 9, I illustrate the consequences of the book for our appreciation of what stands out today as the main neo-Humean view of causation, counterfactual theories. Roughly, I argue that we cannot explain causation in counterfactual terms, but we can explain counterfactuals in causal terms. Chapter 10 then provides a brief summary of similarities and differences between the powerful particulars view I defend, and other recent causal realist accounts. However, before I embark on the issues outlined above, let me offer some background reflections on the philosophy of causation that will facilitate the reading of this book.

1.3 Causation, Metaphysics, and Our Understanding of the Natural World

For a long time, the slogan 'there is a cause to everything' was understood as an expression of the spirit of the scientifically informed mind, and an encapsulation of the belief that everything is *naturally* determined. The epistemic consequence of this belief is that we should seek natural explanations for all phenomena. In other words, the slogan expresses the belief that there is nothing haphazard or magical about the way the world is, and that therefore the world can be rationally understood. The idea that the world can be rationally understood doesn't imply that the world follows the rules of the conceivable. I take it to be fairly clear that our imagination easily outstretches the restrictions imposed by the natural world. We can easily imagine Superman flying at three times the speed of light without having any idea of how to achieve thrust for that kind of speed, or without knowledge of what makes this factually impossible.

We can conceive of time speeding up or slowing down, going backwards or what have you. And although an elucidation of what can and cannot be conceived is an important part of the project of finding out about the world (for instance, it can rule out options), then to find out what the world is actually like cannot only be an exploration of the conceivable. My idea of how to get closer to what the world is actually like is to confront some of the ways we can conceive it to be, with the theories and findings of the empirical sciences. They live under other constraints than the philosophical constraints of what can and cannot be conceived. They are obliged to confront what they conceive with what actually happens when we interact with the world, letting the outcome of those interactions rule out some of the ways we can conceive of the world.

To test philosophical theories against the theories and findings of the empirical sciences is not to let science dictate anything. The various empirical sciences have a narrow focus. Each investigates only a narrow segment of reality. Physics examines the smallest components and how they interact. Chemistry investigates larger compounds and how they react. Biology studies self-replicating compounds. Psychology studies conscious entities. Sociology studies groups of conscious entities. Nobody in these disciplines takes a step back to consider whether there are phenomena that cut across the segments each discipline studies. For instance, no physicist considers whether interactions between the particles they study are causal in the very same way reactions between chemical compounds are causal, or how monkeys peel bananas. That is a task only metaphysicians engage in. Obviously, physicists, chemists, and others can join in, but they can't join in by doing more physics, chemistry, etc. They have to do metaphysics.

Now, however widespread the belief used to be that an understanding of causation is key to understanding the natural world, it is not universally accepted today that causation is a natural phenomenon. Indeed, for a couple of centuries the philosophical community has tended to favour approaches that reduce causation to some or other feature of how the world appears to us in *experience*, largely owing to the way our faculties of representation work, or of how we *think* about those features of experience, while avoiding speculation about what it is about the world that could make it appear to us in that way and therefore make our thoughts about it true. This is one of the senses in which philosophers understand themselves as 'neo-Humeans', i.e. as following the dictum that we should think of things only in terms of how they appear in experience while suspending judgement about how things really are. There is another kind of neo-Humeanism, one that doesn't just suspend judgement but positively affirms that causation in the world is like it appears in experience, notably as a 'mosaic of local matters of fact', between which there are no substantial connections. So, there is both neo-Humean anti-realism

and neo-Humean realism, although the resulting accounts of causation are practically indistinguishable; they only differ with respect to whether or not their proponents suspend judgement about whether their account of causation applies to reality or only to our way of thinking. It can be difficult to tell the difference. Both approaches can be seen in the writings of the same thinker. In his early writings, David Lewis only had the ambition to elucidate the 'ordinary' concept of causation, i.e. the way we ordinarily think (Lewis 1973*a*), while his later writings are more ontologically serious. In *The Plurality of Worlds* (1986*a*), Lewis postulates not only that people ordinarily think in terms of possible worlds (a claim that is up for debate), but that reality is actually made up of a plurality of worlds that are all equally existent and real.

Neo-Humeanism does trace its ancestry back to British empiricism, and we can see the same division in the empiricist tradition between realist and anti-realist empiricists. Hobbes (1656), Locke (1689), and J.S. Mill (1843) all believed that we acquire knowledge of the external world by deliberating on the input of the senses, although Mill suspended judgement about mysterious ties between the observable facts he took to be real. Berkeley (1709) and Hume (1739), however, argued that because all knowledge is confined to what our senses tell us, we can really only know something about the content of the mind. This split in the empiricist tradition is rarely brought out explicitly, and that carries over into contemporary neo-Humeanism. Realist neo-Humeans appropriate arguments from anti-realist empiricists for uses they were not originally meant to be used, assuming this to be of no consequence. As I will show in the next chapter, it does make a difference.

Although, arguably, realism dominated the empiricist thinking in the beginning, the dominant trend in 19th- and 20th-century philosophy has been to treat causation as merely a category of the mind, because it has been believed that the prospect of knowing what objective mind-independent reality is really like is hopeless. As a result, the philosophy of causation has predominantly aimed to reduce the notion of causation to some feature of how we *think* about causation, and/or some *feature of experience* that we accordingly identify with causation, rather than deliberating on whether such thinking is right or wrong in terms of accurately describing an objective feature of mind-independent reality, and/or why the world appears to us in experience in the way that it does. The current Aristotelian revival is a reaction to that trend.

The features of experience or of our thinking to which neo-Humeans have reduced causation so far, are *regularity* (Hume 1748; Graßhoff & May 2001; Baumgartner 2008), *necessary* and/or *sufficient conditions* (Mill 1843; Mackie 1965), *counterfactual dependence* (Lewis 1973*b*; Paul 1998; Hall 2000; Collins 2000), *intervention* (Gasking 1955; von Wright 1974; Menzies & Price 1993; Woodward 2003), and *causal processes* (Salmon 1984), the last often being complemented by an appeal

to *transmission accounts* (Dowe 1992, 2009; Kistler 1998). Below, I'll return to the details of these accounts.

It is not surprising that the shift from understanding causation as a form of natural determination to merely be a feature of cognition had the consequence that many of the features traditionally associated with natural determination have moved to the background. The philosophical community doesn't seem to worry much today about the materialist principle that nothing comes into being out of nothing or is completely annihilated, wherefore there is no longer much focus on answering questions about the origin of the existing features of the world. Neo-Humeans are perfectly willing to postulate the brute existence of every physical state at every spacetime point from the beginning to end of time (and even of a plurality of worlds), without worrying about the question of how these states/worlds came to be. Indeed, even causal realists argue that the rejection of creation *ex nihilo* is just as much in need of justification as the claim that the world is recreated from nothing at every moment (Williams 2019: 213). Nor is there much focus on the idea that changes in the intrinsic properties of things only happen as a result of some *influence* being exerted by one thing on another. Indeed, Lewis can write a paper called 'Causation as Influence' (2000), elaborating on the notion of influence within a neo-Humean framework, without anyone objecting that the paper doesn't say a single word about the influence exerted by anything on anything else. His talk of 'influence' reduces to talk about differences between worlds, which by definition cannot influence each other.

The idea that causation is a category of mind supports the validity of what is sometimes called armchair philosophy, i.e. types of reasoning that stays within the confines of our everyday conceptual scheme, and which takes the theories and findings of the empirical sciences to be irrelevant. Obviously, if we are only dealing with a conceptual category or feature of experience, then the empirical sciences—that profess to be about the way the world really works as opposed to how it appears to work—aren't really relevant; we have all the access we want to the way we think from our armchair.

However, in the last two decades there has been a dramatically renewed interest in an engagement with the sciences, as a result, I believe, of the limitations of armchair philosophy with respect to deciding whether the way we think is right or wrong. As a result, contemporary realists have increasingly sought to re-evaluate the validity of the views of those thinkers of the past that took a naturalist approach and saw causation as a feature of the world. For a naturalist, the natural sciences provide important and relevant input for philosophical views. That kind of naturalist approach to metaphysics goes back to Aristotle, rather than to Plato, wherefore the Aristotelian tradition has naturally become a focal point of this movement. It bears to state already at this point that the main

purpose of a re-evaluation of that kind is not primarily to arrive at a correct understanding of Aristotle. The main point is to consider whether there is something in his ideas that can be brought to bear today on how to understand causation. Some philosophers, Anna Marmodoro being one notable example, do a good job doing exegesis and constructive metaphysics at once (Marmodoro 2007, 2013). This book is an attempt to critically assess the progress of the causal realist revival, and to arrive at a view of causation that stands a better chance of being compatible with the theories and findings of natural science.

1.4   Causation Is Only One of Many Forms of Natural Determination

It is important to note that the slogan 'there is a cause to everything' gives too much credit to causation as the only form of natural determination (Frank 1932; Bunge 1979: Ch. 1). There are many forms of natural determination, some of which are arguably non-causal, at least in light of the narrower sense the term 'cause' has attained in modern times, notably as *efficient* causation (Bunge 1979: Ch. 1.3). For instance, time and space are significant determinants of natural phenomena, and yet are not (typically) regarded as causal determinants. Furthermore, objects in uniform motion continuously change position without being continuously caused to do so. Obviously, a uniformly moving body was at some point acted upon to change their state of motion, but to attribute every change of position from then on to that original act would be to accept action at a temporal and spatial distance.

Indeed, it is really only in the broader, and older, sense of 'cause'— as any natural kind of determination or principle of change—that the phrase 'there is a cause to everything' makes much sense at all. We can then understand it as an expression of the belief that everything is determined in a *natural* manner. In fact, the slogan 'there is a cause to everything' can be read as a paraphrase of Aristotle's claim that 'in all cases of production there is something that is produced, something by which it is produced and something from which it is produced' (*Metaphysics*: 7, 7; 189), which at least doesn't say that everything can be explained in terms of efficient causes alone. His claim includes also the material cause out of which something is produced, and the nature of both Agent and Patient, as the determinants of what comes to be. We should then interpret Aristotle as using the term 'cause' in the wider sense of 'principle of change', which is perfectly in line with the idea that everything that happens in the natural world is determined in a natural manner. I'll return to the details of this older form of causal realism in Chapter 3, but for now let me make it clear that I am not discussing causation as the only form of natural determination, but only as one among many. I assume then that the nature of causation can be conditioned by some of the other kinds of

natural determinations, e.g. that nothing comes into being out of nothing, or is ever completely annihilated.

1.5   Standard Approaches to Causation

The philosophical community has for a long time favoured reductive neo-Humean accounts of causation, so much so that *The Oxford Handbook of Causation* lists them under the heading 'standard approaches' while powers-based accounts are found under 'alternative approaches' (Beebee, Hitchcock, & Menzies 2009). Neo-Humean accounts characterise causation in terms of *regularity*, *necessary* and/or *sufficient conditions*, *counterfactual dependence*, *intervention*, *causal processes* and/or *transmission of conserved quantities*. Below I'll characterise each in turn, using lower case c and e to denote token causes and effects, and upper case C and E for types.

The core idea of regularity accounts is that some phenomenon c is the cause to some other phenomenon e, *iff* phenomena of type C are always followed by phenomena of type E. Now we have an operationalisation of the application of the terms 'cause' and 'effect' and can with a good degree of reliability identify causal connections (provided the characterisation is true).

Accounts in terms of necessary and sufficient conditions really count as regularity accounts (see, for instance, Psillos 2009). They are mere refinements, based on the empirical fact that phenomena of type C are not invariably followed by phenomena of type E, but only when C occurs in the circumstances S. Furthermore, probabilistic accounts only deviate from the above by allowing the regularity to come in degrees. They add statistics as a tool to quantify the degree of the regularity in which C is followed by E on the basis of the available empirical data, and of course to decide whether any given correlated events e_1 and e_2 are correlated strongly enough for us to have reason to suspect that they might be causally related.

The core idea of counterfactual accounts is that c counts as the cause to e *iff* (in worlds closest to ours) phenomena of type E never occur unless C occurs, and the core idea of agency and interventionist accounts is that c counts as the cause to e *iff* E reliably appears and disappears when you manipulate C.

Causal process theories state that causation needs to be understood in terms of causal processes and their interactions, in which conserved quantities are transmitted between causal processes. I will treat causal process/transmission accounts slightly differently than other neo-Humean accounts, because although they are ultimately reductive, it appears their proponents are driven by the same explanatory concerns as the causal realists that push the powers-based agenda. The line between

neo-Humean realists and causal realists is thin at some particular points, and particularly with respect to causal process/transmission accounts.

Now, all reductive accounts suffer from various types of counterexamples which the proponents of each view struggle to resolve by offering modifications. To my mind, the problems will remain as long as no attempt is made to answer the following question: what is it about *C*s and *E*s such that they occur in ordered pairs, that *C*s increase the probability of *E*s, that make *E*s counterfactually dependent on *C*s, and that make intervention a reliable way to identify causal connections? In Chapter 2 I will argue that the standard approaches do not even try to answer this question, and that is why causal realists find them unappealing.

The powerful particulars view, in contrast, is an attempt to offer an explanation of causation, and it may come as a surprise (see Chapter 2) that the powerful particulars explanation of causation does not really contradict the reductive standard accounts, *if* the latter be understood merely as attempts to describe how causation appears in experience. Furthermore, the reductive views are compatible with each other. The alleged incompatibility of neo-Humean and powers-based accounts, I will argue, hinges almost entirely on a meta-philosophical disagreement about the proper method of philosophy in general; speculate vs. don't speculate about what the world is really like such that it appears to be a certain way.

1.6   Powers-Based Approaches

Something *resembling* what is known today as powers-based accounts of causation dominated philosophical thought from Aristotle and well into the Early Modern period. The general idea is that causation involves the production of changes brought about by the exertion of influence by one powerful particular on another. We find it in Aristotle, the Stoics, the Atomists, the Scholastics, and the natural philosophers of the Early Modern period, like Hobbes, and Locke. There was a disagreement on the nature of powers and of the particulars that bore the powers, as well as about the details surrounding their interaction, but agreement on the general idea mentioned above. It is because the powerful particulars view is found in all these traditions—traditions that are in some respects mutually incompatible—that I call it the *standard view* rather than the Aristotelian view.

With the development of empirical science and under the influence of the sceptical philosophy of Hume, philosophy turned ever more suspicious of the core notions of the standard view—especially those whose origins were Aristotelian—and also turned more sceptical toward the ability of philosophy to find out anything about objective reality. It turned sceptical more generally toward anything that could be interpreted as

metaphysical speculation, i.e. anything that went beyond what at any given time was believed to be the limits of our observational powers.

Strange as it would seem, under the *joint* influence of Hume's sceptical philosophy, and Kant's transcendental idealism, we saw the dawn of an era where 'causation' primarily denoted a feature of our innate way of organising our experiences and/or the way we think, rather than a feature of objective reality. Indeed, much of what counts as philosophy of causation in the 20th century can best be understood as an attempt to clarify some or other 'ordinary' concept of causation—something that somehow is emblematic of the way we think—rather than attempts to say something about what the world is really like (Goodman 1954; Mackie 1965; Lewis 1973*a*). I take it that the current rise of powers-based accounts is at least in part an expression of the growing suspicion that further analysis of the way we think is not going anywhere.

Even though mainstream philosophy of the 20th century has predominantly focused on aspects of the way we think about causation, then even before the current revival of causal realism there have been individuals who insisted on treating causation as an ontological category, for instance, Mario Bunge (1979), Roy Bhaskar (1975), Rom Harré & E.H. Madden (1975), Sydney Shoemaker (1980), Dorothy Emmett (1985), Ingvar Johansson (1989/2004), Nancy Cartwright (1989), and Jonathan Lowe (1998). What stands out is that all these thinkers argue that causation is a process of production involving powerful particulars, although they disagree on the details about the nature of the particulars and the production. They endorse some or other version of the standard view. However, although I hope to introduce some interesting ideas from these 'early' contemporary proponents of the powerful particulars view, my focus will be on the development of powers-based accounts in roughly the last 20 years.

The literature reveals a fairly well defined 'discourse' on the powerful particulars view of causation that begins to gain momentum in the mid-90s with the work of C.B. Martin (1994, 1997). From then on there has been a continuous stream of books, papers, and anthologies from philosophers defending the viability of a powerful particulars view, such as Stephen Mumford (1998, 2005), Brian Ellis (2001), Ingthorsson (2002, 2007, 2013, 2015), George Molnar (2003), John Heil (2003, 2012), Anjan Chakravartty (2005, 2017), Mumford & Rani Anjum (2011, 2018), Jonathan Lowe (2006*a*), Anna Marmodoro (2007, 2017), Alexander Bird (2007*a*), Michael Esfeld (2011), Jonathan Jacobs (2011, 2017), Barbara Vetter (2015), John Greco & Ruth Groff (2012), Max Kistler (2012), Nancy Cartwright & John Pemberton (2013), Henry Taylor (2018), Neil Williams (2019), Davis Kuykendall (2019), Lorenzo Azzano (2019), and Joseph Baltimore (2020), just to name a few. Notable critical treatments of the powerful particulars view by those who do not endorse it are Vassilis Livanios (2017) and Travis Dumsday (2019).

The main variation between different powers-based accounts has to do, with first, whether they accept a dispositional analysis of powers according to which powers are dependent for their nature and identity on their manifestation, and consequently accept the categorical/dispositional distinction dictating that properties divide into the exclusive types of powers vs. qualities. The contrasting view is that natural properties are both qualities and powers. I'll discuss this issue in Chapter 8. Second, whether they accept a relational account of causation, in which case they end up with a view of causation that closely approximates to neo-Humean accounts, or reject a relational account to explicitly argue that causation is process of production. The next two chapters are devoted to that latter issue.

What there has emerged some consensus on is the characterisation of causation in terms of *mutual manifestations* of *reciprocal disposition partners*, following C.B. Martin. This characterisation has all but taken over from previous descriptions in terms of powers and stimuli. In fact, it is the nature of mutual manifestations that occupies most of this book, and in particular the implicit assumption that disposition partners are typically of two different kinds—roughly equivalent to the Aristotelian distinction between active and passive powers—whose relation is asymmetric despite what is implied in the terms 'mutual' and 'reciprocal'. Roughly, the idea is that mutual manifestations involve the joint contribution of an active and a passive power in the bringing about of a change in the object possessing the passive power. Sometimes the commitment to this kind of asymmetry, or unidirectionality, present in interactions, is merely implicit, revealed only in the examples chosen to illustrate powers-based causation; say, soluble salt thrown in dissolving water to dissolve the salt. But often enough it comes out explicitly in the characterisation of mutual manifestations, in terms of the coming together of active and passive powers to jointly manifest some particular outcome. An active power is an ability to influence other objects, while a passive power is the ability to receive the influence of an active power and to change in some determinate way.

1.7   The Challenge from Modern Science

The manner of characterising interactions between powerful particulars as involving a *unidirectional* exertion of influence by an Agent on a Patient, which results in a change in the Patient, goes back to Aristotle. One of the main points of this book is to point out that this characterisation is arguably incompatible with what modern science treats as an established fact, notably that all interactions are thoroughly *reciprocal* (Bunge 1979; Ingthorsson 2002, 2007). The fact that interactions are reciprocal is understood in the natural sciences to mean that there are no unidirectional actions, i.e. no instances of one object acting on another

while not itself being equally acted upon by the other object (see, for instance, Resnick, Halliday, & Krane 2002: 83).

To be sure, the term 'reciprocity' occurs frequently in the contemporary literature on powers—it could even be described as a core concept in modern powers-based accounts—but it is then used with a very different meaning than in the context of natural science. The difference is that most powers-based accounts depict powers as reciprocal even though the influence exerted between two objects goes only from the one to the other. The term 'reciprocal' is here only an admission of the fact that at least two powers have to be involved, and that both contribute to the production of the outcome of the interaction. An example would be that the heat of the fire and the ability of a hand to be heated both contribute to the end result of a hot hand. This is the basis of distinguishing between an active power, i.e. the power of an object to influence another object, and a passive power, i.e. the power to be influenced by another object. It is also the basis for distinguishing between Agents and Patients, i.e. objects possessing active and passive powers, respectively. However, according to modern physics, no object acts on another object without being acted upon by the other in return (i) in *the same way*, (ii) to *the same magnitude*, and (iii) at *the same time*. If this is right, we no longer have any objective criterion to distinguish between an Agent and a Patient. Furthermore, if there are no unidirectional actions and the active/passive and Agent/Patient distinctions are mistaken then all notions deriving from or influenced by the idea of unidirectional actions also risk being false by the same measure. This flaw, although serious, is not fatal to the powerful particulars view. In previous work I have sketched a modified powerful particulars view that accommodates the reciprocity of interactions (Ingthorsson 2002, 2007). Indeed, as I will show in Chapter 4, the modification serves to strengthen the powerful particulars view. And, as I have already mentioned, it allows us to understand constitution and persistence of compound entities as causal phenomena, and reconcile (at least to some degree), substance and process ontologies.

1.8   The Empiricist Bias

The influences of empiricism are tangible in much of what counts as mainstream analytic philosophy today, and of course explicitly so in the works of those working within the neo-Humean framework. One aim of this book is to elucidate how empiricist thinking has filtered down to the presuppositional depth-structure of the debate, to the degree that causal realists have now inadvertently assimilated ideas that originate from the empiricist reduction of core ideas that are found in the causal realist traditions. I have already mentioned some ways that this expresses itself, but I'd like to add here that others have expressed similar thoughts. Barry Smith (2005) and Ingvar Johansson (2016) have argued that first-order

predicate logic, advocated by Quine as the language of philosophy, favours a neo-Humean framework. Smith coined the term 'Fantology' to denote this neo-Humean bias. Mumford & Anjum (2015a), and Johanna Seibt (2018), have independently argued for the same conclusion, but focusing on how the acceptance of a certain kind of logic has worked against the acceptance of powers-based and process ontologies, respectively. The question to be asked is of course why first-order predicate logic was accepted so easily by the philosophical community. The answer is, arguably, because philosophy was already geared toward a neo-Humean view of both science and philosophy as primarily descriptive rather than explanatory.

Now, what could really be wrong with the idea that we shouldn't postulate what goes beyond the observable or measurable? Isn't that also a kind of naturalism? Yes, but there is a huge difference between saying (i) that we should *base* our views about the world on observations/measurements and that all our theories must be empirically adequate, and (ii) that we should describe the world *only* in terms of the observable/measurable, and never go beyond that to theorise about what the world is really like such that it appears to us in observations in the way that it does. To begin with, the second option is saying that in hindsight people never ought to have hypothesised about any of the things we now consider observable but weren't when they hypothesised about them.

Another interesting example of when prior commitments affect the appreciation of philosophical views—and which helps to understand why contemporary philosophers are not comfortable with *presentism*, the view that only the present exists, the future not yet, and the past no longer—is that it is difficult to deal with presentism in the language of first-order predicate logic. At least if first-order predicate logic is meant to function like Quine prescribed, notably to specify our existential commitments (for arguments to this effect, see Øhrstrøm & Schärfe 2004; Seibt 2018). Sure, one can introduce temporal operators, but on Quine's understanding, such operators must operate on something existing, wherefore the use of past-tense operators to talk about Chrysippus being in the market place yesterday is still to quantify over existents (there exists an x such that Px). Indeed, as Peter Øhrstrøm and Schärfe argue, it was concerns about Quine's idea about ontology that drove Arthur Prior to develop a temporal logic of a different kind. Ultimately the point is that the philosophical community, arguably, only accepted a logic construed to quantify univocally over entities in a neo-Humean 'mosaic of local matters of fact' because the community was already sympathetic towards such ideas.

1.9   About the Method of My Project

I take it that a naturalist approach is appropriate to the kind of scrutiny I propose to undertake of powers-based accounts of causation. After all, the proponents of such accounts aim to provide an account of

how causation works in the world we inhabit, and by and large assume that their accounts must be empirically adequate; i.e. compatible with the (true) theories and findings of the empirical sciences. I insert '(true)', because it is not naively assumed that the empirical sciences already have all the answers, while metaphysics is doomed to uncritically receive the input from the empirical sciences to try to make the best of whatever it says. It is assumed that metaphysics, as it always has done, is capable of critical scrutiny of the way the empirical sciences make sense of their own theories, and can provide them with very useful feedback that allows them to get closer to the truth (not only the other way around). But I'd better spell out in a little more detail than I did above what I take to be a naturalist approach in metaphysics, i.e. the way the empirical sciences and metaphysics are continuous and yet distinct approaches to the same subject matter.

Allowing myself a healthy dose of oversimplification (more detail in Ingthorsson 2020), the empirical sciences each focus on a particular segment of reality and focus on the understanding of the particular phenomena within that segment. Physics focuses on the elementary particles, their properties, and the particular interactions they engage in, all of which are unobservable to the naked eye. Chemistry focuses on compound entities, constituted by the simpler entities studied by physics and the interactions that occur between those simples; in chemistry they talk about 'reactions' when the compounds interact with other compounds. Often, the reactions studied by chemistry are observable to the naked eye. Biology focuses on even more complicated systems and their interactions. The demarcation between chemistry and biology might perhaps be drawn between systems that have some form of function or causal feedback loops, and which can procreate.

Metaphysics, on the other hand, focuses on broader categories of entities, their ways of being and interacting, that cut across the segments of reality that are the focal points of the individual empirical sciences. Physics deals with electrons and their charges and how charged particles interact. It doesn't deal with cell-division or metabolism in mitochondria, and it certainly does not consider whether there are classes of phenomena that we find in all the segments. That is what naturalist metaphysics does (or can do). It can look at all the different entities studied by the different sciences, and the manner in which they interact, and ask if there is any similarity in the characterisation of widely disparate phenomena. The result of such a scrutiny could be that every empirical discipline postulates entities that can be characterised as persistent bearers of properties (substances), the properties that they bear (properties), the relations they hold (relations), and the events that these bearers enter into (events), and which often bring about a change in them (causation). The empirical sciences do not ask what the most general characteristics of persistent entities are, or of properties, relations, events, causation, etc. The only

discipline that tries to ask and answer such questions is metaphysics. Metaphysics, the kind that I pursue, is the business of asking whether all the phenomena in all the segments fit into a general scheme that all the sciences can incorporate in their scheme of things, or even if they already have incorporated it without realising it.

The fact that I take the findings and theories of the natural sciences to be relevant input to metaphysics—as something we must consider— doesn't mean that I slavishly accept their theories and findings uncritically (a common but misguided criticism of naturalism). They are not the only input of ideas about how things could be. Another input of how things can be comes from our everyday conceptual scheme, from phenomenology, from semantics, and from so-called armchair speculation. None of these individually represent a privileged access to how things really are, but empirical science has the privilege of being the only discipline that explicitly aims to find out something about the mind-independent reality, and to test their hypothesis by strategic interactions with that external reality, and so should be considered to provide important input to our common enterprise in figuring out reality.

In the approach that I take, the findings and theorising of the empirical sciences—at least those parts I can hope to understand—are only taken as valuable input on how the world appears to us in experience, and as providing a plethora of explanations of how things work and a basis for excluding some explanations. I contrast that input with input about how we make sense of the world in our everyday conceptual scheme, and how we talk about them, as well as phenomenological analysis, and I pitch the findings from each of these approaches against the other in an attempt to find some way to bring it all together in one coherent whole. If one approach seems to say something very different from the other, we need to find an explanation as to why that is. Either it is because one is right and the other wrong, or vice versa; we need to figure out which one it is. In doing that, we will be construing some way that it all fits together. I take this to be pretty much the role of metaphysics as described by Johansson (1989/2004), Lowe (1998), Heil (2003), and Simons (2000).

I don't believe we can come to any conclusion about the world merely by scrutinising language, or merely by phenomenological analysis, or merely by *a priori* reflection on the conceptual scheme already in place, or merely by accepting the facts and theories of physics. Metaphysics, in the sense I have sketched it, is a project beyond the scope of an individual. We all take a shot at some part of the overall project, in our own limited way, and in time, over generations, we hopefully move towards greater clarity. Here I take a shot at pushing our understanding of causation, and of the philosophical discussion about causation, just a little bit further.

2 Powers-Based vs. Neo-Humean Approaches

Although powers-based accounts of causation have grown in popularity in recent years to the degree that they are no longer a peripheral view, they cannot be described as mainstream. Indeed, as I have already mentioned, *The Oxford Handbook of Causation* lists neo-Humean accounts as 'standard approaches' while powers-based accounts are among 'alternative approaches' (Beebee, Hitchcock, & Menzies 2009). It is therefore not surprising that many are uneasy with powers-based accounts—they don't know what to make of them—especially those that have been brought up in the dominant neo-Humean framework. And many enough are brought up entirely on a neo-Humean diet. On the other hand, there is a curiosity about powers-based accounts too.

Sara Bernstein (2013) nicely sums up the situation in a review of a volume of papers attempting to bring the new powers-based metaphysics to bear on the problem of mental causation (Gibb, Lowe, & Ingthorsson 2013). Bernstein can see the appeal of the powers-based approach, but worries that it does not relate to the neo-Humean framework with which she is familiar:

> I am optimistic about the potential of the powers-based approach, but I see its major barrier to success to be bridging the gap between itself and other systems, or at least, clearly situating itself with respect to the dominant dialectic. Many advocates of more traditional approaches see the powers-based system as operating within its own philosophical universe and making little contact with the existing framework. This hurts both sides: powers-based theories are only taken seriously by those antecedently friendly to them, and prevailing approaches do not benefit from the theoretical resources of the powers approach. At the same time, using the tools of the more dominant strategies would benefit powers-based theories, as some of their key concepts (properties and substances, to name a few) remain underdeveloped. Clearly connecting powers-based theories to the

traditional neo-Humean framework will open up greater theoretical resources for both sides.

(Sara Bernstein 2013)

I agree that proponents of powers-based approaches must situate themselves with respect to the 'dominant dialectic' but I disagree that the tools of the dominant strategies would benefit friends of powers.

The way to situate powers-based accounts, in my view, is not alongside the neo-Humean framework, in the guise of a rival, or as an admixture to the neo-Humean framework, but as operating on a different level; they should be understood as attempts to *explain* causation, while reductive neo-Humean accounts merely *describe* how causation appears empirically, and how we reason about that appearance in the absence of a deeper explanation of why it appears that way. Now, neo-Humeanism is of course often used as a name for the specific doctrine that everything is 'loose and separate', i.e. that we live in a world of facts between which there are no substantial connections. But it is not obvious to me that every advocate of so-called neo-Humean accounts of causation really are neo-Humeans in that sense. Many enough believe that there are substantial causal connections, but think those connections are beyond our reach for epistemic reasons. This is the explicit view of J.S. Mill whose views on causation are a forbear to modern neo-Humean views.

The interesting question is why proponents of neo-Humean accounts, even such as do not think that everything is loose and separate, resist attempts to explain when the explanations do not obviously disagree with them on the level of description. Indeed, the neo-Humean accounts do not really contradict each other over the issue of how causation appears to us but can be said to disagree about how we popularly reason about causation. We are seeing here a clash between descriptive and reductive approaches (neo-Humean) vs. revisionary and explanatory approaches (powers-based); really, between empiricist and rational realist views. This clash comes to the fore in discussions about causation like it does nowhere else, because causal explanations are the epitome of explanations generally. Operate with a descriptive approach to metaphysics and you will arguably find it difficult to understand the resulting accounts of causal explanations as anything other than further descriptions that call for explanation. Why does the light come on when I flip the switch? Because whenever I flip the switch the light comes on (regularity analysis)? Because when flipping occurs in conditions S, then the light comes on (conditional analysis)? Because in every world closest to ours the light only comes on if the switch is flipped (counterfactual analysis)? Because the repeated manipulation of the switch makes the light reliably come on and off (intervention)? None of these actually answers *why* the flipping of the switch makes the light come on.

2.1   Explaining vs. Characterising Causation

I worry that the philosophical community has for such a long time favoured reductive neo-Humean approaches that the presuppositions that go with that approach have faded away into obscurity, so that they are no longer universally appreciated as reductive. Consequently, neo-Humean accounts are understood as incompatible rivals to powers-based views, for the title of most satisfying *explanation* of causation. I will argue that this is not the case. Neo-Humean accounts describe causation in illuminating and useful ways but do not explain it, while powers-based accounts explain causation in a way that is compatible with neo-Humean characterisations. Let me explain.

Jonathan Lowe, in *The Possibility of Metaphysics* (1998: 108), draws an interesting distinction between 'saying what something consists in' and 'explaining' that something. The distinction I draw is more or less the one Lowe draws, although I cast it in terms of 'characterising' (i.e. describing) and 'explaining'. To *characterise* a phenomenon of interest is to point out some salient feature of the phenomenon as it appears in experience, such that we can recognise it and tell it apart from other things. However, such a characterisation need not really explain the phenomenon, say, why it occurs at all. For instance, we can characterise persistence as when something exists for more than a moment, but that does not explain why anything persists, or why some things persist for a short time and others for a long time. Note that the notions of endurance and perdurance do not explain why something endures or perdures, they are merely characterisations of persistence within the constraints imposed either by an *A* or *B* view of time. We can still ask why something endures or perdures, and neither endurance nor perdurance answers that question.

Explanations of persistence do exist. We need not worry here whether they are good or bad, since the point is merely to illustrate the difference between explaining and characterising. In most materialist ontologies of the past it is postulated that the world ultimately consists of smallest indivisible component parts that persist because they must; they are permanent. The persistence of the fundamental simples is 'explained' by attributing to them a fundamental and primitive property of *permanence*, a property we cannot even in principle establish empirically. The postulation of this property is motivated by the perceived need for permanent simples in order to explain why we experience changes in middle-sized compound objects while also taking for granted that nothing comes into being out of nothing and nothing is ever completely annihilated; it all boils down to the rearrangement of the permanent simples. Unfortunately, permanence only explains, at best, the persistence of simples but not of compound objects. A separate explanation is needed for why the simples sometimes come together to constitute compounds, and why they continue to constitute them over time so that they continue to exist for more

than a moment, i.e. persist. The permanence of the simples cannot help us here; we are short of an explanation. In Chapter 6 I will offer a causal explanation of the constitution and persistence of compounds, but, for now, I hope to have said enough to make sense of the distinction between explaining and characterising in order to apply it to the issue of causation.

Reductive views, as already mentioned, offer an account of causation in terms of some or other feature of our experience or way of thinking about the world, but do not offer an explanation of what gives rise to those features. Indeed, that is the whole point of reductive views; to avoid speculation about what something is really like such that it appears a certain way, and yet say something illuminating about the appearance. Let's have a quick look at the standard views to illustrate that they only characterise but do not explain.

The core idea of regularity theories is that the occurrence of some phenomenon c is the cause to the occurrence of some other phenomenon e, *iff* phenomena of type C are always followed by phenomena of type E. This characterisation offers an operationalisation of the application of the terms 'cause' and 'effect' that seems to allow us a fairly reliable (but not infallible) way to identify causal connections. But if we ask what it is about phenomena of type C such that they always are followed by a phenomenon of type E, the regularity theory doesn't have an answer. It is the regularity that informs us that c is a cause and e an effect, not vice versa. Indeed, the original motivation for developing regularity theories was to avoid speculation about such things because it was believed to inevitably lead into a quagmire of theoretical entities whose nature is forever beyond our grasp. The idea is that we can know what a cause is and what is an effect without knowing on a more fundamental level what it is about the causes such that they bring about effects (or even without knowing whether anything really is brought about at all).

The conditional account is really just a slight improvement on the simple regularity theory, based on the observation that as a matter of fact causal regularities aren't dyadic. C is not invariably followed by E, but only when C occurs in the circumstances S. But again, if we ask what it is about C and S and E such that they stand in that triadic relation, then the conditional account doesn't have an answer. I take it that Mackie's analysis of causation in terms of INUS conditions is an illustration of the fact that we are unable to distinguish between a cause C and circumstances S only on the basis of the empirical grounds the neo-Humean allows, notably whether something occurs or not (Mackie 1965). In the absence of a clear criterion of the difference between C and S, we can equally say that S is sufficient given C, as we can say that C is sufficient given S. In his view, the mere occurrence of any individual factor in a causal nexus is never singularly sufficient for anything, but merely a necessary component of a nexus of conditions that might jointly be sufficient for an outcome.

I allow myself to simplify grossly (for the sake of convenience, since it is of no consequence here) and say that the probability theory only deviates from the regularity theory by allowing the regularity to come in degrees. But, again, probability theory does not explain why the occurrence of C increases the probability of E; it can only quantify the degree to which C does increase the probability of E, on the basis of the available empirical data. The result may give us reason to suspect that C and E actually are related as cause and effect, but further empirical data cannot establish this. We need to also know the causal mechanism that makes C bring E about, and perhaps also understand on the basis of that mechanism how reliable it is in the production of E.

It is important to note that when I ask for an explanation in this context, I am not asking for the empirical explanation of particular phenomena; why a glass breaks when it hits the floor, or why negatively charged electrons repel each other. I am asking for an account of what it is, in the most general sense, that makes *anything* be the cause to *any* other phenomenon (and why the latter is then an effect). In the metaphysics of causation, we look for the most general features of the category of causal connections, and although the input from the empirical sciences is important for that task, they do not do the work for us. No empirical scientist stands back from the multitude of particular causal phenomena in order to ask what all these phenomena have in common such that they count as causal phenomena. Or, rather, if empirical scientists were to do that, they would no longer be doing empirical science but metaphysics. The ambition of the proponents of the standard view is to provide that kind of model (or it should be). Neo-Humeans have no such ambition. They might have the ambition to show that explanations cannot be given, but their accounts of causation cannot support that claim. They need a meta-metaphysical theory about why metaphysics cannot provide explanations. More about that later in my treatment of Hume below.

Getting back to the neo-Humean accounts, then the core idea of counterfactual theories is that *c* counts as the cause to *e iff* (in worlds closest to ours) phenomena of type E never occur unless C occurs. However, we can again ask what it is about E and C in those closest possible worlds that make E counterfactually dependent on C, but the counterfactual accounts don't answer that question. The idea is that we can know what is the cause of something else without knowing on a deeper level how causes in general bring about effects, merely by knowing what would count as a smallest deviation from the actual world and what the consequences would be from that deviation. Since counterfactual accounts are the most widely adopted neo-Humean approach, I will make a special effort in Chapter 9 to question the plausibility of the idea that we could know what counts as a smallest deviation, and the consequences thereof, without first having knowledge of the causal workings of the actual world.

The core idea of agency and interventionist theories is that c counts as the cause to e iff E reliably appears and disappears when you manipulate C. Again, we can ask what it is about C and E that make them be related in that way, and the agency and interventionist theories do not answer that question; we learn to identify Cs and Es by noting what happens in manipulations.

Causal process theories state that causation needs to be understood in terms of causal processes and their interactions. I postpone discussion of them until Chapter 3, because I find them to be a mix of reductive and non-reductive accounts, and which should be understood in contrast to the powerful particulars view that emerges in Chapters 3 and 4. Here I will only say that I find many of them to be reductive despite the explanatory ambitions of their proponents, because they accept the empiricist dogma that real-worldly phenomena must be defined in terms of observable features. For instance, that a causal process is something that carries a 'mark', or that a causal interaction is when conserved quantities are transmitted between causal processes. They are reductive in so far as (i) they remain silent on the causal role of the mark, and (ii) lack an explanation as to why the conserved quantities are transmitted in the way that they are (my interaction model is meant to provide such an explanation).

Now, all reductionist accounts suffer from various types of counterexamples that the proponents of each view struggle to resolve by offering modifications. I suspect that the problems are likely to remain as long as no attempt is made to answer the question I have been asking here: what is it about Cs and Es such that they come in ordered pairs, that Cs increase the probability of Es, that make Es counterfactually dependent on Cs, and that make intervention a reliable way to identify causal connections? I won't try to prove that the problems will forever remain, but I will attempt to show that the powerful particulars view I offer in Chapter 4 answers the question.

2.2 Powers-Based and Neo-Humean Accounts Compatible on the Level of Characterisation

If neo-Humean views are treated merely as characterisations of causation then they do not significantly contradict powers-based accounts, because they neither try to compete on the level of explanation nor do they disagree on the level of characterisation. Indeed, neo-Humean views don't significantly disagree with each other either. Let me illustrate these points using a rough sketch of the standard view and which I will elaborate on in the next chapter. According to the standard view, causation involves the production of changes through interactions between powerful particulars. The characteristics of the changes produced are determined partly by the influence being exerted between the objects in virtue of their so-called active powers (their ability to influence other objects), and

partly by their respective passive powers (ability to change in some way in response to the exertion of influence of an active power). The idea is that, whenever an object possessing certain active powers interacts in a certain way with another object possessing certain passive powers, the outcome is always the same. Whenever you throw the same kind of brick in the same kind of way at a window of a certain type, the window will break. On this view a cause *C* is really a determinate kind of interaction between two determinate kinds of powerful particulars and the effect *E* is the outcome of that interaction. The core idea of this view has been expressed by Mario Bunge (1979: 47) in the following way:

If *C* happens, then (and only then) *E* is always produced by it.

Note that I have yet to explain the nature of 'production' (see subsequent chapters), but this is enough for our present purposes.

The interesting point about the standard view is that it entails that causation will exhibit all the salient features pointed out by the reductive views. It entails that whenever *C* happens *E* will follow, just as the regularity view predicts, but offers a further analysis of *C* that incorporates the distinction between the sufficient and necessary conditions drawn in the conditional analysis; *C* in Bunge's characterisation corresponds to the nexus of *C* and *S* in the conditional analysis, i.e. of the action of the Agent and the passive object or state of affairs upon which the Agent acts. Note that it provides criteria to distinguish between *C* (something 'active') and S (something 'passive'). It also entails that *C* will increase the probability of the occurrence of *E*, although on necessitarian accounts we have to understand the uncertainty implied in 'probably' as due to our uncertain knowledge of the details of any particular instance *c* of the kind *C*.

One of the nicest examples of a dependency between the anti-reductive and reductive views is between the standard view and interventionist accounts. The standard view entails that intervention is an excellent method to identify causal connections. Given that it is true that whenever *C* happens, and only then, *E* invariably follows, it follows that when *C* occurs *E* will appear and when *C* is removed *E* will be removed, *and* that *C* is the only thing that can do that. So, all you have to do to find out the cause to *E* is to manipulate anything that could conceivably be the proximate cause and whenever the presence or absence of anything reliably makes *E* appear or go away, you will have found the cause. Conversely, assume the falsity of the standard view, and it entails that intervention is an unreliable method to identify causal connections. Suppose *C* only sporadically produces *E*, and is not the only thing that produces *E*, then it follows that intervention will be an unreliable method to identify causal connections, and it will be less reliable to the same degree as

the production is less reliable. At least, in the absence of an explanation of why causes reliably bring effects about, the interventionist account doesn't have an explanation as to why manipulation of causes makes the effect appear/disappear.

Galileo Galilei is nowadays often portrayed as an early interventionist. To be more precise, it is argued not only that Galileo developed the scientific method—the use of manipulative experimentation to identify causal connections—but developed a radically new *concept* of causation (Ducheyne 2006; Ducheyne & Weber 2007). Like William Wallace (1988), I disagree with this reading of Galileo. To my mind he is much more reasonably understood as drawing out an epistemic consequence from the standard view of his time, notably that whenever objects of a certain kind interact in a certain way the outcome is always the same. Now, as Wallace points out, Galileo's own texts are probably inconclusive as evidence of his actual views on this particular matter, and I am not really concerned with proving Galileo right or wrong. The main point is that if we assume the standard view, it will work as an explanation of why intervention works, and it does not really contradict interventionist accounts considered as attempts to characterise causation. Contrariwise, if we assume the standard view to be wrong, no explanation exists of why the world should exhibit the features pointed out by the neo-Humean views. Indeed, it is often assumed that these features are brute facts.

It is also interesting to note that the reductive views are compatible with each other. It can be simultaneously true that causation appears to us as a regularity (albeit a triadic rather than a dyadic one), as probabilistic, as counterfactual, and as interventionist. We can however debate which of these characterisations offers the most reliable way to recognise causal connections, but that is surely an epistemic issue.

It seems to me that if proponents of neo-Humean accounts are going to reject powers-based accounts it is not because of any disagreement about how causation appears to us in experience. And it will not be based on disagreement on the level of explanation, because neo-Humean accounts do not offer explanation. The disagreement will revolve around a meta-philosophical disagreement about the proper method of philosophy in general. In other words, the disagreement will revolve around the question of whether or not philosophy (and science) is confined to merely characterise reality as it appears to us in experience.

Of course, there is the issue of necessity, which has been a disagreement between causal realists and neo-Humeans, but more recently also internally between causal realists. But how are we to understand that disagreement? Usually it is understood in terms of neo-Humeans insisting that there is no substantial connection between entities in the world (everything is loose and separate) while causal realists insist there is a substantial connection while perhaps disagreeing on whether that

connection amounts to necessitation or not-quite necessitation. However, there is a vast difference between accepting neo-Humean accounts as simply empirically adequate descriptions of a world that may or may not contain substantial causal connections, or accepting them on a more ontological level as correct accounts of a world in which substantial connections are missing. As I have explained, the epistemic reading of neo-Humean accounts makes them compatible with the standard view, and therefore not of much interest to the subject of this book. From now on, unless otherwise specified, I will predominantly be discussing the ontological understanding of neo-Humean accounts, or, really of neo-Humean *metaphysics*, the doctrine that there are no substantial connections between the entities of our world.

The clash between the descriptive and explanatory approach is nowhere as clear as in Hume's discussion of causation. Indeed, where most contemporary neo-Humeans don't really bother to explicitly motivate their aversion to metaphysical speculation—that is just taken to be the default stance—Hume is very explicit on that point. I gather that is why neo-Humeans often appeal to Hume for support for their position. However, it does not seem to me that Hume offers much to support neo-Humean metaphysics.

2.3   Hume's Discussion of Causation

Much of the discussion about causation and of causal necessity in the modern era has revolved around Hume's dismissal of necessary connections, apparently under the assumption that his discussion is of interest generally. However, as far as I can see, his argument is only relevant for those who accept his sceptical empiricism and his extremely reductive philosophy of mind. Anyone who does not accept those aspects of Hume's philosophy cannot appeal to his discussion of causation in support of their view, nor do they have a reason to disagree with him over causation; they should only disagree with him about his sceptical empiricism and/or philosophy of mind. Let me explain.

Before Hume turns his attention to causation he has already arrived at the conclusion that the content of our minds is limited to sense data (impressions) and the ideas that derive from them, and that all of our ideas are as particular as the sense data they derive from (i.e. no abstract concepts). He then infers from this result, with admirable consistency, that we are unable to conceive of anything that exists independently of perceptions:

> Now since nothing is ever present to the mind but perceptions, and since all ideas are deriv'd from something antecedently present to the mind; it follows, that 'tis impossible for us so much as to conceive

or form an idea of any thing specifically different from ideas and impressions.

(1739: 67)

It immediately follows from this view about the limits of cognition—independently of any concerns about causation—that any putative metaphysical category (such as causation, substance, property, etc.) cannot be conceived of as a feature of mind-independent reality, but only a feature of our understanding. This reading of Hume is perfectly in line with his kind of 'naturalism', notably that we must ground our philosophy on *human* nature (the only nature we are acquainted with in experience).

Ironically, even neo-Humean metaphysics is a non-starter on Hume's view, because it too makes claims about the mind-independent reality that Hume claims we cannot conceive of. Neo-Humean metaphysics is therefore not the kind of sceptical empiricism that Hume represented, nor is it the kind Mill advocated, the one that doesn't doubt the reality of physical facts but suspends judgement on mysterious ties between them. Neo-Humeanism, in so far as it can be understood as a metaphysical position (in contrast to an anti-realist neo-Humeanism), is the 'deliberate affirmation that, as a matter of fact, our world is such that it instantiates no necessary connections between its instances' (Sparber 2009: 15). 'Our world' is meant to be the mind-independent reality we inhabit, not the mind-dependent world of experience. In other words, neo-Humean metaphysics doesn't suspend judgement on mysterious ties but explicitly rejects them.

Hume's position is similar to Berkeley's (1709: §§ 41, 71), notably that the intellect *cannot* move beyond the appearance of things to construe conceptions of what things are like outside the mind, nor can we make judgements about them. Contemporary neo-Humean realists reject this position, and therefore have no reason to accept an analysis that depends on its truth. Indeed, neo-Humeans cannot avail themselves of arguments that presuppose the unintelligibility of their own position. And if causal realists have a quarrel with Hume, it is not with his analysis of causation but the premises on which his analysis rests. Those premises don't explicitly contradict the causal realist account of causation, but of realism generally.

It is important to fully realise the consequences of the limitations Hume puts on cognition. He is not just saying that we cannot conceive of *unobservables* in the usual sense, i.e. entities that do not affect the senses and of which we therefore have no sensible ideas. He is saying that we cannot even conceive of the things we commonly regard as *observable*, i.e. the things we believe affect our senses to give rise to perceptions about them. All we can conceive of are the objects of perception, and they are either impressions or ideas. Hume declares that we are just as much

unable to draw conclusions about external objects from our perceptions, as we are unable to draw inferences from one object of perception to another object of perception: '*there is nothing in any object, consider'd in itself, which can afford us a reason for drawing a conclusion beyond it* [Hume's emphasis]' (1739: 139). The 'objects' of which Hume speaks are the objects of perception, since in his view those are the only objects present to the mind. Furthermore, Hume was not a naïve realist, so he did not think that perception amounts to direct unmediated contact with mind-independent objects; *ergo*, the objects of perception are not identical to the objects that gave rise to those perceptions, and we cannot, Hume claims, form any conception about such objects outside the mind.

It bears mentioning that the so-called 'New Hume' interpretation has no bearing on the issue (for overview, see Read & Richman (eds) 2007). According to that interpretation, Hume was a sceptical realist, that is, he believed there was real causation in nature, but *also* that we could not rationally justify such a belief. Assuming that reading, we still arrive at the conclusion that causation, as a philosophical subject matter, has only to do with what is in the mind, and we are still not able to account for it except in terms of sense data and/or their derivate ideas. For criticism of the New Hume interpretation, see Millican (2009), Crean (2010), and Hakkarainen (2015).

As you can see, it is something of a puzzle how neo-Humean metaphysics—which is a positive claim about the nature of mind-independent reality—is both meant to be somehow derived from Hume's views on properties and causation, and yet manages to completely disregard his views about our inability to justify any belief about anything distinct from perception. I'll be content to suggest that Hume *inadvertently* inspired neo-Humean metaphysics by raising the conceptual possibility of an objective reality without substantial connections. Basically, neo-Humean metaphysics is the hypothesis that the structure of reality outside the mind mirrors the structure Hume described within the mind. Nothing in Hume's philosophy justifies this hypothesis, and, as far as I can tell, the reason philosophers accept it today is that it is a part of the dominant dialectic that they have been brought up to accept. For more detailed criticism of neo-Humeanism, see Bunge (1979: Ch. 3), Armstrong (1983: Part 1), and Maudlin (2007: Ch. 2).

To this it could be objected that neo-Humeans today only have to add to their scheme a form of justification that Hume did not allow, i.e. inference to the best explanation. So they might accept that the content of our minds, in so far as our mind is concerned with matters of fact, is confined to impressions and associated ideas, but that we can still by way of an inference to the best explanation conclude that (most likely) the structure of the world outside the mind mirrors the structure inside the mind. Lewis, for one, clearly appeals to such an inference to the best explanation argument in favour of modal realism. However, the problem

is that contemporary neo-Humeanism is chock-full of ideas that cannot be traced to impressions, and most of its proponents are Platonists about abstract objects, and propositions in general, and many of them accept the reality of negative facts. My critique here is really that neo-Humeanism is inconsistent. It rejects the reality of mysterious ties in spatio-temporal reality, but has no qualms accepting mysterious ties outside spatio-temporal reality, including those holding between distinct spatio-temporal realities (possible worlds). And, is it really the case that the best explanation of the observed features of the world is that there is no explanation? Is it a brute fact that things are the way they are?

Ultimately, the main point is that causal realists have no direct quarrel with Hume about causation, but only with his reductive theory of mind, and that the dispute between causal realism and neo-Humeanism today should be perceived as a dispute between two camps that assume contrary working hypotheses for reasons that today have very little to do with Hume. We have two contrary hypotheses about the best way to make sense of the structure and regularity we observe: (i) that it is the result of substantial connections of some sort yet to be elucidated (causal realism), or (ii) to assume that there are no substantial connections to be elucidated (neo-Humeanism). The dispute will have to be settled by judging their respective success in making sense of the structure and regularity we observe. I will not speculate here about when or how such a judgement is to be made, but I want to make sure that we will be talking about the right kind of causal realism when it is time to make that judgement, and that we understand the objections in the right way.

2.4   The Tools of the Dominant Strategies

We have yet to consider whether Bernstein is right to think that some of the key concepts of powers-based ontologies are underdeveloped, such as properties and substances, and that they would therefore benefit from using the tools of the more dominant strategies. Or, in other words, would a connection to the traditional neo-Humean framework 'open up' greater theoretical resources for both sides? I have doubts about that. Sure, there are problems/controversies about most if not all the core concepts of powers-based accounts. However, those problems are related to an attempt to develop concepts that play some *explanatory* role. The neo-Humean framework avoids those problems by stripping the concepts of their explanatory roles, therefore appearing unproblematic (rather than 'developed'). In most if not all cases, it reduces the problematic concept to some other less problematic concept, but usually at a cost with respect to explanatory power.

Take the paradigmatic neo-Humean conception of property. At the most fundamental level we have categorical properties. They are causally inert qualities, useful for not much more than serving a role as abstract

identifiers in conjunction with Leibniz' law. I say abstract, because although they were once thought to be observable (see, for instance, Goodman 1954: 40) they are now generally assumed to be unknowable. Only neo-Humeans who are also direct realists (if there are any) will disagree. The reason they are supposed to be unknowable is because they are only contingently associated with the second-order powers that affect our senses to inform us about the first-order categorical properties. Since the powers are in turn only contingently related to the sensory qualities that they give rise to, there is no way to tell whether the way a property appears in experience is the way the property really is. This is a strange situation for a type of property that is supposed to ground the identity of objects. The consequence is that we cannot in principle know the identity of anything (not even of properties), and yet we can supposedly know *a priori* that it is necessarily true that if *a* and *b* are identical, they have all properties in common. Leibniz' law is thus turned into a piece of abstract trivia that can at best be applied in purely formal models.

Similar things can be said about the concept of substance. For an Aristotelian, the concept of substance, with the distinction between form and matter, has a certain explanatory role. It explains why there is always something that remains through any kind of change. The reason Chrysippus remains the same as he saunters from the gymnasium to the market is that he changes in accidents (qualitative change) but neither in essence nor matter (numerical identity). On the neo-Humean account of properties and how they ground identity, identity and change are *a priori* incompatible. This point might actually be one the most important things I have ever tried to argue for in print, and I owe it partly to J.M.E. McTaggart. He realised that, if the identity of things is grounded in its qualities, then whenever there is qualitative dissimilarity there is numerical distinctness, and vice versa. Which is why he thought Leibniz' term 'indiscernibility of identicals' was a misnomer; it should be called 'dissimilarity of the diverse' instead. He didn't connect this to the issues we are dealing with now, but the connection is straightforward. And it is very clear that the problem of temporary intrinsics, as formulated by Lewis, is a case where we arrive at the same conclusion; nothing can remain numerically the same unless it remains qualitatively the same. *Ergo*, change is impossible.

I think this should be enough to show that the sentiment that core concepts of powers-based ontologies are underdeveloped can be understood only as the sentiment of those who only want to describe reality but not to explain it, and are bothered by the extra complications that inevitably accompany the explanatory project. Again, the friction between the neo-Humean framework and powers-based accounts (or causal realists generally) revolves largely about meta-philosophical assumptions about the purpose and/or limits of metaphysical thought, and not specifically about the nature of causation, properties, etc.

2.5   Does It Really Matter?

So far I have described causal realism as largely being in agreement with neo-Humeanism about the way causation appears to us; it appears as a regularity in the occurrence of entities of which the effect is counterfactually dependent on the cause, such that if the cause is manipulated the effect will come and go away. Causal realists don't really disagree with anything in this picture, except in degree. For instance, Mumford & Anjum will accept the above description with the qualification that the regularity, counterfactual dependence, and reliability of intervention are not perfect, because they don't think causation involves necessity (only a substantial tendency).

Some neo-Humeans may reject causal realism as an ill-construed attempt to figure out what objective reality is like; or, rather, an attempt to figure out what cannot be figured out. This would be an epistemically motivated neo-Humeanism that opposes powers-based accounts for reasons that have nothing to do with what exactly powers-based accounts suggest about causation. We are talking about people who don't see any appeal in Bernstein's suggestion about a mutual benefit resulting from bridging the gap between the two systems. But they should then also accept that no realism makes any sense, not even neo-Humean metaphysics.

Others may reject causal realism, not because they think nothing can be figured out about mind-independent reality, but because they think powers-based accounts are wrong about mind-independent reality. These would be the proponents of neo-Humean metaphysics, which say that there really are no substantial connections between anything. Now, I find it difficult to discern between realist and anti-realist neo-Humeans, because very few, if anyone at all, explicitly state their position in the matter. However, David Lewis is surely one who explicitly endorses neo-Humean metaphysics, the view that the world is nothing but a mosaic of local matters of fact between which there are no necessary connections. He even explicitly commits to the reality of a plurality of worlds existing in parity.

In sum, the benefits of the powers-based approach is really in the more fine-grained account of what constitutes the cause C which is followed by E, and on which E is counterfactually dependent, and therefore offers a more fine-grained idea of what it is that we can manipulate to produce E or variations of E. The more fine-grained account has consequences for what we would consider to be good scientific practice. So, for instance, Anjum, Copeland, & Rocca (2020) argue, on the basis of Mumford & Anjum's powers-based account of causation, that there is a fundamental flaw in the idea that *only* randomised controlled test studies (RTC) are of high enough quality to ground evidence-based practices. Very roughly, the idea that only RTC are good enough is based on the idea that causation

boils down to a two-place relation between temporally distinct events or states, i.e. 'if *C*, then *E*'. It is acknowledged that the effectiveness of *C* is contingent on the situation in which it occurs, but it is still assumed that there must be some particular factor *C* that is decisively responsible for the outcome *E*. Consequently, it is assumed that any investigation into the causes of some negative outcome *E* has to involve the comparison between two controlled settings in which background conditions are fixed, and that onsly differ with respect to the presence/absence of *C*. As Anjum and colleagues point out, this model is at a disadvantage in discovering, for instance, the aetiology of multifactorial diseases. In that case a more fine-grained powers-based approach has the advantage. The alluring promise of powers-based accounts resides in their potential not only for solving certain philosophical problems, but for improving actual research practices in order to be able to better understand multifactorial illnesses such as chronic fatigue and various other psychosomatic disorders.

3   Causal Realism

Causal realism—in its most general form—is the view that causation is an objective feature of the world and not only a feature of our thinking (see, for instance, Bunge 1979: sect. 1.1.2; Costa 1989; Price 2001: 106; Esfeld 2011). Many enough consider themselves to be causal realists in this sense—they believe in real causal connections—and yet are happy to make do with neo-Humean characterisations of causation, for the sake of convenience, because they think it is at least empirically adequate. Although this book is meant to reveal some of the dangers of that choice of convenience, then in this particular chapter I am concerned more narrowly with more substantial accounts of the nature of causal relations. In fact, I will in the end focus specifically on causal realist accounts that in one way or another appeal to causal powers. This limitation excludes very little since most causal realist accounts involve an appeal to powers. Indeed, Michael Costa divides causal realism into *causal objectivism* and *power realism* (1989: 173), of which the former is merely a commitment to the idea that causal relations would obtain in the world even in the absence of minds perceiving them, while the latter commits to the idea that such relations are to be explained by appeal to powers.

  Having identified the kind of causal realist accounts that are in focus here, which are meant to stand in contrast to neo-Humean accounts that are coupled to neo-Humean metaphysics, I will also introduce an intermediary class of accounts: those that believe causation is a substantial connection but don't want to make a clean break with the traditional neo-Humean framework. I call them *objectivist accounts* (not to be identified with 'causal objectivism', which is just a commitment to the reality of causation). We have then (i) neo-Humean accounts, (ii) objectivist accounts, and (iii) powers-based accounts. I don't mean this division to be exhaustive, but merely good enough for the purposes of this discussion. The neo-Humean and powers-based accounts are here understood as metaphysical contraries, but objectivist accounts have sympathies in both directions. According to the neo-Humean accounts, there is no substantial connection between causes and effects; causation is a relation that supervenes on the distribution of Kim-style events that obtain as

a matter of brute fact. According to powers-based accounts, causation is a substantial connection that can be explained in terms of interactions between powerful particulars. Objectivist accounts are built on the assumption that causation is a substantial connection (contra neo-Humeanism), but their proponents nevertheless either think that the neo-Humean framework (or something very close to it) is the best we have by way of understanding causation, or they take themselves to be developing an alternative explanatory model of causation that is better than power realism. As I will argue, they end up not really being explanatory *because* they stay true to the neo-Humean restrictions on rational speculation.

Another way to explain the difference between neo-Humean and objectivist accounts is that the neo-Humean accounts that are of relevance here are not the ones that merely endorse neo-Humeanism because of epistemic constraints; they believe we can know—if we can know anything—that in objective reality there are no substantial causal connections. Objectivists might on the other hand endorse the neo-Humean characterisations of causation—or settle for them—because they believe an explanation of causation is beyond our reach because of some epistemic constraint. I venture to suggest that they share, or come very close to sharing, the attitude of J.S. Mill with respect to philosophical method and its limits.

3.1 Powers-Based Accounts

If we now focus on powers-based accounts, they do not only commit to the reality of powers, but will typically also commit to, or at least have a stance on, some or all of the following characteristics: *necessity*, *production*, *efficiency*, *uniformity*, and *process* (see, for instance, Bunge 1979: Ch. 1; Dorothy Emmet 1985; Johansson 1989/2004; Ingthorsson 2002, 2007; Huemer & Kovitz 2003; Chakravartty 2005; Maudlin 2007; Esfeld 2011; Mumford & Anjum 2011; Anna Marmodoro 2017). These characteristics are then in one way or another tied to powers; say, that it is the exercise of powers that efficiently produce certain outcomes in a uniform (or more or less uniform) manner, a manner better described as a process than a two-place relation. I think the following passage by Michael Esfeld captures very well the prevailing stance among those causal realists that are looking to explain causation. Note that for purposes of presentation, I have divided into three parts what is originally a single paragraph and edited out an irrelevant parenthesis:

> Causal realism is the view that causation is a real and fundamental feature of the world. That is to say, causation cannot be reduced to other features of the world, such as, for instance, certain patterns of regularities in the distribution of the fundamental physical properties. Causation consists in one event bringing about or producing another event, causation being a relation of production or bringing something into being […]

> I shall take events to be the relata of causal relations, without arguing for this claim in this paper, since this issue is not important for present purposes. More precisely, an event e_1, in virtue of instantiating a property F, brings about another event e_2, instantiating a property G.

> One can therefore characterize causal realism as the view that properties are powers. In short, Fs are the power to produce Gs. Saying that properties are powers means that it is essential for a property to exercise a certain causal role; that is what constitutes its identity. One can therefore characterize causal realism as the view that properties are causal in themselves. To abbreviate this view, I shall speak in terms of causal properties.

> (Esfeld 2011: 157)

To sum up, causal realism is the view that causation is the *production* or *bringing about* of effects by causes, and that this production is accomplished by powers that exercise a causal role. I take this to imply an acknowledgement of the features I have called *production* and *efficiency,* although there is no commitment to *process* in this particular passage. Esfeld's acknowledgement of 'patterns of regularities' in the world may well be interpreted as an implicit commitment to at least some degree of *uniformity* of causal production, but perhaps a uniformity that does not quite add up to necessity.

   The problem facing us now is that further analysis of what it involves for causes to bring about or produce effects are rare, at least in the contemporary debate. Most of those who do discuss causal production, as Esfeld does, stick to the characterisation of an event e_1 bringing about another event e_2, which only deviates from the regularity theory—whenever Fx occurs, Gy follows—by a commitment to the idea that 'brings about' denotes a substantial connection; it doesn't explain how an event brings about another. Indeed, Phyllis Illari has recently complained that no good account of causal production exists and insists that one is needed (Illari 2011). Illari, unfortunately, seems unaware of my contribution (Ingthorsson 2002, 2007), but mentions Wesley Salmon (1980) and Ned Hall (2004), as examples of philosophers that in recent times have to some degree discussed production. I can add Stuart Glennan (2017) to this group, and also the transmission account favoured by Phillip Kitcher (1989), Phil Dowe (1992), and Max Kistler (1998).

   As far as I can see, Salmon, Hall, Glennan, Kitcher, Dowe, and Kistler all take themselves to be causal objectivists and to be developing an explanatory account of causation. However, they all have strong empiricist sympathies and their accounts suffer from similar problems to those I outlined in Chapter 2 for neo-Humean accounts. Below I will briefly address them each in turn and conclude that while they are trying to go beyond the reductive analysis of neo-Humean accounts, they come short

of an explanation. Nevertheless, I think they are interesting because they highlight a number of neglected features that realist accounts should take into account. It is to be noted that *The Oxford Handbook of Causation* lists the transmission account of Salmon, Kitcher, Dowe, and Kistler as a standard approach, but Glennan's mechanistic approach as an alternative approach. To my mind the similarities between them are greater than the differences.

3.2   Salmon: Causal Production as Interaction Between Causal Processes

Salmon suggests that causation be understood in terms of the notions of *propagation* and *production*. Propagation refers to the way persistent things conserve their structure and ability to exert causal influence between the times they interact with other objects, and production refers to the way interactions between things result in modifications in their structure and ability to exert causal influence. In other words, interactions produce changes in persistent entities. Furthermore, persistent entities are to be understood in terms of causal processes. They are causal in the sense that they are able to exert influence, and they are processes because they are not static and unchanging despite the fact that they preserve a certain structure.

Salmon is an empiricist, wherefore he resists any commitment to mysterious causal powers. Instead he appeals to what he calls 'marks', which are basically some observable property of a causal process which are meant to play the same explanatory role that powers do. It is important to note that Salmon explicitly denies that talk of objects as causal processes commits him to an event ontology, i.e. processes are not simply structured series of events, but he doesn't really offer an alternative analysis; he just points to examples (1984: 139ff). I don't take this rejection as a sign of a commitment to substance ontology, but as a rejection of any ontology that tries to go beyond the observable while sticking to the observable fact that things are not static and unchanging even though in many ways they remain the same over time.

Initially, Salmon admits to using the terms 'causal process', 'structure', 'propagation', 'interaction', 'production', etc, as undefined and intuitive notions (1980). It is this intuitive formulation that appeals to causal realists, and which persuades them to think Salmon is saying something interesting. I am one of those, and the view I develop in Chapter 4 builds on these intuitive ideas. But, to make them intelligible within an empiricist framework, Salmon has to find observable correlates. Here is where I find things to worry about.

One of the possibilities Salmon explores is to distinguish genuine causal processes from non-causal processes (say, a shadow) by saying that the former are able to carry a *mark*; some feature that remains stable and which—according to science—is causally relevant for any future

interaction between causal processes (Salmon 1984). Marks are then meant to define what a causal process is, what is being propagated, and what is meant by saying that the process has an ability to influence.

Salmon then defines 'interaction' and 'causal production' in terms of the notion of 'interactive fork'. An interactive fork is when two causal processes intersect, resulting in a modification in their respective marks. Salmon later gave up the mark account due to criticism from Kitcher (1989) and Dowe (1992) and proposed instead to talk of causal processes as entities that propagate conserved quantities, and interactions as the exchange of such quantities between causal processes (Salmon 1984). Salmon then really ends up endorsing the transmission account of causation, i.e. that causation is in some way reducible to the transmission of conserved quantities between cause and effect. More about that later on.

We can illustrate Salmon's account using the example of a batter hitting a baseball with a bat, but striking it in such an unfortunate way that it hits and breaks a window in a neighbouring house. The bat and baseball are causal processes propagating the conserved quantity of momentum. When the bat hits the baseball, the interaction produces a change in the state of motion of the ball; it acquires a different momentum, velocity, and kinetic energy. As the ball travels across the pitch, it preserves its new momentum and propagates it from point *A* to point *B*. When the ball later interacts with the window, it is in virtue of the preserved momentum that it is able to influence that window and modify it (it breaks).

To my mind, this account of a causal process is reductive in the sense that it tries to define causal processes and interactions in terms of observable features they have, rather than in terms of why they have these features. Why are the conserved quantities transmitted in an interaction, and why in the particular way they actually transmitted? We are only told that this is what they do. This point becomes clearer in a discussion of transmission accounts.

3.3   The Transmission Account of Causation

There are two types of transmission accounts; those that take transmission to occur between events, and those that take transmission to occur between what Salmon would call a causal process, i.e. persistent objects. If we take the event-based accounts first, then Kistler presents the core idea in what he calls a reduction statement:

> (T) Two events a and b are causally related in the sense that one is a cause of the other if and only if there exists a conserved quantity Q of which a particular amount P is transmitted between a and b.
>
> (Kistler 1998: 1)

I find this unsatisfying for three interrelated reasons. First, because it offers no criteria to distinguish causal and non-causal instances of transmission.

Second, because it doesn't say what will bring about a transmission of a conserved quantity in the causal instances. Third, because we don't have a generalisable account of the transmission of conserved quantities; it doesn't work for all the cases.

The first problem is that conservation laws entail that any quantity present in any Kim-style event e_1 will be transmitted to the next event e_2, whether or not anything causal occurs between e_1 and e_2. A uniformly moving body will through any given temporal interval constitute a succession of distinct Kim-style events, between which any conserved quantity possessed by any event e_n will be transmitted to a subsequent event e_o. The same holds for two successive states of any arbitrarily chosen material system, regardless of what happens within the system.

The causal process version has the advantage of narrowing things down a little more, since it specifies that the transmission is not just between two distinct events, but between two causal processes, i.e. persistent particulars, when they interact. Dowe (2009: 219) offers the clearest statement of this view, by combining criteria previously put forward by himself and Salmon:

> CQ1. A causal interaction is an intersection of world lines that involves exchange of a conserved quantity.

> CQ2. A causal process is a world line of an object that possesses a conserved quantity. (Dowe 1995: 323)

> Definition 1. A causal interaction is an intersection of world-lines that involves exchange of a conserved quantity.

> Definition 2. A causal process is a world-line of an object that transmits a nonzero amount of a conserved quantity at each moment of its history (each spacetime point of its trajectory).

> Definition 3. A process transmits a conserved quantity between A and B (A ≠ B) if it possesses [a fixed amount of] this quantity at A and at B and at every stage of the process between A and B without any interactions in the open interval (A, B) that involve an exchange of that particular conserved quantity. (Salmon 1997: 462, 468)

Kistler—underestimating how uncharitable critics like me are willing to be—obviously assumes in his reduction statement that everyone realises that the statement is only meant to apply to what happens in interactions between particulars, but Dowe makes this explicit. Dowe's account takes care of the first and second problem. Causal transmissions occur in events involving interactions between two distinct particulars, and it is the interaction between the two particulars that triggers the transmission. It also

brings us closer to the picture offered by physics, which indeed depicts conserved quantities as being transmitted between particles. Physics even postulates virtual particles as carriers for any quantity that appears to be transmitted between spatially discrete entities.

However, the third problem remains. The assumption is that the direction of causation follows the flow of the transmitted quantity—really, that it follows the flow of energy—and this flow is assumed to explain the nature of causation. Why did the ball in motion make the ball at rest move when they collided? Because a quantity of motion was transmitted from one to the other. My problem with this explanation is that it does not generalise to all the relevant cases and will eventually suggest that we need to postulate something more to explain the manner in which the quantities flow, in order to get an account that generalises; we need to postulate that interacting objects always exert an equal and opposite influence on each other, and that this explains the resulting transmission (or distribution) of the conserved quantities. Let me explain this further.

The process version of the transmission account seems to make pretty good intuitive sense only as long as we stick to the standard examples. That is, examples of an apparently active (or more active) thing acting on another more passive thing which suffers a change: a billiard ball in motion encounters another at rest, and changes the state of motion of the latter from rest to motion (Hume 1748: § 36); a leaden ball dropped onto a pillow produces a hollow in the pillow (Kant 1787: A203); a locomotive pulls a truck (Taylor 1973: 35); a baseball is hit by a bat to strike and break a window (Salmon 1980: 50). All these examples fit the standard view, according to which interactions involve an active object influencing a passive object, or, as Aristotle claimed: 'whenever the potential active and the potentially affected items are associated in conditions propitious to the potentiality, the former must of necessity act and the latter must of necessity be affected' (*Metaphysics*: 9, 5; 264). Consider instead two identical billiard balls moving at the same speed, but in opposite directions, that collide head on. They will compress each other and then bounce back in opposite directions, both moving (roughly) at the same speed again but now in opposite directions from before. Which ball transmitted a quantity to the other? Did both transmit to the other, or was any quantity transmitted at all? Perhaps both balls merely changed the direction of their respective quantities without transmitting anything between them? If we are confined to merely describing the manner in which quantities go, we have no explanation, and we have no universally valid account of the direction of causation. To my mind, and which will be elaborated on in Chapter 4, what is missing here is some account of how interacting particulars exert an influence on each other and which explains why the conserved quantities are distributed/changed the way they are when particulars interact. The mere exchange of quantities doesn't answer that question.

Traditionally, it is the concept of 'force' that has explained how interacting particulars influence each other. However, in the friction between reductionistic and anti-reductionist accounts, the concept of force has been just as controversial as the notion of power. Empiricists like Heinrich Hertz wanted to get rid of the concept—to him it represented one of the occult and mysterious entities we postulate only through an invalid inference from observed changes to some imagined invisible cause to those changes—and instead wanted to describe interactions between material entities merely in terms of the changes in the state of motion that they can be observed to suffer. Accordingly, he suggested the second law of motion should not be understood as saying that the force is *proportional in magnitude* to the object's mass times the acceleration it suffers, but as stating an *identity* of force and change in state of motion. It really meant that the third law of motion ($F_{1on2} = -F_{2on1}$) could just as well be expressed by a reconstructed two-way second law; $m_1 \times a_1 = m_2 \times a_2$. In plain English, when two material systems interact, the observed change in the state of motion of the first ($m_1 \times a_1$) is always equal to the observed change in the state of motion of the second *($m_2 \times a_2$)*. Why should we accept a reduction of this kind? The empiricist answer is that we must otherwise appeal to the occult notion of force.

Now, to my mind, the proponents of the transmission account should not really have any qualms about the concept of force, in so far as they seek to ground their view in the notions already in use in the natural sciences. However, to accept it, they have to accept that there is a more fundamental feature to physical interactions than the transmission of conserved quantities. This they do not do, and therefore end up in want of an explanation of why conserved quantities are transmitted in the way they in fact are.

3.4   Ned Hall on the Generative Conception of Causation

Ned Hall aims to spell out the production view of causation, or, as he calls it, the *generative* conception. However, as far as I can see, despite his use of the notion 'pure causal history' his analysis does not go very far beyond the superficial characterisation offered by Bunge, notably that *c* is the cause to *e iff* whenever *C* happens in circumstances *S*, *E* is produced. As usual, it is assumed that *S* is to be defined in such a way as to exclude the presence of any intervening factor, *P*, in the minimal set of components constituting *S*; one that would prevent *C* from producing *E* (these assumptions are what Hall refers to below as 'foregoing constraints'):

> Given some event *e* occurring at time *t'* and given some earlier time *t*, we will say that *e* has a *pure causal history* back to time *t* just in case

there is, at every time between *t* and *t'*, a unique minimally sufficient set for *e*, and the collection of these sets meets the two foregoing constraints. We will call the structure consisting of the members of these sets the "pure causal history" of *e*, back to time *t*.

(Hall 2004: 265)

The problem here, from my perspective, is that Hall's account gives us no suggestion as to what kind of entities would constitute the minimally sufficient set for *e*, or in what manner that set brings about *e*, wherefore we are still left with only a two-place relation between the two events e_1 and e_2 which is claimed to involve 'generation', but whose nature is left unanalysed.

3.5   Stuart Glennan: The Mechanistic Account of Causal Production

Glennan's account is an attempt to say something more about what kind of entity is sufficient for bringing about *e*. As he himself notes, his account may at first blush look similar to previously offered accounts. We can have Esfeld's and Hall's account in mind, even though Glennan does not explicitly mention anyone in particular. One main difference is that we are explicitly told not to understand the event that brings about an effect merely in terms of Kim-style events, i.e. as the exemplification of properties (or relations) by an object (or set of objects) at a time (Kim 1973, 1976), but in terms of an *activity* of the objects involved in the event (Glennan 2017: 177). Glennan's point is to stress the *processual* nature of causation as a natural consequence of taking causal production to consist in the *exertion* of influence between particulars; that this exertion must be understood as an *activity*, which is something that cannot fit within an instantaneous Kim-style event. To be more precise, Glennan thinks of events as activities of particulars and not as entities that themselves are engaged in activities such as 'causing'.

Glennan's account in many ways resembles Salmon's but is sensitive to other details in the causal structure of the world. He suggests we understand causation in terms of *constitutive*, *precipitating*, and *chained* production. Constitutive production corresponds roughly to what Salmon calls 'production', i.e. the 'the ways in which a single event produces changes in its constituent entities' (Glennan 2017: 180), but with added detail. Glennan illustrates constitutive production using the example of searing a steak. The interaction between the steak and the skillet involves the transference of energy to the steak which produces various reactions between the constituents of the steak, i.e. water, proteins, carbohydrates, and fat. Glennan suggests we focus on the changes that matter to the cook, i.e. what happens in the steak. When

the temperature of the steak rises above a certain limit (+140°C), the constituents of the steak—various sugars and amino acids—react with each other to produce a variety of molecules responsible for a range of pleasant flavours and aromas. This is known as the Maillard reaction. The details are not essential to our story, but for those interested see McGee (2004: 778–9). For my purposes, what the reader should remember for the discussion that follows is that Glennan describes production in the following terms: interaction between two particulars (the skillet and steak) is the cause to the changes occurring in the particulars (special focus on the reactions between the constituents of the steak), and that both the interaction and the ensuing changes are processual rather than Kim-style events.

Precipitating production is 'the way in which one or more events produce another event […] by creating start-up conditions for a different mechanism' (2017: 182). The steak being placed on the skillet is given as an example of an event that creates the start-up conditions for the Maillard reaction in the steak.

Finally, *chained production* is a connection between events between which there exists a chain of intermediary precipitating events. We might say that the connection between someone buying a steak and eating it is a series of precipitating events starting with grabbing the steak at the store, putting it in the carrier bag, carrying it home, putting it onto the skillet, browning it, placing it on the plate, and finally eating it.

Is Glennan's account an improvement on previous accounts? I think so. It stresses that causation is at rock bottom a process of production in which propertied objects interact to change each other rather than merely a relation between Kim-style events. In doing so he implicitly takes sides with Salmon (1980) and myself (Ingthorsson 2002), as well as Chakravartty (2005) and Mumford & Anjum (2011), in saying that the kind of relational account offered by Hall (2004) and Esfeld (2011) isn't enough. But in so far as he moves beyond Salmon's account, it is mainly by way of going into greater detail about the changes that happen intrinsically to the objects as a result of their interaction, i.e. what happens in the steak when it interacts with the skillet. That puts increased focus on what happens intrinsically to objects.

In anticipation of what follows in the next chapter, then my complaint is that while Glennan's account—in its abstract form—verges on being identical to the account I offer in 'Causal Production as Interaction' (Ingthorsson 2002), then in the way he concretely illustrates it he appears to be at least implicitly influenced by the idea that causation involves unidirectional influence. He thinks the skillet influences the steak unidirectionally (or at least more than the other way around):

> What makes this a productive activity is, in the first instance, the fact that the activity of searing produces changes in the steak and, to a lesser extent, the pan. Let us focus on the change that matters for the

cook: searing both browns and cooks the steak; it also releases oil, water, and other compounds from the steak.

<div align="right">(Glennan 2017: 180)</div>

It should be admitted that I find the unidirectionality in Glennan's account less explicit than in other accounts, or rather, it can be read as verging on a proper reciprocal understanding. It is mentioned that something does happen to the skillet, but this is somehow a 'lesser' consequence than what happens to the steak, or at least less important to us. However, as I will argue in the next chapter, the risk is that the reason the effects on the skillet are judged to be 'lesser' are anthropocentric. *We* care what happens to the steak, but not too much about what happens to the skillet; but does the universe care more about steaks than skillets? It also ignores that what the reciprocality of physical interactions tells us, as described by physics, is that, in terms of the abstract quantity of influence we call 'force', the skillet is influenced in the same quantity as the steak; the skillet cools down as much as the steak heats up. The difference is the characteristics of the change that this same quantity of influence has on skillet vs. steak; the influence that will change the temperature of the steak to the degree that it becomes seared will also change the temperature of the skillet to the same degree, but will not make the skillet tasty. Now, while the cooling of the skillet is usually less important to us, sometimes it is quite annoying. If we put too much meat on the skillet in one go, the temperature may drop below what is required to produce the Maillard reaction and the meat will cook instead of being seared. That is a disaster in any gourmand's mind (Gordon Ramsey will give any sous-chef a hiding that makes a mistake like that).

The novelty of Glennan's account lies mainly in the characterisation of constitutive production. Without it, precipitating production would be equivalent to the kind of relational realist view offered by Hall (2004) and Esfeld (2011). After all, chained production is really just a succession of precipitating productions, and precipitating production is a succession of constitutive production. Arguably, constitutive production is an elaboration on Salmon's idea of production, but then Glennan focuses not so much on what Salmon calls propagation, i.e. the preservation of a status quo in the object when it is not interacting, but more on the intrinsic *changes* of the particulars when they interact, and then looks at the more coarse-grained structure of the world in terms of precipitating and chained production that is grounded on the close-knit structure of constitutive productions. This is a reasonable thing to do from a pragmatic perspective, in our attempt to make sense of what is happening to us and the world we live in. We need to make a difference between what happens on the scale where nicotine molecules cause the destruction of organic cells in the alveola of the lungs (constitutive production), the scale at which that destruction causes a person to cough (precipitating production), and the scale at which smoking causes cancer (chained production). It is a

shift in perspective from the genuine proximate causes and effects to the distal causes that emerge on a larger scale, and yet it all boils down in the end to the close-knit structure of proximate/constitutive causation that constitutes precipitating and chained production. Indeed, Glennan points out that when we start to talk about chained production, we don't get clear linear chains, but intertwining chains of precipitating production (which are themselves constituted by chains of constitutive production).

Glennan moves in many ways beyond Salmon's account, but I miss the distinction between production and propagation. Production, on Salmon's account, are instances of disruption and change in the causal processes, and Glennan goes into minute details over what happens in one of the causal processes involved. Propagation, on the other hand, are periods during which the object stays the same in the respects relevant for any changes that happen later when it again interacts with another causal process. Propagation is not explicitly accounted for in Glennan's account, either in terms of chained or precipitating causation. Salmon stresses that causal processes, i.e. particulars, are continuously changing even through periods of propagation, but they must still be able to preserve a certain structure through that change. Now, I would not say that Salmon gives an account of propagation, but he at least acknowledges that this seems to be an important feature of the causal structure of the world. In Chapter 6, I will argue that what Salmon describes as propagation is a causal phenomenon, but one that should be cashed out as a combination of *constitution* and *persistence*. Indeed, I will argue that sometimes constitutive causation produces persistence.

To conclude, what is missing from Glennan's account is the same thing missing from the processual version of the transmission theory. How do we explain the exertion of influence or transference of energy between interacting objects? There is no other causal mechanism implied other than what can be understood as transference of energy. As I have already said, that is a mechanism that doesn't generalise to all the cases. I also think Glennan doesn't address the propagation side of things, what happens within an object between interactions when it stays the same.

This completes my discussion of contemporary attempts to say something about causal production while remaining true to the empiricist framework. Now it is time to address those authors that, like Glennan, think that causation must be seen as a continuous process rather than as a relation between events, but who also reject the constraints on speculation imposed by the empiricist tradition; those that talk of powers.

3.6   Causation as a Continuous Process of Production

There are a few friends of powers that, like Glennan, explicitly reject a relational characterisation of causation and suggest instead that causation is a continuous process (for instance, Ingthorsson 2002; Huemer

& Kovitz 2003; Chakravartty 2005; Mumford & Anjum 2011: Ch. 5).
What these writers have in common is that they characterise effects as
beginning at the same time as the cause begins, and both take time to be
completed (wherefore, like Glennan, they deny that the causal relata are
Kim-style events). They also take powers to be central to our understand-
ing of causation (for Huemer & Kovitz the appeal to powers is implicit).
What is different is that Ingthorsson and Huemer & Kovitz, as well as
Chakravartty, consistently talk of *objects* as exerting influence on each
other, while Mumford & Anjum talk of powers doing all the work to
jointly bring about an effect, which in turn is again to be accounted for in
terms of powers (this will turn out to be a decisive difference in Chapter 5).
Huemer & Kovitz, as well as myself, appeal to a greater degree on the
manner in which interactions are described in classical mechanics, while
Chakravartty, Mumford & Anjum, are more prone to appeal to everyday
examples, like sugar dissolving in water. As for the ancestry of such ideas
in the history of philosophy, then Chakravartty and Mumford & Anjum
appeal to individual thinkers that have entertained similar ideas, mainly
to Kant's discussion of simultaneous causation in the Second Analogy of
Experience in *Critique of Pure Reason* (1787), and to Hobbes' discussion
in 'Elements of Philosophy Concerning Body' (1656). Like me, Huemer
and Kovitz appeal to a view we find implicit in classical physics. What I
want to argue now is that the kind of powerful particulars we are now
developing as a novelty was actually the received view across the board
of philosophical traditions well into the Early Modern period.

3.7   The Standard View

It is possible to extract from the causal realist tradition a view of causal
production that has been circulating in philosophy at least since Aristotle,
and about which there was a rough agreement among all the materialist
schools of thought, with slight modifications and variations. I will call it
the standard view. My characterisation of this standard view is extracted
from the way the causal realist tradition is rendered in the works of
Johnson (1924), Bunge (1979, especially Ch. 2), Dorothy Emmett (1985),
Johansson (1989/2004, especially Ch. 12), Dilworth (1996), and Susan
Peppers-Bates (2009), as well as in the works of individual thinkers like
Aristotle, Hobbes (1656), and Locke (1689). This view can even be dis-
cerned in Kant (1787), as an account of how causation appears in the
field of apperception.

   I discern two main variations to the common theme. First, that while
Aristotle listed four different kinds of 'causes'—material, formal, effi-
cient, and final (*Physics*: Bk. 2, 3)—the competing schools, and even later
Aristotelians, tended to be sceptical to final causes. Hobbes, for instance,
argued that there was no intrinsic purpose to the behaviour of inanimate
objects, but allowed that *intentional agents* could initiate actions in order

to achieve a certain goal. Intentions could therefore possibly count as a kind of 'final cause' (Hobbes: 1656: Ch. X, sect. 7). However, intentions are physically inefficacious mental entities, and therefore can only make a difference to the world by accompanying and guiding the physical actions of an Agent. We can intend to drink water all we want, but unless we reach out a hand and lift a glass to our mouths, our intention will not be fulfilled. Hobbes therefore concluded that final causes should be understood as physical actions guided with intention, and therefore count as a sub-class of efficient causes, only present in interactions involving intentional agents. Jonathan Lowe (2006*b*) and Uwe Meixner (2008) have recently expressed similar views about the role of the will in guiding physical actions.

Second, the concept of formal cause arguably changed to exclude the Aristotelian notion of 'essence', to include only the accidental properties of objects, often conceived of as mechanical powers though still described in much the same way that Aristotle often talks of powers as active and passive. For these reasons, I hesitate to attach the name 'Aristotelian' to the standard view, although it is clearly continuous with it in many ways. Indeed, one could argue that much of what today counts as 'neo-Aristotelian' isn't so much a revival of Aristotle's original view but of what could better be described as a rational realist mix of Aristotelian and mechanistic ideas; what I call the 'standard view'. However, I don't mind calling myself neo-Aristotelian, having noted some of the qualifications that go into that term.

While final causes tend to be excluded from the standard view, the other three Aristotelian causes are usually included, albeit in a slightly modified form. The material cause can roughly be understood as the material entity on which the efficient cause acts, and which provides the raw material out of which the effect is produced. The qualitative character of the effect is determined by the properties of the efficient cause and the properties of the material entity on which it acts, those properties together roughly constituting the formal cause. The efficient cause is then the actual exertion of influence of one material entity on another as they interact. These three components are needed to characterise the production of effects in the world of inanimate material objects, while the issue of whether the production is done in order to achieve a goal, or a good, is only relevant in cases where there are intentional agents involved, acting with a purpose.

According to the standard view, then, new states of affairs are brought into existence when an already existing material body, or complex of bodies, changes due to an external influence without which the change would not have occurred, and the new state of affairs never had existed. The kernel of this view comes out clearly in the slogan 'whatever comes to be is necessarily born by the action of a cause'; very likely a paraphrase

of Aristotle's claim that 'in all cases of production there is something that is produced, something by which it is produced and something from which it is produced' (*Metaphysics*: 7, 7; 189), and/or Plato's claim that 'every thing that becomes or changes must do so owing to some cause; for nothing can come to be without a cause' (*Timaeus*: 28a). Typically, the external influence, or cause, is depicted in terms of an 'extrinsic motive Agent' (or, simply, *Agent*), basically some object possessing causal powers, which *acts* upon another object, that object sometimes referred to by the term *Patient* since it lacks a similar active power but is instead able to passively receive the influence exerted by an Agent. Accordingly, a *cause* is the action of some object upon another object, and an *effect* is the change produced in the object acted upon; or, in other words, a cause is the exertion of influence by an Agent upon a Patient and an effect is the resulting change in the Patient.

When I say that a cause is the action of an Agent upon a Patient, then one should not understand the standard view as saying that a cause is merely the action of the Agent, but the *interaction* between Agent and Patient; the passive powers of the Patient are equally important to determine the outcome of the interaction. It is this interaction that I identify with a *process of production*. In other words, the standard view depicts causal production as involving three components: (i) the material causes, which comprise the Agent and Patient conceived of as material bodies, (ii) the formal causes, which are the powers of the Agent and Patient, and (iii) the efficient cause, which is the actual exertion of influence by the Agent on the Patient.

The three components are intimately connected to three fundamental principles about natural reality. The first is the materialistic thesis that nothing comes into being out of nothing or is ever completely annihilated. Craig Dilworth calls it the *principle of (the perpetuity of) substance* (1996: 53), while Bunge uses the term *genetic principle* (1979: 24). The latter captures well what I think is the core idea at play, notably that everything has a *natural origin*; everything comes into being through the alteration of some already existing natural entity from one state to another, due to the influence of some other already existing natural entity. This principle embodies the conviction that there is no magic in the world.

The second principle is the belief that physical action is a distinguishing mark of causal change in the natural world, i.e. any kind of influence exerted by one object upon another. I propose to call it the *principle of action* instead of what it is usually called, notably the *causal principle* (Bunge 1979: 26), or *the principle of causality* (Dilworth 1996: 57). The reason for this change of name is that the principle of causality is today understood very generally as the claim that 'there is a cause to everything', and where it is left entirely open what can count as a cause; even neo-Humean regularity can fit the bill. However, in the causal realist

tradition it is without doubt the exertion of influence that distinguishes causal from non-causal determination.

The third principle expresses the conviction that the world changes in a regular and determinate way, depending on the properties of the objects involved. This is often taken to be equivalent to the belief that the world behaves according to natural laws. That is why Bunge calls it the *principle of lawfulness* (Bunge 1979: 26), but it is also called *the principle of the uniformity of nature* (Dilworth 1996: 55). It is important to point out that this principle does not entail a commitment to the existence of some kind of abstract entities, *the laws*, which somehow rule the behaviour of concrete entities. The principle is equally compatible with the idea—explicitly incorporated into the Aristotelian view—that objects behave in certain 'patterns of being and becoming' in virtue of their intrinsic and universal properties, patterns that we can express in highly generalised law-statements (Bunge 1979: sect. 10.1). That is, the principle is just as compatible with immanent realism—and therefore with powers-based metaphysics—as it is with Platonism about laws. Indeed, when Mumford rejects laws, he is really saying that whatever uniformity there is in nature (and he believes it to be an imperfect uniformity) it is determined by the intrinsic powers of the particulars and not by general laws that directly rule the behaviour of the particulars, or indirectly rule them by determining the causal role of the powers that the particulars bear (Mumford 2005). Anyhow, the three principles described above form the metaphysical framework on which the standard view rests.

Although strictly speaking irrelevant here, it bears mentioning that a Platonist rendering of the principle of the uniformity of nature is only compatible with the genetic principle (everything has a natural origin) *if* we interpret abstract entities as natural entities. But then again, how such entities can meddle with physical interactions in which influence is exerted is a further difficulty. This is a point of disagreement between immanent realists and Platonists.

There are two salient features of the standard view that I would particularly like to emphasise, because they stand in stark contrast to more recent powers-based accounts. First, that it depicts *causal influence* as something that is exerted by an Agent on a Patient. In other words—and this is critical for understanding the main point of this book—the standard view depicts actions as occurring between *persistent objects*, not between *events* or *states*, not even when events are understood Kim-style as consisting of a particular (or particulars) with certain properties at a certain time, i.e. 'a-is-F-at-t_1'. Second, that it does not depict effects as the product solely of the action of the Agent, but of the sum total of material, formal, and efficient causes; the effect is depicted as the product of the way two or more material bodies act on each other in virtue of their powers to produce a change in those very bodies. Let me explain.

3.8 Action or Influence in the Standard View

The idea that influence is exerted between objects runs counter to what has been the mainstream view in philosophy for at least the last century, notably that actions occur between events. It takes some effort to sort out the crooked route by which philosophy converted from what used to be the standard view—that objects act on each other—to the view that a certain kind of event, called the 'cause', acts on a temporally distinct but contiguous event, called the 'effect'. But, as the following quotes from Hobbes show, it is clear that at the beginning of the 17th century the standard view was still alive and well:

> A body is said to work upon or *act*, that is to say, do something to another body, when it either generates or destroys some accident in it: and the body in which an accident is generated or destroyed is said to suffer, that is, to have something done to it by another body; as when one body by putting forwards another body generates motion in it, it is called an AGENT; and the body in which motion is so generated, is called the PATIENT; so fire that warms the hand is the *Agent*, and the hand, which is warmed, is the *Patient*. That accident, which is generated in the Patient, is called the EFFECT.
>
> (Hobbes 1656: Part II, Ch. IX, sect. 1)

What Hobbes says entails that a Kim-style event e_1 existing at t_1 *cannot possibly* act on an effect e_2 existing at t_2, because that effect does not exist until the Agent has worked its influence on the Patient to provoke a change in the Patient, thus bringing the effect into existence. Whatever influence is transmitted from cause to effect must be mediated through the Patient. Indeed, Hobbes continues to say 'when an Agent and Patient are contiguous to one another, their action and passion are then said to be immediate, otherwise, mediate' (Hobbes 1656: Part II, Ch. IX, sect. 2). The contiguity of which Hobbes speaks is *spatial* contiguity, not temporal contiguity; it is the spatial contiguity of two continuously coexistent objects. This is still the basic idea behind the so-called principle of locality, the conviction that objects can only interact with their immediate surroundings.

 Arguably, the reason philosophy stopped talking of influence being exerted between objects and began talking about influence being exerted between events has to do with the empiricist reduction of material, final, and formal causes (including the notion of powers). That shift had only just begun when Hobbes was writing. Indeed, Hobbes is contemporary with Galileo and Bacon, and earlier than Newton, wherefore we can place his works in the very early stages of the scientific revolution (for reference, see Shapin 1996). Hobbes is critical to final causes and reduces

formal causes to the accidents (which he takes to be powers) of the things involved, but retains the idea of causation involving the exertion of influence between an Agent and Patient. The development of modern science and the rise of a corresponding empirically minded philosophy in the years following Hobbes' work challenged the standard view even more, leaving in the end only the notion of efficient cause, but in a modified sense.

Following Bunge, the empiricist reduction of the Aristotelian causes to just 'efficient cause' was motivated on the following grounds: (i) that the efficient cause was the only clearly conceived cause, (ii) that therefore only the efficient cause was mathematically expressible, (iii) that only the efficient cause could be assigned an observable correlate, notably, an event (usually a motion) related to another event (another motion) in accordance with fixed rules (i.e. laws of motion), and (iv) that as a consequence only the efficient cause was controllable in a manner that made possible the harnessing of nature, which was the sole aim of science according to Bacon's instrumentalism (Bunge 1979: 32–3).

In the process, the empiricists not only reduced the notion of 'cause' to 'efficient cause', but also substantially modified the notion of efficient cause from the one we see in Hobbes. To later empiricists, the transmission of influence from an Agent to a Patient, and the powers that were meant to determine the nature of such influence, were mysterious because they were unobservable and hence a mere metaphysical fiction. So, for instance, Hume argued we have no impression of the force exerted between two objects, but we can observe the changes in the state of motion from before to after the objects become contiguous. Similarly, Hertz defined the notion of *Force*, not as influence exerted by one object on another, but as the observed effect of such (alleged) exertion of influence, notably as 'the independently conceived effect which one of two coupled systems […] exerts upon the motion of the other' (1956: 185). Compare this to Newton's definition of 'impressed force' in the *Scholium* to the *Principia*: 'An impressed force is an action exerted upon a body, in order to change its state, either of rest, or of moving uniformly forward in right line' (1686: Bk. I, Definition IV). Newton clearly thought of force not as the resulting change in the state of motion, but as the cause to that change, and he thought of the force as being exerted on a body, not on an event. The event later thinkers identify with the effect, i.e. the change of state of motion, alternatively the resultant state of motion, is here conceived of as the *result* of an impressed force between bodies, but not as the proximate *recipient* of that force.

Following the empiricist reduction of influence to just lawfully connected changes, the efficient cause—formerly the action of Agent on Patient—came to be understood merely in terms of some observable correlate, usually the motion of an object *preceding a* contact with another object. Effectually, the notion of efficient cause was thus transformed into an observable behaviour (e_1) that was predictably followed by another

behaviour (e_2) according to fixed laws of motion. This is a conception of efficiency in which the notion of 'influence' is at best implicit in the manner in which the laws of motion predict changes (or observed differences) in some observed quantities. However, we are getting closer to a conception of causation as an event e_1 being followed by another event e_2; one in which it is perhaps implicitly assumed that the relation between them somehow involves the exertion of influence of something on something else, but where that component is not spelled out. Since it is not explicitly specified how that influence is exerted, and any speculation about it is sure to be judged as metaphysical speculation, then we can easily see how people may have come to think of the relation between e_1 and e_2 as somehow involving the exertion of influence by e_1 on e_2, even if that is not really spelled out. I take it to be the task of causal realists to reintroduce a substantial account of causal influence, but they have to put it in its right place; between objects, not events.

3.9 Two Types of Empiricism

It is worth noting that empiricism comes in two very different flavours: realist vs. anti-realist empiricism. On the one hand we have realists like Locke and Hobbes who clearly think we can know about the world through experience, and only through experience (see, for instance, Locke 1689: Ch. 26; Hobbes 1656: Ch. 1). For them causation is a connection in nature that involves the production of effects by causes, although they believe also that our *understanding* of this connection must be inferred in a rational manner from the content of the ideas of sensation, and the manner in which they were presented to our minds as a result of the influence of the objects on our senses.

Anti-realists like Berkeley (1709: §§ 41, 71) and Hume (1739: 8–13) deny that experience allows us knowledge of an external reality, and deny that rational reflection can make up for the deficiencies of our senses, and therefore take causation to be a *mere* psychological link between impressions/ideas in the mind from which nothing can be inferred about objects that caused those impressions/ideas. As Bunge observes, Berkeley and Hume gave birth to the idea—still widely present in the empiricist tradition—that causation is not an ontological but a purely epistemological category (1979: § 1.1.2). Indeed, some neo-Humeans argue it is an epistemic/conceptual category that we project on a world in which the corresponding ontological category is absent.

However influential the empirical anti-realists have been, they haven't persuaded every empiricist that empirical knowledge of the world is impossible. J.S. Mill, like Locke, clearly thinks of causation as an ontological category involving production, although he too wants to make sense of that category in empiricist terms (Mill 1843: Ch. 5). However, I worry that the conflict between these different flavours of empiricism is not yet resolved, with the result that arguments that really depend on the

Berkeleyan/Humean rejection of experience as a way of knowing about the world—for instance, Hume's discussion of causation and properties (see §3 in the previous chapter)—are used and accepted by empiricists who do not accept the rejection of the possibility of knowledge of an external reality.

To appreciate the difference between realist and anti-realist empiricism, consider the role played by the notion of 'contiguity' in Hume's analysis of causation, compared to anyone thinking of causation as a connection in nature. Contiguity, in the ontological sense, is a constraint on the exertion of influence in accordance with what is known as the *principle of locality*, the idea that things only act on their immediate surroundings because influence has to be mediated by some material substrate; no object can exert an influence on nothing, nor can influence be transmitted by nothing across empty space. Originally—in accordance with the idea that actions occur between *bodies*—this constraint had to do with *spatial* contiguity, as we already discussed in relation to Hobbes. However, according to Hume the notion of action or force has no empirical content and, in his analysis of causation, contiguity therefore only serves as a condition for the psychological association of two events or bodies *in perception*; it is a mere psychological trigger. It just so happens that we are equipped with a principle of association that creates a conceptual link between those objects of perception that *appear* to be spatially and temporally contiguous. For Hume it is entirely beside the point whether they *really* are contiguous because their contiguity has nothing to do with influence passing between them.

3.10 A Return to the Empiricist Reduction of Causes

To repeat, under the influence of empiricism the expression 'efficient cause' no longer came to mean 'influence exerted by Agent on Patient on contact' but instead came to mean more specifically 'observable behaviour of object *prior* to becoming contiguous with another object'. Likewise, the effect was no longer understood as 'change in Patient as result of the exertion of influence' but instead as 'observable behaviour of object *after* becoming contiguous with another object'. Consequently, realist empiricists that wanted to retain the notion of 'action' as an intelligible notion were required to treat 'action' in terms of the precedent motion of one object (the observable correlate of influence) and 'effect' as the subsequent motion of another object (the observable correlate of the reception of influence). In other words, efficient causation came to be considered either as the passing over from cause to effect of some observable quantity, such as motion, or as a mere relation between motions. This idea is at the core of contemporary transmission accounts of causation (Kitcher 1989; Dowe 1992; Kistler 1998).

Arguably, whereas in philosophy the reduction of the idea that objects act on each other to the idea that causation is just a succession of observable behaviour was fairly successful, it didn't impact as strongly on the scientific community. Just consider the following passage from James Clerk Maxwell in 1877:

> The mutual action between two portions of matter receives different names according to the aspect under which it is studied, and this aspect depends on the extent of the material system which forms the subject of our attention. […] if we confine our attention to one of the portions of matter, we see, as it were, only one side of the transaction—namely, that which affects the portion of matter under our consideration—and we call this aspect of the phenomenon, with reference to its effect, an External Force acting on that portion of matter. The other aspect of the stress is called the Reaction on the other portion of matter.
>
> (1877: 26–7)

To Maxwell, influence is clearly exerted between two portions of matter, and not between events.

Everything is not always so clear-cut, though. For instance, we find that in classical mechanics the term 'action' is used in two different ways, one corresponding to the empiricist idea that action is the behaviour of an object prior to an interaction, and the other that action is the influence exerted by one object on another during an interaction. According to the former, 'action' denotes the sum of the kinetic energy of an object over an interval of time, whether or not that object actually influences anything else. This use complies with the idea that actions are to be identified with the motion of an object prior to an interaction. It is in this sense physicists may sometimes talk of pure motion as an 'action', but as Hertz observed:

> the name 'action' for the integral in the text has often been condemned as unsuitable […] these names suggest conceptions which have nothing to do with the objects they denote [*in mechanics, RI*]. It is difficult to see how the summation of the energies existing at different times could yield anything else than a quantity for calculation.
>
> (1956: 228)

The other understanding of 'action'—the way Maxwell uses the term— refers to the exertion of force that occurs on contact between objects. Basically, the first sense of 'action' is of the energy *acquired* by an object as the result of a change in its state of motion and which is then *preserved* in the object over time as long as nothing acts on it. Salmon would call this propagation. The second sense of 'action' is of the influence *exerted* to

produce a change in the state of motion of another object, which complies more clearly with the earlier standard conception of efficient cause. This is what Salmon calls production and Glennan calls constitutive production.

In any case, it is the empiricists' refusal to deal with anything other than observable events—this holds for realist and anti-realist empiricists alike—that motivated the shift in conception of efficient causation, such that actions came to be intuitively thought of as the exertion of influence by an *event* on another *event* (of one motion on another, as it were) rather than the exertion of influence by an *object* on another *object*. In light of this development it is ironic that in physics today—as it always has, really—actions (in the latter sense of 'action' discussed above) are still taken to occur between bodies, not events (see Resnick, Halliday, & Krane 2002: Ch. 3), while philosophers, even those who argue in favour of the reality of forces, will insist that talk of forces acting between bodies is a derivative manner of speaking. This comes out clearly in the following passage by Bigelow, Ellis, & Pargetter, in a paper where they aim to develop a realist conception of forces as causes:

> Often, we speak of forces as operating between, not events or states, but objects. On our view, this is a legitimate but derivative mode of expression. The primary relation holds between events or states; and in virtue of this primary relation there will be various derivative, indirect relations holding between the various salient individuals which are involved in the events.
>
> (Bigelow, Ellis, & Pargetter 1988: 624)

The authors are actually claiming that physics is wrong to speak of forces operating between bodies. However, as far as I can see, philosophy is the only intellectual discipline that insists on forces operating between events rather than bodies.

There are other things to worry about in this brief passage, which relates to what I have previously discussed. For instance, they claim that the relationship holding between spatially contiguous bodies at t_1 is derivative of a relation that holds between the event of their becoming contiguous at t_1 and an ensuing event or state at t_2. However, since these authors agree that causation involves the bringing into existence of effects by causes, then at t_1 there is no future state at t_2 to which the contiguous objects at t_1 could stand in a causal relation or which they could exert an influence on. Are they saying that the relationships between the contiguous bodies existing at t_1 only become causal once they have ceased to exist, notably at t_2?

To my mind, the idea that the relationships holding between interacting particulars is derivative of relationships between events makes little sense unless one also assumes that events are the fundamental building blocks of reality and therefore the primary bearers of powers. On the

basis of the citation above from Esfeld, seen in isolation from the rest of his writing, it would be possible to interpret him as thinking of events as the fundamental building blocks and the bearers of powers, but this would conflict with other passages in which he talks about charge and mass as powers of particles and not of the events in which those particles partake. Arguably, Esfeld resorts to talk about events exerting powers when discussing causation because it is a convenient choice in the current philosophical discourse, where events are by default treated as the relata of causation. However, he reverts back to talk of powers being the properties of things when discussing the nature of properties, because that is the received way of thinking about properties and their bearers. The most reasonable conclusion, from the perspective of substance ontology, is that events are epiphenomenal entities; they only exist in so far as substances are doing one thing or another. I'll qualify that claim in Chapter 7.

I find it difficult to see that the idea that forces operate between events, or that events exert influence on each other, is anything other than a philosophical eccentricity that is the consequence of a temporary phase in the philosophical tradition, during which it was believed that sensory data was the only intelligible content of thoughts. As far as I can tell, none of the major schools of thought before the rise of empiricism conceived of actions occurring between events. The Aristotelians, Stoics, and Scholastics took actions to occur between two material objects whose relation is synchronous. Similarly, the Atomists, both ancient and modern, assumed that the atoms impinge on each other and thus change each other's state of motion, and this is still a standard understanding in particle physics. In the *Large Hadron Collider* at CERN, they are not trying to accelerate events to smash them into other events; they try to accelerate high-energy particles to make them smash together. In chemistry the assumption is that various substances react with each other. Oxygen reacts with some fuel to combust; oxygen does not react with combustion. Furthermore, the common sense conception is that bodies act on each other: the leaden ball dropped upon a pillow acts on the pillow to make a hollow (the falling doesn't act upon the forming of a hollow); the horse pulls the cart (not: the motion of the horse pulls the motion of the cart); the brick hits the window (not: the motion of the brick hits the breaking of the window). I find no other source of the idea that events influence other events than the reduction of the Aristotelian notion of efficient causation to what was believed to be the observable correlates of cause and effect, i.e. motion before and after two objects become contiguous. But, this means that the idea that actions occur between events clearly originates from a tradition that *wanted to remove the notion of influence* and substitute it with observed regularity. Causal realists should better avoid this conception of causal influence.

To illustrate fully how the idea that influence is exerted between events has infiltrated philosophy, consider that the scholastic term 'transeunt

causation' is used today, in certain circles at least, to denote event causation and then in explicit contrast to Agent causation, in the belief that this follows the medieval use. The opposite is the case. Malebranche presented Suarez's notion of transeunt causation as when 'the action is in some way distinct in reality from the material cause that receives the effect', and immanent cause as an action 'received in the operating thing itself … received in the same faculty by which it is elicited' (from *Oeuvres Complètes de Malebranche*, translated citation from Susan Peppers-Bates 2009: 108). W.E. Johnson drew the distinction between transeunt/immanent, in the correct medieval form, as action between continuants vs. within a continuant (1924: xxiii–xxiv), as did C.D. Broad following Johnson (1968: 78), and also Dorothy Emmet who claims that medieval philosophers derived the distinction from Aristotle (1985: 77). Emmett traces the origin of the misconception about transeunt causation to Chisholm's decision to use the term to denote event causation in contrast with agent causation (Chisholm 1966). To be fair, the blame is on his readers, because Chisholm admits that he may not be using the term in the original medieval sense. However, today, due to Chisholm's influence, transeunt causation is often understood as event causation.

According to the standard view, while a cause can in some way be considered an event, the event in question is not itself an entity that exerts an action on the effect; the event *is* the exertion of influence by an Agent on a Patient. In fact, it is not really possible to conceive of the efficient cause as an entity distinct from the material cause, nor of them as wholly distinct from the formal cause, because the efficient cause is the influence exerted by one material object on another, and the characteristics of this influence are determined by the powers of the objects involved; it all comes together as one unity in an interaction, every component dependent on the other. It is in this sense that the standard view is a *total cause conception*; any effect is always the product of the nexus of efficient, material, and formal causes. Consequently, the necessary connection—if there is one—should be looked for in the connection between the total cause and its product, not between the efficient cause alone and the effect. Indeed, that is exactly where Hobbes thinks a necessary connection is to be found:

> an entire cause, is the aggregate of all the accidents both of the Agents how many soever they be, and of the Patient, put together; which when they are all supposed to be present, it cannot be understood but that the effect is produced at the same instant; and if any one of them be wanting, it cannot be understood but that the effect is not produced.
>
> (Hobbes 1656: Ch. X, § 3)

The conclusion is that the standard view characterises influence as occurring between objects, not events, and the necessitating cause as the sum total of causal factors present in an interaction between an Agent on a Patient. There is, however, a very serious problem with this view, first clearly identified by Bunge (1979) and later elaborated on by myself (Ingthorsson 2002, 2007, 2019), notably that it is in a certain sense incompatible with modern physics. This problem is the subject of the next chapter.

4 Causal Production

As I argue in Chapter 3, although contemporary causal realists often appeal to the notion of causal production, they don't really say too much about what exactly it is. They list what they think are the ingredients, but I find there are gaps and faults in the description of the process leading from input to output. Furthermore, they seem unaware of the detailed account of causal production—the standard view—that has been around at least since Aristotle. However, the standard view has its problems. In this chapter I will illustrate these problems and develop my ideas on how the standard view can be modified to overcome them.

4.1 The Standard View: A Reminder

The core idea of what I call the standard view is that new states of affairs are brought into existence when an already existing material substance changes due to an external influence, without which the change would not have occurred, and the new state of affairs had never existed. Typically, the external influence, or cause, is depicted in terms of an object possessing active causal powers, the Agent, which acts upon an object possessing passive powers, that object being called the Patient. Accordingly, a cause is the action of an Agent on a Patient, and an effect is the change produced in the Patient.

I emphasised three features of the standard view. First, that it depicts causal influence as something that is exerted between *persistent particulars*, not between *events* or *states*. Second, it depicts the exertion of influence as being *unidirectional*; it goes one-way from Agent to Patient. Finally, it does not identify causation with the action of the Agent alone, but with the way two or more material objects *interact* in virtue of their powers to produce a change in those very objects. In Hobbes' words, 'the efficient and material causes are both but partial causes, or parts of that cause, which in the next precedent article I called an entire cause' (1656: Ch. IX, 4). It is this interaction between Agent and Patient—which is at the same time an exertion of influence between the two objects and a

change in those objects—that I identify with a *process of production*, and which I believe we should recognise as the instances comprising the ontological category of causation. However, since the standard view is partly mistaken about the exact nature of that interaction, it needs to be modified.

4.2   A Fundamental Flaw in the Standard View

As far as I can see, the standard view is incompatible with one particular finding of modern science—now accepted as fact—that unidirectional actions *do not exist* and that all physical interactions are instead thoroughly reciprocal; whenever one object exerts an influence on another, the other simultaneously exerts an influence of the same kind and of the same magnitude on the first. If this is true, we should infer that the standard view is false and that all notions deriving from or influenced by the idea of unidirectional actions—such as the concept of *mutual manifestation* and *reciprocal disposition partners*—risk being false by the same measure, in so far as they incorporate any aspect of the active/passive or Agent/Patient distinction.

Mario Bunge, in his book *Causality,* first published in 1959, was the first to note the above-mentioned incompatibility of the standard view with modern science. Later editions bore the title *Causality and Modern Science*, and all references in this book are to the third edition published in 1979. However, while Bunge came to the conclusion that the standard view is ontologically inadequate, he did not think a concept of causation based on interaction was viable. He suggested we keep the standard view as a methodologically useful approximation. Below, I will go through the details of Bunge's critique of the standard view, and then argue that he is wrong to dismiss interaction as the fundamental feature of causation. The end result is a total cause conception based on interaction as the fundamental mechanism of causal production. I believe this view is compatible with the theories and findings of natural science, establishes causal connections as necessary in two different ways (see Chapter 5), offers a way to understand constitution and persistence as causal phenomena (see Chapter 6), and which reduces the incompatibility of substance and process ontologies (see Chapter 7).

The idea that interactions are reciprocal is not new. It came to be established as fact in conjunction with the acceptance of Newtonian mechanics generally, and it is one of the features of classical physics that remains unchallenged by relativity and quantum physics. The exact sense in which interactions are reciprocal is expressed most clearly in the third law of motion:

$$\text{Third Law of Motion}: F_{1on2} = -F_{2on1}$$

The novelty of my view is therefore not in the characterisation of interactions as reciprocal, but in the suggestion that we can, and should, understand causation in terms of interaction thus defined. To be more precise, it is new that such an idea is defended. Bunge considers the possibility but dismisses it.

The concept of interaction, as defined by classical mechanics, is one that philosophers constantly misunderstand, because of the connotations that the terms 'action', 'reaction', and 'interaction' have in ordinary language. It requires something of a shift of perspective to get a full grasp of the concepts. The claim 'for every action an equal and opposite reaction'—which is the essence of the third law of motion expressed in plain English—is familiar enough, but it is often misunderstood to mean that the terms 'action' and 'reaction' denote two ontologically different phenomena, of which the action comes first and provokes the reaction. In fact, it is often understood as saying that 'for every influence transmitted from A to B, A must lose as much as B gains', which is really a question of conflating classical mechanics with what the Scholastics called 'transeunt causation', a forerunner to the transmission theory (see, for instance, Bennett 1974: 59; Le Poidevin 1988). In other words, the 'reaction' is often understood as the effect of the action, and therefore it is assumed that 'for every action a reaction' states that effects react back on their causes. On this understanding, causes and effects interact, which is not at all how the third law is understood in science.

4.3   Colloquial Language vs. Science: 'Action', 'Reaction', and 'Interaction'

To begin with, let us become clear about what is meant by 'action' in this context. We are not talking about the behaviour of intentional Agents, i.e. our 'actions', nor are we talking generally about the motion of inanimate objects; we are only talking of the influence exerted by one object on another. The colloquial use of action is much broader. For instance, when a tennis player returns a ball across the net, we may describe everything the player does from the moment she initiates the swing until the racket connects with the ball as an 'action'. However, if by 'action' we mean 'influence exerted by one object upon another', then the racket only begins to act on the ball when they become contiguous. We unreflectively think of the whole movement as an action upon the ball, because the swing is initiated and guided by the intention to strike the ball and because we think the power put into the entire swing is relevant for the outcome. The latter is in a sense true, but we have combined under one description what Salmon calls propagation *and* production.

Indeed, in an unreflective stance we are prone to think of the movement of a brick flying through the air, prior to its collision with a window, as a part of its 'action' upon the window. That is, as if that previous motion

involves a similar kind of effort as the muscle work of the tennis player. However, prior to actually touching the window, the brick exerts no influence upon it and exerts no effort to move through the air. Unaccelerated motion is not an action according to the law of inertia, which says that material bodies continue in their motion in the absence of forces. A moving object, unaffected by anything else, continues in its motion without having to 'do' anything about it. Uniform motion is really a state, a state of motion. We may say that the brick we threw 'flies through the air', as if it was actively engaged in some activity, but of course this is just a figure of speech. The brick only *acts* on the window in so far as it is exerting an influence on it, and that influence first begins on contact and ends as soon as they lose contact.

As already mentioned in Section 3.10, the above-mentioned ambiguity about action is actually found in the jargon of classical mechanics, not just in ordinary language. The term 'action' is used to denote the exertion of force in interactions between objects, and for the time integral of the kinetic energy of a material object, as taken between two times, whether or not that object exerts a force on anything. This may allow talk of uniform motion as an 'action'. Hertz objected to this use of the term (1956: 228), but with no noticeable effect. It is still used to denote the sum of the kinetic energy of an object. However, his objection is still valid and here I use the term 'action' only to denote the exertion of influence between objects in the former sense.

Let us turn to 'interaction' and 'reaction'. In colloquial use, 'interaction' refers to various types of communication where the parts take turns at affecting each other, e.g. in conversation when one person greets the other and the other 'reacts' with an appropriate response. Or, to use a somewhat construed example of communication, when boxers take turns delivering blows. This is not the kind of interaction I am talking about here. I am talking about the mutual exertion of influence between two objects as they come into connection with each other; when brick hits window, or glove meets face. The third law states that the window acts on the brick with a force of the same magnitude as the brick acts on the window, and the same goes for the face exerting a force on the gloved hand.

As for 'reaction', then it is popularly used to refer to the ensuing response to some influence, e.g. when someone 'reacts' to a greeting by greeting back or reacts to a blow to the head by wincing. In this use, 'reaction' means 'effect' or 'response'. In physics 'reaction' refers to the resistance immediately offered by a thing when it encounters another thing, e.g. the resistance offered by the window being hit. That is, 'reaction' does not refer to the *breaking* of the window when it is hit by a brick; the breaking is an effect. It is the resistance offered by the window *before, or while, it breaks* that is a reaction. In the same way, wincing from a blow to the head is not a reaction but an effect. The immediate resistance offered by the face to the glove is a reaction, and it is felt in the

hand of the boxer and may well feel as painful as the blow to the head, say, if a bone breaks in the hand on impact.

4.4   Reciprocity of Interactions

Having clarified the relevant terms, let me explain in exactly what sense interactions are understood to be reciprocal, highlighting the most important points by stating them as numbered proposals at the end of the argument for that particular point. I will then continue assuming the truth of that proposal. The first point, already established, is the scientific fact that whenever a material thing acts upon another thing, the thing acted upon reacts to that action:

  (P1): There are no actions without reactions.

The fact that the apparently passive recipient of an action does not merely 'react' by suffering a change but by exerting an influence on the object that acts on it, is often neglected in discussions about causation. And yet it has been noted—most notably by Bunge—that the fact that there is always a reaction seriously threatens the standard view of causality:

> A severe shortcoming of the strict doctrine of causality is that it disregards the fact that all known actions are accompanied or followed by reactions […] In other words, the polarization of interaction into cause and effect, and the correlative polarization of interacting objects into Agents and Patients, is ontologically inadequate.
>
> (Bunge 1979: 170–1)

The standard view is undermined because the ever-present reaction shows that there are no genuinely passive objects; there are no objects that only receive influence but do not in turn influence the object that influences them, nor are there substances that influence other things without being themselves affected in any way. This fact contradicts the supposed unidirectionality of actions.

  If P1 is correct, the standard view must be modified in some way to incorporate the reaction of the Patient. But will this decisively show that the relation between action and reaction is symmetrical? Can't it be argued that, even if there is a reaction to every action, the reaction is produced or provoked by the action, and therefore counts as part of the effect produced by the action? Let me spell this objection out in concrete detail. According to P1, when a brick collides with a window, the brick acts upon the window and the window reacts to the brick, and both objects suffer a change: (i) the window breaks, and (ii) the brick changes its state of motion (i.e. its momentum, velocity, and kinetic energy). But is it possible to argue that it is the action of the brick that initiates everything, i.e.

provokes the reaction and so in effect produces both the breaking of the window and the change in the state of motion of the brick?

The action of the Agent has been considered to have causal priority mainly for two different but interrelated reasons: (i) because of the assumed distinction between active and passive substances, and (ii) because the action has been thought to be temporally prior to the reaction. I find that when it has been pointed out that there are no genuinely 'passive' substances, i.e. substances that merely receive influence but do not in turn influence other things, then the causal priority of actions over reactions depends on showing that the action is in some way *temporally* prior to the reaction, either because the action begins prior to contact, or because the reaction is temporally retarded in relation to contact. How are you otherwise going to tell which one acts and which one reacts?

The idea that actions begin prior to contact presupposes that the motion of the Agent prior to the encounter is a part of its action upon the Patient, and this, as I argued above, is either to see it acting with a purpose that inanimate objects are incapable of having, or to conflate the two notions of 'action' described above: (i) action as exertion of influence, and (ii) action as a sum of the kinetic energy of an object over a period of time. Whatever uniform motion through space is, it is something different than the influence exerted by a thing upon another when they encounter one another. No influence is being exerted prior to contact.

The idea that the reaction is temporally retarded with respect to contact requires that the thing acted upon initially 'gives way' to the intrusion of the other without offering any resistance. One could perhaps appeal to the way a rubber band initially offers little or no resistance when it is stretched, but successively the resistance increases. While examples of this kind may have some intuitive appeal when considering a rigid object in motion colliding with a soft, elastic body at rest, they don't when considering a soft elastic body in motion colliding with a rigid body at rest. In the latter case it is obvious that the resting body resists the moving body from the first instant they gain contact with each other. And, on closer inspection, the former example is not as convincing as it may first appear to be. In order for the reaction to be temporally retarded, it is not enough that it is initially very small, but must be entirely absent, because if the reaction is initially very small, so is the action. It requires no large effort to begin to stretch a rubber band.

We may perhaps not arrive at decisive answers on this issue by considering common-sense examples, so let us consult the science. I am afraid the received scientific understanding leaves no leeway for an interpretation in which the reaction is provoked by the action. According to classical mechanics, action and reaction occur *simultaneously* in opposite directions, and both are of equal magnitude. This relationship is clearly expressed in the third law of motion, which says that whenever any

object whatsoever exerts a certain force on any other object, the latter always acts on the former with a force of equal magnitude and at the same time. The law depicts the action and reaction, or force and counterforce, as being thoroughly reciprocal to the degree that 'we are free to consider either of them as the force or the counterforce' (Hertz 1956: 185). We are free to do that because we lack any objective means to give either of them priority. Both happen simultaneously and are of the same magnitude and of the same kind. The explicitly expressed consequence of all this is that physics does not consider interactions as being composed of ontologically different kinds of influences, an 'action' and a 'reaction', of which one kind is only a response to the other. The terms 'action' and 'reaction' are only taken to reflect the subjective aspect under which the scientist considers the interacting objects, i.e. according to which changes the scientist is interested in: the changes in the window, or in the brick. The point comes out beautifully in the previously cited passage by James Clerk Maxwell, where he says that the mutual action between two objects 'receives different names according to the aspect under which it is studied, and this aspect depends on the extent of the material system which forms the subject of our attention' (1877: 26). However, as he points out, when we 'take into account the whole phenomenon of the action between the two portions of matter we call it Stress' (1877: 26). Physics really treats action and reaction as 'two aspects of a single phenomenon of reciprocal action' (Bunge 1979: 153). That phenomenon is called 'stress'.

Perhaps an even better way to illustrate the point is to consider that the idea that the action provokes the reaction only has intuitive appeal in a subset of all interactions, while the idea that neither action nor reaction can be given priority generalises to all the cases. In other words, the standard view only has intuitive appeal when applied to cases when one object is in motion while the other is at rest, and when one object appears to us as more active, bigger, sturdier, etc. It doesn't work if you consider cases such as two identical billiard balls, moving at equal speed in opposite directions, colliding head on. Or, two cars (or any object you like) of the same make, moving at equal speed in opposite directions, colliding head on. In those cases, it appears clear that each acts on the other simultaneously, to the same magnitude, but in opposite directions. I will now present as the second proposal what classical mechanics takes to be an empirical fact that:

(P2): Interactions are thoroughly reciprocal.

It bears mentioning that, while the first and second laws of motion are known to fail in the domain of very fast-moving and massive objects (i.e. where relativity deviates from classical mechanics) as well as in the quantum realm, the third law is still assumed to hold good.

4.5   Bunge's Rejection of Interaction as the Basis of Causation

Bunge acknowledges the ontological significance of the fact that there are no actions without reactions and takes it to show that the standard view cannot be anything but an approximation of causality. Nevertheless, Bunge rejects the idea that causality could be explained in terms of inter-action, 'if only for the simple reason that material objects are in a state of flux, so that generally the action has over the reaction the definite "advantage"—to use an anthropomorphic expression—of priority in time' (1979: 162). He adds:

> The frequent asymmetry of interactions, as well as the fact that processes in which the antecedent disappears altogether cannot be described as interactions (although they involve reactions upon different objects), renders interactionism inadequate as a universal doctrine. Causation cannot be regarded as a particular case of inter-action because the latter lacks the essential component of irreversible productivity.
>
> (1979: 170–1)

To sum up, Bunge has four objections against interaction as a universal doctrine of causality: (i) that when we consider the fact that things are in a state of flux, actions are temporally prior to the reaction, (ii) that interactions are often so asymmetric that the reaction and its effect can be quantitatively neglected, (iii) that some processes, e.g. the spontaneous decay of various kinds of micro-particles, cannot be described as interactions, and (iv) that if the action does not give rise to the reaction then interaction contains no element of productivity, which he rightly considers an essential component of causality. I am afraid that these objections really show that Bunge misunderstands the nature of interactions.

I will discuss the objections in a different order than they appear above but begin with the first objection that the action is temporally prior to the reaction. This objection is based on not cleanly separating reaction from effect. When talking generally about interaction Bunge says that 'the effect always reacts back on the input' (1979: 170), and when discussing gravitational interaction he says that '[e]very change produced by m_1 on m_2 reacts back on m_1' (1979: 150), and when discussing the sense in which interactions are reciprocal, he only states that there is a reaction to every action, and that the reaction is equal to the action, but does not say that they occur simultaneously. In fact, he says that every action is accompanied or *followed* by a reaction (1979: 170). But reactions are not effects, they are exertion of influence, and that influence is exerted simultaneously with the action.

Indeed, Bunge's third objection, that interaction cannot handle processes where the antecedent ceases to exist, is only intelligible if one

assumes that he does identify reaction with effect. Bunge uses an example of spontaneous decay of particles to illustrate his point, notably of a pion into a muon with the emission of a neutrino:

> In this case, the parent particle (π) is unstable; it decays spontaneously (that is, without any known extrinsic cause, though presumably as a result of an inner process) […] This is an irreversible, typically genetic process, the thing furthest from interaction—despite which meson theories usually treat this process as if it were a mutual action between coexistents. More exactly, the parent-child connection existing between the pion and its descendants is described as an interaction eliciting that very transition—despite the fact that the products are not yet born.
>
> (1979: 163)

If the muon and neutrino only exist after the decay of the pion, then obviously there cannot be talk of interaction *between* them and the pion. The pion cannot act on something that does not yet exist, and the muon and neutrino cannot react back on something that no longer exists. Note that Bunge does not object generally to the possibility of products reacting back on what produced them, but only in cases where the 'producer' ceases to exist in the process of producing the effect. I agree that the decay cannot be described as an interaction between the pion and its descendants, because, as I believe, interactions do not occur between cause and effect, but only between two objects whose mutual action is arbitrarily called 'action' vs. 'reaction'.

  I think that the example is really a worse anomaly for the standard view. If the decay occurs in the absence of an external cause, then it isn't a case of causation at all, because the standard view defines causation in terms of changes produced by an external influence. Then we are really dealing with a non-causal change, and the example falls outside the category of phenomena being discussed, namely causal production. However, if the decay is a result of an inner process, as Bunge suggests, then it can be treated as a product of the interaction between the constituents of the unstable pion—a quark and antiquark—which ultimately results in the emission of a muon and neutrino. My conclusion is that Bunge identifies the reaction with the effect and therefore thinks that if the Agent ceases to exist in the process of producing the effect, then there is nothing for the effect to react back on. Secondly, he cannot see a process internal to the pion as being the cause of its decay, because he restricts the meaning of the term cause to 'extrinsic motive Agent' in accordance with the standard view:

> Now, by definition, of all kinds of cause, the efficient cause is the motive or active one, it is, moreover, an Agent acting on things *ab*

extrinseco and one that cannot act on itself. The efficient cause is, in short, an *external* compulsion, hence, an essential mark of (efficient) causation is externality.

(1979: 173–4)

Let us now turn to Bunge's second objection, that interactions are often so asymmetrical that the reaction can be neglected. This objection is an argument in favour of the standard view, but it is not really an objection to interaction. It is a justification of the application of what Bunge calls the *causal approximation*. He makes a point of the fact that interacting things always affect each other mutually with equal force (despite *also* talking about it as the effect reacting back on the cause), but then points out that, when there are large quantitative differences between the things, the larger thing will hardly be affected at all by the interaction. Bunge uses gravitation to illustrate his point: 'Only if one of the masses is much smaller than the other (for example, a stone as compared with the whole Earth), can the greater mass be regarded as the *cause* of the acceleration of the smaller one, and the reaction of the latter's motion be *quantitatively* neglected' (1979: 150). That is, even though the interaction really is reciprocal, sometimes the effect produced by the reaction is so small that it can be *neglected*, and the interaction be treated as being approximately causal in the standard sense. Bunge argues that '[i]n some cases this involves no error at all' (1979: 150).

It is justified to ask in accordance with which standards is the reaction negligible? The answer is 'in accordance with the explanatory interests of the observer': 'by a suitable choice of the reference system (change from laboratory system to center-of-mass system) [...] an interaction problem is thereby transformed into an ideal causal problem' (1979, 151). That is, Bunge argues that, when the effect on one of the things is for all practical purposes negligible, we are methodologically justified to do the kind of aspect-shift that Maxwell described so well, i.e. sometimes we are justified to neglect those aspects of reality that we have no explanatory interest in.

I will not argue that the application of the standard view is *never* methodologically justified in scientific practice. But the fact that interactions can often be approximated to fit the standard view, by a suitable choice of what we think is negligible, is not an *ontologically* valid argument against interaction. Indeed, as Maxwell noted, we are often interested in both sides of an interaction, in which case it is inappropriate to apply the approximation. For instance, in playing billiards we need to be equally concerned about putting the cue ball in a good position for the next shot as with putting a colour ball in a pocket. Consequently, it appears we are only sometimes justified in applying the standard view, but we are always justified in regarding the interaction as reciprocal. Interaction generalises, the standard view doesn't.

There is clearly a tension between Bunge's ontological and methodological considerations. He makes a point of the ontological inadequacy of the standard view, regarding the relation between the action and reaction, but prefers the standard view for its methodological utility in cases where there is a large quantitative difference between the effects produced in an interaction, and because he thinks interaction does not involve production. While I acknowledge the methodological concerns, my primary purpose is to find an ontologically adequate account of causal production, and I think, contrary to Bunge, that interaction, when properly understood, involves production.

Bengt Molander (1982, 140–3) has also considered the suggestion that the standard conception of causality should be replaced by a conception based on interaction but dismisses it on similar grounds as Bunge. He thinks that even if interactions are reciprocal it will still be true that the brick breaks the window, although at the same time the window causes the brick to lose momentum. My response is that such statements are not inappropriate or false as explanatory statements in relation to our explanatory interests. But it does make the individuation of causes and effects dependent on a subjective choice of reference, wherefore it is ontologically inadequate.

Bunge's objection that interaction does not involve production is the most serious objection to interaction, but also one that is easy to counter. His objection is based on the mistaken view that the symmetry we find between interacting objects translates to a symmetry between cause and effect. In Bunge's words: 'Let us agree to call interactionism, or functionalism, the view according to which causes and effects must be treated on the same footing, in a symmetrical way excluding both predominant aspects and definitely genetic, hence irreversible, connections' (1979: 162). This view, he claims, may be regarded as a 'hasty extrapolation of the mechanical principle of the equality of the action and the reaction' (1979: 162). I on the other hand think he has wrongly understood the mechanical principle as a statement about the relationship between cause and effect, when it should really be understood as merely a characterisation of the cause; a cause is reciprocal action.

4.6   Interaction Involves Production

Instead of taking the symmetry of action and reaction to show that causation cannot be modelled in terms of interaction—because it depicts the relation between cause and effect as symmetric—I propose we take it to show that the reaction cannot really be an effect, and that the action cannot really be a cause in its own right. Really, I propose we take it to show that we must re-examine the standard conception of 'cause'. By abandoning the standard view that causes must be understood in terms of 'extrinsic motive Agents', an idea arguably originating from outmoded pre-Newtonian physics, we are free to conceive of the interaction as a

whole as the cause, and the change in the compound whole of interacting things as the effect. According to this view, the relation between action and reaction is symmetrical, but it does not follow from this that the relation between the cause (i.e. the interaction) and the change it produces is symmetrical. Bunge overlooks this possibility because he takes causes to be *by definition* external to the changing entity.

It is in fact difficult to conceive of interactions without thinking of them as the production of changes in the interacting things, when, as here, interaction refers to the mutual influence of two things upon each other. The notion of force has always been understood in terms of production of changes. Newton defined it as 'an action exerted upon a body in order to change its state, either of rest or of uniform motion in a straight line' (1686: 13), and Hertz defined it as 'the independently conceived effect which one of two coupled systems […] exerts upon the motion of the other' (1956: 185). If force, meaning the action of one thing upon another, cannot be separated from the changes it produces, then neither can interaction be separated from the production of changes, since an interaction consists in two reciprocal forces. That the forces are reciprocal means that neither can exist without the other, and that, therefore, neither can produce the other. However, together they can produce a new state of affairs.

It may be objected that static interactions are counterexamples to the claim that interactions always result in changes (see, for instance, Baltimore 2020: sect. 4.2). Two bricks leaning on each other exert a force on each other, and yet nothing changes. Baltimore suggests we instead use the term 'result' to describe the outcome of interactions. Now, while nothing important hinges on this point—the use of 'result' rather than 'change' still admits that interactions are productive—I still am inclined to insist that the result be understood as 'change'. Many of the 'results' we would *prima facie* identify as non-changes, because no change turns up in our experience, seem to me to be changes. Consider the two bricks leaning on each other. According to physics, each brick is in a state of downwards motion; they are falling, just like any object resting on the surface of the Earth. However, that falling is continuously thwarted by the other brick. From a human observer's point of view nothing appears to change, but in physical terms, their state of downwards motion is continuously being changed to not falling. Now, whether or not this account of static interactions holds true, I take it still to be the case that:

(P3): Interaction involves production.

4.7 Unidirectionality Due to Agency Bias

I have so far argued that inanimate material objects cannot be objectively distinguished into Agent and Patient, and thereby that the distinction between causes and effects *in terms of* 'actions of an Agent' and 'changes in a Patient', respectively, is ontologically inadequate. It involves a neglect

of the reaction of the Patient and change produced in the Agent. The causal production of a state of affairs involves the interaction of things, and this interaction is reciprocal. That is, in explaining the causal relation between events, there is always a story to be told about the interaction between the things involved. However, this account does not sit easily with our everyday view of the world, which appears largely to consist in asymmetries.

I do not deny that we often experience interactions as being asymmetrical, but I suggest that this experience of asymmetry is subjective. On the one hand, it boils down to the kind of choice of perspective that Maxwell describes so well, and which is guided by what we consider important from the perspective of our explanatory interests. It does appear to be more 'fatal' for a window to be smashed to smithereens than for a brick to lose momentum. But, surely, the sense in which the breaking or destruction of objects is 'fatal' is a subjective evaluation. A smashed window is inconvenient to a house-owner in a way that the loss of momentum by a brick is not. That is why we attend to the breaking of the window as an important effect of the interaction, while the loss of momentum by the brick is a negligible side-effect.

To be sure, different things change in different ways when entering into interactions, and these differences can be quite dramatic, but does not justify giving one of the interacting things, e.g. 'the flying brick', the privilege of being the sole producer of the subsequent change, e.g. that the window breaks and that the brick loses momentum and velocity. The objective character of the subsequent state of affairs as a whole is determined jointly, and reciprocally, by all the things involved. We can easily explain the difference in how things change in interactions by appeal to the differences in the intrinsic properties of the things themselves as they enter into an interaction, instead of to an asymmetry between the influences they are subject to. Different things change differently when they are influenced in the same way, not because of an asymmetry in the way they influence each other but because they are different before being exposed to the influence. If you exert a certain force on a window it will break. If you exert the same force on a brick it will only change its state of motion.

On the other hand, our tendency to single out objects in motion as particularly active and important in the production of an effect can be explained by appeal to the view of F.P. Ramsey (1929), R.G. Collingwood (1940), D. Gasking (1955), and G.H. von Wright (1974), that the conception of cause in terms of 'extrinsic motive Agent' is biased by our conception of agency. The agency view states that we could not *know* causality, or could not form the *concept* of causality, unless we knew from ourselves and our actions what it is to act as a cause. The agency view might then be taken to claim that the concept of *efficient cause* is derived from our knowledge of ourselves as active Agents. This is indeed the view of Evan Fales. Fales accepts Hume's thesis that knowledge of

necessity and *forces* are not given in outer experience but claims against Hume that knowledge of *action* and *force* is given in our inner experiences of the effort exercised by our bodies (1990: 11–14). The question is: to what extent does our nature as intentional Agents, with all the cognitive capacities we have, memory, prediction, etc., bias our view of the objective workings of causality in the world of inanimate material objects? Is our conception of ourselves as 'efficient causes' really applicable to how changes are produced in inanimate material objects by other inanimate material objects? What I have argued so far suggests that we cannot divide inanimate objects into Agents and Patients, and I have tried to abstract the component of intention from our notion of efficient cause to arrive at a more objective notion.

4.8   Interactions Are Causally Fundamental

If it is correct that causal production always involves interaction between coexisting things, then it follows that interaction is ontologically prior to the causal relations believed to hold between temporally distinct events. I think this is something we can observe in our everyday lives, when we reflect upon it. We know we cannot accomplish anything without acting upon things in some way or another, and we always feel their resistance (reaction) when we do. We always suffer a change ourselves when accomplishing a change in something else, but as long as this change is negligible to our purposes, then it will go by unnoticed. It is indeed because we feel the resistance of the things we interact with, that we can adjust the effort we make to their resistance.

We also know we value different consequences of interactions in various ways. Indeed, different people value the same consequences differently, and we know that the mechanical forces involved have little or nothing to do with their value. A tiny little push on the edge of a cliff can have dramatic consequences, while a full body tackle in the ice-hockey rink does not matter a jot. We even disagree on what is to be considered as the Agent, i.e. the efficient cause, e.g. in deciding questions of responsibility. It may not be intuitively given at first glance that interactions are thoroughly reciprocal, especially when it comes to intentional acts, but I think it is convincing on second reflection in light of everything that has been said.

(P4): Interaction between coexisting objects is ontologically prior to the one-sided existential dependence relation between two temporally distinct events.

I am prepared to leave it open whether intentional beings could in some sense be examples of 'extrinsic motive Agents', in accordance with the standard view. That is, whether individuals with the capacity to a)

perceive themselves in relation to other objects, b) have preferences and desires, c) predict various consequences of various actions before they are performed, d) choose the action that is perceived to lead to a desired effect, e) adjust the effort and direction of the action to the reaction of other objects, may perhaps be capable of *initiating* actions, and not just interact with the environment in the sense described above. This has in fact been suggested by Lowe (2006*b*), and Meixner (2008), and Hobbes before them (1656: ch. X, sect. 7). The list of Agential capacities is not complete, nor do the distinctions aim to represent logically, or otherwise, independent capacities, but rather interdependent capacities. I just want to make a sharp distinction between how *we* calculate, and perhaps initiate, our actions in the world, and how physical objects in general actually interact in the course of real events. To what extent intentional actions differ from interactions is beyond the scope of this paper, but I do think our cognitive abilities bias our understanding of the nature of causal production in the world of inanimate material objects.

4.9   Necessary Connections

In Chapter 5 I will address more thoroughly the question of causal necessity, arguing that if the character of an interaction is entirely determined by the properties of the interacting things, and the properties are universal powers, then whenever two objects of a certain kind interact in a certain way, the very same kind of effect is always produced. We would then have, in principle, constant and invariable type-type relations, i.e. a necessary connection. Here I will argue, however, that there is another kind of necessary connection to be found, a genetic token-token link.

Assuming, as I do, what appears to be an empirical fact, that nothing can be produced *ex nihilo,* then we must consider *out of what* a new state of affairs is produced. A potter cannot produce pots without clay, factories need raw material out of which they can produce ready products. Without raw material that can be altered into a new shape, nothing can be produced. What is the raw material out of which a new state of affairs is produced, and where does it come from? I propose that a new state of affairs is produced by the interaction of things, out of the very substance those interacting things are made of, i.e. the state which is produced is made of the same substances as were involved in the interaction. It is the same substance that constitutes a window and then a pile of broken glass. It is the same brick that comes flying through the air and then lies at rest in that very pile of broken glass.

Since interaction requires at least two things, then the production of a new state of affairs requires a unity of things or substances, and we can speak of that unity as a *compound substance*. My understanding of a compound substance is roughly that of a system, as this term is used in physics. If the state of affairs produced by the interaction is produced out

of the substance of the interacting things, then the state of affairs produced by an interaction consists in the very same compound substance that was involved in the interaction. On this account, the relationship between two different but in this sense causally related states of affairs is necessary in the sense that they are necessarily constituted by the very same substance. Causally produced change is then an alteration in the state of a compound substance, brought about by that very substance itself, through the interaction of its parts. Note that this also involves that the production of a state of affairs is always at the same time a destruction of an existent state of affairs. The coming into existence of a state of affairs, that does not come into existence through the transformation of a system from one state to another—the first going out of being as the second comes into being—must be creation *ex nihilo*.

   (P5): Two causally related states of affairs are necessarily constituted by the very same compound substance.

4.10   Production Requires Endurance and an *A* View of Time

If causation involves production, then things must endure rather than perdure, because perdurance is incompatible with production, if creation *ex nihilo* is ruled out. Also, this kind of production is incompatible with the *B* view of time. The argument for these two points is spelled out in greater detail in (Ingthorsson 2001, 2009, 2016: Ch. 7). Very briefly, if it were assumed that things persist by perduring, i.e. by having temporal parts, and all times exist in parity, then each part must be a distinct and independent substantial entity, each existing at that time without ever coming into or going out of being. If that is the case, then I fail to see what could be produced, by what, and out of what. If one temporal part of a thing is broken and an earlier part is whole, without the part being whole having changed into being broken, then out of what was the broken part produced? The broken part must either (i) always have existed, in which case it was not produced, (ii) be produced by being brought into reality from outside reality, (iii) be produced out of the substance of the Agent, in which case a brick would have to be able to change into a pile of glass and the problem would shift to the production of the brick lying in the pile of glass, or (iv) be produced *ex nihilo*.

   Indeed, most perdurantists would agree and say that, since things perdure, there is no production, everything simply exists, albeit at different times. Here we have the neo-Humean idea of the world as a mosaic of local matters of fact, where everything is loose and separate and nothing originates from anything else. In such a world there can be no generic connections, since they would constitute a substantial connection. I want to make as good sense as I can of the idea that causality involves production, but reject the possibility of creation *ex nihilo*, and will then have to

conclude that if causality involves production then substances persist by enduring. I take myself to be in agreement with Haslanger (1989), who argues that the perdurantist position commits to a neo-Humean interpretation of causality, i.e. to a correlation view of causality, in which production has no place (Ingthorsson 2001). Indeed, it is widely argued that change generally speaking, not just causally produced change, necessarily requires endurance of the substances involved (*Metaphysics*: Bk. 12, 2; Lowe 1998; Mellor 1998; Simons 2000).

Again, if it is assumed that all times exist in parity, which is the basic tenet of the *B* view of time, then it follows that every state of every physical system exists in parity with each other, and this requires a perdurantist view of persistence. Perdurance, as I have argued, is incompatible with production, and if the *B* view is committed to perdurance, then the *B* view is incompatible with production. This is a simple consequence of Lewis' problem of temporary intrinsics (1986*a*: 202*ff*). The only view of time that will allow a state of a system to cease to exist as it gives rise to a new state is the *A* view of time (my favoured version is *presentism*). Now, since the focus of this book is on causation and not time, I refer to my other publications regarding the connection between different views about time and different views about persistence, and their connection to different views on causation (Ingthorsson 2001, 2002, 2009, 2016).

Two states of affairs that are causally related, in the sense given above, will be different states of the very same compound substance, which has changed from one state to the other due to the interaction of its parts. The production of changes cannot then really be construed as involving an external influence on the changing entity (although the parts of that entity act mutually and externally on each other), but to influences *internal* to a changing compound, or aggregate of substances. That is, on this account, causally produced change is always a change within a system.

(P6): Causally produced change is always internal to a compound substance.

4.11   Simultaneous Causation

To characterise causation in accordance with the third law of motion not only challenges the ontological priority of the action. It also requires us to think of causes and effects as occurring simultaneously. The idea is not new, but is controversial even though recently it has been gaining in popularity. Kant accepted that most effects are simultaneous with their causes, which caused him some worries about how to understand the ontological priority of causes (1787: B233–56), and today this view is championed by authors like myself, Michael Huemer & Ben Kovitz (2003), and Mumford & Anjum (2011: Ch. 5). Like myself, Huemer &

Kovitz appeal to the way physics describes interactions, and Mumford & Anjum partly appeal to common-sense examples of the kind Kant suggests, and partly to Huemer & Kovitz's interpretation of physics. I think we are thus in very good agreement. There are plenty of objectors though. Bunge's objection to interaction, previously discussed, is an objection to simultaneous causation, and more recently we find Robin Le Poidevin (1988) and Tobias Hansson Wahlberg (2017) attempting to prove that causes and effects cannot be simultaneous, because it would entail that things will be in two incompatible states, and/or violate the conservation of momentum. Hansson Wahlberg explicitly relates to myself, Mumford & Anjum, and Huemer & Kovitz, as well as to Le Poidevin.

Le Poidevin appeals to what he calls *the principle of reciprocity*, 'a necessary part of any cause is itself affected as a direct result of that cause's bringing about its effect' (1988: 152), which Jonathan Bennett calls the 'Balance Principle', i.e. 'in any fully intelligible causal transaction something gains what something else loses' (1974: 59), and says that this has been part and parcel of the pre-Newtonian mechanics of Leibniz and Descartes. Le Poidevin illustrates the consequences of the principle for interactions between two objects A and B: 'For any A and B, if A's being F at time t causes B's being G at t′ then A is no longer F at t′' (1988: 155). The core idea at play is really that causation is the transmission of something from an Agent to a Patient, and that the Patient's gains must be equal to what the Agent loses, otherwise something has been added out of nothing or completely annihilated (depending on whether B gains more or less than A loses). Having that understanding in mind, Le Poidevin attempts to show that a contradiction ensues from the idea that causes and effects are simultaneous:

> Consider the following example of a moving billiard ball A's striking another billiard ball B, which until then was stationary. As a result, both balls are in motion. At the moment of the encounter, A's momentum was M_A. After the encounter, A's momentum is M_A* and B's is M_B. Clearly, M_A* must be less than M_A: A cannot have brought about movement in B without a reduction in its own momentum (otherwise momentum is not conserved). Now A's momentum at the moment of the encounter is a necessary part of the cause of B's subsequent movement. So if cause and effect are simultaneous, then A's having momentum M_A must be simultaneous with B's having momentum M_B. But if B has momentum M_B, then A's momentum must be M_A*. So A's momentum, at the time of the effect, must be both M_A and not M_A. Since this is absurd, causation cannot be simultaneous in this case. What happens rather is that at the time of the encounter, A's momentum is M_A, and at times thereafter, M_A*.
>
> (Le Poidevin 1988: 157–8)

In short, the argument is that if causes and effects are simultaneous, then A must have M_A simultaneously with B having M_B, yet must also simultaneously have lost M_A to have caused M_B.

Hansson Wahlberg argues in exactly the same way, but instead of describing the consequence as a contradiction he argues that it entails the doubling of momentum, which is a violation of the conservation of momentum:

> Now, assume first, for simplicity, that the billiard balls are perfectly rigid and that the collision is instantaneous, occurring at an instant of time *t*. Then, if a's momentum is passed to ball *b* at *t – b* acquiring the momentum (the effect) simultaneously with *a* being, for the last time, in possession of its momentum – the sum total of the objects' momenta is doubled at *t*. The consequence is that the law of conservation of system momentum is violated at *t*.
>
> (2017: 113–14)

My objection is that both are using an example clearly intended to conform to classical mechanics, but describe it in terms of the philosophical two-place model involving a conception of causes and events as instantaneous Kim-style events in which one transmits a quantity to the other, which I don't think is suited to the Newtonian understanding of what happens in interactions. That is, we are given M_A as the cause, i.e. the state of A when it *begins* to exert an influence on B and are supposed to accept that the exertion of influence can be completed instantaneously. And we are given M_B as the effect, i.e. the *result* of a change in momentum instead of the *change* of momentum itself. It is true that the quantity of momentum can be calculated for an infinitesimal timepoint, but that doesn't mean that we should understand the exercise of an influence or a change in momentum in terms of instantaneous entities. Here I am with Glennan (2017) that causal realists must assume causes and effects to be activities that take time rather than instantaneous states.

I propose we should understand the interaction as follows. Assume A and B have momenta M_A and M_B at the instant t at which A and B come into contact. From that moment on A and B *begin* to mutually exert a force on each other, and simultaneously their respective momenta *begin* to change and continue to do so as long as A and B mutually exert an influence on each other. At no point in time is it the case that A and/or B have *and* do not have a given momentum M_A*/M_B*.

The key thing is to realise that the claim 'causes and effects are simultaneous' does not mean that the cause and effect are completed instantaneously (that only follows if you treat causes and effects as Kim-style events). The idea is instead that the exertion of an influence and the

production of a change occur simultaneously but over a temporal inter-val, producing a succession of states (see also Huemer & Kovitz 2003).

Now, while physicists never forget to mention that interacting objects mutually exert forces on each other, they primarily explain interactions in terms of unidirectional transferences of quantities from one object to the other. In accordance with my earlier criticism of the transmission theory in Chapter 3, I believe this understanding is ontologically inadequate, even though it offers a useful method for calculation that usually isn't misleading. The problem is that it doesn't generalise to all the cases. It only works in cases that intuitively fit to an Agent-Patient interpretation but fails in symmetric interactions, such as when two identical billiard balls, moving with equal speed in opposite directions, collide head on. They compress on collision and are then pushed off again in the opposite direction to the one they had before, but neither loses momentum while the other gains. Both equally change their direction of motion. How are we to understand the change in terms of transference of momentum in these cases? We have a choice between thinking that each transfers equally to the other (momentum swap) or we assume the interaction only results in a change in the direction of respective momenta (or change in the quantity in asymmetric cases), but not really that momentum is transferred. I think we should let the concept of force do the work it was initially meant to do, notably to explain why colliding bodies change the momentum of each other, and not offer a double explanation (force + transference). My reason to favour that explanation is that it is sim-pler and works for the full range of cases, while the transference model only works for asymmetric interactions. Now, whether we accept this suggestion or not has no consequences for the point that simultaneous causation does not really lead to contradictions as long as we accept that exertion of influence and changes in state of motion are temporally extended phenomena.

Hansson Wahlberg develops another argument. Assuming that causal influence cannot be transmitted faster than light, he argues, then it takes time for the momentum to transfer across the spatial extension of the two combined bodies, wherefore causes cannot be simultaneous with their effects, at least if we are dealing with anything larger than point particles. So, even if we assume bodies to be rigid and that there is only action on contact, then because the momentum must be distributed across the whole spatial extension of each object, some of the momentum must have some distance to travel—however small—and this cannot happen instantaneously. My response is that he is again assuming that causation is transference, and that transfer cannot occur gradually. His argument poses no problem if momentum is not transferred but only changed as I have suggested. But it is not a problem even if we assume transference,

if the influence and the change are treated as activities that can occur gradually over time. If influence takes time rather than happens instantaneously, then there is plenty of time for the interaction to start slow and gradually increase as the action is propagated within each object, which is actually what happens in non-rigid objects, i.e. why they are compressed on contact. The conclusion is that on the interaction view I propose, and even on a transmission account that allows influence and change to occur gradually, it follows that:

(P7): Causes are simultaneous with their effects.

4.12   Causation a Form of One-Sided Existential Dependence

It should be noted that in (Ingthorsson 2002) I claimed that causes preceded their effects in time, even though I also argued that an influence and the change it produces happen in the way I describe above, wherefore (P7) represents a change in my account. In 2002 I was too conformist and wanted to find something that could roughly resemble the cause-effect relation depicted by the traditional two-place model. But the result was that I did not adequately distinguish my understanding of the mechanism of production (interaction) from my understanding of the relationship between the products of that mechanism, i.e. between the successive states of affairs that interactions produce. I now turn to describe the latter, i.e. what is the relation between the state destroyed in an interaction and the state produced by that very same interaction. I believe it is a kind of one-sided existential dependence and that it is this relation that has traditionally been identified as the causal relation.

Strictly speaking, it is the interaction between the parts of the compound substance that produces a change in the compound from one state to another, the states themselves are always products. That is, the interaction is the process in which the effect is produced. But, I suggest, the relation between any state *from* which the compound substance changes and the state *to* which it changes, due to the interaction, nevertheless be characterised as being one of producer to product, keeping in mind that it should only be understood as a relation of origin; one state comes into being out of the other. Any given state of a substance, although merely a temporary form of the substance, cannot be separated from the substance itself; the state cannot exist without the substance. If, as I have argued, the state to which the compound substance changes is necessarily constituted by the very same substance as the state from which it changes, then the state to which the compound changes is for its existence dependent on the state from which the compound changes, because the state to which it changes 'inherits' both the substance, and the character, from the state the compound changes from.

The temporally distinct states of any such compound, related as producer to product through the interactions of the parts of the compound, will then hold a relation of one-sided existential dependence; the state to which a compound substance changes is dependent for its existence upon the state from which the compound changes, but the state from which it changes is not dependent for its existence upon the state to which it changes.

> (P8): The state produced by an interaction stands in a relation of one-sided existential dependence to the state destroyed by that very same interaction.

4.13   The Asymmetry of Causation

In light of (P1)–(P8), the so-called causal asymmetry between a cause and an effect, which is traditionally formulated as 'it does not follow from "*x* caused *y*", that "*y* caused *x*"', can be given in the following terms: *if x* produced *y*, *y* could not have produced *x*, because that would have required *y* to produce *x* prior to its own production, which is impossible.

This is a good time to point out a significant difference between my view and the *perspectival view* advocated by Huw Price (1996), who also argues that the standard view is biased by agency. I argue that the relation between interacting entities should be considered to be symmetrical, but the relation between the successive stages of the process of production to be asymmetrical. Price argues that the relation between earlier and later stages of a process hold a symmetrical relation, i.e. he argues that the state which a system appears to have changed from, and the state which it appears to have changed to, when its parts interact, really hold a symmetrical relation. Or perhaps better, he thinks that such processes have no intrinsic direction. Price's argument is based on the fact that the mathematical formulations of the laws of nature in physics are time invariant. That is, because it makes no difference to exchange the time variable *t* with its contrary *-t*, in the fundamental laws of physics, any process can be *described* as going either backwards or forwards in time, without violating those laws.

Perhaps accounts of causality that rely *only* on the aspects of lawful connections between *types* of events in light of the most general formulations of the laws of physics will have to accept Price's conclusion, but then, I think, those accounts are incompatible with the conception of causality as involving production in accordance with the genetic principle. If the relation between two states of one and the same system is genuinely symmetrical, neither could have produced the other. Admittedly, the genetic principle imposes no special direction to certain *types* of processes—a system can evolve from a state of type A to a state of type B, or vice

versa—but it excludes that the relation between two *token* stages of one and the same causal process is symmetric. One will have to have come into being out of the other, or neither has. They cannot both come into being out of the other, because then both would have to exist before the other to allow the production of the other. Or, in other words, if a system evolves from A to B and back to A, say, when a caesium-133 atom oscillates between two energy states, then we still have a linear succession of token states sA_1, sB_2, sA_3, sB_4…, of which each state produces the next.

Furthermore, it seems to me that the time symmetry of the general formulation of the laws of physics disappears when you apply them to concrete systems, and therefore doesn't really imply that time is reversible or symmetric. In the general formulas, the symbols for quantities are really just placeholders for whatever real worldly quantity you want to insert when you describe the behaviour of an actual system, and they therefore don't have a direction. Once you insert the quantities representing the actual state of the system, they acquire a direction (they are vector quantities). As far as I know, the direction of these quantities is not affected by the reversal of the sign of the time variable. Imagine that you describe the state of a system at a given time whose parts are moving outwards from a centre, and then reverse the sign of the time variable in the equation you have, and then predict how that system will develop towards the 'past'. It will expand towards the 'past', because your reversal of the time variable does not reverse the direction of the state of motion of the parts of the system. This means that the actual process you are describing will not turn into its mirror reflection if you reverse the time variable.

In other words, the time symmetry of the laws of physics only means that one physical system can evolve from state A to B, that another system can evolve from state B to A, and that a system can evolve from state A to B, and then back to state A, but it will then really evolve through a linear order of successive states 1, 2, 3… An epistemic consequence of this is that we cannot tell in which direction the world is evolving merely by looking at the internal structure of physical processes. However, it seems to me that it is still the case that no matter in which temporal direction the world is as a matter of fact going, there is no way to reverse that direction. If it is evolving in what we at present identify as 'backwards', that is the only way it can be evolving (and therefore really is the forward direction).

It is then another question of how plausible it is to think that we could be wrong about the direction in which the world is evolving. What do we have to assume about the function of our brain to allow it to falsely represent the world as developing in the opposite direction from how it is actually developing? We can easily imagine such a scenario on a purely conceptual level; Martin Amis describes it in the novel *Time's Arrow* (1992). But can we explain how it really happens on the basis of our current knowledge of physical reality? A naturalist should conclude that our current physics rules out time reversal.

4.14   Conclusion

The reciprocity of interactions, as described by modern science, contradicts the assumed unidirectionality of interactions in the standard view, and removes the basis for making a distinction between active and passive objects. If, instead, causal production of change is taken to be the result of reciprocal action between parts of an aggregate (or system), a relation of one-sided existential dependence can be found to hold between the state produced in the aggregate by the interaction of its own parts, and the state destroyed by that same interaction. In other words, the interaction between the parts of a system destroys a state of affairs at the same time as it produces a new state of affairs, the produced state of affairs being for its existence one-sidedly dependent upon the state that was destroyed.

Thus conceived, the two states of affairs may be *logically* independent, in terms of each being of a certain type, but they cannot be conceived as distinct and independent substances. If it is assumed that they are distinct compound substances, then the question out of what they are produced is left unanswered, or it is answered that they were not at all produced, or it is answered that this effect was at least not produced by this cause. I suggest that in order to explain production of changes, and the existential dependence relation between the change and what produced the change, it is necessary to assume that things, or the substance they are made of, persist by enduring. Conversely it means that, if endurance is abandoned, production must be abandoned too.

When cause and effect are seen to be made of the same substance, it is impossible to think that any particular effect could just as well have been produced by some other cause than it in fact was. Any attempt to think of the effect as having been produced by some other cause will necessarily involve thinking of it as having been produced out of some altogether different substance, and therefore as being a different effect. It is still possible to think that a certain *type* of effect could have been produced by a number of different *token* causes, but not that a particular *token* effect could have been produced by any other *token* cause than the one that actually produced it.

I hope I have made a good case for the claims that interaction, as I have described it, (i) *can* be conceived to involve production, (ii) involves a necessary generic connection between an interaction and its product, and (iii) should be taken seriously as a possible hypothesis of the factual nature of causality. It is a hypothesis that fits most of the components of the common notion of causality, i.e. that causality involves *production* due to *causal influence* such that the producer and product hold a relation of *one-sided existential dependence*, although it combines these components in a somewhat different way than is usual. It also fits certain widely accepted metaphysical principles, i.e. the genetic principle and the

principle of lawfulness, but requires that the principle of action be substituted for what may be called the principle of reciprocity. This, I believe, is only a minor modification, however, and serves to correct the agency bias of the standard view, resulting in a conception of causation that fits to current scientific explanations of how changes are produced. My final suggestion is thus:

(P9): Interaction and necessary causal production are the same process.

5   Causal Necessity

The general sentiment in contemporary metaphysics is that there is a very strong case to be had against the necessity of causal connections ('CASE', for short). Indeed, even causal realists tend to accept that causation—although a real and substantial connection—falls short of being necessary (Schrenk 2009; Mumford & Anjum 2011, 2018). To my mind CASE consists largely of a collection of straw man arguments that only address a conception of causation—which I will call 'relational realism'—that to my knowledge has never been put forward in defence of the necessity of causal connections. Indeed, I don't know of any arguments that address the standard view, as I have presented it. In this chapter I will first show that the main arguments now comprising CASE cut no ice against the standard view, and then argue that the argument developed by Mumford & Anjum, which does *prima facie* appear to address a view much like my own, ultimately does not work.

I will not attempt to address every argument ever put forward against causal necessity, but enough to cover the main bulk of CASE. As a consequence, I can't claim to completely vindicate causal necessity, but only reveal how overestimated the strength of the arguments against it have been. The main bulk of CASE consists of five main objections: (i) Hume's discussion of causation (1739, 1748), which I have already addressed at length in Chapter 2, (ii) Anscombe's criticism of a logically/conceptually necessary connection, (iii) Russell's (1912) *problem of action at a temporal distance*, (iv) the family of counterarguments sometimes denoted collectively as *the problem of interference and prevention*, which has been used to criticise both the kind of conditional account of causation that J.S. Mill (1843) defended and more recent powers-based conceptions of causation (Mumford & Anjum 2011), and (v) the objection that even if everything in the actual world happens in accordance with causal laws or in virtue of dispositional essences, causation would still fail to be necessary because the laws of nature and/or the dispositional essences are themselves contingent (Armstrong 1983; Lange 2004).

I have already discussed Hume's argument in Chapter 2, arguing that it really only argues that there is no logical/conceptual necessary connection

between impressions and/or ideas. The question therefore is whether any-one has argued that this is what causal necessity is, which is what (ii) is all about. I won't address Hume's own argument any further. I will also ignore (v) because its rebuttal requires a more serious discussion of the possible worlds metaphysics than is possible here. I will however go some way to criticise its applicability to explain causation in Chapter 9, but that discussion is not really geared to the problem of causal necessity, but the more general question of whether possible worlds metaphysics can offer any account at all of causation. That I don't discuss (v) has no serious impact on the discussion of the other objections, since they don't address views that claim validity beyond the actual world.

5.1   Causal Necessity as Logical Necessity

The idea that causation is a kind of logical/conceptual connection has been defended by Kant (1787: A189/B233), Hegel (Hegel & Wallace 1874: 215ff), and McTaggart (1915). At least Kant defends that the con-nection involves necessity. However, none of them are causal realists, so to refute their arguments is not to refute causal realism. Nothing more needs be said about that.

It is more relevant to note that Elizabeth Anscombe identifies Hobbes as someone who claims that causal necessity is a logical connection, and claims that it was his argument that Hume 'overthrew' (Anscombe 1971). Anscombe bases her appraisal of Hobbes on his claim that when a certain total cause is present 'it cannot be understood but that the effect is pro-duced at the same instant; and if any one of them be wanting, it cannot be understood but that the effect is not produced' (Hobbes 1656: Ch. X, §3). I think Anscombe is too quick to infer from Hobbes' use of the phrase 'cannot be understood' that the connection is *purely* logical or conceptual. Hobbes' epistemology is empiricist, albeit with clear ratio-nalist overtones. In his view, the senses provide us with knowledge about the powers of material objects, and thus knowledge of causes, because powers are causes (Hobbes 1656: Ch. I, §4). From our empirical knowl-edge of these causes we can rationally calculate the effects they produce, and, vice versa, the causes from the effects. In light of this, I think it would be more charitable to interpret Hobbes as making claims about what can/cannot be thought without violating the nature of the external material world as it is empirically known to us (or assumed). If we find fault with his view, it is with his epistemology and not with his logic.

Someone having read Newton's *Principia* in the early 18th century could argue like Hobbes that if Newtonian mechanics is accepted as a true description of the world, then it cannot be understood, on pain of contradicting Newton's mechanics, but that an object acted upon by an external force will change its state of motion in proportion to the force applied. We then have a valid nomological-deductive argument,

moving from premises established empirically, leading to a conclusion that therefore is *a posteriori*. This is a thoroughly realist approach to metaphysical explanation, which has very recently been described by Naomi Thompson (2019). In her view, metaphysical explanations are subject to epistemic constraints imposed by the context in which a question is asked; they are not explanations of what must be the case without regard to any concerns except what can or cannot be conceptualised. This is the kind of natural necessity, and nothing more, that I have hopes of defending, and it seems that this is the kind of necessity that Russell, as well as Mumford & Anjum, is criticising.

Hume may well have had Hobbes' view in mind, but he cannot be read as overthrowing it merely by showing a flaw in the logic (the logic is valid). He first had to deny Hobbes' premise that the senses give us knowledge of the nature of external objects, thus turning the question of causation into a mere conceivability issue; one unrestrained by epistemic concerns.

It is noteworthy that Anscombe doesn't dwell on Mill's account of causation (1843: Bk. III, Ch. V), because he also argues, like Hobbes, that from a total cause an effect follows of necessity. Perhaps she realises that it isn't easy to portray Mill as a causal realist, since he explicitly claims to 'make no research into the ultimate or ontological cause of anything' and said that the causes he concerns himself with 'are not efficient, but physical causes' (1843: Bk. III, Ch. V, §2). A physical cause, in Mill's view, is an observable physical state of the world; a physical fact. One may suspect that Mill believes there are real causal connections, in the same way Hume accepts there probably are causal connections as soon as he turns his mind away from philosophy. But that doesn't make them causal realists, since they positively deny that causal realism is a viable philosophical project. Basically, Mill suspends judgement about the reality of 'mysterious and powerful ties' that may or may not connect the physical facts. To be specific: (i) unlike Hume but like Hobbes, Mill believes experience provides knowledge of physical facts, (ii) unlike Hobbes but like Hume, Mill believes experience does not provide knowledge of powers, (iii) unlike Hume, Mill believes (his version of) the law of universal causation can be inductively justified (Mill 1843: Bk. III). Importantly, Mill's belief in the validity of the law of universal causation is not based on any belief about the nature of substantial connections in nature. That disqualifies him from being a causal realist, and I am happy to concede that Mumford & Anjum's argument is problematic for Mill.

The bottom line is that none of the thinkers that claim that causation is a logical/conceptual connection are causal realists—at least none of those that appear in the debate—and so arguments professing to show that there is no logical necessity involved do not address causal realism. It is a straw man argument.

5.2   The Problem of Action at a Temporal Distance

In his famous paper 'On the Notion of Cause' (1912), Russell attempts to deliver a decisive blow to causal realism, and especially its endorsement of the *law of causality*, i.e. the view that causes necessitate their effects. Russell thinks the law is incoherent and that the notion of cause should even be abandoned. Indeed, he argues that the notion of cause hardly ever occurs in the natural sciences. His attack on causal realism has been very influential in 20th-century philosophy of causation, perhaps mostly by provoking various critical responses to it. Nancy Cartwright (1983) criticises him for assuming a much too general understanding of law, and Patrick Suppes (1970) points out that it is simply false that the word 'cause' never occurs in the natural sciences.

Anjan Chakravartty offers a particularly interesting criticism (2005). According to Chakravartty, Russell is attacking a particular view that he clearly thinks is the standard realist account of causation—and his argument is indeed a valid refutation of that view—but, Chakravartty argues, the view he attacks simply isn't causal realism proper. According to Chakravartty, the 'standard realist account' that Russell addresses is the account favoured by philosophers in the 20th century who take the statements of ordinary language to be the unit of analysis. In ordinary language, causal claims have the general form 'C causes E', which seems to suggest an understanding of causation as a two-place relation between events and/or states. But, as Chakravartty points out, 'everyday descriptions of causal phenomena are ambiguous with respect to the precise details of causation' (2005: 15). Once you get down to the precise details, it is all too obvious that the two-place relation characterisation is not much more than a shorthand description that skims the surface of an underlying *causal process* that gives rise to and connects events (Chakravartty 2005).

I believe Russell's view was *received* by his contemporary analytical philosophers, for exactly the reasons Chakravartty outlines. It fitted like hand to glove to the linguistic turn of 20th-century philosophy. However, I think more can be said about the way Russell himself *arrives* at that account. I believe it is the result of a particularly ill-conceived fusion of the neo-Humean constant conjunction formula and the causal realist notion of efficient causation, which Russell falsely believes is the standard view.

Russell thinks the philosophical concept of 'cause' is a 'relic of a bygone age' (1912: 1), and that it is unclear what to make of the 'law of causality', i.e. the idea of a necessary connection between causes and their effects. However, remarkably, in his treatment of these ideas, Russell does not consult anyone of the bygone age who has explicitly said anything about the concept of cause or the kind of necessity meant to connect it to its effect. Instead he proposes to find out what philosophers 'commonly

understand by "cause"' (1912: 2) by looking it up in Baldwin's *Dictionary* (1901), specifically the entry 'Cause and Effect', written by G.E. Moore. Russell cites this passage:

> Cause and effect [...] are correlative terms denoting any two distinguishable things, phases, or aspects of reality, which are so related to each other, that whenever the first ceases to exist, the second comes into existence immediately after, and whenever the second comes into existence, the first has ceased to exist immediately before.
>
> (Russell 1912: 2)

The passage bears no sign of anything we can distinctly associate with the kind of causal realism I discussed in Chapter 3. Instead we find only something that looks like a paraphrase of Hume's constant conjunction formula, which had arguably imposed itself on the way 'philosophers in general' write about causation at the time Russell and Moore were writing, i.e. on philosophers focused on philosophy as analysis of the statements of ordinary language. That Russell does indeed think this characterisation captures the essence of causal realism comes out clearly when he later notes that 'it would seem that only the later parts [of a cause] can be relevant to the effect, since the earlier parts are not contiguous to the effect, and therefore (by the definition) cannot *influence* the effect' (1912: 5; italics are mine). However, the definition he starts from doesn't say anything about influence and neither does the rest of Moore's entry. Russell simply assumes that the temporal contiguity described must have to do with the principle of locality as a constraint on efficient causation, because he knows causal realism is all about efficient causation.

As I think, Russell starts his characterisation of causal realism using a paraphrase of the empiricist conception of causation as invariant regularity between temporally successive events and has already then started on the wrong foot with respect to what kind of contiguity is relevant for causal realism. He then assumes that causal realists intend influence to be exerted from the first event to the second, just as Bigelow, Ellis, & Pargetter later assume (1988). As I mentioned earlier, the importance of temporal contiguity in Hume's view of things has nothing to do with causal efficiency or influence in any sense relevant for causal realism. Hume declared the idea of objects exerting influence on each other empirically empty; we cannot trace the idea to any impression. For Hume, contiguity has only to do with our natural inclination to associate in the mind those things that *appear* contiguous. Causal realism instead appeals to spatial contiguity between temporally coinciding objects.

Ultimately, the view Russell attacks is the idea that causation is a two-place relation between temporally distinct entities, of which the former exerts an influence on the latter. I will call this view 'relational realism'. To my knowledge nobody has ever argued in favour of this view before

Russell, but as we have seen in Chapter 3, this became the received realist view of the 20th century. In slightly better detail, this is Russell's argument:

(1) Causes and effects are temporally successive and distinct events (no overlap).
(2) Causes exert their influence on the effect.
(3) Cause and effect must be temporally contiguous to allow the cause to exert its influence, and to ward off the possibility of anything happening in the temporal gap between cause and effect that could prevent the influence from reaching the effect.
(4) If the structure of time is continuous, cause and effect cannot be contiguous; between the time the cause ends and the time the effect begins, no matter how close they are, there will always be an infinity of time-points in between.
(5) The Law of causation is incoherent.

This is what is now known as the *problem of action at a temporal distance*, but it should be clear that it is only a problem for the view that causation involves the exertion of influence between temporally distinct events, which is arguably a view Russell constructs for himself out of pieces that do not really belong together. It is a straw man argument.

In Chapters 3 and 4, I have already argued that causal influence has always been construed as something exerted between interacting and continuously coexisting objects, i.e. whose relation is continuously synchronic, and that view is immune to Russell's objection because there is no temporal gap between the object that acts and the object acted upon. Indeed, one of the main reasons Mumford & Anjum favour simultaneous causation is because they think it is the only way to avoid the problem of action at a temporal distance. My argument here is really that no one had really thought of actions as occurring between events, before Russell invents 'relational realism'.

Now, while it is true that the recent resurgence of causal realism in the philosophy of causation is meant to represent a return to the older standard view, most powers-based accounts still tend to represent causation as a relation between temporally distinct relata; between powers and their manifestations (for instance, Molnar 2003; Bird 2007*b*), or between events composed of particulars with powers (Esfeld 2011). Their views are therefore still susceptible to Russell's argument, and, for the same reasons, to the problem of interference and prevention.

5.3   The Problem of Interference and Prevention

Russell's problem of action at a temporal distance is actually a special case of the problem of interference and prevention. His argument is designed to show that the general model of causation in terms of temporally

successive entities, of which the first exerts an influence on the second, cannot in principle exclude that an interference or prevention occurs in the unavoidable 'gap' between the end of a cause and beginning of an effect. The only difference really is that Russell's argument isn't formulated in terms of a concrete example, only on a highly abstract temporal characterisation of causal connections.

However, the problem of interference and prevention, as it is usually formulated, typically refers to a family of concrete counterexamples that share the following general form. Given the occurrence of some kind of cause *C*, whose occurrence is supposed to reliably yield an effect *E,* then it seems always possible to assume the presence of some additional item ϕ such that it would be the case that *E* would not follow even if *C* occurred, or, in the case of pre-emption, that even if *E* actually does follow *C* then it isn't as a result of *C* but instead of the additional item ϕ. The argument presupposes a certain understanding of necessity, notably the kind of *unconditional* necessity of which J.S. Mill spoke:

> If there be any meaning which confessedly belongs to the term necessity, it is unconditionalness. That which is necessary, that which must be, means that which will be whatever supposition we make with regard to other things.
>
> (1843: 339)

This is the type of necessity that Mumford & Anjum take to be relevant for causal necessity, and which they use to construe what they call 'the test of antecedent strengthening'; a criterion they think any successful account of causal necessity must fulfil:

> If *A* necessitates *B*, then: If *A* plus ϕ, for any ϕ, then *B*.
>
> (2011: 57)

It is important to note that the test is to be applied to particular theories. That is, it should be understood as saying that whatever a certain theory specifies as a cause *C* that necessitates a specified effect *E*, then it had better be the case according to that theory that if *C* happens we always get *E* 'no matter whatever supposition we make with regard to other things'; i.e. the test must be true to the theory being tested both in the kind of cause and effect posited, and the kind of interference ϕ added to the situation. In other words, an argument operating with neo-Humean causes and effects, as well as interfering factors, will be irrelevant as a critique of a powerful particulars view. And, of course, it will not do to postulate as interferences/preventions anything that is actually forbidden by the theory in question, say, interferences by a deity, wizards, or other occult phenomena (which are forbidden on any naturalist account). A theory

can hardly be expected to be impervious to objections invoking phenomena that the theory excludes as unreal.

So, in sum, the problem of causal necessity is this: is there an account of causation that can establish that whenever a certain kind of cause C occurs, then whatever suppositions we make with regard to anything else not already excluded by the theory, the effect E always and invariably follows? Today it is taken for granted that for any putative cause C ever suggested in the history of philosophy, it has been all too easy to construe counterexamples involving some interfering factor ϕ that prevents the putative effect E from following.

Over the years different types of counterexamples involving prevention and/or interference have come to be roughly divided into four main categories: (i) prevention, (ii) pre-emption, (iii) finks and (iv) antidotes. Prevention occurs when some factor ϕ in the circumstances in which C occurs prevents E from occurring. For instance, suppose that an assassin fires a shot and the bullet travels the distance to the victim but is deflected out of harm's way on the very moment of impact, by a second bullet fired by another assassin; leaving only a scratch, as a mark of them having achieved contiguity. This is supposed to illustrate that C can occur—the assassin fires a bullet, the bullet travels to the victim and becomes contiguous with it—and yet something intervenes to prevent the intended effect from occurring.

Pre-emption is when ϕ does not prevent E from following C, but prevents C from being the actual cause to E simply by producing E instead of C. In that case, ϕ is really another token cause of the same kind as C, which competes to produce E, and wins. Consider assassins Duke and Drew shooting at the same victim and both hitting the victim in the heart, except that Drew is a tad closer wherefore her bullet kills the victim an instant before Duke's bullet arrives (alternatively, renders the victim as good as dead). The assumption is that a cause C occurs to completion but its designated effect, the death of the victim, is pre-empted by ϕ; the victim is already dead or as good as dead. There is a counterfactual version of this in which both bullets hit the heart simultaneously, but each pre-empts the other in the sense that even if one wouldn't have hit, the other would.

In the case of prevention and pre-emption, the interfering factor is *external* both to the cause and the effect, or, rather, external to the object that acts (Agent) and the object that is acted upon (Patient). Finks and antidotes, on the other hand, are instead peculiar kinds of interfering powers that are intrinsic to the Agent and Patient, respectively. Finks are located in the kind of object that counts as an Agent in the standard view, and they work their interfering magic by defusing the Agent's own supposed power to produce a certain effect. Using a suitably modified assassin scenario, consider a bullet fired at a victim and reaching the victim's heart but causing no damage because in addition to having momentum,

impenetrability, and velocity (powers to produce death), the bullet has the 'finkish' power to instantaneously evaporate when coming into contact with striated muscle cells, such as the heart is made of. C.B. Martin's famous finkish live wire—deadly until touched—is another example of the same kind (1994). Consequently, the bullet penetrates the outer layers of tissue, but evaporates as soon as it touches the heart; the victim is maimed but doesn't die.

Antidotes are powers of the Patient either to defuse its own liability to be affected by the action of the Agent or defuse the Agent's power to produce an effect in the Patient. Thus, if the victim is disposed to instantaneously become perfectly permeable (i.e. ghost-like) when coming into contact with a particular metal present in the bullet, then on contact the victim becomes ghost-like, and the bullet passes through the victim without causing damage. Antidotes to poisons work in a similar way, either by directly neutralising the toxic abilities of the poison (e.g. active charcoal) or affecting the ability of the Patient to be affected by the poison (e.g. calcium channel blockers). Of course, real-life antidotes are extrinsic to the Patient and so really preventive factors.

It is indeed very difficult to come up with realistic cases of 'finks' and 'antidotes', since normally we would consider anything that has an intrinsic 'fink' or 'antidote' to already be immune to the action of the Agent in question, or unable to act on the Patient in the way required to produce the intended effect. An assassination-victim who is intrinsically immune to poison will simply not have the power to be poisoned that is prevented by a second power of immunity. This is why discussions about interference and prevention are full of oddly contrived powers, such as becoming perfectly permeable. Sometimes they even invoke magic (e.g. Manley & Wasserman 2008: 62) or divine intervention (Reeder 1995: 143). The supernatural is already excluded by causal realist commitment to the genetic principle but let us see how the standard view might respond to the more realistic counterexamples.

In sum, arguments invoking cases of prevention and interference aim to show that some cause C occurs to completion and yet does not lead to E, because some ϕ interferes to stop E from happening, or to make C superfluous to the production of E, as in the case of prevention and preemption, or cancel the occurrence of E by making the Agent impotent or the Patient impervious.

5.4   The Standard View Response

To my mind, no counterexample ever used to illustrate interference and prevention really presents a case where a cause C *of the kind accepted by the standard view* occurs to completion but is being prevented from having its effect because of some interfering factor ϕ. Instead they are cases where an effect is prevented because the *cause* is prevented from

ever taking place. Indeed, on closer inspection they simply illustrate what Hobbes thought was obvious, notably that only the total cause can necessitate anything at all. In either case, the outcome of whatever scenario is being described turns out to be the inevitable and necessary result of what actually happens. Ergo, every counterexample really turns out to be a *confirmation* of the general schema 'whenever an Agent *A* acts in a certain way *C*, on a Patient *P*, then the outcome is invariably *E*', rather than a *refutation* of it.

Take the double-assassin situation as example. Duke's bullet being deflected out of harm's way by Drew's bullet is really a situation where Duke's bullet is prevented from actually exerting the kind of influence on the victim that would be required to produce a death; we do not really have a case of a completed action by *A* on *P* being prevented from having its intended effect *E*. It is a case of *A* being prevented from acting on *P*, and thus it is a prevention of a particular cause *C* (i.e. the action of *A* on *P*) from being realised rather than preventing an effect *E* from being realised once *C* is complete.

In fact, the 'counterargument' seems designed to confirm rather than refute the standard view, because it really says that whenever a bullet *A*, on collision course with person *P*, collides with a second bullet *B* in the manner *C**, then, necessarily, *A* and *B* are deflected away from *P* in manner *E**. Pre-emption, finks, and antidotes can be dealt with in roughly the same way.

Consider pre-emption. Assassins Duke and Drew simultaneously fire their bullets and both hit the heart of the victim, except that Drew is a tad closer to the victim wherefore her bullet hits a fraction of a second before Duke's bullet arrives, and kills the victim before Duke's bullet has the chance to do it. I can again object and say that the example assumes that the nature of the object acted upon, or the character of the influence exerted on it, makes no difference regarding the type of cause being realised. But, surely, it is one thing for a bullet to act on a living body and another for it to act on a dead body (or one that is as good as dead). In fact, in the case at hand, the bullet fired by Duke has, of necessity, exactly the effect that it would be expected to have if it hit a dead body; it does all the damage to dead tissue that it should do. Basically, Duke's bullet never realises the action of hitting a living body, and so never is a bullet hitting a living body that is prevented from producing a death. Instead, it is a bullet hitting a dead body and accordingly produces a certain kind of damage. The only thing that is prevented from being fulfilled is someone's intention or expectation of 'death-by-the-hand-of-Duke', which arguably is a type of prevention that might be relevant to some form of popular narrative of how things unfold, but not to physical causation in objective reality. On the standard view I have presented, cases of prevention or pre-emption do not provide genuine cases of a cause being realised and yet failing to produce the effect.

Someone might worry about overdetermination and the resulting threat to the necessity of *A*'s and *B*'s action on *P*. That is, doesn't the fact that *P*'s death is overdetermined by the actions of *A* and *B* mean that neither was necessary for the production of death, although each was sufficient? I don't worry about this. To begin with, overdetermination only applies to causal reasoning that involves common-sense concepts like 'hit by a bullet' and 'death'. If we think of the case on the level of determinate physical reality, then one bullet produces one amount of physical damage and two bullets produce double that amount; any real-life scenario that didn't comply to this characterisation would be a violation of conservation laws. Whatever changes are produced by the first bullet will not be a prevention of the effect produced by the latter bullet; those changes merely serve to change the state of the Patient in time for its interaction with the second bullet and the outcome of that interaction will be necessary.

Consider now the finkish case in which a bullet 'hits' the victim's heart but causes no damage because the bullet is such that it evaporates instantaneously when coming into contact with striated muscle cells. Again, if it is assumed, in accordance with the standard view, that causation is a matter of one object exerting an influence on other objects, and that the outcome is determined by the particular combination of powers present, then I can't see how finks pose any threat at all. A fink is simply something that makes an Agent *A* unable to exert on the Patient *P* the kind of influence required to produce a given effect *E*, and thus prevents (or, rather, excludes the possibility of) the realisation of a cause *C* required to produce *E* rather than preventing the realisation of *E* once *C* is realised. It is only if we assume that the world is an aggregate of local matters of fact where nothing exerts an influence on anything else that a bullet being fired and becoming spatio-temporally contiguous with striated muscle cells can count as a cause running its course to completion. But that is a peculiarly neo-Humean conception of causation that is irrelevant to causal realist concerns.

What about antidotes? If the victim is disposed to instantaneously become perfectly permeable when coming into contact with the particular alloy of which the bullet is made, and then does actually become permeable when they touch, then this is not a case where the exertion of the type of influence that typically produces a death is prevented from having its intended effect; it is a case of *A* being unable to exert that particular type of mechanical influence on *P* that produces death, and instead, necessarily, exerts the kind of 'magical' influence that makes *P* perfectly permeable. Given that *P* is such that it becomes perfectly permeable when coming into contact with the particular alloy of which *A* is made, then, necessarily, when *A* touches *P* it makes *P* become perfectly permeable and therefore unaffected by the mechanical powers possessed by *A*.

Both finkish and antidotal cases obviously presuppose that it is *individual* powers that necessitate, i.e. that the necessary causal connection is assumed to hold between exertion of individual powers and their particular manifestations. As we have seen, the standard view does not say this; it insists it is only the sum total of powers in the interaction that can necessitate. To look for causal necessity in the connection between individual powers and their manifestations is arguably to confuse the issue of causal necessity with the question of whether powers have dispositional essences. I deal with that problem in Chapter 8.

Some readers may worry that surely two Agents can act in exactly the same way on the same kind of Patient and yet produce different results. One example would be Duke punching a sandbag to make it swing, and Drew punching the sandbag in exactly the same way without making it swing, because only in the latter case is Mr T there to steady the bag. However, there is no problem here unless you require—like the empiricists do—that the cause results in a particular spatio-temporally observable behaviour or event; the observable correlate of bag swinging from position A to B. To require such a thing is a peculiarly empiricist prerequisite; it is a refusal to engage in talk about unobservables, or theoretical entities generally. Rationalists can accept that both punches produce identical effects even though that may not be obvious to an observer. In physical terms, both punches produce an identical change in the sandbags' linear momentum—which is typically represented in classical mechanics by a vector quantity using the symbol $\mathbf{p}$—from $\mathbf{p}$ to $\mathbf{p}^*$. The difference between the two scenarios is that immediately after Drew hits her sandbag and changes the sandbag's momentum from $\mathbf{p}$ to $\mathbf{p}^*$, there is a *subsequent* exertion of an oppositely directed force by Mr T, that changes the sandbag's momentum again (from $\mathbf{p}^*$ to $\mathbf{p}$). In Duke's case, on the other hand, there is no such subsequent action and the change in linear momentum is therefore realised in the form of motion from A to B. To repeat, in both cases the linear momentum changed from $\mathbf{p}$ to $\mathbf{p}^*$, which means that in both cases a punch resulted in the same effect, even though it was not visible for an observer.

We can use Newton's cradle to illustrate further the real but invisible manifestation of change. The cradle contains five metal balls suspended by strings to lie contiguously in a straight horizontal plane (and we assume they are perfectly aligned and balanced). If the ball at one end of the line is lifted up and released, it will swing down to hit the adjoining ball but the only visible change in the cradle is the movement of the ball on the opposite end, which will swing up and back to again hit the line of balls, of which the only visible result is again a movement of the ball on the opposite end. A causal realist will say that even though no change is perceived in the three balls in the middle, they still undergo a change in their intrinsic properties, notably in their linear momentum (from $\mathbf{p}$ to $\mathbf{p}^*$), and that a succession of interactions takes place along the

line of balls. The only way to interfere with this succession of interaction is to introduce another Agent that influences the cradle in such a way as to prevent some other influence from taking place. But it is impossible to prevent an effect from an already exerted influence, without violating conservation laws.

To my mind, then, the standard view can deal with the counterexamples provided under the heading 'prevention and interference'; provided we think that (i) the necessary link holds between, on the one hand, an exertion of a particular type of influence of a particular type of Agent on a particular type of Patient, and, (ii) the resulting change be allowed to include unobservable changes in the intrinsic properties of things, such as a change in the linear momentum of an object in the absence of an observable movement.

5.5   Interference and Prevention in Powers-Based Causation

I have so far argued that the standard view can deal with counterexamples where it is assumed that individual external objects (simple prevention and pre-emption), or individual factors, or powers, intrinsic to the objects involved (finks and antidotes), interfere and/or prevent the efficacy of the actions of individual Agents or powers. However, Mumford & Anjum have also addressed the question of whether the total set of powers involved in an instance of causation could necessitate an outcome (2011: Ch. 3). At first blush, their argument may seem to address a total cause conception of the kind I am promoting. But this is not the case. They characterise the total cause merely in terms of the *obtaining* of a set of powers, in a way that ignores the manner in which the powers are exerted in an interaction between powerful particulars. That makes quite a difference.

Mumford & Anjum argue that if a given effect E is assumed to be the necessary result of a combination of a set of powers, say $[p_1, p_2, p_3, p_4]$, then—for it to pass the test of necessity—whenever powers $[p_1–p_4]$ obtain we should always get E no matter what other factor ϕ obtains with them. I acknowledge that, as a matter of empirical fact, this is not the case. If $[p_1–p_4]$ obtains we get E, but if $[p_1–p_4 +\phi]$ obtains, we get E^*. Their argument is logically valid. I just find that it doesn't address the standard view, and as far as I know, the view they address has never been defended.

Admittedly, Hobbes writes that an entire cause is the 'aggregate' of all the powers of the interacting objects, which could be interpreted as a characterisation of causes merely as a certain combination of powers. However, one will then have to neglect Hobbes' repeated claims that it is the *body* of the Agent that acts upon the *body* of the Patient, albeit in virtue of possessing various powers (Hobbes 1656: Ch. IX, §§1–4, 8). Obviously, what Hobbes has or hasn't said isn't really paramount here. We can just appeal to the empirical fact that the way

powers are distributed between interacting entities is important for the production of an outcome, not just the collected presence of powers at a place and time.

The standard view states that causes are interactions between powerful particulars; not the mere obtaining of a set of powers. It matters for the characterisation of the cause at hand to which particulars the powers $[p_1-p_4]$ belong. An interaction between particulars A and B where the powers $[p_1-p_4]$ are distributed as $A[p_1, p_2]$ vs. $B[p_3, p_4]$ is a different kind of cause than an interaction where the same set of powers $[p_1-p_4]$ are distributed as $A[p_1, p_3]$ vs. $B[p_2, p_4]$ (A has p_3 instead of p_2, and B has p_2 instead of p_3). Therefore, one cannot argue without further qualifications that, for powers-based causation to be necessary, whenever powers $[p_1-p_4]$ are present one must always get E. One has to consider how the powers are distributed between the particulars.

Let us go through this in detail. Consider an interaction ($\Leftrightarrow$) between two particulars A & B with each a set of powers $[p_1, p_2...]$, which then produces ($\supset$) an outcome E:

$$A[p_1, p_2] \Leftrightarrow B[p_3, p_4] \supset E. \tag{1}$$

According to the standard view, this is one kind of interaction and so a particular type of cause leading to a certain effect, E. And it is different from the following interaction in which the same powers are present, but distributed differently between the particulars to produce the effect E^* (A has p_3 instead of p_2, and B has p_2 instead of p_3):

$$A[p_1, p_3] \Leftrightarrow B[p_2, p_4] \supset E^*. \tag{2}$$

So even in different cases where we have the same powers present, we may not get the same effect if they be distributed between the particulars in different ways. Furthermore, neither (1) nor (2) are instantiated in the following interaction where we have $[p_1-p_4]$ + an additional power ϕ:

$$A[p_1, p_2] \Leftrightarrow B[p_3, p_4, \phi] \supset E^{**}. \tag{3}$$

The last interaction is not an instance in which either (1) or (2) occurs, but are somehow interfered with by ϕ; it is a different interaction, and so a different cause producing a different effect. Either ϕ is a part of the interaction, in which case the only thing that is prevented is the occurrence of causes of kind (1) and (2), or ϕ is introduced after or during the interaction, and in either case it comes too late to prevent the manifestation of an exertion of influence that has already taken place. Remember that the standard view takes causation to be a continuous process of mutations that continues as long as the objects are exerting an influence on each other. At any point in that process an exertion of influence will

have an immediate result, but interfere at any given point and the process morphs into another process through the addition of a participating powerful particular; one that will necessarily evolve a certain way.

Really, if we take into account the distribution of powers among interacting particulars, then the only way to introduce an additional power ϕ to the set of powers $[p_1–p_4]$ is either by introducing an additional external particular, C, bearing the power ϕ, as in the case of simple prevention and pre-emption discussed earlier, or we introduce ϕ as already belonging to the Agent and/or Patient, in which case we have finkish and antidotal cases all over again.

I can't see a way to interfere or prevent anything on this model. The only thing that could make interactions of this kind anything less than necessary is if we already postulate that the individual powers are slightly indeterminate in their contribution to an interaction. This seems to be what Mumford & Anjum mean in calling powers *tendencies*. However, their argument in favour of treating powers as tendencies is the argument that I have just discussed, notably that sets of powers do not necessitate any particular outcome because something can interfere with their production of effects.

It is possible that Mumford & Anjum do not think of their view as motivated by the interference and prevention argument alone, but rather see a theoretical advantage in the idea of a tendential modality in itself. After all, they claim that even in situations where 'there is nothing that prevents the effect from occurring, but still, it need not occur' (2018: 18). As Stathis Psillos remarks in his review of Mumford & Anjum's most recent book, *What Tends to Be*, this move 'seems to pile a new mystery on the old' (Psillos 2019), notably of explaining how it could be that an influence is being exerted but nothing happens even if nothing is interfering.

Psillos speculates whether they are modelling the modality of tendencies on phenomena like radioactive decay; at any moment the atom has the tendency to decay—and nothing is there to prevent it from decaying— and yet it may not decay. Psillos is worried that idea turns on a problematic view of the probabilistic nature of decay, which I think is a valid worry. Indeed, one can worry about more than the probabilistic aspect of decay. If one thinks decay is a spontaneous non-causal phenomenon, because it is not triggered by an interaction between the atom and any other powerful particular, then it isn't causal at all. If one instead thinks that we are dealing with an atom that has the power to decay, but doesn't because it isn't exercising the power, then we don't have a case where a power is being exerted without producing an effect, but a case where a power isn't exerted.

Anyway, if we confine ourselves to interactions between powerful particulars, then Mumford & Anjum's tendential modality seems to allow for the possibility that two billiard balls could collide and fail to exert

their powers on each other, or, that even if they do exert them, they don't change each other's state of motion because they fail to exert their passive power to change state of motion. To my mind, that would be rather mysterious.

5.6   Causal Necessity Without Ceteris Paribus Clause

I think it is important to point out that the powerful particulars view really amounts to an account of unconditional necessity without appeal to *ceteris paribus* clauses. This is because it includes in the notion of cause the components that others take to be mere conditions; the conditions that must stay the same to secure the connection between *C* and *E*. The *ceteris paribus* clause is meant to cover all these allegedly non-causal conditions. I think it follows from my argument above, that we don't have to add a *ceteris paribus* clause to exclude the presence of any additional item ϕ that could interfere or prevent, because the presence of any such additional item would only constitute a failure of the obtaining of a certain kind of cause *C* rather than the prevention of an effect, and constitute the obtaining of a certain other kind of cause *C**, which inevitably gives *E**.

5.7   Conclusion

In this chapter I have argued (i) no causal realist has argued that causal connections are logically necessary, (ii) the problem of action at a temporal distance addresses a view that has never been defended, but was merely Russell's mistaken conception of causal realism, and that the powerful particulars view is immune to the problem, (iii) that the problem of interference and prevention similarly addresses only relational realism while the powerful particulars view is immune to it, and (iv) that Mumford & Anjum's version of the problem of interference and prevention addresses a view of causes as sets of powers that has never been used to defend causal necessity, and that the powerful particulars view is immune to their arguments too. The conclusion must be that CASE does not amount to a serious challenge to the necessity of causal connections.

6 Constitution and Persistence

In Chapter 2 I introduced the distinction between *characterising* and *explaining* features of reality, arguing that neo-Humean accounts of causation only strive to characterise, while causal realist accounts strive to explain, all the while being compatible with neo-Humean characterisations on the level of description. Causal realist and neo-Humean accounts only disagree in so far as proponents of neo-Humean accounts of causation also explicitly endorse neo-Humean metaphysics, i.e. the view that there are no substantial connections in reality that could explain how causation appears in experience. I also claimed that extant accounts of the persistence of compound entities only characterise and do not explain. The only explanatory account of persistence I know of—that the world ultimately consists of permanent simples—only explains the persistence of simples. Now it is time to deliver on the promise of a causal account of the persistence of compound entities. Indeed, I think it also offers an explanation of the *constitution* of compound entities in contrast to mainstream accounts that merely characterise the relationships between the parts of constituted compounds (for an overview, see Wasserman 2018; Varzi 2019).

I think there are very obvious causal aspects to the constitution and continued existence of compound entities, especially in light of the scientific image of the world. Just consider the explanations given in secondary education physics and chemistry of the physical bonds that hold compound objects together, say, how elementary particles constitute an atom through continuous interaction. Physics, chemistry, and biology simply do not depict the entities they study as things that passively continue to exist as long as nothing destroys them; they continuously and actively preserve themselves through the interaction of their constituents. However, in philosophy it is rare to see anything more than a passing reference to the possibility that there might after all be something causal about persistence, except among proponents of process ontology. However, it is assumed that substance and process ontology are rival and incompatible views. I'll return to that question in Chapter 7.

6.1   The Incompatibility of Causation, Constitution, and Persistence

Constitution is typically regarded as a non-causal phenomenon. So, for instance, there is no reference to anything causal in the entry 'Material Constitution' in *The Stanford Encyclopedia of Philosophy* (Wasserman 2018). One rare exception is Karen Bennett (2017). Bennett objects to the received view that what she calls 'building relations' are non-causal (such as grounding, composition, determination, realisation, and constitution). According to her 'there are particular building relations that are partially defined in causal terms, and there are even building relations that only obtain diachronically, in virtue of causal facts' (2017: 2). I agree with this claim, but I also think that she doesn't specify the kind of causal account needed to substantiate that claim.

The idea that *persistence* might be causal occurs more frequently, but, again, usually in process ontology. Whitehead takes persistent entities to be processes, i.e. a sequence of temporally ordered stages, each of which undergoes 'concrescence'—a transition from potentiality to actuality— and the concrescence of each stage causally triggers the concrescence of the next (Emmet 1992). Peter Simons observes that the persistence of things seems to be dependent on the processes taking place in the thing and suggests we should think of persistent entities (or *continuants*, as he calls them) as dependent on a base of occurrents that stand in causal relations to each other (Simons 2000). Simons intends his view to be an alternative to the endurance and perdurance views, which he thinks are flawed because neither addresses the explanatory gap between the permanence of simples and the constitution and continued existence of compounds. Now, while I agree with Simons about the flaws of the endurance and perdurance views, I don't like the idea of process as a structured series of stages or occurrences either. As I have previously argued, the stage view of process reduces change either to *variation* between stages— a problem it shares with perdurantism (see Ingthorsson 2001, 2009)—or to *concrescence* of stages, i.e. transitions from potentiality to actuality, which I think inevitably violates the genetic principle. I will discuss the nature of processes in Chapter 7.

Salmon also thinks persistent objects are causal processes but denies that this commits him to an event ontology of process (1984: 139ff). However, Salmon really doesn't give an account of what a process is, nor does it seem possible to apply his account of causation in terms of propagation and interaction to the continued existence of processes. It rather seems that persistence, i.e. the ability to propagate a mark, is constitutive of causation rather than vice versa.

The general lack of attempts to make something of the causal aspects of constitution and persistence—outside process ontology—can plausibly be explained by the fact that the mainstream definitions of the notions

of causation, persistence, and constitution present them as *prima facie* incompatible. Causation is understood as a *diachronic* relation between *distinct relata*—one event being the cause of another—and thus as an *external* relation. Indeed, the concept of non-causal spontaneous change is typically understood as one that happens in the absence of an external cause, like the decay of unstable particles. Persistence is also understood as a diachronic relation, but one holding between an object and itself, and thus *intrinsic* to the persistent thing. This may be the main obstacle to a causal account of persistence. If we are looking for the cause for the persistence of an object, we are not looking for an external compulsion; if an object persists, it does so of its own accord. Finally, constitution is typically understood as a relation between a whole and its parts and is traditionally taken to be a relation that can hold at any given time, i.e. *synchronically*. Consequently, it is a relation that is intrinsic to the constituted object but still not a relation of identity. In sum, tradition dictates that causation is a diachronic relation between distinct relata, persistence a diachronic relation between an object and itself, and constitution a synchronic part-whole relation. Against this background, to offer a causal account of constitution and persistence will require us to think differently about one or more of these notions. Here I explore what happens if we accept the interaction view presented in Chapter 4.

6.2 Characterisations

Persistence is typically characterised as something existing for more than a moment, which isn't an explanation of why anything persists at all, or why things persist for a short or long time. The notions of endurance and perdurance do not explain either why something endures or perdures; they just characterise persistence in ways conceptually compatible with the constraints of the *A* vs *B* view of time. According to the *A* view, all objects are only in the present but it allows what is present to change, wherefore things can come to exist at many times by passing from one time to another. According to the *B* view, time does not pass, and all times exist in parity, wherefore objects cannot pass through time and must therefore extend through time, having different parts existing at different times (Ingthorsson 2009). Similarly, mereology doesn't tell us what makes p be a constitutive part of some whole w, but only illustrates the formal relations between p and w and any other parts of w, once it is assumed that p_1, p_2, p_3... are constituents of w (for an overview, see Varzi 2019).

It bears mentioning that Kit Fine's 'Aggregative' and 'Monster' objections to mainstream accounts of constitution (1999: 62–3) boil down to the claim that they don't explain constitution; that they treat compound objects as mere aggregates, and therefore can't really tell true unities from mere aggregates. This allows us to create 'monstruous' entities

like the sum of Cleopatra, a slice of ham, and the sum of all objects that merely existed before or after the existence of a particular ham sandwich. Indeed, Fine's account in terms of what he calls 'embodiments' is in some ways an attempt to provide an explanatory account, and he thinks it is roughly Aristotelian (Fine 1999). However, since Fine's 'embodiments' are primitive *sui generis* relations between a whole and its parts, they are only explanatory in the same sense the attribution of 'permanence' to fundamental particles is an explanation of persistence. The fact that constitution applies to compounds and not simples, i.e. not to fundamental entities, raises the question of whether this strategy really is warranted for compounds. Furthermore, it is an account that doesn't obviously connect 'embodiments' with the scientific explanation of how elementary particles make up atoms, atoms make molecules, *etcetera*. If not for anything else, if an account of constitution in terms of interaction works, it has the advantage over Fine's account of not being *sui generis*, and clearly links our philosophical views with the theories and findings of science. For a critical discussion of Fine's account, see Kathrin Koslicki (2008: 75–89).

The distinction between unities and aggregates is central for the issue of constitution. A unity is a collection of entities that bear a substantial relation to each other, such as the particles in an atom, atoms in a molecule, molecules in an alloy, or even the parts of an automobile. An aggregate, on the other hand, is a collection of entities that do not stand in any substantial relation to each other, such as the various deserts on the planet, or the cars parked in a given car park at a certain time. They amount to no more than the mereological sum of its parts. The existence of an aggregate is dependent upon the existence of each and every part, but the existence, qualities, and behaviour of the parts are not dependent upon the existence of the collection. In a unity, the parts are qualitatively modified as a result of becoming a part of the unity, and their behaviour is partly determined by the whole of which they are parts, but they also contribute to determine the behaviour of the whole. An example is the way the rotation of the Earth along its axis and around the sun is continuously influenced by the sun, moon, and other planets, and, vice versa, the presence of the Earth in the solar system affecting the behaviour of the system. Obviously, one cannot hope to explain the persistence of aggregates in terms of interaction, since by definition there are no substantial connections between the parts of an aggregate, and interaction is a substantial connection. We may have better luck with unities.

The objects we ordinary identify as objects, like tables, billiard balls, and persons, not only appear to us as unified wholes, but are accounted for scientifically as collections of parts bound together by physical bonds that all count as interactions in the way I describe them in Chapter 4. That is, ordinary middle-sized objects are compound substances constituted by parts bound together by continuous interaction. A brick is a unity of parts and it can in turn be a part of an aggregate whose parts

interact, e.g. when the brick collides with a window. In this example, the unity of the brick is preserved but the unity of the window is destroyed. In 'Causal Production as Interaction' (Ingthorsson 2002) I maintain that the aggregate whole of brick and window also persists through an interaction, but this is *not* something it does because of the interaction, rather in spite of the interaction. The aggregate whole persists because of the 'existential inertia' it borrows from the simple substances that are its fundamental parts. Whether any unity contained in that aggregate survives as well—say, the brick—is decided by how well the unity succeeds in preserving its structure.

6.3   A Causal Account of Constitution and Persistence

To accept interaction as a fundamental feature of causation is to admit that the ontological category of causation is bigger than causal realists previously thought. Most importantly, it allows us to include in the category of causation all the interactions that take place between the constituent parts of objects, and which are already understood as the bonds that glue those parts together to form a unity. Interactions explain constitution and persistence as two sides of the same coin; they bond the parts together to constitute an object, and they preserve the unity over time. Persistence basically is the temporal dimension of constitution.

Interactions have not been considered causal before, because they were perceived to lack what the standard view postulates to be an essential feature of causation: *external compulsion*. Now, in a sense, interactions do not really lack external compulsion since interacting entities do act on other entities. It is only from our 'middle-sized dry goods' perspective that whatever happens invisibly between distinct parts inside an object strikes us as happening in the absence of external compulsion, and indeed as happening in the absence of any visible compulsion.

It actually turns out that there is really very little for me to do, *vis-à-vis* a causal account of constitution and persistence, once interactions are accepted as causal, except to point out that this is how objects are already understood to be constituted by their parts and why they continue to exist over time, i.e. persist. The standard model of physics describes how our physical reality is built up of simple elementary particles that interact with each other to constitute compound entities (for reference, see Donoghue, Golowich, & Holstein 2014). Quarks interact strongly to constitute protons and neutrons. Protons and neutrons in turn interact via the nuclear force (derivative of strong interaction) to constitute atomic nuclei, and the nuclei interact electromagnetically with electrons to constitute atoms. None of these compounds—protons, neutrons, nuclei, atoms—are static and unchanging, but dependent for their stability, i.e. continued existence, on continuous and dynamic interactions between the component parts. Atoms are more complicated than

presented here (including various weak interactions between the particles of the standard model), but this is enough for our present purposes.

Going up to the level of molecules, they are constituted by chemical bonds. There are ionic bonds, which involve electrostatic attraction between charged ions, and covalent bonds which involve two atoms sharing electrons. Again, these bonds—just like those between the constituent parts of the atoms—are thoroughly dynamic and yet, provided they are stable, still maintain a certain status quo of whatever system they work to constitute. I will not attempt to characterise in more detail the nature of this dynamic equilibrium on the atomic level. I don't think I can provide a concise enough description that is both theoretically correct and easily understood by those readers who, like me, don't have an expert knowledge of either physics or chemistry. Instead I will illustrate the dynamic equilibrium using an object we are all roughly familiar with, namely the solar system.

The solar system is a compound constituted by a collection of bodies, held together by the interactions taking place between them. I'll confine the model to just sun and planets for the sake of simplicity. It has persisted for millions of years and will continue to persist for millions more. However, the solar system is never the same between any two points in time; it is continuously changing in a way that serves to preserve the system. Unless the planets were moving in their orbits as they currently are, the solar system would quickly collapse; too fast and they will gradually escape the gravitational pull of the sun, too slow and the sun will swallow them. Indeed, the current equilibrium is actually only so stable as to put off the inevitable end beyond any timescale we need worry about. The planets are ever so slightly moving away from the sun, and the moon moving away from Earth.

In Chapter 7, I will argue that objects like the solar system, as well as any relatively stable material system exhibiting the same dynamic equilibrium, count as a process, if we by 'process' mean any persistent entity for which change is essential for its continued existence. However, I will argue that this does not detract from their status as substances in the Aristotelian sense (to be distinguished from the modern analytic conception of substance).

6.4   Problems?

There are two questions that arise here. First, how can objects be differentiated, if, as physics dictates, everything is always interacting gravitationally with everything else in the universe? The answer is that we get a differentiation in terms of the strength of unity of a compound, which varies by the strength of the type of interaction involved and the overall stability of the system. The four fundamental types of interactions that obtain in the world are the strong and weak interaction, which obtain

at close range between the constituents of an atom, and gravitation and electromagnetic interaction which are more long range. Each of these are of different strengths, and thus allow us to differentiate between different physical systems that will display a different kind of unity depending on the interactions by which they are constituted. The gravitational interaction between the Earth and the moon is strong enough to cause tides, but not enough to really disrupt the cohesion of the Earth as a unitary whole. Then again, the ionic and covalent bonds between water molecules are strong enough to explain the cohesion of water, but at room temperature water has little or no effect on the cohesion of our bodies since the cells of our skin are bonded together much more strongly than the molecules of water. I take this differentiation to fit the way science already differentiates between entities. To be sure, this way of differentiating between unities is messy. We really get a very intricate hierarchy of partly overlapping systems, but that is a conclusion that seems to fit our observations of the world we live in.

Second, does my account answer the special composition question, i.e. does it specify under what conditions something composes something else? I won't delve deep into that question. I don't actually think I provide a new account of constitution, except in so far as I offer a way to understand the way constitution is already understood in the sciences, as being a causal phenomenon. What I can say is that I think this is an account of constitution that has an advantage over Fine's embodiment account, and one that provides a more fine-grained account of what Bennett calls 'diachronic building'.

6.5   Concluding Remarks

I think I have now given at least the outlines of a causal theory of constitution and persistence. If causation is interaction, then every stable compound constituted by interactions between its parts is causally constituted. Persistence merely represents another side of the same coin, notably how long the system thus constituted is preserved by the same interactions that constitute it. The length will partly be decided by how intrinsically stable the system is, i.e. how well it is able to uphold itself over a period of time in which it does not interact with other systems, and in part by the absence of interactions with other systems. For instance, while a crystal glass is intrinsically stable, it is easily shattered by interactions with other things. Such a glass may last for a shorter time than something less stable if it is unfortunate enough to interact in a destructive way with something, but longer than sturdier things if it is lucky enough to avoid a tumble and a fall. We then have an explanation not only of how compound entities are causally constituted but also of why they persist for longer or shorter periods of time.

7   Substance and Process

In Chapter 6 I stated that the conception of constitution and persistence that I present there implies that compound objects are to be considered processes, if by 'process' we mean any entity for which change is essential for its continued existence. Furthermore, I claim that they can still be considered substances. This contradicts received wisdom according to which substance and process ontologies are contrary and incompatible views. Here, in Nicholas Rescher's words:

> Process philosophy diametrically opposes the view—as old as Parmenides and Zeno and the Atomists of Pre Socratic Greece—that denies processes or downgrades them in the order of being or of understanding by subordinating them to substantial things.
>
> (Rescher 2012)

However, like Albert William Levi (1958), I find it less than perfectly clear just what the relevant opposites are. For one thing, as Rescher himself points out, what unites philosophers of process is more an opposition to substance ontology rather than a common view on the nature of process (Rescher 2012). And, for another, it is not clear that all metaphysical systems that make use of the notion of substance share the features process ontologists ascribe to substance ontology. Levi indeed says, 'it is not easy to find in either Bergson or Whitehead criticisms which specifically controvert the relevant sections of Aristotle's *Metaphysics*' (1957: 750). Rescher implicitly recognises this as well, because even though he characterises process philosophy as diametrically opposed to substance philosophy, he includes Aristotle as a 'key figure in the history of process philosophizing' (1957: 11). So, what exactly is this conception of substance to which process ontology is opposed?

7.1   Substance Ontology

The term 'substance' is a philosophical term of art, which covers all kinds of sins. Here I will only delve as far into the details of the issue of substance as is necessary to clarify the contrast between substance and process ontology, and in particular for identifying which versions of each view stand a chance of being compatible with each other. For a more complete overview, see Robinson (2020).

One sense of understanding 'substance', which is possibly obsolete, is of whatever is considered the most fundamental entity out of which everything else is made. According to this use, processes are the substances of process ontology. It seems clear that the dispute between process and substance ontology is about the nature of the most fundamental entities, in which case this sense of 'substance' is irrelevant to elucidate the kind of opposition process ontologists have in mind.

Contemporary views on 'substance' are of a certain kind of fundamental entity, and they are all popularly believed to be closely connected to the kind of entities that Aristotle picked out using the term 'ousia', i.e. persistent entities that exhibit a certain unity. These are what we ordinarily call 'things' like sticks, stones, and teacups. We can ignore the complication to do with the distinction between natural unities and artefacts, which, although important to understand Aristotelian metaphysics, is not a concern in modern metaphysics (although it should be). It is with this focus on 'things' that process ontologists think we can begin to see a contrast between substance and process ontology. The idea is that substance ontology assumes that the fundamental constituents are whatever remains fixed and unchanged in an ever-changing world, while processes are ever-changing. Two questions arise. First, is it fair to think of Aristotelian hylomorphism in terms of our ordinary notion of 'things'? Second, do substance and process ontologists really disagree about which of the entities that appear to us in our everyday conceptual scheme are more fundamental than the other? I think not.

If we begin with the latter question, then the answer is clearly no. Everyone agrees that teacups are more fundamental than a tea party, atoms are more fundamental than teacups, and elementary particles more fundamental than atoms. Process ontologists just insist that teacups and atoms are processes, and that even when we get down to the elementary particles of the standard model, they too are processes rather than substances. So, the disagreement is not about *which* entities are fundamental, but about the nature of what everybody regards as fundamental. The real disagreement is whether the most fundamental entities are dynamic or static, and process ontologists think that substance ontology makes these entities out as being inherently static. But are substances really static on all accounts? A good place to start answering that question is to have

a look at what Johanna Seibt calls the *paradigm of substance ontology* (1996), which I think is the clearest statement to date of what contemporary substance ontology amounts to, from the perspective of contemporary process ontology.

7.2   The Paradigm of Substance Ontology

What Seibt dubs the *paradigm of substance ontology* is the result of an attempt to break up the seemingly irresolvable dispute between the endurantist and perdurantist approaches to persistence (2000). She finds both approaches to be inadequate but also that proponents of each view seem incapable of fruitfully addressing each other's concerns in order to resolve the issue. Seibt follows Ramsey's advice in solving these kinds of situations by searching for implicit assumptions that might hamper the treatment of the issue. Such implicit assumptions can become embedded in the presuppositional depth-structure of the debate and go by unnoticed. The idea is that if these presuppositions are brought to light, it becomes clear how they govern and constrain the construction of ontological theories. Seibt concludes that the debate on persistence is marred by a number of presuppositions derived from substance ontology, and which prevent us from considering other alternatives:

> Even those contemporary ontologists who drop the traditional talk about substances still import the presuppositions of the traditional paradigm; even in ontologies whose basic entities are events, tropes, or state of affairs, the presuppositions of substance ontology are still operative. Substance-ontological presuppositions enter into the very formulations of ontological problems and restrict the space of possible solutions.
>
> (1996: 143)

Seibt proposes that, within the regional ontology of existence in time (i.e. persistence), there are at least five such presuppositions, which I paraphrase in the following way:

(1) *A theory of persistence is a theory of the existence in time of things or persons*; things and persons are the paradigm examples of persistent entities, and therefore such entities as processes, events, heaps, masses, and collections are neglected.

(2) *What exists is exhausted by concrete individuals and/or abstract entities*; abstract entities are not changes, and do not change, but individuals persist and undergo change.

(3) *The idea of enduring entities is not a theoretical construct*; it is the way we 'naturally' intuit persistent entities, as opposed to the

theoretical construct of perdurance, wherefore the burden of proof lies heavy on the perdurantist.

(4) *The principle of subjecthood*; persistence through change requires a logical subject of which it can be said that it has a property before a change which it has not after; *a*-is-*F*-at-t_1 and *a*-is-*G*-at-t_2 represent a change *iff a*-at-t_1 is identical to *a*-at-t_2.

(5) *The state analysis of change*; change is to be analysed in terms of differences between times in the states of a logical subject; a change consists in *a*-being-*F*-at-t_1 and *a*-being-*G*-at-t_2, respectively.

Seibt argues that it is the presupposition that only substances can be the logical subjects of persistence and change that is the obstacle preventing the discussion moving on. For instance, she notes, all changes that involve the annihilation of the logical subject have traditionally been rejected as real changes, because it has been thought that changes require the persistence of the logical subject. Seibt then proposes a process ontology as a possible way to move on.

Now, I agree with Seibt that these presuppositions are at work in *many* contemporary discussions about the nature of change and persistence. For instance, Hugh Mellor argues that if an animal is blown to smithereens that event is not a change, because the animal, the logical subject of change, doesn't survive (1998: 97). However, I don't think this paradigm of substance ontology really captures the essence of Aristotelian substance ontology, nor do I think process ontology offers a solution to the problem of persistence and change. Consider that Aristotle does allow creation and destruction of logical subjects to be a change; it is a change in the underlying substance that composes the object that modern thinkers identify as the logical subject, say, the biological material that constitutes the animal. As I see it, her critique is only relevant to those that think of substance more broadly as something that sustains properties but which has no intrinsic structure or nature. Let me call that the *modern analytic view* of substance. It is a conception of substance from which most things Aristotelian have been abstracted—in particular, the structure so typical of hylomorphism is absent—but is still popularly believed to have its origin in Aristotle for reasons probably lost in history. In its most extreme form this modern analytic view of substance is expressed as: 'Something must exist, then, and have qualities, without being itself either a quality or a relation. And this is Substance' (McTaggart 1921: §67).

Within the modern analytic view, there is then a controversy about whether substance should be thought of as a 'thin' particular, i.e. as something that sustains the properties but which doesn't include the properties, or a 'thick' particular, i.e. as the property bearer with its properties included. Neither view tells us how the substance unites the properties, and both views assume that the identity of the particular is grounded in the qualities that it sustains. It is the latter, as we will see, that leads to

such conundrums as the problem of temporary intrinsics (Ingthorsson 2001, 2009, 2016), and which leaves us with a static conception of substance whether or not we embrace endurantism or perdurantism, or the *A* or *B* view of time. So, I actually think (3) plays very little part, and I think a sixth presupposition is at play; that the identity of a substance is grounded in its qualities.

Now, it is true that not everyone assumes that the identity of a substance is grounded in its qualities. There is the option of thinking that particulars have a primitive haecceity, a *thisness*, that defines their identity regardless of which properties it has. This allows us to think of the particular as persisting through changes remaining numerically the same, thus avoiding the problem of temporary intrinsics. However, I find that so few endorse a haecceitic conception of identity today that it is unnecessary to dwell on the option. Furthermore, the haecceitic conception does not explain either how a collection of parts constitutes a unity, or why that unity continues to exist over time and to preserve a structure through continuous changes, wherefore it has little interest to anyone whose concern is to explain the dynamic features of reality. With that said, I put the haecceitic view to one side. I will not argue it is unintelligible but try only to present an alternative that is more appealing.

Let us return to the conception of substance as something that sustains properties and whose identity is determined by those properties, and deal with the choice between a 'thin' vs. 'thick' conception. For my present concerns, the distinction is of no consequence as long as the identity of the particular is still assumed to be grounded in its qualities. Arguably, to reject that assumption is to opt for the haecceitic conception that I have just taken off the table. For the sake of simplicity, I therefore proceed assuming that whatever I say about the conception of a 'thick' particular applies equally to the 'thin' conception.

The notion of a thick particular fits very well with the five presuppositions that make up the paradigm of substance ontology, especially the state analysis of change; a particular is something existing at a certain time having a certain property, i.e. 'a-is-F-at-t_1'. And it is intuitively appealing as long as we avoid thinking about the thick particular as a persistent and changing entity. The problem is that the principle of subjecthood and the state analysis of change together with Leibniz' Law (representing the idea that identity is grounded in the qualities) inevitably generate a contradiction (in what follows F and G represent incompatible properties):

(1) *The principle of subjecthood*; a-is-F-at-t_1 and a-is-G-at-t_2 represent a change *iff* a-at-t_1 is identical to a-at-t_2.

(2) *The state analysis of change*; a-is-F-at-t_1 and a-is-G-at-t_2, respectively.

(3) *Leibniz' law*; if a ($a=a$-at-t_1) and b ($b=a$-at-t_2) are identical ($a=b$), they have all properties in common.

(4) *Problem of temporary intrinsics (from (1)–(3))*; a is both F and G.

David Lewis' solution to the problem of temporary intrinsics is to adopt a perdurance account of persistence, i.e. to assume that 'a-at-t_1' and 'a-at-t_2' are distinct temporal parts of a temporally extended compound a, which *is* both F and G, but only in so far as one part of a is F and another part is G. The argument doesn't really show that nothing can endure. It just shows that enduring things cannot change on pain of contradiction, *given* the truth of the premises. However, once you have (allegedly) shown that endurance is incompatible with change, endurance has really lost its *raison d'être* and perdurance comes across like the most generalisable account of persistence in light of the empirical observation that things appear to have different properties at different times.

Now, as I have argued elsewhere (2001, 2009, 2016), it is important to point out that the argument spelled out above actually needs another premise, which is the assumption that all times exist in parity:

> *Principle of temporal parity*; no time is ontologically privileged, everything at all times is equally existent and real.

This principle, which is the cornerstone of the *B* view of time, blocks the possibility of resolving the problem of temporary intrinsics in the way Aristotle does, notably to argue that a can well be F at one time and G at another, because the difference between a-being-F-at-t_1 and a-being-G-at-t_2 is not just a question of difference between two qualitative states, but a going out of being of a-being-F and coming into being of a-being-G. Indeed, Lewis accepts that presentism, the view that only the present exists, avoids the problem of temporary intrinsics, but he also dismisses presentism as a view he just cannot take seriously (1986*a*: 202*ff*); the principle of temporal parity is part of the presuppositional depth-structure that he takes for granted.

The upshot of all of this is that if we accept (1)–(4) + temporal parity then a can only endure if it remains unchanged, and every change has to be reduced to variation between parts. Either way, we have just arrived at a concept of substance that is static and unchanging, i.e. either as the unchanging enduring particular, or as the unchanging collection of temporal parts of a perduring particular.

I think we are now in a position to better understand Rescher's claim that substance ontology either denies processes or degrades them by subordinating them to substances. There are two views that deny processes and both consider change to be illusory, or at least deny that anything comes into or goes out of being; everything simply exists. On the one hand there is the Parmenidean view that there is only one thing (or substance), and that all diversity and difference in time and space, not just change, is mere appearance. On that view there is no change, hence no process. However, as far as I know, no one defends this view today.

On the other hand, there is the view that there is diversity (many substances dispersed in space and time) and difference in their qualitative states (including between times), but no real change. No entity comes into being or goes out of being, nor do they change the qualities they have or relations they bear to other substances. This includes all ontologies that combine a *B* view of time with a single category trope ontology or any kind of perdurantism (including stage theory and Kim-style event ontology). In other words, this is neo-Humean metaphysics.

The view that accepts the reality of processes but degrades them by making them subordinate to substances is the view that processes are changes in persistent things, i.e. they are changes in substances or activities of substances (Munsat 1969; O'Shaughnessy 1971–72; Emmet 1992). However, although I think this view is fairly widely held, its proponents hold it in a state of ignorance concerning the implications of the presuppositions that go into the modern analytic view of substance, and often without taking any explicit position on the nature of time or whether they hold an endurantist or perdurantist view of persistence. They just assume it makes good sense to think of every successive state of an object as an 'event', roughly a Kim-style construction of a particular having a certain property at a certain time, but then an entity that is dependent on the substance that constitutes the state. Emmet is an exception. She realises the dangers of that kind of event ontology.

I find it difficult to offer any meaningful criticism of the view that takes processes to be changes in substances, because I find that the proponents of this view are typically assuming a modern analytic conception of substance as just something that sustains properties and seem unaware of any of the problems I outlined above. I will instead simply ask: is there no alternative? Aristotelian hylomorphism seems to be a promising option.

7.3   Aristotelian Substance Ontology

The Aristotelian view is clearly the best candidate for a reconciliation of substance and process, once we dispel the myth that the modern analytic view, even in its 'thick' version, is Aristotelian. Aristotelian scholars will tell you that the closest one can come to a concept of something that sustains properties in Aristotle is that which the Scholastics called *prima materia*, which is something that can only potentially exist without properties (Johansson 1989/2004: 27; Byrne 2018: 120–31: Sentesy 2020: 300). The idea that substance can be identified with *prima materia* ignores the fact that *prima materia* is a substance only in virtue of some unifying principle that gives it a certain function; only a unity of form and matter is properly speaking a *being* (ousia).

Now, this is not the time to go into the intricacies of Aristotelian hylomorphism, i.e. the details of how Aristotle portrayed the inner structure and workings of a substance. That is a subject that requires its own

book-length study (or an entire series of books). Let it suffice to say that several modern authors claim to revive hylomorphism, for instance, Johnston (2006), Koslicki (2008), and Lowe (2011), while Aristotle-scholars like Marmodoro (2013) are questioning the Aristotelian lineage of their ideas. Perhaps the only thing that these authors have in common with each other and with Aristotle is the rough idea that substances have a unifying structure. However, as Johnston observes, the challenge has been to provide an account of what kind of structure gives the kind of unity that is characteristic of the things we identify as substances:

> *What it is for* this hydrochloric acid molecule *to be is* for this positive hydrogen ion and this negative chlorine ion *to* be bonded together. […] The idea that each complex item will have some such canonical statement true of it might be fairly called 'Hylomorphism'. For it is the idea that each complex item admits of a real definition, or statement of its essence, in terms of its matter, understood as parts or components, and its form, understood as a principle of unity.
>
> (Johnston 2006: 658)

What I have argued is that, in the scientific image, all bonds between every component part of any compound are characterised as interactions, wherefore it seems natural to suggest that interaction is the principle of unity we have been looking for. I have explained how we can think of interactions as a causal phenomenon and pointed out that the unities that are thus causally constituted persist in a thoroughly dynamic fashion. Now, obviously my suggestion does not converge with Aristotle's ideas, because I am abstracting from scientific theories he did not have access to, but the dynamic aspect is at least compatible with what both Kosman (2013) and Sentesy (2020) argue, notably that Aristotle is moving towards an account of being as activity in *Metaphysics*.

I think we can see the dynamic aspects of Aristotelian substances without going into the details of his hylomorphism merely by looking at the way the Aristotelian tradition has typically *characterised* substances (as opposed to *explained*). The following seven claims represents a more or less complete list of these characteristics (for reference, see Loux 1998:107; Hoffman & Rosenkrantz 1997: 2–3):

1. Substances have attributes without themselves being attributes.
2. Substances are contingent beings.
3. Substances are spatio-temporal beings.
4. Substances are capable of change.
5. Substances are persistent beings.
6. Substances are independent beings.
7. Substances can affect each other.

These claims require some clarification. The first can be interpreted as saying that substances sustain properties but are separate from them, but only if one ignores everything said about the inner structure that provides the unity of form and matter. It could perhaps be formulated more precisely as: substances have attributes but cannot be reduced to *mere* attribute.

The second claim relates to the idea that things come into and go out of being, indicating that their existence is by no means a necessary affair. This may appear to contradict the sixth claim, that they are independent entities, i.e. the claim that every substance exists independently of any other substance, because if something comes into being it must be brought into being by something else that preceded it. However, it is relevant to distinguish between what could be called synchronic and diachronic dependence. Any particular substance is contingent in the sense that it comes into and goes out of being, and as such it is diachronically dependent on (i) a previously existing 'material cause', i.e. a portion of matter capable of alteration, and (ii) an 'efficient cause' which provokes a change in that portion of matter, thus bringing a new unity of form and matter into being. However, once in existence, substances have an intrinsic nature independently of the nature of other synchronically existing substances (which, according to the *A* view of time, are the only substances that exist). A person can continue to exist and be the person she is even if her entire family is destroyed (especially if she is unaware of this fact). A statue can survive the levelling of the city it has adorned. However, substances are not so autonomous as to the existence of other substances as to be unaffected by them when they interact. The contingency of substances is reflected not just in their coming into being but also in their destruction. Any substance can be destroyed by the influence of other substances. The independence of substances is thus clearly conditioned by causality; no substance comes into being without a cause, and no substance is immune to the causal powers of other substances that can change and/or destroy them. This conditionality of independence by causality is what constitutes the contingency of being. Independence is further related to persistence, because persistence is on the one hand a consequence of existential independence, but on the other is similarly conditioned by causality. The persistence of an object begins when it is created by a cause and ends by its destruction by a cause. Change is then related to most other claims. Change is an alteration in the properties of a persistent substance, produced by the influence of another persistent being. It is also the creation and/or destruction of a substance. The spatio-temporality of substances is a claim that is less obviously connected to other claims. It has mainly to do with the idea that time and space are boundary conditions upon all being but is also related to the idea that substances are contingent beings that can affect each other and thus share a common realm of existence. The spatio-temporality of substance is thus indirectly related to all other claims.

It is at this point we must observe that, while substances are contingent, *prima materia* is not. Well, it is controversial whether Aristotle really meant to attribute permanence to *prima materia* itself, since it is something that cannot exist on its own, or only to say that the total amount of substance can neither increase nor diminish. Particular unities of form and matter can be created and destroyed, but they can only be created by altering an already existing substance, and only be destroyed by being transformed to another kind of substance. It is in the persistence of substances when their attributes go out of existence, that makes substance more basic than attributes. It is because of the permanence of substance that change is not creation out of nothing, nor the annihilation of things.

Now, while the above characterisation of substances clearly implies that substances are dynamic in the sense that they *can* change, it is not as clear that they *must* continuously change. Mark Sentesy (2020) makes a convincing case for saying that Aristotle was at least moving towards a conception of substances as essentially activities. That is, not merely as objects possessing active powers, but as entities in which their potential is continuously being actualised. That is all well and good, but it is an account that still doesn't bring Aristotle into harmony with the theories and findings of modern science.

Ultimately what I want to take away from this section is that the Aristotelian conception of substance is at least not obviously of something static and unchanging that merely sustains properties, but of something that can both change and survive that change. It is designed to answer the question of how things can always be different and yet be the same, thus fusing the Heraclitan and Parmenidean points of view. The modern analytic view ends up saying that they cannot be different and yet the same, which is kind of Parmenidean. And, as I will now explain, extant views on process end up saying that everything is always different and never the same, which is the Heraclitan opposite.

7.4   Process Ontology

The literature on process philosophy covers a wide field, not all of which is pertinent to the discussion in this chapter, which focuses on the objective nature of persistent material objects. First, there are discussions that don't revolve so much about the place of process/substance in the order of being but only in the order of understanding. I exclude from this discussion views/arguments that profess to be making ontological claims *solely* on the basis of claims about the status of our conceptual scheme or of the language in which this conceptual scheme is expressed. If our knowledge of objective reality is fallible, as it is generally assumed to be, the actual status of our current conceptual scheme, and the language we use to express that scheme, is not a good guide to how reality actually is.

Alternatively, our conceptual scheme may be the product of irrational and random evolution. There is no guarantee that a product of evolution correctly captures the fundamental structure of reality, as opposed to simply being instrumentally useful in an anthropocentric way. Therefore, to show, say, that the concept of substance is at present basic to our understanding of the world does not constitute a decisive ontological argument in favour of the view that material objects are as a matter of fact basically substance-like, not process-like. It is possible that the notion of process may be conceptually subordinate to the notion of substance, given the present status of our conceptual scheme, and of the language in which this conceptual scheme is expressed, while material objects may in actual reality basically be processes.

For similar reasons I also exclude discussions about processes within the philosophy of mind, in so far as they promote process as a basic entity for reasons having to do *solely* with the characteristics of thinking and reasoning. Thinking and reasoning, in fact anything mental, is generally assumed to be of a very different kind than anything physical. It is at least generally assumed to be very difficult, even impossible, to draw any conclusions about the mental from considerations about anything physical, and vice versa. Such conclusions tend to be supported by arguments to the point that there really are no genuinely mental entities (that they are reducible to physical entities), or that there really are no physical entities (physical reality merely being a figment of the mind).

Process philosophy, then, is a doctrine that promotes process in a broad sense but cannot be treated as a single theory about the nature of processes. After all, it is considered to include ideas of process as basically physical (Whitehead 1929), as basically organic (Bergson 1910), and as basically psychological (James 1890). And, besides the mainstream view of processes as a series of events or occurrences (which on its own harbours different accounts), there are those that deny that the notion of process commits to an ontology of events or occurrences (Salmon 1984: 139ff; Seibt 2000). Rescher, indeed, admits that process philosophy is not so much a developed doctrine at all, but rather a projected programme (2000: 21). He nevertheless presents process philosophy, very tentatively, as a doctrine *inclined* towards certain basic propositions:

1. Time and change are principal categories of metaphysical understanding,
2. Process is a principal category of ontological description,
3. Processes are more fundamental, or at any rate not less fundamental, than things for the purpose of an ontological theory
4. Several, if not all, of the major elements of the ontological repertoire (God, Nature as a whole, persons, material substances) are best understood in process terms.

5. Contingency, emergence, novelty, and creativity are among the funda-mental categories of metaphysical understanding.

(Rescher 2000: 5–6)

Note, firstly, that none of the propositions concern the nature of processes; they declare (i) what are the basic categories of metaphysical understanding, according to process ontology, i.e. time, change, contingency, emergence, novelty, creativity, and process, or (ii) declare the explanatory advantages of process ontology. Only the latter may appear to be in opposition to Aristotelian substance ontology, but only if we find that a particular conception of process actually contradicts Aristotelianism. Aristotelianism also includes time and change as essential to our understanding of the fundamental structure of the world, and holds that the world contains contingency, emergence, novelty, and creativity. The question is, do process philosophers account for these notions in a different, or a better way?

Process philosophers do, for instance, claim that a process continuously creates itself, i.e. that there are internal causal relations between the stages of a process. While the persistence of substances is not predominantly explained in causal terms, substance ontology is no stranger to the idea, notably what W.E. Johnson calls *immanent causation* (1924). Dorothy Emmet acknowledges internal causation (1985), and Hoffman & Rosenkrantz claim that the relationship between the parts of compound objects is causal (1997). So, writers on both sides acknowledge this feature too.

What is clearly disputed is whether entities that have hitherto been understood in substance terms are better understood in process terms, which is of course the central issue in settling the question as to whether processes are more fundamental than substances. But, in addressing this issue, we need to get clearer on the sense in which things should be understood as processes.

I think we can express the standard account of process in a very similar manner to that in which I spelled out the Aristotelian view of substance earlier on:

(1) A process is a *complex* of temporally ordered stages or phases.
(2) The stages or phases involve *becoming*—a transition from potentiality to actuality.
(3) The stages or phases are *distinct*.
(4) A process is mereologically *homeomerus*—each stage is itself also a process.
(5) Has a *temporal structure*—each stage has connections with the future and the past.
(6) The stages are *systematically* linked to each other, either causally or functionally.
(7) The stages are *generically* linked—each stage originates from a previous stage and is the source to a later stage.

It is striking that only one of these features clearly contradicts the account of process that substance ontologists like Dorothy Emmet offer (1992), notably that the stages of the process are distinct. She thinks that processes are changes in substances or activities of substances, which manifest themselves like a complex of temporally ordered stages that each involve becoming and are in a way homeomerous in so far as the change in the substance is continuous. Furthermore, they have a temporal and systematic structure and are generically linked. Indeed, the reason they are not distinct is because they are different changes in the *same* substance, which is what explains the generic link. The claim about distinctness is also one of the main features of the standard view of process that Salmon (1984), Seibt (2000), and Nicholson & Dupré (2018) reject.

The question for me is whether process ontologists are able to plausibly explain the generic link between stages without resorting to the notion of a common constituent, of something being preserved from one stage to the other, i.e. of something resembling the concept of substance. Without a notion of some kind of common constituent, process philosophy seems to be committed to claim that the stages are absolutely distinct. Christopher Austin makes a similar point, when discussing recent attempts to develop a process ontology for biology, notably that their rejection of everything 'substantial' jeopardises their viability (2020).

Now, it is a worry that, with the exception of the standard view I present above, explicit alternative accounts of process are actually very scarce. The alternative accounts amount to no more than rejections of the notion of substance—well, a rejection of the modern analytic view—and an appeal to empirical facts about the dynamic nature of the phenomena studied by the natural sciences. In the recent volume *Everything Flows*, the closest we get to a characterisation of what a process is amounts to an inclusion or exclusion of theses (1)–(7) above (Nicholson & Dupré 2018: 11–13). They say that processes 'are extended in time: they have temporal parts' and 'depend[s] on change for [their] occurrence' and that it is 'a mistake to suppose that processes require underlying things, or substances'. Instead, they take reality to be 'constituted by processes all the way down', which is to say that they hold 'change, or better dynamicity, as fundamental or primitive' and that 'like time itself, it is continuous' (which is why they reject the Whiteheadian account). What emerges very clearly is that the motivation for thinking about the world in terms of processes is the empirical observation that all the entities of the biological realm are in continuous change. This is a naturalism which I approve of very much. However, I think there is a risk that the opposition to substance is based on a flawed dichotomy, which takes its departure from a mistaken identification of the notion of substance with the modern analytic view.

7.5   The Problem of Process Ontology

I'll begin by outlining the problems of process philosophy by address-
ing Whitehead's view, and then ask whether alternative views fare any
better. I will answer no. It should be noted that I rely on the exposition
of Whitehead-expert Dorothy Emmet (1992), and not Whitehead's own
writing which is notoriously obscure. According to Emmet, Whitehead
portrays a process as a succession of distinct stages, phases, or events,
each of which involve becoming, or 'concrescence', as Whitehead dubs it;
a transition from potentiality to actuality. The concrescence of each stage
triggers the concrescence of the next.

The problem is that if the stages of a process are strictly speaking dis-
tinct entities, i.e. contain no common constituent, every new concres-
cence is a transition from potentiality to actuality of a distinct entity. We
must then wonder about the ontological ancestry of each stage before
it undergoes concrescence. It cannot exist as a potentiality in the pre-
ceding stage, because that would mean they are not, properly speaking,
distinct. The options I can see are (i) that each stage comes into being
out of nothing, i.e. invokes absolute becoming, or (ii) requires that all
the stages in the history of the universe exist as potential beings from the
beginning of time, and which then become actual in a temporally ordered
fashion (much in the fashion of a Leibnizian predetermined order), the
concrescence of one triggering the concrescence of the next. The latter
view could be illustrated, by analogy, to the sport of lining up dominoes
in intricate patterns and then flipping one of them to watch them all
fall down in an ordered sequence. The finished construction of lined up
dominoes represents reality at the dawn of history, entirely made up of
pure potentiality; the potentiality of the falling of each piece. The falling
down of a domino would correspond to the moving from potentiality to
actuality of the state of lying down, the *concrescence* in Whiteheadian
terms, while the push of each falling domino (of each concrescence)
on the next, would be the causal/generic link, corresponding to what
Whitehead called *transition.*

The domino analogy is flawed, because in our imagination we are giv-
ing ourselves actual dominoes that are actually standing up, containing
the potentiality of falling, and therefore the conception of one falling
domino acting on another is really a conception of an actuality acting
on another actuality. We have to imagine the standing dominoes as pure
potentiality, and that once something becomes actual—i.e. the domino is
lying down—it can influence a pure potentiality. If every stage is causally
connected to the next, but distinct from it—the concrescence of one stage
causing the concrescence of another distinct stage—then a causal realist
must be prepared to accept that an actuality can exert an influence on

a mere potentiality that is not grounded in any actuality. Furthermore, there is the difficulty that the concrescence of a stage partially overlaps the transition relation between stages (the domino still falling while it is beginning to push the next one to it), and therefore the stages are not strictly speaking temporally distinct, and the difficulty of what it is that pushes and what is being pushed when it is claimed that the process is the pushing and being pushed itself.

The Aristotelian view, on the other hand, depicts the transition from potentiality to actuality as occurring from a potentiality which is latent in an existent substance that at that time actually has a determinate and occurrent nature. A brick hits an actually existing window to provoke a transition from an actual and occurrently determinate state of the window to a state that previously was only potential in the actual state of the window: the state of breaking. On this view there are no pure potentialities, only potentialities inhering in actualities. And if we move on to my interaction model, then the parts of a unity are continuously acting on each other to provoke a transition from potentiality to actuality in each other, but a potentiality residing in entities that already exist as actualities. The planets of the solar system are at any given time actually striving outward, but their course is continuously being adjusted by the gravitational pull and away from that outward motion. The potential for changing their state of motion resides in the actual state of motion they already have. The result is a continuous change in the state of motion that results in an orbit that serves to preserve the unity of the solar system. We can trace that kind of continuous modification of the behaviour of every part of everything down to the level of elementary particles. And yet the Aristotelian account allows us also to say that still something remains the same through all these changes. In the end we are of course forced to give up the notion of the preservation of the logical subject and say that, ultimately, 'something' is conserved. But that is already allowed in the original Aristotelian view. And typically this 'something' is something we are unable to individuate in the ordinary sense, much like we are unable to individuate electrons or energy. What reasons can we have for accepting this view? In the past the motivation has been that the alternative is to accept creation *ex nihilo*, which for millennia has been considered tantamount to magic. That is motivation enough for me.

It worries me that philosophers seem all too willing today to accept creation *ex nihilo*. For instance, Neil Williams responds to Judith Jarvis Thomson's complaint against creation *ex nihilo* in the following way:

> Not only is it not *obviously false*, it does not seem false at all. What is wrong with a picture of the world as recreated at every moment? What can one possibly point to as evidence that this is not the true story?

(Williams 2019: 2014)

The most straightforward answer is that the arguments in favour of accepting what Mario Bunge calls the *genetic principle* (1979: §1.5.2), and Craig Dilworth calls the *principle of the perpetuity of substance* (1996: Ch. 2) is that every observation and theory in the natural sciences confirm it, and that it has therefore become a methodological principle. The conservation laws of physics are the most recent manifestations of these principles, and they act as methodological principles in research. Without them, physics would be fumbling in the dark. In high energy physics (and elsewhere), it is essential to keep track of the conserved quantities, in what is described as a 'tiresome book keeping exercise' (Mishra 2013: 116) in order to be able to justify the validity of your findings. Also, to be able to backtrack to mistakes and errors that can have occurred. In contrast, creation *ex nihilo* is typically regarded as an untestable notion. If you accept it, why wouldn't you accept magic and/or more generally the supernatural?

7.6   A World of Processes/Substances

The main argument today in favour of process ontology comes from the observation that every persistent unitary entity that we scrutinise dissolves on closer inspection into a myriad of processes and reveals itself to have vague borders. The sun is a giant mass of churning hydrogen, continuously undergoing fusion reactions to produce helium with the emission of tremendous amounts of energy. Indeed, those reactions are part of what makes the sun (relatively) stable. Furthermore, there is no clear border as to where the sun begins and ends.

On the scale of ordinary middle-sized dry goods, things that to the naked eye appear completely inert, say, a billiard ball, is made up of atoms that essentially are tiny dynamic systems of interacting parts. The dynamic nature of things is more out in the open in the domain of biology. Our bodies are self-regenerating systems in which a multitude of processes are continuously in progress. We may think that digestion is a process carried out by the intestines, which to us look more like 'substances' (in the modern analytic sense), but when we look at the intestines, they are a churning mass of cells and enzymes and capillaries, etc. The cell is in turn a similar system of various parts with different functions, and they too turn out on closer inspection to be systems, etc.

However, to decide whether the change or the persistence of a system is more fundamental we need to carry this reduction through to the fundamental level, and then the empirical buck stops at a dead end. The standard model of physics is our best empirical theory yet about the most fundamental entities of the universe, and it is typically represented as containing 28 different elementary particles. However, 'elementary' is taken to mean that they have no *measurable* internal structure, which in turn is taken to mean that it is *unknown* if they are composed of further

particles (Braibant, Giacomelli, & Spurio 2012). The ultimate point is that empirical science is at any given point unable to definitely determine that we have reached rock bottom. We may, for instance, wonder whether the fact that physics has calculated (and for some, confirmed) the decay rate of these elementary particles can be a reason to think that they cannot really be 'elementary' in the philosophical sense of 'indivisible'; they seem to be able, in theory, to transform into something else. However, given what the standard model says today, does it convey a message of the elementary particles being nothing but activity? It seems to me that the answer is indecisive. Plenty of elementary particles have a very short lifetime. The muon has a decay rate of approx. 2.1×10^{-6} seconds (Beringer et al. 2012). However, the proton, which is a compound composed of three quarks, has a calculated lifetime of 1.67×10^{34} years (Bajc, Hisano, Kuwahara, & Omura 2016). In other words, it has a lifetime of approximately 16 billion trillion trillion years. In comparison, the universe is currently at 13.8 billion years. That is pretty stable.

Now, what to make of all of this? My own methodological position is that we should take the findings and theories of the empirical sciences seriously *and* should strive to construe metaphysical models that are not only compatible with them in the sense of 'do not contradict them', but also clearly connect with them. I think my account of causation in terms of interaction delivers on that goal. I think it also explains why every compound entity we pretend to understand scientifically is an entity for which change is essential, resulting in a temporally structured series of temporary stages that are causally connected. I take this to tick enough of the process ontological boxes to substantiate the belief that we should take 'change, or better dynamicity, as fundamental or primitive' (Nicholson & Dupré 2018: 13) and that it is compatible with the claim 'that the living world is a hierarchy of processes, stabilized and actively maintained at different timescales' (Nicholson & Dupré 2018: 3). However, the explanatory value of also assuming that through the continuous change there is something preserved, strikes me as being too great to give up on. Also, I take the existence of entities that can persist for 16 billion trillion trillion years as quite a decent anomaly for the claim that nothing is stable. Given that pretty much everything in this world contains a sizeable quantity of protons that remain stable over time, do we not have reason to assume that at least something in the world is pretty darn stable?

Is the position I sketch ultimately a substance or process ontological position? I am happy to leave that open for further discussion, but it is at least not a modern analytic view of substance, nor is it a view of process that denies every notion of a common constituent that in some sense remains stable (NB, *stable*, not *static*) over time. Indeed, there are plenty of philosophers out there who have come to the conclusion that the traditional dichotomy between substance and process, and endurance and perdurance, is outdated and that we need to take a fresh look at the nature of the basic constituents of nature (see, for instance, Meincke 2019).

8   Powers

The philosophical dispute about the nature of powers is best understood against the background of what has for some time been the dominant view of properties generally. Armstrong calls it the 'British empiricist' or 'British analytic' view (2005: 312), but I will call it the *modern analytic view of properties* (MAP) because it belongs to the same conceptual framework as the modern analytic view of substance that I introduced in Chapter 7. According to MAP the determinate state of reality is grounded—on the most fundamental level—in causally inert and self-contained qualities (Black 2000), while powers are, at best, second-order properties that either supervene directly on regularities in the pattern of fundamental qualities, or have their basis in regularities determined by laws of nature. The idea that properties divide into two mutually exclusive types—non-dispositional qualities (sometimes called 'categorical properties') vs. non-qualitative dispositions—of which the qualities are more fundamental than dispositions, is what is known as the *categorical/dispositional distinction* (C/D distinction).

Today, neo-Aristotelians are challenging MAP, arguing that powers can be fundamental but disagreeing about whether or not to reject the C/D distinction. Some friends of powers indeed insist that 'any decent theory of dispositions should preserve this fundamental distinction' (Ellis & Lierse 1994: 34) and many enough find the view that properties can be both quality and power 'totally incredible' (Armstrong 2005: 315). My goal is to explain why the distinction should not be preserved, and why the identity of quality and power is totally believable.

8.1   Preliminary Remarks About My Treatment of Powers

I have so far avoided any discussion about the nature of powers, even though powers are fundamental to the view I propose. This is because my discussion about causation doesn't really hinge on any particular conception of powers, but also because a discussion of the nature of powers requires a book-length study of its own. In fact, I doubt a discussion of powers really adds much at all to what this book has already said, but

it would be an omission not to say at least what kind of powers I think fits with the view of causation I advocate. However, my discussion won't really engage in the subtler details such as you find in Molnar (2003), Bird (2007*a*), and Livanios (2017).

In fact, perhaps the main reason I discuss powers is that I think it would be a missed opportunity not to strike the iron while it is hot and continue my illustration of how neo-Humean thinking has influenced contemporary causal realism, not only with respect to the nature of causation but also the nature of powers. For instance, the reason so many friends of powers stick to the dispositional analysis of powers is to a great extent due to the empiricist conviction that the nature of unobservable powers must be accounted for in terms of how they indirectly turn up in experience, in the form of observable changes in observable/measurable qualities, or in our everyday lived experience. I believe that once those influences are laid bare, they will no longer appeal (as much) to the causal realist as reasons to maintain the *categorical/dispositional distinction* and it will be easier to accept that properties can be both qualities and powers; i.e. powerful qualities. Now, I have no hopes about proving that properties actually are both qualities and powers, but only to strengthen my previous claim that we *can* think of them in that way (Ingthorsson 2013).

8.2   Different Views of Powers

Now, while an understanding of the contemporary debate about powers ultimately hinges on a historic retrospect of the rise of empiricism which ultimately gives us neo-Humeanism, it is best to first offer a brief overview of the disagreement among causal realists today. On the one hand there is the *pure powers* conception, according to which powers and qualities are mutually exclusive types of properties—as tradition dictates—but which denies that powers must be grounded in such qualities. The proponents of this view can be called *purists*. Purists take the nature of powers to be exhausted by their 'dispositional essence', something they construe as a relationship to whatever the power is able to bring about, their 'manifestation'. That is, they accept the dispositional characterisation of powers. On this view, powers are really nothing but a relation to something that has not yet come to be (wherefore powers are often thought of as a mere potentiality that has not quite yet come to be). Qualities, on the other hand, are construed as having a self-contained and intrinsic nature that is not essentially related to anything in particular. Consequently, powers are 'pure' by only having a dispositional essence while qualities are 'pure' by lacking a dispositional essence (Ellis & Lierse 1994; Molnar 2003: Ch. 3; Bird 2007*a*: Ch. 5). Purism divides again into *dualism* and *monism*. Dualists accept the existence of pure qualities alongside pure powers on the fundamental level (Ellis & Lierse 1994; Molnar 2003: Ch. 10), while monists hold that all fundamental properties are pure powers (Bird 2007*b*).

The dissenting view is that we can make sense of fundamental properties neither as pure powers nor as pure qualities, but must understand them as both, i.e. as *powerful qualities* (Martin 1997; Heil 2003; Jacobs 2011; Ingthorsson 2013). However, it is admitted that the ways we *conceptually* understand properties as powers and qualities, respectively, aren't obviously compatible. The challenge then is to explain why they must be both, despite appearing incompatible, or to find a way to conceptually reconcile them. My focus is on the latter.

Very briefly, I will first challenge three ideas supposed to support MAP. First, the idea that qualities are observable and therefore knowable, while powers are mysterious theoretical posits. Second, the idea that Hume's treatment of causation really shows that what an object is like is contingently connected to what it can do. Third, the idea that Locke's view about properties supports the idea that only qualities are properly speaking in the objects while powers are only in the mind, which then is taken to support the idea that powers can at best be higher-order properties grounded in the categorical qualities. Finally, I will argue that our grasp of geometrical properties like shape is a rare example of where our grasp of quality and power goes hand in hand. The trick is then to explain why our grasp of geometrical properties is exceptional, and why we should still infer from that exceptional case that all natural properties are both quality and power.

8.3 Are Qualities Observable?

Nelson Goodman gives a very eloquent statement of what used to be the received view about properties, notably that only qualities are observable and thus knowable, and consequently ontologically basic because they can easily be thought of as intrinsic and fundamental properties of objects:

> 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 no less important to us than its overt behaviour, 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.
>
> (1954: 40)

Once powers are accepted as epistemically dependent on qualities and can only be characterised in terms of changes in the observable qualities, it invites the thought that they are also ontologically dependent and therefore properties that cannot be fundamental.

However, problems have emerged in this popular view to the degree that a survey of the literature reveals an array of conflicting statements:

(i) qualities are observable and thus knowable (Goodman 1954: 40), (ii) qualities are unobservable and in principle unknowable (Lewis 2007; Sparber 2009: 141–2), (iii) qualities are needed to make sense of powers (Ellis & Lierse 1994; Armstrong 1997; Ellis 2010), and (iv) we can only make sense of qualities in contrast to powers, as properties that do not have a dispositional essence (Armstrong 2005; Bird 2007*a*: Ch. 4.1). What is going on?

The reason for this confusion, I believe, is that 'observable' means different things in different conceptual frameworks. Qualities may come across as observable if we take our unit of analysis to be our everyday lived experience, in particular if the focus is on the explanation of statements in natural languages, and naively understand whatever appears in our awareness as 'what we see'. Also, if we endorse direct realism and assume that in perception we are directly acquainted with the object itself; that it is not mediated by phenomenal qualities. But if we believe conscious experience is a product of the senses in response to a stimulation by external objects, there is a problem. The problem can be formulated as a combination of what George Sparber calls *the argument from humility* and *the argument from quidditism* (Sparber 2009: 141–9), which work against MAP and in favour of fundamental powers.

It is important to note that I make no effort to prove indirect realism. I am merely pointing out consequences of holding certain views in an attempt to elucidate a framework in which it is *possible* to think of properties as both qualities and powers. I don't aim to prove that everyone *must* think of them that way. However, I also believe that to point out such consequences has considerable force in the controversy I am addressing, because most of those engaged in the issue do accept some form of indirect realism. If you, dear reader, are an indirect realist, you have reason to pay attention.

Returning to the arguments Sparber identifies, if experience is produced by the senses in response to external stimuli, then powers are the only natural properties that can stimulate the senses to give rise to experiences. We now have two ways to argue for the same conclusion, first, from humility:

> *The argument from humility*. If phenomenal qualities represent the properties that gave rise to them, then the only kind of properties represented in experience are powers. If the only properties we can know are those represented in perception, the only properties we can know are powers.

The argument aims to show that neo-Humeans, in so far as they accept indirect realism, put the cart before the horse; it is the concept of causally inert quality that is a theoretical posit whose nature cannot be known, which is inferred from observable powers. Alternatively, that

neo-Humeans are assuming that natural qualities of external physical objects resemble the phenomenal qualities they are acquainted with in perception, only to realise they cannot explain how such qualities affect the senses. In this light, an ontology of powers appears to have an advantage, at least from an indirect realist perspective. Indeed, Lewis accepts the force of the argument and endorses *Humean humility*; he admits that the nature of the properties that make up the mosaic of local matters of fact cannot even in principle be known (Lewis 2007).

Second, from quidditism:

> *The argument from quidditism.* If phenomenal qualities do not represent the powers that immediately gave rise to them, but instead the qualities that ground the power that can affect the senses, then the contingency of the connection between the causally inert qualities and the powers they ground means we cannot know whether the phenomenal quality that is the product of the exercise of powers correctly represents the quality.

The conclusion really is that the nature of neither qualities nor powers can be known.

Now, given this result, one may wonder why the notion of pure quality still has such a strong hold on the philosophical community. I think it is because the argument from observability isn't the only reason qualities are found to be ontologically fundamental. There is an *a priori* argument that appeals to everyone, but it works its magic implicitly because the philosophical community seem oblivious of the influence it exerts.

8.4   *A Priori* Reason for Qualities Being Fundamental

Lurking in the presuppositional depth-structure of most ontologies, I think we can find what could be described as a proto-idea of what it is to be a fundamental property and which explains the popularity of the purist distinction between qualities and powers, and why so many believe that powers cannot be fundamental. We are dealing with a cluster of very general characteristics popularly used to describe fundamental properties (seldom all at once), notably being *simple, first-order*, *independent*, *intrinsic*, *monadic*, *objective, actual* or *occurrent,* and *determinate.* Indeed, dispositional essentialists like Molnar believe that pure powers, if fundamental, must have these characteristics (Molnar 2003: Ch. 2.4). I am sure every reader recognises them, but let me spell out the interconnections between them:

> *Simple*—Fundamental properties must be simple, because if a property is a combination of properties it is dependent on those other properties and hence not fundamental.

First-order—Fundamental properties must be first-order (they must be properties of a particular rather than properties of a property) because if they were a property of a property, they would be dependent on other properties and hence not fundamental.

Independent—Simple, first-order, fundamental properties must be ontologically independent, i.e. can be instantiated independently of the instantiation of other property instances. If they could not be instantiated independently of other property instances, they would be dependent on other properties and hence not fundamental.

Intrinsic—Fundamental properties that are simple, first-order, and independent will also be intrinsic to their bearers; they will be instantiated in an object independently of anything external.

Monadic—Fundamental properties that are simple, first-order, independent, and intrinsic must be monadic or non-relational, that is, they can be instantiated in 'one place only' (which is arguably just a different way to express independence).

Objective—If fundamental properties are ontologically independent of other things then their nature and existence must also be independent of what we happen to believe about them; they are *objective*.

Actual/Occurrent—If an object is the bearer of a fundamental property then there are no two ways about it, it *actually* has it right then and there, which is arguably the same as saying that it has it *occurrently*.

Determinate—Things have at any given time (actually/occurrently) a fully *determinate* nature. A red object is never indeterminately red, but surely always has a surface colour with a determinate hue, lightness, and chroma.

I will use the term BASIC as shorthand for properties that are simple, first-order, independent, intrinsic, monadic, objective, actual/occurrent, and determinate. Remember that the point is not to establish that BASIC is a true characterisation of fundamental properties. The point is to illustrate how this proto-idea leads philosophers to the conclusion that powers cannot be fundamental properties, i.e. BASIC. It is because powers have been understood as relations to manifestation properties in accordance with the dispositional analysis, and thus as neither monadic, intrinsic, nor independent.

We can now appreciate that proponents of pure powers vs. powerful qualities face different challenges. The purist needs to show how a property whose nature is exhausted by a relation to a manifestation property can be BASIC. In particular, how can they be independent of other property instances? Indeed, as I have argued elsewhere, this is what creates an unresolved tension in Molnar (2003) and Bird's (2007*a*) characterisation of powers (Ingthorsson 2015), because although they want powers to be a relation to other properties they also want them to be intrinsic to their

bearers. Molnar's solution is to argue that powers have a kind of physical directedness that is analogous to the directedness of intentional states—it doesn't require the existence of what it is directed towards—while Bird argues that they are related to *possibilia*, which do not need to be instantiated for the power to be directed to them.

The proponents of powerful qualities face the challenge of how to portray a property that plays a distinct causal role without its nature thereby being *constituted* by a relation to a manifestation or determined by the nature of the manifestation. This is a challenge when the philosophical community predominantly conceives of the very idea of a causal role in terms of a relation to a manifestation, wherefore the lack of a relation to an outcome is understood as a lack of a nature.

Now, I have elsewhere offered a more fine-grained discussion of the concept of BASIC properties and its implications for the regress of pure powers (Ingthorsson 2015). Now it is time to further undermine the conceptual integrity of MAP, by arguing that in addition to the problem with the observability of qualities, contemporary realists, whether neo-Humean or anti-Humean, cannot really appeal to Hume to support the separation of quality from its causal role.

8.5   Hume's Separation of Quality and Causal Role

The idea that qualities are causally inert ultimately derives from Hume's analysis of causation (1739: 69–176). At the heart of the analysis is what Hume considers a principle, namely that *'there is nothing in any object, consider'd in itself, which can afford us a reason for drawing a conclusion beyond it'* (1739: 139 [Hume's emphasis]). The objects he refers to are the objects of perception, which are the only things he thinks are present to the mind. Indeed, he thinks the content of our minds reduces to impressions and the ideas derived from impressions. There are no abstract ideas, nor ideas about how things really are, as opposed to what they appear to be like. He is pretty explicit about this:

> Now since nothing is ever present to the mind but perceptions, and since all ideas are deriv'd from something antecedently present to the mind; it follows, that 'tis impossible for us so much as to conceive or form an idea of any thing specifically different from ideas and impressions.
>
> (1739: 67)

The upshot of all this is that there is nothing in the content of a perception of an object that allows us to infer how that object will appear in future perceptions; for instance, how it will change if becoming contiguous with another object. That is, there is no *conceptual connection* between an impression/idea of how an object of perception occurrently *is*

and how it may appear in future (what it may *do*). Furthermore, Hume appeals to the problem of induction to show that we cannot compensate for that absent conceptual connection by appealing to our past experience that whenever property *F* has so far appeared, then *G* has followed. In his view, '*even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experiences*' (1739: 139 [Hume's emphasis]). It is from this argument we are meant to draw the conclusion that qualities (what the object is like) are only contingently connected to a causal role (what the object can do). Such a conclusion will follow if we apply what can be called the 'test of separability', the idea that what can be separated in thought is also separated in reality.

The realist must be careful here. Hume didn't argue that we could separate the qualitative characteristics of *natural* properties from their natural causal role, did he? He said that we cannot infer from the characteristics of the *phenomenal qualities* we are acquainted with in perception anything about *other phenomenal qualities* we are not perceiving. So, what Hume argues is that all phenomenal qualities are conceptually 'loose and separate' and we cannot even form the idea of natural qualities that are different and separate from these phenomenal qualities. Hence, we cannot really conclude on the basis of this argument that, because the phenomenal qualities present to us at different times are loose and separate, then somehow this establishes that natural qualities are only contingently connected to a causal role. We should rather say that we cannot at all conceive of natural qualities and their causal roles in nature. In other words, the argument is useless for anyone wanting to make claims about objective reality, because it really says such claims are null and void.

In order not to be unfair to empiricism generally, it is important to make a sharp distinction between empiricists like Hobbes (1656) and Locke (1689) who thought experience gives knowledge about an external reality, and anti-realists like Berkeley (1709) and Hume (1739), who thought we can only know what is in the head. A similar distinction is found between realist and anti-realist neo-Humeans. Realist empiricists only insist that all our knowledge about the external world be based on experience, while anti-realist empiricists insist that because our knowledge can only be built from the material provided by the senses we cannot know anything about the external world. This is a distinction that isn't consistently drawn in the philosophical community, wherefore realist empiricists make use of arguments that really only work for the anti-realists, and likewise within the neo-Humean community.

The result that what an object looks like in perception is only contingently related to our conception of what it is capable of doing should not be a surprising result for any indirect realist who accepts that the content of perception does not as a general rule represent the true nature of the property as it is in the object. And they don't arrive at this conclusion

from reading Hume, but from accepting empirical findings about perception. Indeed, I take it that we generally rely on science to provide knowledge about what objects are really like, as opposed to what they appear to be like. Indirect realists believe that, when we see an object as red, our senses are representing a property to us in a way that doesn't really do justice to the surface properties of the object. To begin with, a sensation of red does not tell us anything about the ability of the surface to absorb and reflect light, which is what science tells us is responsible for the object appearing red. But the indirect realist has therefore no reason to think that the actual properties of the atoms on the surface are only contingently related to their ability to absorb and reflect light of certain wavelengths. It is only when we have come to some understanding of what the properties of the object are really like as opposed to what they appear to be like that we can consider whether the object's qualitative state is distinct from what it can do. Here is where we look to science for guidance, and it seems to me that the way the sciences describe the qualities of objects identify those qualities with what the objects can do. As far as I can tell, this is, in extreme paraphrase, the reason Alexander Bird denies that powers are quiddities (Bird 2007*a*). I agree wholeheartedly.

8.6   The Powerful Qualities View Implicit in the Scientific Image

Let us consider an example of how science characterises a basic property of a basic entity, i.e. a simple property of a part that has no further parts: the negative charge of an electron. Charge comes in multiples of smallest units (it is quantised), and the electron carries exactly one such unit, e^-. Now, is charge a quality or power? It is clear that physics describes the nature of charge only in terms of what its bearer can do. Charge is understood as a property that makes the bearer experience (in a non-phenomenal sense) either an attractive or repulsive force in the vicinity of other bearers of either negative or positive charge. We have no phenomenal qualities that represent the nature of this property to guide our understanding. We can think of two billiard balls repelling each other and scale it down in our imagination to fit interactions between electrons. But that would not be a conception of what it is to be charged, only of the consequence of an interaction between charged particles. Consequently, we cannot conceive in such terms what it would be to be negatively charged when the bearer is not repelling or attracting.

   Indeed, the potential behaviours (or manifestations) that science attributes to any charged particle are heterogeneous. Any charged particle has the potential to change the state of motion of itself and any interacting particle in a nearly indefinite number of determinate ways, depending on the sign of the charge of the other particle, as well as the kinetic energy and direction of motion of both particles. This is why Coulomb's law,

which describes electric charge, has a scalar and a vector form; the scalar form giving only the magnitude of the charge individually, while the vector form gives both the magnitude and the direction of the change in state of motion depending on the charge and initial values of the physical properties of the particles involved (Griffiths 1998: 59). Still, charge is clearly described as a BASIC property of its bearer even at times when the bearer is not exerting it. It is indeed a relativistic invariant.

Is it then a quality or a power? Charge is understood only in terms of its power-bestowing features, and it is still characterised as a simple, first-order, independent, intrinsic, monadic, objective, actual/occurrent, and determinate property of any particle, even when it isn't interacting with any other particle. Charge fits the formal characteristics associated with BASIC properties, even though its unique essence—that which distinguishes it from other intrinsic properties—is clearly specified in terms of what the bearer can do. But what it can do is not one thing, it is the totality of mutations allowed by Coulomb's law, given the determinate values of the variables of the interacting particles. Indeed, the heterogeneity of what the bearer can do is clearly down to the heterogeneity of interactions that it can partake in, not to a heterogeneity of its intrinsic nature.

Should we say the electron has an infinity of powers, each one corresponding to the possible determinate manifestation of a change in the state of motion of an interacting particle? Bird seems to suggest something like this, when he insists that each fundamental power must have a single manifestation (2007a: 21–4). I think that is extravagant, in the Occamian sense. Why postulate many, when one is enough? My suggestion is that we treat charge as a way of being that we cannot *fully* conceive of as it is in itself, but allow that we can come close to such an understanding by charting what difference it can make to have such a property, and then we denote that way of being with a sign with a symbolic function: e^-. The most important thing to consider is that the specifications of what the bearer of charge can do (the variations of its ability to attract vs. repel) are not constitutive of the unique essence of that property but are instead specifications of what we know about the consequences of interactions between particles with that nature. Like Martin (2008) and Heil (2003), I think these specifications have an epistemological import only.

Esfeld & Sachse have made a more concrete suggestion of how to understand negative charge which is interesting to relate to, namely that the manifestation of the dispositionality of electric charge is not the attraction or repulsion of the opposite or same charge, but the spontaneous generation of an electromagnetic field (2011: Ch.2). It is an interesting suggestion, but the notion of single powers spontaneously manifesting something is problematic in its own right, and then there is the problem of how to generalise the suggestion, both to other examples of possible spontaneous generation of a manifestation by a single power, such as the decay of a muon into a gluon and neutrino, respectively, and to the case

of shape. A ball manifests shape, and in virtue of that shape interacts with its surroundings in various ways, similar to the way an electron interacts with its surroundings to produce a variety of outcomes. But should we say a ball has the power to spontaneously generate shape? It is possible that, in the fact that charge is fundamental while shape of a ball is not, we have a relevant difference.

However, it is a matter of controversy whether these phenomena really involve spontaneous production of a manifestation by a single power. The answer arguably depends on the theory you choose to describe the phenomena. For instance, the decay of a muon can only be treated as spontaneous in so far as the muon is treated as a simple particle without any internal structure up until it decays. The fact that the muon decays, producing three particles, makes you wonder if it really is an elementary particle exempt in the epistemic sense that its internal structure is unknown. If it is treated as an unstable compound, then the decay can be interpreted as the outcome of the interaction between its parts. In the case of fields, the question is whether to treat charge as a simple property of an electron (classical model), or as a power of a particle that spontaneously generates a field (a mix of classical and quantum models), or to treat the particle as an excitation in a field (quantum field theory). In the latter case, the field is not generated by a power, it is the bearer of the power to attract or repel. So, the question of whether there are particular powers that spontaneously manifest something on their own and not mutually with other powers has to do with questions that have to be settled within physics. One could argue in a similar vein to Galen Strawson (2008: 271–2) that since, in this case, no matter what model we chose, we cannot understand the electron, the charge, and the ability to attract/repel, as separate from each other, then the distinction between them is only conceptual and doesn't reflect an ontological distinction. This would fit nicely in with my suggestion that we have to accept that we cannot fully conceive of the electron as it is in itself, but allow that we can come close to such an understanding by charting the various aspects that seem inseparable from it.

Esfeld & Sache's suggestion is, I think, very close to this in spirit, but does actually cater to the idea that the nature of a power must be thought of as identical to a separate manifestation. The idea is that charge is a power that spontaneously manifests a field, but is still different from the field. Aren't we again back in the conception of powers as 'nothing but a relation to a manifestation'?

Now, as I said, I am not out to decisively prove one way or another, only to suggest a way in which we can think of quality and powers as one. For a more detailed discussion of the problem, see Livanios (2017: Ch. 2). And there are in fact a great number of properties that very plausibly are wholly characterised or picked out in terms of what they do and yet are considered to be determinate properties that characterise the object as it

is in itself. Indeed, for the type of theoretical entities discussed above, this is the rule rather than the exception (Thompson 1988). A particle can be described as having a determinate rest mass that remains invariant between frames of reference in relativistic physics. You cannot get much closer to describing something as it is in itself at a given time. And yet, there is no understanding of what mass is, of its essence, except in terms of what an object with mass can do. But the point is that the consequences of the exertion of the power in some particular circumstances must not be made equivalent to the unique essence of that power, any more than we should make the nature of an effect constitutive of the nature of the cause.

Interestingly, I don't think I am presenting you with a new position on the subject. I think this is exactly the view Locke spelled out for us, contrary to popular belief. The popular belief is based on Berkeley and Hume's misrepresentation of what Hume calls 'modern philosophy', for which Locke is a main figure, notably that secondary qualities are only in the mind.

> They who assert that Figure, Motion, and the rest of the Primary or Original Qualities do exist without the Mind, in unthinking Substances, do at the same time acknowledge that Colours, Sounds, Heat, Cold, and suchlike secondary Qualities, do not.
>
> (Berkeley 1709: §10)

Hume similarly claims that modern philosophy takes secondary qualities to be 'nothing but impressions in the mind, deriv'd from the operation of external objects, and without any resemblance to the qualities of the objects' (1739: 226).

8.7   Locke's Powerful Qualities View

Today, the acceptance of the C/D distinction is generally associated with the rise of British empiricism, which is sort of right, but it shouldn't be traced to Locke. I think it is clear that Locke had no notion of a distinction between causally inert qualities and non-qualitative dispositions. Rather he assumed properties were both qualities and powers in one sweet package. I think his view is of interest to the modern reader, which I paraphrase in the following way:

(1) The material world is made up of corpuscles that individually have primary qualities, and which constitute compound objects that in addition acquire secondary and tertiary qualities that are grounded in the primary qualities of the simple corpuscles.
(2) Both primary and secondary qualities are in the object.
(3) Both primary and secondary qualities can be sensible, i.e. both give rise to *ideas of sensation* when they affect the senses.

(4) *All* qualities, whether primary, secondary, or tertiary, are powers.
(5) The distinction between primary and secondary qualities is not between objective and subjective qualities, but between fundamental and higher-order qualities in objects.
(6) The objective vs. subjective distinction is not a distinction between qualities in the object and ideas of sensation in the subject, but between those ideas of sensation that correctly vs. incorrectly represent the qualities of the object.
(7) Ideas of sensation of the primary qualities are *epistemically objective* because they represent those qualities as they really are in the object, while the ideas of secondary qualities are *epistemically subjective* because they do not represent the secondary qualities as they really are in the object.
(8) The intellect can infer the presence of a third kind of *insensible* power/quality, on the basis of the observed changes in the sensible qualities but cannot conceive of it independently of the changes in the sensible qualities.
(9) Insensible tertiary qualities really are in the object.

Claim (i) is uncontroversial, but (ii) contradicts Berkeley and Hume, and today the philosophical community is divided on this point. I take it that Berkeley and Hume's opinion about Locke's views on primary and secondary properties is currently the popular view. It is how the primary-secondary distinction is typically presented in textbooks (for instance, Baggini & Fosl 2010: 124). However, Locke scholars agree with my claim (for instance, see Heil 2003:80; Uzgalis 2020: §2.2). Admittedly, Locke says that the secondary qualities are 'nothing in the objects themselves but power to produce various sensations in us by their primary qualities' and goes on to say about the third kind of powers that they are 'barely powers' (1689: 49). The modern reader who endorses the categorical/dispositional distinction may be tempted to interpret these claims as reductive. However, to read them in that way contradicts the following claims in which Locke explicitly claims that primary, secondary, *and* tertiary qualities are really in the object (NB. when Locke writes 'subject' he is talking about the subject of predication, i.e. the object, and not the perceiving subject):

> Whatsoever the mind perceives in itself, or is the immediate object of perception, thought, or understanding, that I call idea; and the power to produce any idea in our mind, I call quality of the subject wherein that power is.
>
> (1689: 48)

> Thus a snowball having the power to produce in us the *ideas of white*, *cold* [secondaries, RI], and *round* [primary, RI], the power to produce those *ideas* in us, as they are in the snowball, I call *qualities*;

and as they are sensations or perceptions in our understandings, I call them *ideas*: which *ideas,* if I speak of sometimes, as if in the things themselves, I would be understood to mean those qualities in the objects which produce them in us.

(1689: 49)

To these may be added a third sort which are allowed to be barely powers though they are as much real qualities of the subject, as those which I to comply with the common way of speaking call *qualities*, but for distinction *secondary qualities*.

(1689: 49)

[porphyry] has, indeed, such a configuration of particles, both night and day, as are apt from the rays of light rebounding from some parts of that hard stone, to produce in us the *idea* of redness, and from others the idea of whiteness: but whiteness or redness are not in it at any time, but such a texture, that has the power to produce such a sensation in us.

(1689: 53)

As far as textual evidence goes, this is pretty clear. Locke clearly takes primary, secondary, and tertiary qualities to be real qualities of objects that are present in them even when they are not actively producing any effects or sensations, albeit qualities that are grounded in a base of corpuscles possessing primary qualities.

Locke's claim that 'primary qualities of body, which I think we may observe to produce simple ideas in us, viz. solidity, extension, motion, or rest and number' (1689: 49) shows that he takes primary qualities to be sensible, i.e. powers that give rise to *ideas of sensation*. That is, both primary and secondary qualities are powers. Combined, the cited passages show that Locke didn't think the distinction between primary and secondary qualities is a distinction between objective and subjective qualities, but between fundamental and higher-order qualities. The secondary qualities are structural properties, properties the corpuscles cannot have individually, and thus cannot be reduced to mere primary properties (even though they can only exist in so far as corpuscles form a compound).

If we turn to the objective vs. subjective distinction, then this appears very clearly as a contrast between different sensible ideas of quality, not between property in object vs. idea of property in the mind. In Locke's view, the sensible ideas of primary qualities resemble the qualities and are thus objective in an *epistemic* sense (1689: 51). The content of the idea of sphericity represents a spherical object like it actually is, and similarly the idea of a solid. These ideas, while being *in* the mind, are not *biased* by the working of our senses; they do not misrepresent what the object

is really like. However, the sensible ideas of secondary qualities do not resemble the secondary qualities they represent. An object appears red but has in reality a surface texture that absorbs some wavelengths of light and reflects others. The sensible idea of red is in the mind, and it does not represent the object as it is independently of the mind. Ergo, it is subjectively *biased*. But it represents a 'texture' of the object that it has even in the dark (like porphyry), but which the senses leave us ignorant about. The objective vs. subjective distinction applies only to the epistemic status of ideas in the mind.

Admittedly, when Locke says that whiteness or redness are not in the object, it may seem like a rejection of secondary properties. However, Locke is arguably saying that the phenomenal quality of the sensible idea that represents the quality does not resemble what the object is really like, and nevertheless the idea *represents* the quality. Representation works without resemblance, as is generally the case with symbols. The word 'banana' symbolises and denotes the familiar fruit, but the word does not resemble the fruit; the word 'ndizi' does the same job for those who know Swahili. Similarly, the mathematical symbols of the laws of physics represent certain properties and relationships without bearing any resemblance to the phenomena they represent; mass doesn't look like '*m*'. The phenomenal quality of red symbolically represents and denotes a certain surface structure, but without informing us what that surface is actually like. Consider that no experience of red informs us that a red surface is made out of atoms that absorb and emit light of certain wavelengths.

Now, even though our sensible ideas of secondary qualities do not resemble the quality itself, they are still 'adequate' in Locke's view. That is, to see that one object is red and another blue tells us they are different, and that is of enormous practical value even when we don't know what the objects are actually like. It allows us to distinguish edible juniper berries from poisonous yew berries.

What arguments can be mustered for the epistemic objectivity of our ideas of sensation of primary properties? In the eyes of modern readers, Locke only gives the rudiments of an explanation (1689: 54–6). It boils down to saying that the content of our ideas about the primary qualities allows us to understand how they influence the senses to produce the corresponding sensible ideas about them. We can understand how an object which is solid makes an impression on our senses to produce an idea of solidity, but we cannot understand how a distant yellow object affects our senses to produce a sensible idea of yellow. Locke's rudimentary idea points in the right direction but needs to be adapted to our current knowledge of the world. The explanation we can give today of why an object looks yellow is that the object emits photons and they impinge on photoreceptors in the retina, and … wait! Haven't we embarked on an explanation of how the sensible idea of yellow is produced, invoking

primary properties such as solidity, motion, etc. but never the quality of being yellow?

The idea is really that if we ask the sciences why an object appears to be yellow, then the phenomenal quality of yellow doesn't enter the explanation. There is a discrepancy between the appearance of an object as yellow in experience, and the scientific explanation of why the object appears yellow. In fact, it isn't clear how science actually explains why the causal chain of influences ends up as yellow, or green, or red, or some other colour. Nevertheless, we are able to objectively discern, using measuring equipment, the refraction pattern of surfaces and conclude that it doesn't coincide with what we subjectively report. Hence, we have reasons to doubt the objectivity of the sensible idea. However, if you ask the sciences why a billiard ball appears to be spherical, they will answer that, while we do have to take into consideration how our sensory apparatus works, ultimately it appears thus because it *is* spherical. We can even support that conclusion with measurements with radar, ultrasound, laser, and more, to show that in this case the senses are not misleading us. All available means to determine the shape of the billiard ball will come to the same conclusion: the ball is spherical. Perception and science agree in the case of shape, and there is no conflicting evidence to make us doubt the accuracy of the perception. Well, obviously we sometimes get shape wrong, but we can also understand under what conditions we get it wrong and how we deviated from the truth. But with colour we are wrong (or rather ignorant) all of the time about the surface properties of objects, even in the most ideal circumstances. I think we have here a plausible example of a conception of quality in which quality and power fit perfectly well together. If this is right, we have a counterexample to the idea that, while we can understand the nature of qualities in a single act of perception, we can only understand their power in terms of conditionals that say what the object will do in certain conditions.

8.8   Properties and Conditionals

The distinction between qualities and powers is often drawn by saying that ascriptions of powers, but not of qualities, entail certain kinds of counterfactual conditionals (Carnap 1936)–; Goodman 1954; Mumford 1998; Choi 2008). Hugh Mellor (1974) objects, saying that ascriptions of qualities also entail conditionals. For instance, '*a* is a triangle' entails 'if its corners were counted the answer would be three'. However, even though it is admitted that qualities also entail conditionals, our understanding of the quality in question seems to be independent of our understanding of any such entailment, while our understanding of a power appears to be dependent on our understanding of the entailment. We are able to understand what it is to be a triangle without knowing that its angle sum, if measured, is 180°. But we cannot understand what it is

to be fragile without understanding that a fragile object would break if struck smartly. In this case, the conditional appears to provide the defining features of the power—what it *is*—while for qualities the conditional only spells out a consequence implied by the defining features that we are already acquainted with and understand. We understand *a priori* of actually counting that a triangle has three corners, but we do not understand what it is for something to be fragile *a priori* of experiencing something breaking easily. This implies that we somehow understand powers only through the conditionals, but that we have some other independent means to understand qualities. Note now, that the *only* qualities we can pretend to understand in this independent way are the ones we can perceive. It is only those features of reality of which we have *sensible ideas*, such as colours, smells, shapes, etc., that we can *pretend* to understand independently of any conditional that might be given. None of the properties discovered and described by science, but which we have no corresponding perception of, can be understood independently of conditionals of some sort.

However, it is also the case that of all the different sensible ideas we have, very few of them allow us really to infer much about the behaviour of the object we perceive to have the quality. We can perceive the smell of a wine, but what can we infer from that about the behaviour of the wine except how it will affect us? Spatial properties, the kind of properties on which perception and science agree, are an interesting exception. Shape is the best case there is for the identity of qualities and powers. We can arguably understand from our grasp of shapes, *a priori* of testing any particular shape, that a key fits in the appropriate lock in a way that no key with a different shape will do; we understand that a leaden ball dropped onto a pillow will make a hollow corresponding to whatever shape the ball happens to have. Indeed, Ingvar Johansson argues that we can 'intuit [...] a kinematics of mechanisms in the same way as we can intuit some truths in Euclidean geometry' (1997: 411). His point, I take it, is that Archimedes didn't have to build an Archimedean screw to know it would work. He could see that it had to work merely on the basis of the visual representation of the geometry of the mechanism. If this is true, then we have in our grasp of shapes, a grasp of how quality and power go hand in hand. That is perhaps why engineering is such a successful and practical discipline.

Counterexamples have been given to show that shapes do not have a causal role, but they are inconclusive. For instance, Jonathan Lowe has suggested in conversation that a spherical soap bubble does not thereby have the power to roll on an inclined plane, which it should if its sphericity bestowed on it a distinctive causal role. Also, that regions of empty space have a shape but do not thereby influence anything in any distinctive ways.

The soap bubble is problematic for the purist who thinks powers are necessarily tied to particular manifestation properties, but not for the

powerful qualities view. Adherents of the latter will say that, while the property contributes in the same way always, it will produce different outcomes depending on what it is interacting with on a given occasion. Take a standard ball from a pinball machine and submit it to a range of experimental conditions, say, testing it on a range of different automated pinball machines with a range of surfaces, including magnetic and sticky toffee surfaces. The shape of the ball will in each of these conditions contribute to the outcome, but the ball will not behave the same because different powers will be jointly contributing in each case. Furthermore, it will behave differently in each condition compared to balls with a different shape (all other properties being equal); the difference will be down to the difference in shape. Replace the pinballs with soap bubbles with different shapes and the soap bubbles will behave very differently from the pinballs, mostly for reasons having nothing to do with their shapes, but the difference between the behaviour of spherical and non-spherical soap bubbles would be down to shape if that were the only difference in properties.

As for the empty sub-regions of space, they simply are not distinct objects that bear property instances of any kind *independently* of the rest of space, and so should not be expected to have any distinct causal powers of their own.

8.9   The Determinate Nature of Unmanifesting Powers

Although an answer is already hinted at in §6, I should address explicitly the idea that, while an object always has determinate and occurrent qualities, its powers are somehow indeterminate until the moment they manifest some particular outcome. This is a consequence of the idea that the nature of a power is determined by the particular manifestation it brings about and it isn't until the manifestation actually occurs that the power reveals its determinate nature. The 'reveal' is then identified with the instantiation of the power. So, how can we think of unmanifesting powers as determinate property instances? Exactly like we think of an electron as having a particular charge even when it is not exercising that property to change the state of motion of anything. And exactly like we think of a ball as being spherical and having the ability to roll, even though it is not actually rolling down an incline. The trick is merely to accept, as realists have always done, that properties/objects have a determinate nature independently of what we see or measure. Admittedly there are problems with this view on the quantum physical level. I propose we withhold judgement on whether the indeterminacy postulated to obtain on the quantum level really applies on any level above that until we understand the implications of quantum physics better. Indeed, philosophers of the past who have advocated the C/D distinction have not appealed to the indeterminacy of quantum physics. And, if what I have argued above is on the right track, then the trouble of attributing a

determinate nature to powers also applies to qualities. Either it applies to both or neither. As far as I see, we have no reason to regard the perceived manifestation of a power as being constitutive of its nature, any more than we have a reason to regard phenomenal qualities of experience to be constitutive of the nature of the physical qualities they represent.

Martin and Heil admit that we might need to understand a power in terms of the particular manifestations it contributes to produce—they call them 'typifying manifestations' (Heil 2003: 81)—but they warn us against identifying the power with them or taking them to be constitutive of the BASIC nature of the power. In the end we must consider every manifestation as an indirect indication of the presence of a property whose nature we must try to grasp by abstraction from its various manifestations.

8.10   Active and Passive Powers

Having addressed the influences of neo-Humeanism at some length, I return to questions about the neo-Aristotelian conception of powers, notably that some powers are active while others are passive, and that powers collaborate only in particular types of active-passive pairs, to mutually manifest something. The power to heat teams up with the power to be heated to heat up the object that can be heated. The power to dissolve teams up with the power to dissolve to dissolve the object with the latter property. This idea of active and passive powers fits to the description of interacting objects as active and passive.

Now, I have already addressed the flaw in the idea that there are objects that only act but are not acted upon (Agents) and those that only receive an influence but do not act back (Patients). My argumentation there appealed to the scientific image, and I think that is also pertinent here. To my mind, our ordinary everyday view of the world as divided into active and passive components dissolves under the microscope of science. Let us consider the case of water dissolving salt, which Anna Marmodoro has explicitly used to criticise my characterisation of interactions as symmetrical (2017). According to her, there is reciprocity between water and salt, both influence each other, but there is also asymmetry, because while the water dissolves the salt, the salt does not dissolve the water. My reply is that this only works if we stay in the confines of our everyday conceptual scheme, and even then it is problematic. So, for instance, we are typically faced with situations in which we add small amounts of salt to a large amount of water, say, adding salt to the pasta water. But someone who has worked in a saltfish factory (like I have) and seen rain pouring down for hours on mountains of salt to no visible effect, has to wonder what happened to the water. Was it perhaps separated molecule by molecule to bind to the salt molecules? Would we perhaps have a concept of 'salt dissolving water' if we had experiences like that more frequently?

Furthermore, the asymmetry of the interaction between water and salt only makes sense if we attribute the power to dissolve/be dissolved to large quantities of water or salt. However, if we consult science on what this power really is, it tells us it has to do with the properties of individual molecules and not water or salt in bulk. Water molecules are polarised and therefore interact either with the Na^+ or Cl^- in salt crystals, depending on which way it is 'pointing' on contact. Whether the interaction results in the dissociation of the Na^+ and Cl^- depends on whether the ionic bonds in the crystal are stronger or weaker than the attraction to the water molecules. The dissociation will continue only as long as individual water molecules can interact with individual NaCl molecules. These interactions are each individually perfectly reciprocal. However, they are also invisible to the eye, wherefore the only thing we see is salt disappearing from sight. Arguably, how we construe what actually happens is pretty much due to the limitation of our senses.

Does this mean that there is no distinction between active and passive powers? My thoughts on this subject have not taken a definite shape. It seems clear that my criticism of the distinction between active and passive substances does not change the fact that we do have to make a distinction between how an object affects other objects, and how it reacts to the influence of others, but I am less certain whether this requires two distinct categories of properties, or whether we can allow a single property to be an ability both to act and be acted upon. I lean towards the second option. Hence, I do not fully agree with Hansson-Wahlberg (2019) who argues that the category of passive power is superfluous, in so far as he attempts to get rid of the category of passive power without saying what serves the role passive powers have traditionally played. He argues that once we have agreed that water has the power to dissolve sugar it adds nothing to postulate that the sugar possesses water-solubility. I disagree, first, because he assumes that we attribute such properties as 'sugar-dissolving' and 'water-soluble' to water and sugar, which would mean that water has distinct properties for dissolving sugar, salt, bread, etc. Surely there is but one property the water has, namely that the water molecules are polarised and this is what makes it such an effective solvent for a range of substances. Note that ice does not dissolve as well, because in ice the water molecules are too strongly bonded to each other to effectively interact with salt, sugar, bread, etc. So, something has to explain why water does not have the same effect on everything, and that role is played by the properties of the object that is subject to the influence. Admittedly, Hansson-Wahlberg doesn't reject categorical properties, so it is possible that he thinks that all we need are active powers and then categorical properties. For someone like me who thinks all natural properties are powerful qualities, this option is not available.

As I said, my thoughts on this subject are only beginning to take shape but let me illustrate how one could possibly argue that all natural

properties, once we have properly understood them, are at the same time active and passive. Consider inertia, which is typically understood as a passive power, i.e. a body's ability to resist changes in the state of motion. Interestingly, Newton thought it 'may be considered both as resistance and impulse' (1686: Def. III). It is a passive resistance 'in so far as the body, for maintaining its present state, withstands the force impressed', but it is an active power to influence other objects 'in so far as the body, by not easily giving way to the impressed force of another, endeavours to change the state of that other' (1686: Def III). We have to consider that inertia is an intuitive concept, which is typically quantified in physics in terms of mass and/or momentum. And the more you think about it, the less clear are the distinctions between the three concepts. The American *National Institute of Standards and Technology* (2020) explains that the mass of an object is the 'measure of its inertial property or how much matter it contains'. One can wonder whether 'inertia' and 'matter' represents the power vs. quality of 'materiality', in accordance with how that distinction is drawn in MAP, while momentum is a quantification of the amount of power in relation to amount of matter + state of motion. After all, momentum is defined, roughly, as mass x velocity. The interesting thing, however, is that the momentum of an object a both determines the force a exerts on another object b, when they interact, and is a property that changes in proportion to the quantity of the force b exerts on a, which in turn is proportional to the quantity of force a exerts on b. Momentum is active and passive at the same time, i.e. something that simultaneously affects and is affected. We have a similar situation with respect to other properties tied to mass, such as gravitation. All massive bodies experience and exert gravitational forces, and that results in continuous changes in the momenta of the gravitationally interacting bodies, if only in their direction.

Whether it is feasible to generalise this suggestion across the board, so that every fundamental property be understood as something that can affect and be affected, in some way or another, is not something I can answer in this book. We can consider it a matter for further research that will have to include physicists. However, if it is feasible, I can anticipate it having some implications for what is called *the problem of fit*.

8.11   The Problem of Fit

The problem of fit is the puzzle of why powers only seem to team up with particular other powers and not with every power. Traditionally, it appeared that the fit was always between an active power with a passive power. However, I think we need to seriously consider whether this is how things really work or whether it is merely a feature of our ordinary conceptual scheme, generated by our fairly superficial perception of the world of ordinary middle-sized dry goods. Because, in the disparity

between our everyday conceptual scheme and the scientific image, it is not just that the apparently *asymmetric* interactions between *different kinds* of powers are explained by science to be *symmetric* interactions, but arguably also, to a great extent at least, interactions between *the same kind of powers*. The water and salt molecules do not interact in terms of different types of powers, but in terms of interactions that are all electrostatic. Billiard balls exert forces on each other in virtue of their respective momenta. Particles repel or attract in virtue of charge. In the scientific image, it appears that *like power responds to like*. If that is the case, the answer to the problem of fit is that objects interact in so far as they possess the same powers. However, again we have come to a point in the development of this view that is in its infancy, and I leave it at that.

8.12   What Is Doing the Work: Powers or Particulars?

As I have argued, it is the powerful particular that exercises influence and which is influenced, rather than the events in which the particular is involved. I could therefore be interpreted as taking side with substance causation over event causation (but see qualification below). However, contemporary powers-based accounts are typically either silent on the role of the particular bearing the powers, or explicitly argue that it is the powers and not the particulars that are doing all the work (Mumford & Anjum 2011; Buckareff 2017). My position is the reverse; that it is the powerful *particular*, rather than the *powers* of the particular that do the work, even though of course the particulars exert different kinds of influences depending on what kinds of powers they have. The difference may seem minuscule but matters for our understanding of the inner structure of a persistent entity, and for any attempt to tie together our understanding of causation, persistence, and change.

   In my view, a power—like any property—is just the way an object *is*, and powers aren't the kind of entity that can be independently engaged in an activity, nor are they entities that can change. To say that an object's momentum is doing something to another object's momentum, while the object that has the momentum isn't doing anything, seems like a category mistake. In saying this, I don't think I differ much from other writers that have argued in favour of *substance causation*, i.e. the position that particulars *can* be causes, such as Lowe (2008), Helen Steward (2012), and Ann Whittle (2016). We agree the particulars and not their powers are doing the work. However, my view is that *interactions*—not particulars— are causes although I view interactions as mutual exertion of influence between *particulars* rather than between powers. To clarify, Lowe, Steward, and Whittle understand causation more along the standard view and therefore understand interactions as unidirectional actions, and

consequently think that one of the interacting objects is doing the causing, while the other changes.

The argument that contemporary writers appeal to for the primacy of powers is based on the same observation that Hobbes once made, namely that it appears that particulars exert an influence not merely because they are particulars but because they have certain powers. In Hobbes' words: 'the agent hath its effect precisely such, not because it is a body, but because such a body, or so moved' (1656: Ch. IX, 3). However, we would do well to observe that Hobbes doesn't really draw the conclusion that we must therefore regard powers as entities that can in themselves exert an influence. He consistently continues to write that it is the Agent that acts on the Patient and not that a power acts, even though they act and are acted upon in virtue of possessing certain powers. To insist that powers are somehow separate and independent from their bearers seems to rely on the acceptance of the modern analytic conception of substance, and the deficits of that view are spelled out in Chapter 7. On the Aristotelian view of substance, properties are the way an object is, rather than a separate entity that is somehow coupled to the object. So, even if an object acts in a certain way not merely in virtue of being a bearer of powers, but because it has this or that power, it doesn't automatically mean that the power is the right kind of entity to act.

Another problem with powers-based accounts that ignore the role of the particulars is that they fail to connect explicitly to questions about the proper subject of change and to issues about persistence. This is admirably put into focus by Williams (2019: Ch. 8), although I do not agree with Williams' favourable views on perdurantism and even creation *ex nihilo*. But I do agree with him that, to properly connect powers-based accounts to traditional disputes about change and persistence, we have to address the role played by particulars.

The fact that the role of the particular is largely ignored in the literature has serious repercussions. For instance, it leads C.B. Martin to use a flawed analogy—the two triangles example—to illustrate his view, which in turn allows Mumford & Anjum to interpret that analogy as proof that Martin is not presenting us with an alternative view *of* causation but as an alternative *to* causation, i.e. not a view of causation at all (Mumford & Anjum 2015*b*). Martin's example of two paper triangles that together 'mutually manifest' a square when put side by side is meant to illustrate the contrast between his and other views on mutual manifestations:

> You should not think of disposition partners jointly causing the manifestation. Instead, the coming together of the disposition partners is the mutual manifestation: the partnering and the manifestation are identical. This partneringmanifestation identity is seen most clearly

with cases such as the following. You have two triangle-shaped slips of paper that, when placed together appropriately, form a square. It is not that the partnering of the triangles causes the manifestation of the square, but rather that the partnering is the manifestation.

<div align="right">(Martin 2008: 51)</div>

To be sure, it is an ill-chosen example. The triangles can form a square without exerting any influence on each other at all, and without producing any intrinsic changes at all in each other, and thus without becoming a unitary entity. It is an example of a mere Cambridge change. If Martin is not guilty of simply choosing a bad analogy, Mumford & Anjum are right to understand him as advancing what they call a 'compositional view', in which powers come together to constitute a mutual manifestation rather than producing genuine changes in anything (Mumford & Anjum 2015*b*).

Now, I find it difficult to reconcile this 'compositional' understanding with Martin's overall view but admit that it is not easy to decisively refute it. But it does at least seem reasonable to say that *if* the bearers of powers had a more protruding role in the discussion about powers, it is unlikely that Martin would have been tempted to use the example. At the very least, he would have realised that his readers would be wondering whether he actually thinks causation is a mere Cambridge change, and explicitly address that worry. The fact that he uses the example without addressing the worry leads me to think that the triangle example only signifies a bad day at work. All he wants to say, I think, is that powerful particulars come together to jointly influence each other and that as they influence each other they change to give rise to the effect; that is one sense in which you can say that the coming together *is* the mutual manifestation. The following passage can support that interpretation:

In coming to a new account of cause–effect, one should put the matter in the active voice in terms of the manifestings of the manifestations of the disposition in its reciprocatings with its reciprocal disposition partners as active partnerings with mutual manifestings. *That* is cause–effect redesigned and improved.

<div align="right">(Martin 2008: 4)</div>

The passage is a mouthful, I'll give you that, but it is not easily reconciled with Mumford & Anjum's interpretation of the two triangles passively merging to manifest a square.

Indeed, I find that Mumford & Anjum are also guilty of not taking into account the role of the particular, wherefore they too become guilty of using a flawed analogy—the vector model—to illustrate their view of causation. They suggest we can illustrate powers-based causation using vector arrows situated in what they call a 'quality space' (2011: Ch. 2),

building on a suggestion by Lombard (1986). The term 'quality space' is meant to capture the idea that a change in an object is always with respect to some or other property, and that an object always changes from one determinate qualitative state to another determinate state by passing continuously through the intermediary states. An object which accelerates from an initial velocity v_1 to a final velocity v_n will accelerate through all the intermediary velocities v_2, v_3, v_4 … before reaching v_n. We can thus think of any specific qualitative state as occupying a position on a continuum of adjacent determinate states, that continuum being thought of as the dimension of change for that quality (how it can potentially change). The vector model thus represents the dimensions of change of the sum total of (causally relevant) properties located at a given spacetime point during a causal interaction.

The quality space can be either one, two, or three-dimensional; each dimension representing a spatial dimension. In a one-dimensional quality space, the locus of the powers—their position in spacetime—is represented with a vertical line, and vector arrows pointing either left or right represent the potential dimension of change of each power, towards some or other manifestation property in the given situation. In a three-dimensional quality space, the locus of the powers in spacetime is represented with a dot in a 3D coordinate system, and the power-vectors point outwards in all directions from the dot.

The flaw of the one-dimensional model is that it cannot be used to illustrate that in a real-world interaction the powers belong to at least two different objects. Consequently, it depicts causal interactions merely as a nexus of co-present powers that can change along a given dimension regardless of how the powers are distributed between the objects involved. Really, like the parallelogram used in classical mechanics, the one-dimensional vector model can only *accurately* represent the forces operating on one object at a time. It is therefore not really a model of an interaction, but only of one side of an interaction.

The three-dimensional model could possibly be used, because it allows the introduction of multiple loci of powers from which arrows emanate representing the dimensions of potential change. Mumford & Anjum never do that, and explicitly express that they think it makes no practical difference to use the one-dimensional model for the sake of simplicity (2011: 45). Since they only use the one-dimensional model, my point about the flaw in the analogy is still valid.

The moral is: don't forget about the particular in your account of causal interactions. After all, any account of causation that aims to explain how causes produce effects in a manner that accommodates the moral of the genetic principle—that production is never coming into being out of nothing or complete annihilation—must clarify the role of the continuants, because only they are able to persist through an interaction to provide the raw material out of which a new state of affairs comes into being.

8.13   Concluding Remark

The upshot of it all is that I think the conviction with which the philo-
sophical community, by and large, stick to the C/D distinction and MAP
is built on a rather shaky ground, and that power and quality really aren't
as conceptually incompatible as tradition dictates. The case for MAP is
shaky because it relies on the conviction that qualities but not powers
are observable, that Hume's argument can be used to support a realist
acceptance of MAP, and that Locke, and therefore the united front of
British empiricism, stands behind MAP. If I am right, the only thing that
really supports MAP is the incompatibility of the proto-idea of funda-
mental properties as BASIC and the popular relational characterisation
of powers, and the fact that the relationship between the objective nature
of objects and the way they are represented in experience is complicated.
However, it also seems to me that the indirect realist, who regards the
manifest image with scepticism and contrasts it to the scientific image,
can arrive at a conception of properties as powerful qualities. I have
argued that there is one class of properties on which the manifest and
scientific images coincide—shape—and that here we can conceptually
reconcile the powerful and the qualitative. My suggestion is that we can,
by analogy, assume that the same really holds for other properties but
that we are confined to a more abstract and symbolic understanding of
them because of the limitations of our perception.

Now, readers may wonder whether I really have touched upon what
many believe is the core subject of the debate about powers: do powers
have dispositional essences? They will find it difficult to connect my dis-
cussion, for instance, to what Livanios writes on the issues (2017: Ch. 6).
Well, it should however be clear that I don't believe we should think of
such essences as determined by a single manifestation, or even by the
variety of manifestations that a power can mutually manifest. My sug-
gestion is simply that they have powerful qualities. Such qualities are in
some sense essences, i.e. there is something it is like to have the quality,
and that essence will determine the way the object will affect and be
affected. But it is not the same kind of dispositional essence that has so
far been discussed within the neo-Humean framework.

9   A Critique of Counterfactual Theories of Causation

In Chapter 2, I discussed neo-Humean accounts of causation, arguing that they cannot aim to explain causation but only to characterise it. Furthermore, I argued that the standard view explains causation in such a way that it agrees with all the neo-Humean accounts about how causation appears to us (and they with each other). That is, there is overall agreement on the level of characterisation. The point should be further elaborated but I find little space to deal with the range of neo-Humean views in a single book. I have therefore singled out *counterfactual theories of causation* (CTCSs) for further scrutiny, because, I believe, they enjoy the widest following in the philosophical community generally. The point is to show that philosophers find CTCs appealing partly because, like all neo-Humean accounts, they agree on the level of characterisation with the standard view, and partly because the followers of CTCs just aren't clear on the implications of the various ontologies that CTCs are supposed to be compatible with, not even with the implications of neo-Humean metaphysics.

9.1   The Appeal of CTCs

Philosophers seem to find CTCs appealing for three main reasons. First, many find a brute intuitive appeal in the idea that counterfactuals are connected to causation in a way that makes them apt to explain causation. Second, CTCs are perceived to offer the promise of explaining the unclear notion of *causation*, in terms of the (allegedly) clearer and better understood notion of *counterfactual dependence*. Third, CTCs are popularly believed to be usefully applicable pretty much regardless of one's ontological commitments. The latter really boils down to the perceived utility of possible world *semantics* as a 'powerful conceptual tool for constructing philosophical arguments and for analyzing and developing solutions to philosophical problems' (Menzel 2017: §2.2.4). The idea is that this tool works for everyone regardless of their ontological commitments.

I disagree on all three scores and find the philosophical discussion about the professed advantages of CTCs to be strangely one-dimensional. First, I cannot identify any *prima facie* reason to think that causation can be

explained in counterfactual terms, but plenty to think that causation can explain counterfactuals. A handful of very simple examples are enough to show that our acceptance of counterfactual claims is typically determined by our (presumed) causal knowledge and not the other way around. Indeed, for millennia, the dominant philosophical view has been that the nature of causation entails the counterfactual dependence of effects on their causes. Indeed, it still part and parcel of the scientific view of the world that natural phenomena do not happen haphazardly but are determined by the properties of whatever entities are involved. A uniformly moving body will change its state of motion *if*, and *only if*, it is acted upon by another body, from which it follows that 'if it hadn't been for the action of *A* upon *B*, *B* wouldn't have changed its state of motion'. This way of thinking about natural changes goes back to Aristotle and has arguably entered our everyday conceptual framework, and that explains why we now think there is an intuitive link between causal and counterfactual claims. However, it implies that causation explains counterfactual dependence rather than the other way around.

As for the promise of explaining causation in terms of an allegedly much better understood notion of counterfactual dependence, this hinges on the conviction that the construct of 'possible worlds' really is the powerful conceptual tool that it is hyped up to be, and that it really does deliver an explanation of counterfactual dependence that is independent of, or prior to, our understanding of causation. Here we find many dissenting voices. It has been argued that most counterfactuals are in fact false, because the world just isn't determinate enough to make such definitive statements as '*E* only if *C*' true (Hájek 2014). Others claim the construct of possible worlds is hopelessly unclear because it relies on primitive, undefined, and, more importantly, *indefinable* notions such as 'accessibility' (Sowa 2006) and 'overall comparative similarity' (Morreau 2010); notions that are even less clear than causation. Stephen Barker has argued that counterfactuals cannot be about possible worlds (2011), and Jonathan Jacobs has argued that to flesh out the meaning of causal claims about what happens in this world by talking about other possible worlds is just to change the subject (2010). Finally, Marcus Arvan has argued that the Lewis-Stalnaker analysis of counterfactuals is flawed because it cannot handle certain metaphysical and/or epistemological possibilities that it should be able to handle, such as the possibility that this world is just a computer simulation (Arvan 2016).

Indeed, far from being ontologically versatile or neutral, I find that very few ontologies, if any at all, make sense of the idea that counterfactual dependence is more fundamental than causation. For instance, no ontology that posits substantial connections between cause and effect seems to have any reason to make use of the construct of possible worlds, except perhaps as a useful heuristic for making explicit the consequences of what we already presume to know about those substantial connections.

However, in that case it is the substantial connection that explains/grounds the counterfactual claim, and not the other way around, wherefore the heuristic value has nothing to do with ontology. Indeed, like Hájek, I even find reasons to doubt that neo-Humean metaphysics can ground causal claims on counterfactual dependence (see Sections 9.5–9.7).

It is important to keep in mind throughout that CTCs pull in different directions as to what it involves to 'explain causation'; directions that do not necessarily coincide but are rarely kept cleanly apart (see Hájek 2014, for a more detailed discussion). Some proponents of CTCs only aim for a 'good old-fashioned conceptual analysis of causation—a detailed explanation, that is, of how our *ordinary* concept works' (Collins, Hall, & Paul 2004: 30), while others are more ontologically ambitious and offer a substantial ontology of possible worlds in which counterfactual dependence is more fundamental than causation. There is no guarantee that our ordinary concept is true, or even ontologically deep, wherefore we must evaluate separately the purported fruitfulness of CTCs for the elucidation of ordinary concepts (which I discuss in Sections 9.2 and 9.3), and as an ontological theory (which I discuss in Sections 9.4–9.7). To be ontologically serious, CTCs must present a model of reality—some possible worlds construct—whose fundamental principles make counterfactual dependence more fundamental than causation.

It is not easy to separate the two directions in which CTCs are pulling, especially since some thinkers go sometimes in one direction and sometimes in the other. For instance, Lewis only aims for an account of the ordinary concept in (1973*a*), but his modal realism is surely ontologically serious (1986*a*). However, as Brian Weatherson observes (2014), since hardly anybody accepts Lewis' modal realism, the question arises as to what other kinds of ontologies would serve the same purpose for those thinkers who still want to be ontologically serious. I argue that they are at least much fewer than popularly believed.

There is arguably a third direction, also not cleanly kept apart from the other two, namely the study or construction of a purely formal language of possible worlds. Indeed, the definitions of 'counterfactual dependence', 'comparative similarity', and 'nomic dependence' are often enough treated as mere stipulations about what they mean in the formal language, and criticism of them—guided by concerns about their application to ordinary language and to the world—are then shrugged off as misunderstandings of the conceptual framework. Such responses are in place *if* the goal is simply to construe a formal language without *any* concern for its relevance for the way we think and speak or to what the world is really like. However, if it is meant to explain the ordinary concept or some feature of reality, then such criticism cannot be shrugged off so easily. I only address whether CTCs can be deployed to clarify the ordinary concept *or* for more ontologically serious tasks, which is why I ignore the purely formal approach.

9.2   Counterfactuals Explain Causation vs. Causation Explains Counterfactuals?

Proponents of CTCs often appeal to an *intuition* that causation and counterfactuals are connected in such a way as to make counterfactuals apt to explain causation:

> The philosophical importance of counterfactuals stems from the fact that they seem to be closely connected to the concept of causation. Thus it seems that the truth of the above conditional [if Suzy hadn't thrown the rock, then the bottle wouldn't have shattered] is just what is required for Suzy's throw to count as a cause of the bottle's shattering.
>
> (Collins 2007: 1127)

The discovery of that intuitive link, thus suggesting CTCs, is popularly attributed to Hume, following a suggestion by Lewis (1973*a*: 556). Here is the relevant passage from Hume:

> we may define a cause to be an object followed by another, and where all the objects, similar to the first are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed.
>
> (Hume 1748: §VII)

Why did nobody pick up on Hume's alleged counterfactual idea for 200 years? According to Collins, it simply wasn't feasible until Stalnaker and Lewis made certain advances in our understanding of counterfactual conditionals (Collins 2007: 1127). One can question the characterisation of what Lewis and Stalnaker did as advances in our understanding. What they arguably did was to create an entirely new conceptual framework where the term 'counterfactual dependence' was defined internally to that framework but whose application to the ordinary conceptual framework or to reality is entirely unclear.

The attribution of the discovery to Hume is so widespread today that even causal realists accept it (for instance, see Maudlin 2004; Broadbent 2007; Mumford & Anjum 2011: 149). And yet it is plainly wrong. The connection between causation and counterfactual claims has been at the heart of the causal realist tradition for a very long time but then in a manner that grounds counterfactual claims in causation, not the other way around. Consider Galileo, Hobbes, and Spinoza a century before Hume:

> A cause is that which put [placed], the effect follows; and removed, the effect is removed [sic].
>
> (1612: cited in Drake 1981: 217)

> An entire cause, is an aggregate of all the accidents both of the Agents […] and of the Patients, put together: which when they are supposed to be present, it cannot be understood but that the effect is produced at the same instant; and if any of them be wanting, it cannot be understood but that the effect is not produced.
>
> (Hobbes 1656: Ch. X, §3)

> From a given definite cause an effect necessarily follows; and, on the other hand, if no definite cause be granted, it is impossible that an effect can follow.
>
> (Spinoza 1677: I, axiom III)

They all present causation as entailing the counterfactual dependence of effects on their causes: no effect but for the cause. Indeed, the main difference is that Galileo, Hobbes, and Spinoza have a semicolon before completing their statements while Hume puts a full stop in the middle.

A more plausible reading is that Hume was stating the contemporary view before he went on to criticise it but made a blunder regarding punctuation. This explains why he thinks he is describing one view in two different ways. Lewis' reading only raises questions. Why would Hume offer two definitions of causation only to present an account that has little to do with either of them? Hume's own view is not that causation *is* a *real* constant conjunction, from which it follows that '*E* only if *C*'. He argues we are psychologically primed to draw an *invalid inference* about a constant conjunction from a limited number of observations. The further inference that '*E* only if *C*' would be equally invalid. As far as we know, he said, anything can follow from anything.

My reading also explains why Hume never gave a counterfactual analysis himself. The lack of a necessary connection between cause and effect entails the lack of a counterfactual connection between them. For Hume, what is popularly perceived as a counterfactual relation would really boil down to the irrational belief that some generalisations are true. Such generalisations can even be true as a matter of brute fact, if indeed *C* is always followed by *E,* and *E* never happens unless *C* does. However, nothing is needed to make it be so in a neo-Humean reality.

The objection has been raised that the passages from Galileo, Hobbes, and Spinoza do not really contain any counterfactuals and so imply nothing about a causal theory of counterfactuals. My answer is that we are talking about a link between causation and a kind of dependence that is today popularly labelled 'counterfactual dependence'. The linguistic construct we call 'the counterfactual conditional', i.e. 'if *C* had not occurred, *E* would not have occurred' is one handy way to express the nature of that dependence, but it is not the only way. So, I am not claiming that you find a causal theory of the linguistic construct, but that there has existed a

causal theory of the existential dependence of effects on their causes that the linguistic construct is meant to express. The only phenomenon relevant for CTCs is the relation between cause and effect which proponents of CTCs say should be understood as Hume stated it, namely 'if the first object had not been, the second never had existed' (Hume 1748, §VII).

The 'New Hume' interpretation, i.e. that Hume is a realist about causal powers and causal necessity (for overview, see Read & Richman 2007), changes nothing for my argument. If he is, it is still false that Hume suggested that counterfactuals could explain causation.

The fact that the connection between causation and counterfactuals is explicit and clear in the causal realist tradition shouldn't be a surprise. Bunge paraphrases the core idea of causal realism—which he traces back to Aristotle—as: 'if C happens, then (and only then) E is always produced by it' (Bunge 1979: 47), which entails the counterfactual claim 'if C had not occurred, E would not have occurred'. Now, I do not doubt that counterfactual reasoning is an important part of causal reasoning. However, the question here is whether such reasoning really *clarifies* anything about the fundamental nature of causation or whether it *presupposes* an understanding of causation. In the latter case, counterfactual reasoning only serves to spell out our causal knowledge but does not add to it.

Obviously, the truth of a causal theory of counterfactuals hinges on the truth of causal realism. The above only amounts to saying that *if* causal realism is true, it offers a clear understanding of the connection between causation and counterfactuals; it entails that counterfactual dependence is grounded in causation. What remains to discuss is the contrary claim that if it is assumed that casual realism is *false*—that causal connections are contingent—is it possible to ground counterfactual dependence on an ontology of possible worlds? I think not.

It should be noted that the term 'counterfactual dependence' is widely used in philosophy but rarely defined, and the definitions that do exist are problematic. Lewis, for instance, states that E is counterfactually dependent on C *iff* the following two sentences are true: (i) if C were to occur, then E would occur, and (ii) if C were not to occur, then E would not occur (Lewis 1986*b*: 242). The problem is that both sentences represent some fact as dependent on some other fact, and yet neither sentence is referring to an existent fact. Consequently, while the definition may appear intuitively to make good sense as specifications of the circumstances under which a counterfactual is true, on closer reflection the ontology of those circumstances is deeply problematic, because they involve a commitment to negative facts. Lewis' modal realism does not contain negative facts. He would have to say that the surface grammar of 'were' and 'would' in the definition above is misleading; they don't really indicate what wouldn't happen if something else doesn't happen but must somehow be indicative of facts in a reality constituted by a plurality of

worlds. The circumstances that have to obtain in such a reality, to make the above-mentioned sentences true, are the following: a) it is the case in all possible worlds closest to ours, that when *C* occurs *E* also occurs, and b) it is not the case in any possible world closest to ours that *E* occurs without *C* occurring also. I remain uncertain whether b) truly is a non-negative fact, but what is certain is that Lewis' ontology cannot allow expressions like 'were' and 'would' to be understood as statements about mere possibilities. They must be statements about facts obtaining in some subset of worlds that exist in parity with ours. On this understanding, he is not committed to negative facts but only to a plurality of worlds.

Although many proponents of CTCs are averse to modal realism, not everyone is so toward negative facts. Marc Lange talks about the truthmakers for counterfactual conditionals as *contrary-to-fact 'facts'* (Lange 2009: 47) but doesn't explicitly specify what they are. They can't be what he calls 'categorical facts' because they are concrete instantiations of categorical properties. And they can't be concrete instantiations of categorical properties in a possible world since Lange rejects modal realism. Plausibly, they are abstract primitives that are meant to serve the same truthmaking function as categorical facts do in regard to statements about concrete matters of fact. However, since such truthmakers are abstract entities manifesting a relationship between what doesn't happen if something else doesn't happen—i.e. relationships that by definition are never concretely manifested—they do look like negative facts. If this is right, Lange understands 'counterfactual dependence' as a relation between abstract primitives. *Ergo*, if causation is grounded in counterfactual dependence, then concrete causal phenomena are grounded in abstract entities.

Having now presented the causal explanation of counterfactual dependence—and drawn attention to the ambiguity of the term 'counterfactual dependence'—we should ask whether the connection between causation and counterfactual dependence really presents itself to people in a way to make counterfactuals seem particularly apt to explain causation rather than the other way around.

9.3 The Ordinary Concepts of Cause and Counterfactual Dependence

How intuitively plausible is it that we ordinarily understand the meaning of causal claims in terms of the truth-conditions of counterfactual claims? For my part, I doubt that anyone will accept *any* counterfactual as true unless they believe they know the underlying causality. The claim 'I wouldn't have gotten a jackpot on the one-armed bandit if I hadn't tapped it three times on the side and stomped hard once' will be considered mere superstition, *unless* we know there is a mechanism inside the machine that has to be triggered by tapping and stomping to

produce three aces in a row. The same goes for claims involving the healing potency of homeopathic concoctions, say: 'I wouldn't have survived cancer if I hadn't drunk extremely diluted amounts of ginger'. Only those who accept the causal powers of homeopathic concoctions will accept the claim. Someone who feels ignorant about the connection will remain agnostic.

Consider a plausible counterfactual: 'if there hadn't been lightning there wouldn't have been thunder'. Do we first believe that there would be no thunder without lightning and therefore infer that lightning is the cause of thunder, or do we believe that lightning is the cause of thunder and therefore there will be no thunder without lightning? Note that the counterfactual is only true if we first stipulate that the thunder of which we speak is the sound produced by lightning, or, in other words, if we *define* thunder as the sound *caused* by lightning. The counterfactual is false if we allow 'thunder' to also include sounds from explosions, landslides, and earthquakes. Conceptually, the connection depends on the stipulation of a causal link.

Bear in mind that it was not always believed that there would be no thunder without lightning. It was a scientific discovery that lightning is an atmospheric electrical discharge that produces a shockwave in the air, which we in turn perceive as thunder. The pagan Scandinavians believed that lightning and thunder were two correlated effects of a common cause: the spark and sound, respectively, produced when Thor went berserk with his hammer. On that understanding there may well have been thunder without lightning, since Thor's hammer may not always have ignited a spark even though it always produced a sound. Here it is our understanding of the causal connection that determines whether we accept or reject the counterfactual, and, as far as I can tell, the same is true for every causal counterfactual. This argument is completely open to refutation. Just come up with a counterfactual we accept while being ignorant of a causal connection, and which establishes that there is indeed a causal connection. But in the absence of a plausible example, consider the idea refuted that CTCs can clarify the content of the ordinary concept of causation.

I have demonstrated that you can make sense of counterfactuals if you presuppose the causal realist understanding of causation paraphrased by Bunge. On the other hand, I can see no discernible *prima facie* reason for thinking that counterfactuals can explain causation. Of course, if one is already persuaded by philosophical arguments that the causal realist conception is false—that causation is contingent and thus cannot ground counterfactual dependence no matter what our ordinary concept says— one might be tempted to find another basis for counterfactual claims, such as similarity between possible worlds. But will any other ontology do the trick, in particular an ontology of possible worlds? If it is not at all clear which ontological model of possible worlds will ground the truth of counterfactual claims, how optimistic can we be about the promise of

explaining an unclear notion of causation by appeal to a clearer notion of counterfactual dependence?

9.4   Possible Worlds: Truth-Conditions or Truthmakers?

Before moving on to discuss what kind of possible worlds ontology can ground the truth of counterfactual and causal claims, I should address the fact that talk of 'possible worlds' is often understood to be void of ontological commitment, or ontologically neutral:

> Like most contemporary counterfactual theories, Lewis's theory employs a possible world semantics for counterfactuals. Such a semantics states truth conditions for counterfactuals in terms of similarity relations between possible worlds. Lewis famously espouses a realism about possible worlds, according to which non-actual possible worlds are real concrete entities on a par with the actual world. However, most contemporary philosophers would seek to deploy the explanatorily fruitful possible worlds framework while distancing themselves from full-blown realism about possible worlds themselves. For example, many would propose to understand possible worlds as maximally consistent sets of propositions; or even to treat them instrumentally as useful theoretical entities having no independent reality.
>
> (Menzies 2017: §2.1)

Like Martin (2000), I worry that if possible worlds talk is merely a means to state truth-conditions for counterfactuals, but without ontological commitment, it has no explanatory value in ontology.

Truth-conditions, as Menzies says, are conceptual entities, typically said to represent what *would* have to obtain in order for a given claim to be true. However, false claims also have truth-conditions; *ergo*, truth-conditions are not truthmakers. To elucidate the meaning of causal claims in terms of the truth-conditions of counterfactuals has very limited value in ontology, because it will not tell us which truthmakers actually obtain. Can it perhaps tell us what the world *could* be like? Well, if as I have argued, to elucidate the truth-conditions of counterfactuals is just to draw out the implications of the causal knowledge we think we already possess, then the possible worlds that are the result of such deliberation do not tell us anything new about the world. It only makes explicit the consequences of what we already presume to know.

Of course, if we were able to infer from a given claim *one and only one* truth-condition, i.e. what the world *must* be like to make it true, then that would narrow the scope of our ontological investigations enormously. However, the situation is not so simple for the kind of claims that figure in philosophical controversies, such as what it is for some event to be the

cause of another. Our beliefs about what must be the case in order for some claim or other to be true varies according to the philosophical theory we are committed to, and hence there are as many truth-conditions for causal or counterfactual claims as there are theories that profess to provide an account of the claim. That is why I consider it naïve to think that the mere elucidation of the truth-conditions of counterfactual claims will clarify anything about causation or counterfactual dependence.

It is also naïve to think that spelling out truth-conditions of causal and counterfactual claims in terms of possible worlds is ontologically neutral. If truth-conditions only represent what we think the world must be like to make the claim true, *given* some already assumed knowledge of the world, then to spell them out in terms of possible worlds reveals that you have already assumed that our reality is one accurately represented on the model of possible worlds. This means the approach is not ontologically innocuous, not even as a way to spell out how we ordinarily think about the world. To give the truth-conditions of causal claims in terms of possible worlds really is *prescriptive* of how we should think instead of *descriptive* of how we actually think.

What exactly is the philosophical value of elucidating the truth-conditions of causal claims? There is, arguably, a truth-conditional analogue to the underdetermination thesis for scientific theories. For any given set of observations there is always more than one theory compatible with those observations, wherefore the observations are unable to ultimately determine which theory is true (but can at times rule out one or other theory). By analogy, for any causal claim there is always more than one idea about what would make that claim true, i.e. there will be more than one truth-condition for every claim. If this is right, then to clarify truth-conditions is, at best, to clarify what a particular theory is claiming but that does not constitute a theory in its own right or support a theory. Having clarified the truth-conditions of a particular claim in light of various theories, we have yet to address the question of which of those many truth-conditions are most likely to represent what the world is really like. We cannot resolve the latter question by further elucidation of truth-conditions.

Consider again the claim 'lightning causes thunder'. According to CTCs the meaning of this claim is to be given in terms the truth-conditions of the counterfactual 'thunder would not have occurred if lightning had not occurred'; i.e. the claim is true *iff* in all the worlds closest to ours, thunder only occurs if lightning does. Have we dispelled all controversies about what it is that makes causal and counterfactual claims true? We have not. The causal realist will still insist that the causal claim is true *iff* lightning is the invariant and unique *producer* of thunder; the regularity theorist will say the claim is true *iff* every lightning is as a matter of brute fact invariantly followed by thunder, and so on, *mutatis mutandis*, for every theory of causation. But then, surely, to clarify the truth-conditions (plural) of a claim (singular) is only the first step in any ontological

inquiry; it is an inventory of conceivable alternatives. It is then a further step, and ontologically weightier, to discuss which of the many conceivable types of truth-conditions mutually exclude one another and which of these we have best reason to believe actually obtains in reality in the form of truthmakers.

To sum up, possible worlds construed as truth-conditions may be regarded as hypothetical or theory-dependent ideas about what the world might be, but do not justify any claim about the world; they are only as justified as the theory on which they depend. This could be the sense in which talk of possible worlds is meant to be a useful heuristic for talking about causation and counterfactuals within some or other framework, but it can hardly be understood as a theory of causation in its own right and doesn't resolve any controversies about the nature of causation.

Now, obviously, if it isn't in the power of ontology to even address the question of which truth-conditions actually are true—only to make an inventory of all the conceivable truth-conditional models—then ontology only deals with truth-conditions and never in truthmakers. This book is written on the assumption that ontology is concerned with truthmakers. Let us have a look at the kind of truthmakers an ontology of possible worlds can muster.

9.5   Possible Worlds as Truthmakers: Concrete or Abstract?

There are two main ways in which possible worlds are construed as ontological models of reality: concrete or abstract models. Menzel lists a third, 'combinatorialism' (2017: §2.3), but I ignore it here, because it isn't clear to me that all versions of it are ontologically serious, and because any problems found with the concrete and/or abstract constructs will threaten combinatorialism as well.

What Menzel calls 'concretism'—i.e. modal realism—construes possible worlds as concretely existing worlds existing in parity with the world we inhabit but not sharing our spatio-temporal boundary. Some opponents of CTCs argue that this is the only ontologically serious (but absurd) view of possible worlds (Martin 2000; Heil 2003). Indeed, even proponents of CTCs admit that hardly anybody accepts modal realism (e.g. Weatherson 2014).

The problem I have with concretism is that, even if we accept the reality of a plurality of concretely existing worlds—despite misgivings about empirically inaccessible worlds—then the viability of CTCs will vary depending on the ontology of a given particular world, and will at best only appear to be viable if it is assumed that all worlds are contingent in accordance with neo-Humean metaphysics. If a world is governed by primitive natural laws that are not reducible to patterns in the distribution of facts but make things happen in determinate ways, or is a world in which the causal powers of particulars make things happen

in determinate ways, *then other worlds are irrelevant* to what happens in those worlds. It does not matter whether the laws are contingent in the sense that there could be other laws at other worlds. That comparison is irrelevant to decide if the facts in this world could have been distributed differently in this world.

What I am arguing here contradicts the received view. For instance, Collins, Hall, & Paul believe that the counterfactual account of causation is available also to non-Humeans:

> Can facts about the laws themselves be reduced to the totality of categorical facts? Some—notably, Lewis (e.g., 1986*b*)—will say 'yes,' others 'no.' We wish merely to note that with respect to the aims of the essays in this volume, almost nothing hangs on this dispute. (For example, Lewis's account of causation could be adopted wholesale by one who disagreed with him about whether the laws themselves reduce to categorical facts.)
>
> (Collins, Hall, & Paul 2004: 12)

Like Tim Maudlin (2004), I disagree with this claim. For someone like Maudlin who thinks that laws are *sui generis* primitives that secure that nature evolves in certain patterns, then other possible worlds are ontologically irrelevant. Maudlin has no reason at all to adopt Lewis' account of causation. Possible worlds in his conceptual scheme represent at best possible ways the world could evolve within the boundaries allowed by the laws, i.e. possible worlds are determined by the world, and not vice versa. Indeed, *all* ontologies that include a commitment to substantial lawful or causal connections that govern what happens have thereby grounded counterfactual dependence on those substantial connections, in which case other possible worlds become redundant, or merely consequential. Whoever thinks that the laws of a world do not supervene on the contingent matters of facts of that world, but instead connect these matters of fact in determinate ways, has no reason to appeal to similarities between possible worlds as the basis for either causal or counterfactual claims. In such worlds 'C caused E' is not true in virtue of similarities between closest possible worlds, but in virtue of a lawful connection between Cs and Es in the world in question. My conclusion is that it is not clear at all whether anyone other than neo-Humeans will find any reason to adopt Lewis' account of causation. Indeed, it is not even clear that neo-Humeans can make use of CTCs (see Section 9.6).

It bears mentioning that Marcus Arvan offers an analogous argument when criticising the Lewis-Stalnaker analysis of counterfactuals for not being able to handle the possibility that possible worlds are computer simulations. In his view, the primitive causal features of 'necessitarian' simulations/worlds make other worlds irrelevant (2016: §2.3). In fact, he argues that the same goes for neo-Humean simulations (2016: §2.2).

According to what Menzel calls 'abstractionism', possible worlds are 'abstract entities of a certain sort: They are states or conditions, of varying detail and complexity, that a concrete world could be in — they are ways that things, as a whole, could be' (2017: §2.2.). Now, a prerequisite to make this work as an ontological theory, these abstract entities must be the truthmakers of the counterfactuals whose truth-conditions are meant to give the meaning of causal claims. They must be considered to be some kind of contrary-to-fact truthmakers, or contrary-to-fact 'facts' that make our beliefs true. It is not enough that they represent truth-conditions, since they would then just be the conceptual consequences of a certain theory; consequences that would be just as much in need of truthmakers as the original claims we are considering.

I find that an ontology of abstract truthmakers threatens to make concrete reality irrelevant for the truth of our claims about concrete reality (see also Joshua Mozersky's criticism of ersatz presentism (2015: 44)). If counterfactuals are made true in terms of abstract truthmakers, and causal claims are true derivatively of the truth of counterfactuals, then it seems that no causal claim of the kind 'C caused E' is ever made true by concretely occurring instances of causation but solely by some kind of abstract reality. To be sure, it may be true that the events C and E occur merely in virtue of their concrete occurrence, but that does not make true 'C caused E', since this is true only in virtue of the abstract contrary-to-fact truthmaker that 'if C hadn't occurred, E would not have occurred'. And, observe, in particular, that on this account the abstract truthmaker is not grounded in any concrete truthmaker, because there is no concrete state of affairs manifesting what wouldn't have happened if something else hadn't happened. The contrary-to-fact truthmaker that the bottle would not have shattered had not Suzy thrown the rock must exist as *abstractum* whether or not Suzy actually throws a rock, or else we would have to conclude, contrary to abstractionism, that the actual goings-on in concrete reality ground the abstract entities, which would undermine the fundamentality of the contrary-to-fact truthmakers. In sum, it appears to be a consequence of abstractionism that the claims whose meaning we are supposed to elucidate in terms of the truth-conditions of counterfactuals, never are claims about concrete reality but only about an abstract reality whose instantiation is contingent. I think this would come as a surprise to most competent speakers of English, which further undermines the claim that CTCs offer a plausible account of the ordinary concept of causation.

Does anything of the above show that abstractionism is false? No, it just shows that to accept it comes at an intellectual cost that is rarely brought out; a cost that is relevant for making an educated evaluation of the appeal of CTCs. In fact, most of what I have said could be interpreted merely as taking sides with some form of concretism (see below). However, it still seems to me that within the domain of abstractionism we

see the same problems being repeated. First, that CTCs based on abstract possible worlds are implausible as an account of the ordinary concept of causation for the same reasons given above; we ordinarily accept counterfactuals on the basis of our causal knowledge, not the other way around. Second, that even in abstract ontologies where there are substantial connections, those connections will ground the counterfactual dependence between antecedent and consequent in counterfactual worlds, and then comparative similarity between worlds in an abstract plurality of worlds becomes irrelevant. Again, only a small subset of abstractionist ontologies will suit CTCs, and those assumed to be contingent will be open to the same objections I develop in Section 9.6.

In sum, abstractionism threatens to make concrete reality irrelevant for the truth of our causal and counterfactual claims, and abstractionism and concretism alike seem to imply that CTCs could at best be applied to that subset of worlds that are neo-Humean, i.e. where laws/substantial connections supervene on the distribution of local matters of fact. However, I worry there are not enough constraints on the manner in which such worlds can differ—whether concrete or abstract—to provide the structure required to ground counterfactual dependency on comparative similarity.

9.6   Neo-Humean Metaphysics and CTCs

I have suggested that only neo-Humeans should find CTCs appealing, because they reject the substantial connections that others appeal to as the basis of counterfactual dependence, and therefore find themselves in need of something else to ground the truth of counterfactuals. The suggestion is that a plurality of possible worlds will do the trick. Now, from the outset one can wonder why neo-Humeans should think counterfactual dependence is so important that they need to find an ontological basis for it. Shouldn't they just accept that if the world is contingent—i.e. anything can follow from anything—then counterfactuals just are false? Causal realists that give up on necessary connections, Mumford & Anjum being a case in point (2011: 148–55), realise that they are unable to ground the truth of counterfactual conditionals and accept that they are strictly speaking false. Similarly, Hájek argues that most counterfactuals are false because they concern outcomes from chance processes (2014). I will not pursue this worry further here. It simply is a fact that many take CTCs and possible worlds to provide ontological grounds for counterfactuals, so let us consider the options within a neo-Humean framework. Neo-Humean metaphysics, as I understand it, is the view that reality is nothing but the contingent distribution of 'local matters of fact' on which everything else supervenes:

> Humean metaphysics is the deliberate affirmation that, as a matter of fact, our world is such that it instantiates no necessary connections

between its instances. Nothing follows in a metaphysical sense from something else. There is only contingent co-existence. This is the simple and explicit framework of neo-Humean ontologies. The whole but only the distribution of fundamental contingent entities in the world is to be taken as a primitive ontological given. According to the neo-Humean, this distribution is enough to determine all the other features of our world. There could not be a difference in the world without a difference in the distribution of fundamental contingent properties. In other words, this distribution is the basis on which everything there is in the world supervenes.

(Sparber 2009: 15)

Lewis himself describes it as the view that all truths supervene on the distribution of fundamental categorical properties over points or regions of spacetime, and where the instantiation of a categorical property entails nothing about the instantiation of properties outside that region (Lewis 1986*a*: ix–xi). I take it, then, that Lewis accepts the standard view on supervenience:

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 & Bennett 2018: first paragraph)

My worry is that if every world is contingent and the laws supervene on the distribution of facts in the above-mentioned sense, there is nothing that prevents two worlds from being near perfect duplicates whose only difference is in respect of the entities *C* and *E* of any arbitrary chosen counterfactual 'if it hadn't been for *C* then *E* would never have happened', *C* is always followed by *E* in one world while in the other *E* occurs in the absence of *C*. The only worlds that could be closer than that are those that are perfect duplicates. To base counterfactual dependence on perfect duplicates would make counterfactual dependence so trivial as to mean absolutely nothing for our causal deliberations.

   If worlds are contingent in the manner spelled out above, there can be a world where I invariably sneeze when my nose tickles (and only then), and then a near-perfect duplicate world where my nose doesn't tickle before I sneeze. Remember that we are dealing with perfectly contingent worlds where the presence/absence of a fundamental fact implies nothing for anything else. Such regularities are brute regularities, as the neo-Humean supposes every regularity to be. As a consequence, the counterfactual conditional 'you would not have sneezed if your nose hadn't tickled' is true and false, respectively, in worlds that differ in nothing except for the fact that in one a tickling nose is followed by a sneeze, but not in the other.

Many readers will not be convinced about the intelligibility of the example above, because they can't help thinking that a sneeze in the absence of a tickle must imply a lot of other differences, say about the neurology of my nose. But that kind of thinking only reveals either an implicit rejection of neo-Humeanism (i.e. implies that one fact does imply something about other facts), or a failure to understand how extreme neo-Humeanism really is. The bottom line is that if everything is contingent, in the manner described above, the presence or absence of a particular quality makes no difference to any other quality.

Perhaps it is better to formulate the point in a more abstract manner, to avoid letting our intuitive causal understanding of the world get in the way of understanding what neo-Humean metaphysics entails. Assume that every world is constituted by a certain distribution of fundamental categorical properties across spacetime, and that the instantiation of one property entails nothing about instantiation of other properties. This means that for any world, w_1, you can have a set of near duplicate worlds [w_2, w_3, w_4... w_n,] each of which is different from w_1 by the absence of one single type of fact C corresponding to a particular consequent of an arbitrary counterfactual conditional accepted as true in w_1. We can formulate that on the level of the individual instances of the categorical properties that are supposed to constitute that world. Can there be a smaller difference between worlds? In fact, we don't even have to assume that C is always absent when E occurs. It is enough that C is absent on one occasion for it to be false in w_2 that 'E only if C', and similarly for it to be false that it takes less of a departure from the actual world to make the antecedent true along with the consequent than to make the antecedent true without the consequent. In a contingent reality, no other departure from the actual world is needed for the antecedent to be true *without* the consequent. If you don't accept this conclusion and believe that the absence of one thing implies the absence or presence of something else, you clearly do not believe that the world is completely contingent; you deny that the instantiation of a property entails nothing about the instantiation of other properties.

Interestingly, in a discussion of the counterfactual 'if kangaroos had no tail they would topple over', Lewis appears to address the concern I have raised and deny the consequence:

> We might think it best to confine our attention to worlds where kangaroos have no tails and everything else is as it actually is; but there are no such worlds. Are we to suppose that kangaroos have no tails but that their tracks in the sand are as they actually are? Then we shall have to suppose that these tracks are produced in a way quite different from the actual way. [...] Are we to suppose that kangaroos have no tails but that their genetic makeup is as it actually is? Then

we shall have to suppose that genes control growth in a way quite different from the actual way (or else that there is something, unlike anything there actually is, that removes the tails).

<div style="text-align: right">(Lewis 1973b: 9)</div>

Ty my mind it appears as if Lewis is denying his neo-Humean convictions; things are *intimately* connected with each other. Consider that he does not merely deny that a world where kangaroos don't have tails and still don't topple over would count as a closest possible world; he *denies there could be any such world at all*. Clearly, there are constraints on how things combine to form worlds, in which case neo-Humean metaphysics is false. My problem with neo-Humean metaphysics, in the form I have presented it, is that it excludes nothing and allows everything; except possibly violations of the law of contradiction and Leibniz' Law, but these are violations that only are relevant within a world and so would be irrelevant to CTCs. That Lewis is straight in one world while his counterpart is bent in another is neither a contradiction nor violation of Leibniz' Law; it is only a sign of the qualitative distance between those worlds.

Now, several anonymous referees have objected to my argument in this section. Two objections stand out. The first is that Lewis didn't exclude the reality of causal powers, and hence didn't exclude the reality of substantial connections; he just didn't think there were any powers in our world. I think the rationale behind this objection is that one shouldn't think of neo-Humeanism as the view that all worlds are neo-Humean; that it is enough that some worlds are neo-Humean for it to be correct to say that reality is contingent. So even if there are worlds that are governed by laws, then those laws are contingent because worlds are not necessarily governed by laws, and therefore reality (i.e. the totality of possible worlds) is contingent.

To my mind, this objection doesn't do CTCs any favours. It basically turns possible worlds ontology into the philosophical analogue of Douglas Adams' plentiful universe depicted in *The Hitchhiker's Guide to the Galaxy*. Adams' universe is so plentiful that there is no point of producing any commodities because somewhere there is a planet where they grow naturally, even earmuffs, mattresses, and space-rockets. Possible worlds ontology is so plentiful that even causal powers are real, somewhere. Indeed, if it admits that causal powers are real in some world, it surely must also admit that there are worlds where laws do not supervene on the facts but instead necessarily connect the facts in determinate ways (*qua* Maudlin 2004). We could then have two worlds that are exactly alike in the distribution of matters of fact, differing only in the fact that in one world the laws supervene on the facts and in the other the laws necessitate the distribution of the facts; we couldn't tell the difference. That would be manifestly absurd. Possible worlds ontology appears to be

plentiful enough to allow every philosophical theory to be true in some world or other, and that is why I cannot consider it an ontologically serious theory. It admits everything and forbids nothing.

Furthermore, if there are worlds where there are causal powers and/or lawful connections, then they are worlds in which CTCs is false, because in that world 'E only if C' is true because C produces E and is the only thing to do so, regardless of any similarities to any closest possible world. Which shows that those who endorse substantial connections have no reason to adopt CTCs.

The other objection is that I am wrong to conclude that if laws supervene upon local matters of fact then laws cannot affect independently the overall similarity and difference between worlds. I'll address that objection below.

9.7   Comparative Similarity and Supervenience

It has been objected that my comparison between worlds with respect to a single matter of fact is out of place, because comparisons between worlds are rarely so simple. Lewis emphasises that the 'respects of similarity and difference that enter into the overall similarity of worlds are many and varied' and in particular that 'similarities in matters of particular fact trade off against similarities of law' (Lewis 1973*a*: 560). I am struggling to understand the relevance of this objection, just as much as I am struggling to see how there can be trade-offs of the kind Lewis mentions, given the basic ontology he is committed to. Let's first focus on the relevance. Whether or not comparisons between worlds can be, or even often are, 'many and varied' does not really undermine my claim that there is nothing in the neo-Humean ontology that prevents two worlds from being near-exact duplicates that differ only with respect to the minimal requirements to make a counterfactual true in one world and false in the other. If there is nothing in the ontology that shows this to be *impossible*, then the objection holds good.

I detect two ways in which it is argued that laws contribute to overall similarity independently of the distribution of facts. The first relates to an early definition of 'nomic dependence' offered by Lewis, in which he claims that the set F of true propositions of particular fact does not imply on its own every material conditional, but that we need in addition a set of law-propositions L:

> The family $C_1, C_2, \ldots$ of propositions depends nomically on the family $A_1, A_2, \ldots$ iff there are a nonempty set L of true law-propositions and a set F of true propositions of particular fact such that L and F jointly imply (but F alone does not imply) all the material conditionals $A_1 \supset$

C_1, $A_2 \supset C_2$, ... between the corresponding propositions in the two families.

The suggestion here is that *F* and *L* contribute independently of each other to the set of material conditionals true in a given world. I find that this conflicts with Lewis' understanding of the laws as supervening on the distribution of particular fact, which is the standard sense often paraphrased as 'there cannot be an *A*-difference without a *B*-difference' (McLaughlin & Bennett 2018). Indeed, Lewis says quite explicitly that: 'we have an arrangement of qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that' (Lewis 1986*a*: ix). This makes laws epiphenomenal consequences of the distribution of qualities, and thus we should be able to infer all the material conditionals from *F* alone, i.e. from the set of true propositions of fact. There is a one-to-one correspondence between the local matters of fact—on which the truth of the law-propositions in *L* supervene—and the true propositions in *F*, in which case *F* alone surely implies all the material conditionals *and* the law-propositions in *L*. After all, the regularities in a world are not there because of the laws in that world, the laws are there because the regularities are there; that is what is meant by the laws supervening on the facts. Basically, when all worlds reduce to a mosaic of local matters of fact there really is only one respect of similarity and difference: distribution of local matters of fact.

The second way in which it is argued that even if laws supervene on the facts they can still contribute individually to similarity between worlds is by analogy to other kinds of supervening phenomena. So, for instance, that even if the moral supervenes upon the physical, there could be a big overall difference between worlds that were physically very alike, since a tiny physical difference can make a big moral difference. The argument here is that a tiny difference in base properties *can* make a big difference in higher-order properties. Two responses come to the fore. First, I am finding it difficult to construe an example where a tiny physical difference gives a big moral difference, except those that translate into examples of tiny physical differences leading to big physical differences on which big moral differences supervene. A tiny nudge in the ice-hockey rink may not amount to much morally, but on the edge of a cliff it may have serious moral implications. However, there is a moral difference only if the nudge leads in one case but not the other to a physical fall and physical death or injury. The big supervening moral difference then reduces to a big physical difference. Second, for CTCs to work it is not enough to show that there *can* be tiny base differences that give big higher-order differences. They must show that *any* tiny difference in the distribution of

base qualities that contradicts *any* given counterfactual *must* make such a big difference with respect to overall similarity that there will always be a comparatively closer world where the tiny differences in base qualities do not contradict the counterfactual. It is the latter claim that I am challenging.

Although I have nowhere come across anyone arguing this, one could consider whether the distinction made between strong and weak supervenience could support the case of CTCs. However, as McLaughlin & Bennett point out (2018: §4.3.4) when the domain of base properties is restricted to intrinsic properties—as is the case with neo-Humean worlds—then weak and strong supervenience coincides. And yes, there is talk of supervenience relations of varying modal force, but the variation is between nomological and logical necessity (McLaughlin & Bennett 2018: §3.1). None of these distinctions will support the idea that in a neo-Humean reality laws can vary somehow independently of the facts.

Is there still a sense of 'supervene' that could be used to make sense of overall comparative similarity and consequently motivate the intelligibility of CTCs? I am happy to leave that open. My aim is only to show that on closer inspection there is nothing particularly clear about the notion of counterfactual dependence to make CTCs look more appealing or promising than causal accounts of counterfactuals (or to simply give up counterfactuals), and to show that the scope of ontologies for which CTCs are apt, or relevant, is considerably smaller than popularly believed.

It bears mentioning that I am not the first to question the intelligibility of the concept of *comparative overall similarity*. Michael Morreau argues that 'there can be no combining of the various similarities and differences of things into useful comparisons of overall similarity' (Morreau 2010: 469). However, while I question that there can be but one kind of similarity/difference, Morreau argues that even if worlds could differ in many different respects, there is no useful way to combine these different respects into one measure of overall comparative similarity.

9.8   Conclusion

Above I have tried to undermine the appeal and promise of CTCs, initially by appeal to relatively simple considerations that I think have been overlooked. For instance, that nobody accepts counterfactual claims about phenomena whose causal connection they don't think they understand, and that this understanding agrees with a longstanding philosophical tradition according to which counterfactual dependence is grounded in necessary causal connections. These considerations do not prove that CTCs are completely untenable but should show that their intuitive appeal has been exaggerated and that they have yet to address fairly basic objections.

The thrust of my argument is that, contrary to popular belief, CTCs are not equally appealing to everyone regardless of their ontological commitments. In particular, I argue that an ontology of possible worlds is irrelevant for everyone who is committed to substantial connections of some form. Indeed, philosophers who deny both neo-Humean contingency and necessary connections do not thereby automatically turn to CTCs but instead simply accept that counterfactuals are strictly speaking false. Finally, I have argued that CTCs should not even be appealing to the proponents of neo-Humean metaphysics. Or, to be more precise, I have argued that if reality is contingent in the way neo-Humeans claim it is, there is nothing to constrain reality in the way required to reinstate counterfactual dependency in terms of comparative similarity. Neo-Humeans, in giving up substantial connections, should give up counterfactual dependence.

10 The Contrast to Alternative Views

We have now arrived at the end of rather long and winding argumentation in which I have at once tried to draw up a contrast between causal realist, objectivist, and neo-Humean approaches—and between the range of accounts within each approach—and introduce into the mix a novel causal realist account; the view that causation is at rock bottom a question of reciprocal interaction between powerful particulars. In order to help the reader sort out the resulting chaos, I will here briefly summarise the contrasts I have drawn, as well as point out some of the more important writers from the 20th century that I have neglected. I start with the causal realist vs. neo-Humean contrast. Please note that I do not offer any repetition here of the contrast between different views on substance, process, constitution, and persistence, because the discussion on those subjects is pretty much confined to the particular chapters in which they are discussed.

10.1 The Neo-Humean Contrast

I have argued that the various accounts within the dominant neo-Humean framework—the regularity, probabilistic, conditional, counterfactual, and interventionist accounts—can at best be considered characterisations of how causation appears to us in experience/measurements. They cannot explain what it is about the world that makes it appear that way. I have also pointed out that there is a difference between endorsing neo-Humean accounts *with* or *without* the commitment to Humean supervenience and the principle of recombination, i.e. to neo-Humean metaphysics; the view that objective mind-independent reality is a mosaic of local matters of fact between which there are no substantial connections. First, one can endorse some or other (or all) neo-Humean account as an adequate account of how causation appears to us, and/or of how we think about causation, without making a judgement on whether they really explain what causation is all about. That is an approach to neo-Humean accounts that is irrelevant to metaphysics, but I think Chapter 9 provides some reasons to doubt that this is how we actually think.

Second, one can treat neo-Humean accounts as the best we can get because of some or other epistemic constraint, e.g. that we are unable to form concepts that go beyond the content provided by the senses. But such a conclusion, it seems to me, relies on a radical scepticism that will also entail not only that metaphysics is a non-starter—including neo-Humean metaphysics—but also that empirical science is confined to merely characterising reality. I have said nothing to decisively exclude that position, but I have pointed out the commitments that come with that view, and I think they amount to a pretty decent objection to it, at least to everyone who thinks that metaphysics is an intelligible enterprise.

Third, one can treat neo-Humean accounts as the best accounts we have for what objective mind-independent reality is like, i.e. that reality is a mosaic of local matters of facts between which there are no substantial connections. For the neo-Humean that explicitly endorses neo-Humean metaphysics, i.e. doesn't just suspend judgement but positively affirms there are no substantial connections, the situation is different. First of all, I am not sure I can identify any motivation for taking that view at all, once it has been pointed out that Hume's philosophy only supports epistemic/sceptical neo-Humeanism and renders neo-Humean metaphysics just as much a non-starter as other realisms. Indeed, like Hájek (2014), I find that proponents of the realist neo-Humean position are striving in directions in which I see no reason for them to strive. If everything is loose and separate, as they claim it is, why not just accept that probability is all we can get? But let's be clear. All we can get is *brute* probability. After all, this is a question of adopting probabilistic accounts because of a commitment to the idea that there are no substantial causal connections. To adopt this view will have repercussion on the prevailing view in the sciences that correlation does not entail causation. Or, rather, it entails that it is correct that correlation does not entail causation, but for very different reasons than is currently the case. The current reading is epistemic—we can't know about causation on the basis of correlation—but on a neo-Humean view it will have to give way to an ontological reading; correlation is all there is. In reality, smoking doesn't cause cancer, it is only correlated to it.

10.2   The Causal Objectivist Contrast

What I have called causal objectivist accounts are those accounts whose proponents believe there are substantial causal connections but accept what I have described as the realist empiricist approach to making sense of such connections. It basically means they think we should seek to explain the nature of causal connections but in terms of what the sciences take to be observable. Here we have, to mention a few, Salmon's account of propagation and production, Kitcher, Dowe, and Kistler's account of transmission of conserved quantities, and Glennan's mechanistic account.

I consider these accounts to be very interesting, and indeed, I build my own account on the works of these authors, but cannot deny that, ultimately, I find them wanting. The problem is that the accounts they propose only work for the same paradigmatic examples that proponents of the standard view focus on, i.e. where there is an 'active' object that exerts an influence on a 'passive' recipient of the action, and which consequently 'suffers' a change. None of these accounts generalise to all the cases, in particular the ones where the interaction not just actually is thoroughly reciprocal, but also appears to be. This is a very serious objection in the face of the fact that the natural sciences assume that every interaction is thoroughly reciprocal.

10.3   The Contrast to the Standard View

According to the standard view, causation is the action of an Agent upon a Patient, which brings about a change in the Patient. Like Bunge, I think the standard view is the best approximation to the way we ordinarily think about causation and have done for millennia. We think causation involves an exertion of influence, and we think that objects in motion act on objects at rest, causing them to change. It is Suzy's throw that breaks the bottle, the horse that pulls the carriage, the falling leaden ball that makes a hollow in the pillow, etc. What I have pointed out, picking up on Bunge's suggestion, is that this very intuitive view contradicts what is today accepted as fact in the natural sciences, that interactions are not unidirectional but reciprocal. No object acts on another object without the second acting in like measure on the first.

Now, even though the standard view falsely depicts interactions as involving unidirectional action, it does admit that the outcome of an interaction is determined by all the interacting objects alike; it agrees that effects are what powers-based accounts dub 'mutual manifestations'. For this reason, the view I propose is the result of simply accepting the reciprocity of interactions and consequently abandoning the Agent vs. Patient distinction, wherefore we can no longer talk of the contribution of each as ontologically different types of causes. Indeed, the suggestion is that we consider interactions as a single unitary phenomenon from which neither 'action' nor 'reaction' can be separated.

Notable modern advocates of the standard view are Dorothy Emmet (1985) and Ingvar Johansson (1989/2004). Of the two, Emmet comes closest to holding a traditional standard view and the arguments that I develop against assumption of unidirectional action and distinction between Agent and Patient apply to her position. Johansson, while accepting the distinction between power and quality, and active and passive powers, develops an interesting case for analysing efficient causality in terms of what he calls 'action by mixture' (1989/2004: 190–1), which comes very close to my analysis of causation in terms of interaction. He

describes efficient causation as exertion of influence between entities that in a sense become one during their interaction, and that cause and effect occur simultaneously, while still the cause produces the effect. Pretty much the only difference is that I explicitly spell out the sense in which interactions are reciprocal to the degree that the Agent vs. Patient distinction is shown not to apply, whereas he doesn't.

10.4   Contrast to Powers-Based Accounts I: Relational Realism

I have argued that contemporary powers-based accounts can be understood as attempts to revive the standard view, and in so far as they assume the Agent vs. Patient distinction, they are subject to the problems associated with that distinction. That provides one of very few contrasts between myself and Martin. However, many powers-based accounts are also implicitly influenced by the neo-Humean framework. This comes out most clearly in the acceptance/use of the dispositional analysis of powers, and in the assumption that no harm will come from the continued characterisation of causation as a two-place relation, albeit between the exercise/instantiation of a power and the realisation of a manifestation. This is what I have called relational realism. Relational realism, I have argued, is the result of a conceptual shift that occurred when empiricism rose to power and strove to eliminate everything judged to be unobservable. The result was that the concept of influence was reduced to mere uniform regularity, even by those who still believed that there are substantial causal connections that involve the exercise of influences. I have argued that while the effort to reintroduce a concept of causal influence and/or production is admirable, it just won't work if one sticks to relational realism. In fact, although I have mostly complained about Russell's problem of action at a temporal distance—arguing it is a straw man argument—it is still a very effective argument against relational realism. Russell's argument still misses the point, because relational realism isn't usually promoted in defence of causal necessity.

The main problem, however, is that relational realism really merges influence and production, or, rather, it merges the relation between the entities that exercise influence on each other, with the relation between what produces and what is produced. They cannot really be treated as the same thing. Nothing can exert an influence on something and at the same time produce that same something. Influence must be exercised between coexistent entities (neither produces the other), and it is the change that this influence brings about that is the effect (mutual exercise of influence produces a change). Writers that I take to at least implicitly endorse relational realism, and to which my criticism applies, are for instance, Bigelow, Ellis, & Pargetter (1988), Esfeld & Sachse (2011), Bird (2007a), and Tugby (2012), just to name a few.

10.5   Contrast to Powers-Based Accounts II: Process Realism

Writers like Mumford & Anjum (2011) and Chakravartty (2005) reject relational realism, and argue in favour of treating causation as a continuous process of production. Indeed, they favour simultaneous causation. However, they don't really provide an in-depth analysis of what this process involves. Mumford & Anjum's vector model can at best be thought of as a characterisation of the consequences of the process, but not as an explanation of why those consequences obtain.

It seems to me that Chakravartty more or less accepts the Agent vs. Patient distinction, but it is an interesting question whether Mumford & Anjum endorse or reject it. As far as I know, they don't address any problems with it, and sometimes they use examples that imply unidirectional influence, such as a stove heating a room, depicting this roughly in terms of the transmission accounts. However, sometimes they use examples that are more reciprocal. For instance, they don't describe the dissolution of sugar in water in terms of water acting unidirectionally on the sugar. They describe it in terms of water and sugar both being parts of the cause that produces a sweet solution (2011: 121–5). It is difficult to decidedly tell whether this is endorsement of reciprocity in my intended sense, or more in the way Marmodoro (2017) describes interactions as reciprocal and yet asymmetrical, i.e. that while water dissolves salt, salt also influences the water by salting it. The fact that Mumford & Anjum's works are ambiguous on this point at least shows that they are not completely aware of the problems I have highlighted in this book. But it is clear that, of all existing accounts, I agree with theirs the most.

The main contrasts to Mumford & Anjum's view is that, while I (attempt to) provide an explanation of the process of production, they merely characterise it in a way that is largely in agreement with the implications of that explanation, except that they deny that causation involves necessity, even though it is a substantial connection. These two differences are intimately linked, because it seems to me that their argument against causal necessity will not work if the process of production is modelled in terms of interaction. I hope to have introduced new information that may at least persuade them to think that they may have to modify some detail of their account. Alternatively, that it helps them identify where I might have gone wrong.

10.6   Contrast to Powers-Based Accounts III: Structural Accounts

There are various types of structural powers-based accounts, of which it is enough to mention two in order to draw up a contrast that arguably generalises to them all. On the one hand we have Alexander Bird's

causal structuralism (2007*a*), in which powers are intrinsic properties whose nature nevertheless is nothing but a relation to a manifestation property. He thinks they can be both intrinsic and relational because the standard view of 'intrinsic' is of something whose nature is independent of the nature of other substances. Bird's powers are properties whose nature is exhausted by a relation to a manifestation property, even when the object is not exerting that power, and so is not dependent on standing in a relation to an instantiated property in any other object. They do however depend on there being an abstract realm of possibilia, that form an interconnected structure. Now, I have argued that this doesn't really make powers intrinsic, because possibilia are still external to the power, and so there remain doubts about the reconciliation of 'intrinsic' and 'relational' (Ingthorsson 2015). But the main point of departure is that Bird does accept a distinction between active and passive possibilia and doesn't characterise interactions as reciprocal. Hence his view is liable to the criticism directed at the standard view.

Neil Williams has for some time been developing what he calls *blueprint holism* (2019), which in many ways reminds me of Bird's view. The idea is that powers are intrinsic properties which contain in themselves what he describes as a 'blueprint' of the specific types of interactions the bearer can engage in. It differs from Bird's view in that Williams insists that the potentiality inheres in the object and does not require the existence of an abstract structure of possibilia. The criticism I direct against Bird, that it cannot reconcile the intrinsic/relational incompatibility, does not obviously apply here. The crucial question will therefore be whether Williams provides an account of how powers produce effects.

It is interesting to note that Williams explicitly declares that those who talk of active and passive powers, and Agents and Patients, are the 'worst offenders' with respect to falsely singling out individual powers as causes unto their own, or at least as particularly important in bringing about an effect (2019: 125). Instead he argues in favour of a total cause conception and insists that the relationship between powers is reciprocal. He talks about 'constellations of power', which when they obtain bring about an effect. Furthermore, he stresses that such constellations are not simply a collection of powers, but—like I argued against Mumford & Anjum's version of the interference and prevention argument (Section 5.5)—that it really matters how the powers are 'arranged' in the constellation.

Now, I can complain that, while Williams rejects the Agent vs. Patient, and active vs. passive distinction, and endorses reciprocity, the examples he uses are exactly the same as those used by those who endorse those distinctions, such as water dissolving salt, and that the reciprocity involved is merely that of equal importance. So, he clearly imagines that the blueprint present in the solubility of salt contains the interaction with the dissolving power of water. He does not take it a step further to argue

that, under the microscope of science, the dissolution of salt in water is in the end a matter of molecules mutually influencing each other with respect to the same power and not two different powers.

What is more interesting though is that Williams explicitly means for his model of constellations to be ontological rock bottom: 'Instances of various constellations obtain, and they give rise to their manifestations. That is the full picture; it is not a metaphor designed to give way to some deeper analysis' (Williams 2019: 129). Although he doesn't refer to 'Causal Production as Interaction' (Ingthorsson 2002), it is clear that he thinks that analyses of the kind I present there, and defend again in this book, are bound to fail. His argument is that accounts that invoke constellations of objects that exert influence on each other, say energies or forces, only mean that the powers aren't powers at all. I take him to mean that if manifestations are the results of forces and not, say, of the momenta of the objects involved, then momentum is not a power.

I have three worries. The first is that Williams' account ends up being only a characterisation. It says—adapting the example I used in Section 5.5—that when constellation $A[p_1, p_2] + B[p_3, p_4]$ obtains, it brings about E, but it doesn't explain why that constellation brings it about (or why any constellation brings anything about). It is really just a brute fact that p_1, p_2, p_3, and p_4 contain the blueprint that make that particular constellation bring E about, but there is no further story to be told about how they actually do it.

Second, I don't see how his account relates to the theories and findings of the natural sciences, which do talk about forces and energies in an attempt to explain why certain constellations bring something about. Finally, I don't see how talk of forces implies that the properties of objects are not powers. Surely the idea is that momentum is a property that objects have, and which make the object able to exert a force. The force isn't a power or distinct property that somehow arises to exert a force. It *is* the influence that the object exerts in an interaction in virtue of having momentum. Otherwise we have an infinite regress of the kind Psillos thinks the conception of pure powers leads to (Psillos 2006), namely that powers need powers, that need powers, that need powers … etc. *ad infinitum*. For a critique of Psillos' position, see Marmodoro (2010).

10.7   Contrast to 20th-Century Friends of Powers

What this book doesn't do, and which a complete discussion of the philosophy of powers should do, is to discuss the input of all those 20th-century thinkers that arguably kept powers ontology alive and which had a decisive influence on those who initiated the 21st-century realist revival. I am talking about philosophers like Roy Bhaskar (1975), Rom Harré

& E.H. Madden (1975), Sydney Shoemaker (1980), Dorothy Emmet (1985), Nancy Cartwright (1989), and Ingvar Johansson (1989/2004). I have mentioned some of them, and discussed some of their views to a certain degree, but not really given their views much space. This is the result of a conscious choice. An effort to simplify the discussion to some little degree, by focusing on 21st-century thinkers. However, it is appropriate to acknowledge their importance.

What I can add, by way of indicating some similarities and contrast, is that Shoemaker, and Cartwright tend to think of the role of powers in event-causal terms, i.e. in terms of Kim-style events constituted by powers, and their view of powers is pretty close to the standard view. In Bhaskar's views I find more similarity than contrast to my own views, also with respect to his naturalism and how he relates to science, which is also a characteristic of Cartwright. Harré & Madden, Emmet and Johansson formulate themselves more in substance-causal terms, yet, like me, stress the processual nature of causation, and Johansson comes pretty close to the interaction view I develop here. That is no fluke. Johansson is probably the person whose works and 25 years of personal/professional communication have influenced me the most.

10.8   Concluding Remark

What I have so far said about alternative powers-based accounts really amounts to the following. Notably, that while proponents of powers-based accounts usually aim to explain causation rather than merely characterising it, I find that most of them don't in the end fully reach that aim. But they all represent a step in the right direction. I believe my account has the advantage of providing a general model that very plausibly comes across as a generalisation from the theories and findings of the natural sciences, and therefore is a philosophical account that aligns well with the explanations the sciences already offer for a range of different phenomena. To show it to be wrong will involve arguing that, despite my claims to the contrary, it isn't a good generalisation from the sciences. Alternatively, one will have to show that, even though my account works on the more fundamental levels on which we find particles, atoms, and chemical compounds, it doesn't apply to the world of ordinary middle-sized dry goods, because objects on that level have emergent properties/powers that allow unidirectional influence. I am ready to admit already now that I find intentional agency difficult to reconcile with a world in which all interactions are reciprocal. However, the solution to that problem will hinge on a better understanding of agency.

What I definitely think I have provided, never mind whether it is actually true or false, is a novel perspective on a very complicated issue.

I hope it will be received as just that; not as a claim to ultimate truth, but as an attempt to push the issue just a little bit further by providing a fresh, albeit provocative, perspective. I don't believe in what appears to be a fashionable approach to metaphysics, namely to be reconciliatory and cater to what everyone can agree. If everyone agrees, despite differences, we have only decided on the minimal common denominator and are ready to address what really matters: which one is the true account?

Bibliography

Anjum, R., Copeland, S., & Rocca, E. (2020), *Rethinking Causality, Complexity and Evidence for the Unique Patient: A CauseHealth Resource for Health Professionals and the Clinical Encounter*, Springer Open Access. Available online: https://link.springer.com/content/pdf/10.1007%2F978-3-030-41239-5.pdf.

Anscombe, G.E.M. (1971), 'Causality and Determination', in E. Sosa & M. Tooley (eds), *Causation*, Oxford: Oxford University Press, 88–104, 1993.

Aristotle (1957), 'Physics', in P.H. Wickstead & F.M. Cornford (trans.), *Aristotle*, vol. IV, Cambridge, MA: Harvard University Press.

Aristotle (1998) *Metaphysics*, H. Lawson-Tancred (trans.), London: Penguin.

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. (2005), 'Four Disputes About Properties', *Synthese* 144(3): 309–320.

Arvan, M. (2016), 'A Refutation of the Lewis-Stalnaker Analysis of Counterfactuals', *Metaphysica* 17(1): 109–129.

Austin, C.J. (2020), 'Organisms, Activity, and Being: On the Substance of Process Ontology', *European Journal for Philosophy of Science* 10: 13. Available online: https://doi.org/10.1007/s13194-020-0278-0.

Azzano, L. (2019), 'The Question of Realism for Powers', *Synthese* 196(3): 29–54.

Baggini, J. & Fosl, P.S. (2010), *The Philosopher's Toolkit*, Chichester: Wiley-Blackwell.

Bajc, B., Hisano, J., Kuwahara, T., & Omura, Y. (2016), 'Threshold Corrections to Dimension-Six Proton Decay Operators in Non-Minimal SUSY SU(5) GUTs', *Nuclear Physics B* 910: 1–22.

Baldwin, J.M. (ed.) (1901), *Dictionary of Philosophy and Psychology*, New York: The Macmillan Company.

Baltimore, J.A. (2020), 'Dispositionalism, Causation, and the Interaction Gap', *Erkenntnis* [online first]: https://doi.org/10.1007/s10670-019-00213-3.

Barker, S. (2011), 'Can Counterfactuals Really Be About Possible Worlds?' *Nôus* 45(3), 557–576.

Baumgartner, M. (2008), 'Regularity Theories Reassessed', *Philosophia* 36: 327–354.

Beebee, H., Hitchcock, C., & Menzies, P. (eds) (2009), *The Oxford Handbook of Causation*. Oxford: Oxford University Press.

Bennett, J. (1974), *Kant's Dialectic*, Cambridge: Cambridge University Press.

Bennett, K. (2017), *Making Things Up*, Oxford: Oxford University Press.

Bergson, H. (1910), *Time and Free Will: An Essay on the Immediate Data of Consciousness*, F.L. Pogson (trans.), Montana: Kessinger Publishing Company.

Beringer, J. et al. (2012), Particle Data Group, PR D86: 010001. Available online: http://pdg.lbl.gov.

Berkeley, G. (1709), *A Treatise Concerning the Principles of Human Knowledge*, T.J. McCormack (ed.), Dover: Dover Publications, 1940.

Bernstein, S. (2013), 'Review of Sophie Gibb, E. J. Lowe, & R. D. Ingthorsson (eds.), Mental Causation and Ontology', *Notre Dame Philosophical Reviews* 1: 1. Available online: https://ndpr.nd.edu/news/mental-causation-and-ontology.

Bhaskar, R. (1975), *A Realist Theory of Science*, Leeds: Leeds Books.

Bigelow, J., Ellis, B., & Pargetter, R. (1988), 'Forces', *Philosophy of Science* 55(4): 614–630.

Bird, A. (2007a), *Nature's Metaphysics*, Oxford: Clarendon Press.

Bird, A. (2007b), 'The Regress of Pure Powers?' *Philosophical Quarterly* 57(229): 513–534.

Black, R. (2000). 'Against Quidditism', *Australasian Journal of Philosophy* 78: 87–104.

Braibant, S., Giacomelli, G., & Spurio, M. (2012), *Particles and Fundamental Interactions: An Introduction to Particle Physics*, Dordrecht: Springer.

Broad, C.D. (1968), *Induction, Probability, and Causation*, Dordrecht: Springer.

Broadbent, A. (2007), 'Reversing the Counterfactual Analysis of Causation', *International Journal of Philosophical Studies* 15(2): 169–189.

Buckareff, A. (2017), 'A Critique of Substance Causation', *Philosophia* 45(3): 1019–1026.

Bunge, M. (1979), *Causality and Modern Science*, New York: Dover.

Byrne, C. (2018), *Aristotle's Science of Matter and Motion*, Toronto: University of Toronto Press.

Campbell, J. (1993), 'A Simple View of Colour', in J. Haldane & C. Wright (eds), *Reality, Representation, and Projection*, Oxford: Oxford University Press, 257–268.

Campbell, K. (1976), *Metaphysics: An Introduction*, Encino, CA: Dickenson.

Carnap, R. (1936), 'Testability and Meaning', *Philosophy of Science* 3(4): 419–471.

Cartwright, N.D. (1983), *How the Laws of Physics Lie*, Oxford: Oxford University Press.

Cartwright, N.D. (1989), *Nature's Capacities and Their Measurement*, Oxford: Oxford University Press.

Cartwright, N.D. & Pemberton, J. (2013), 'Aristotelian Powers: Without Them, What Would Modern Science Do?' in J. Greco & R. Groff (eds), *Powers and Capacities in Philosophy: The New Aristotelianism*, New York: Routledge, 93–112.

Chakravartty, A. (2005), 'Causal Realism: Events and Processes', *Erkenntnis* 63(1): 7–31.

Chakravartty, A. (2017), *Scientific Ontology: Integrating Naturalized Metaphysics and Voluntarist Epistemology*, Oxford: Oxford University Press.

Chisholm, R.M. (1966), 'Freedom and Action', in K. Lehrer (ed.), *Freedom and Determinism*, New York: Random House.

Choi, S. (2008), 'Dispositional Properties and Counterfactual Conditionals', *Mind* 117: 795–841.

Choi, S. (2012), 'Intrinsic Finks and Dispositional/Categorical Distinction', *Noûs* 46(2): 289–325.

Collingwood, R.G. (1940), *An Essay in Metaphysics*, Oxford: Oxford University Press.

Collins, J. (2000), 'Preemptive Preemption', *Journal of Philosophy* 97: 223–234.

Collins, J. (2007), 'Counterfactuals, Causation, and Preemption', in D. Jacquette (ed.), *Handbook of the Philosophy of Science: Philosophy of Logic*, Amsterdam: Elsevier, 1127–1143.

Collins, J., Hall, N., & Paul, L.A. (2004), 'Counterfactuals and Causation: History, Problems, and Prospects', in H. Collins & T. Paul (eds), *Causation and Counterfactuals*, Cambridge, MA: MIT Press, 1–58.

Costa, M.J. (1989), 'Hume and Causal Realism', *Australasian Journal of Philosophy* 67(2): 172–190.

Crean, A. (2010), 'Humean Humility', *Logical Analysis and History of Philosophy* 13: Special Issue, H. Beebee & M. Schrenk (eds), David Hume's Epistemology and Metaphysics, 17–37.

Dilworth, C. (1996), *The Metaphysics of Science*, Dordrecht: Klüwer Academic Publishers.

Donoghue, J.F., Golowich, E., & Holstein B.R. (2014), *Dynamics of the Standard Model* (2nd ed.), Cambridge: Cambridge University Press.

Dowe, P. (1992), 'Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory', *Philosophy of Science* 59: 195–216.

Dowe, P. (1995), 'Causality and Conserved Quantities: A Reply to Salmon', *Philosophy of Science* 62(2): 321–333.

Dowe, P. (2009), 'Causal Process Theories', in H. Beebee, C. Hitchcock & P. Menzies (eds), *The Oxford Handbook of Causation*, Oxford: Oxford University Press.

Drake, S. (1981), *Cause, Experiment and Science: Galilean Dialogue Incorporating a New English Translation of Galileo's "Bodies That Stay Atop Water, or Move in it"*, Chicago: University of Chicago Press.

Ducheyne, S. (2006), 'Galileo's Interventionist Notion of "Cause"', *Journal of the History of Ideas* 67(3): 443–464.

Ducheyne, S. & Weber, E. (2007), 'The Concept of Causation in Newton's Mechanical and Optical Work', *Logic and Logical Philosophy* 16: 265–288.

Dumsday, T. (2019), *Dispositionalism and the Metaphysics of Science*, Cambridge: Cambridge University Press.

Ellis, B. (2001), *Scientific Essentialism*, Cambridge: Cambridge University Press.

Ellis, B. (2010), 'Causal Powers and Categorical Properties', in A. Marmodoro (ed.), *The Metaphysics of Powers: Their Grounding and Their Manifestations*, London: Routledge, 133–142.

Ellis, B. & Lierse, C. (1994), 'Dispositional Essentialism', *Australasian Journal of Philosophy* 72: 27–45.

Emmet, D. (1985), *The Effectiveness of Causes*, Albany: SUNY Press.

Emmet, D. (1992), *The Passage of Nature*, London: Palgrave Macmillan.

Engelhard, K. (2010), 'Categories and the Ontology of Powers: A Vindication of the Identity Theory of Properties', in A. Marmodoro (ed.), *The Metaphysics*

of Powers: Their Grounding and Their Manifestations, New York: Routledge, 41–57.

Esfeld, M. (2011), 'Causal Realism', in D. Dieks, W.J. González, S. Hartman, & M. Stöltzner (eds), *Probabilities, Laws, and Structures*, Dordrecht: Springer: 157–168.

Esfeld, M. & Sachse, C. (2011), *Conservative Reductionism*, New York: Routledge.

Fales, E. (1990), *Causation and Universals*, London: Routledge.

Fine, K. (1999), 'Things and Their Parts', *Midwest Studies in Philosophy* 23: 6–74.

Foster, J. (1982), *The Case for Idealism*, London: Routledge & Kegan Paul.

Frank, P. (1932), *The Law of Causality and its Limits*, S. Cohen (ed.), Dordrecht: Kluwer Academic Print, 1998.

Gasking, D. (1955), 'Causation and Recipes', *Mind* 64: 479–487.

Gibb, S.C., Lowe, E.J., & Ingthorsson, R.D. (eds) (2013), *Mental Causation and Ontology*, Oxford: Oxford University Press.

Glennan, S. (2017), *The New Mechanical Philosophy*, Oxford: Oxford University Press.

Goodman, N. (1954), *Fact, Fiction and Forecast*, London: Athlone Press.

Graßhoff, G. & May, M. (2001), 'Causal Regularities', in W. Spohn, M. Ledwig, & M. Esfeld (eds), *Current Issues in Causation*, Paderborn: Mentis Verlag, 85–114.

Greco, J. & Groff, R. (eds) (2012), *Powers and Capacities in Philosophy: The New Aristotelianism*, New York: Routledge.

Griffiths, D.J. (1998), *Introduction to Electrodynamics* (3rd ed.), Englewood Cliffs, NJ: Prentice Hall.

Hájek, A. (2014), 'Most Counterfactuals Are False', PhilArchive copy v1. Available online: https://philarchive.org/archive/HJEMCAv1.

Hakkarainen, J. (2012), 'A Third Type Distinction in the Treatise', *Hume Studies* 38(1): 55–78.

Hakkarainen, J. (2015), 'Why Hume Cannot Be a Realist', *Journal of Scottish Philosophy* 10(2): 143–161.

Hall, N. (2000), 'Causation and the Price of Transitivity', *Journal of Philosophy* 97: 198–222.

Hall, N. (2004), 'Two Concepts of Causation', in J. Collins, N. Hall, & L.A. Paul (eds), *Causation and Counterfactuals*, Cambridge, MA: MIT Press, 225–276.

Hansson Wahlberg, T. (2017), 'Meso-level Objects, Powers, and Simultaneous Causation', *Metaphysica* 18(1): 107–125.

Hansson-Wahlberg, T. (2019), 'Active Powers and Passive Powers – Do Causal Interactions Require Both?' *Philosophia* 47: 1603–1612.

Harré, R. & Madden, E.H. (1975), *Causal Powers: A Theory of Natural Necessity*, Oxford: Blackwell.

Haslanger, S. (1989), 'Persistence, Change, and Explanation', *Philosophical Studies* 56: 1–28.

Hegel, G.W.F. & Wallace, W. (1874), *The Logic of Hegel*, Oxford: Clarendon Press.

Heil, J. (2003), *From an Ontological Point of View*, Oxford: Clarendon Press.

Heil, J. (2012), *The Universe as We Find It*, Oxford: Oxford University Press.

Hertz, H. (1956), *The Principles of Mechanics*, New York: Dover.

Hobbes, T. (1656), 'Elements of Philosophy Concerning Body', in W. Molesworth (ed.), *The English Works of Thomas Hobbes of Malmesbury*, London: Bohn, 1839.

Hoffman, J. & Rosenkrantz, G.S. (1997), *Substance: Its Nature and Existence*, London: Routledge.

Huemer, M. & Kovitz, B. (2003), 'Causation as Simultaneous and Continuous', *The Philosophical Quarterly* 53: 556–565.

Hume, D. (1739), *Treatise of Human Nature*, L.A. Selby–Bigge (ed.), Oxford: Clarendon Press, 1896.

Hume, D. (1748), *Enquiry Concerning Human Understanding*, Harvard Classics, vol. 37, Boston, MA: P.F. Collier & Son, 1910.

Illari, P. (2011), 'Why Theories of Causality Need Production: An Information-Transmission Account', *Philosophy and Technology* 24: 95–114.

Ingthorsson, R.D. (2001), 'Temporal Parity and the Problem of Change', *SATS–Northern European Journal of Philosophy* 2(2): 60–79.

Ingthorsson, R.D. (2002), 'Causal Production as Interaction', *Metaphysica* 3: 87–119.

Ingthorsson, R.D. (2007), 'Is There a Problem of Action at a Temporal Distance?' *SATS–Northern European Journal of Philosophy* 8(1): 138–154.

Ingthorsson, R.D. (2009), 'Can Things Endure in Tenseless Time?' *SATS–Northern European Journal of Philosophy* 10(1): 79–99.

Ingthorsson, R.D. (2013), 'Properties: Qualities, Powers, or Both?' *Dialectica* 67(1): 55–80.

Ingthorsson, R.D. (2015), 'The Regress of Pure Powers Revisited', *The European Journal of Philosophy* 23(3): 529–541.

Ingthorsson, R.D. (2016), *McTaggart's Paradox*, New York: Routledge.

Ingthorsson, R.D. (2019), 'Mario Bunge and the Current Revival of Causal Realism', in M.R. Matthews (ed.), *Mario Bunge: Centenary Festschrift*, Cham: Springer Verlag, 205–217.

Ingthorsson, R.D. (2020), 'Ontological and Methodological Reflections on the Virtues of Unification', *The Journal of Physics: Conference Series* 1466: 012006.

Jacobs, J.D. (2010), 'A Powers Theory of Modality: Or, How I Learned to Stop Worrying and Reject Possible Worlds', *Philosophical Studies* 151(2): 227–248.

Jacobs, J.D. (2011), 'Powerful Qualities, Not Pure Powers', *The Monist* 94(1): 81–102.

Jacobs, J.D. (ed.) (2017), *Causal Powers*, Oxford: Oxford University Press.

James, W. (1890), *The Principles of Psychology*, Cambridge, MA: Harvard University Press, 1981.

Johansson, I. (1989/2004), *Ontological Investigations: An Inquiry into the Categories of Nature, Man and Society*, Reprint Philosophy 4, Modern Classics of Analytic Philosophy, vol. 4, Frankfurt: Ontos Verlag. First published 1989, by Routledge. Available online: https://www.degruyter.com/downloadpdf/title/303559.

Johansson, I. (1997), 'The Unnoticed Regional Ontology of Mechanisms', *Axiomathes* 8: 411–428

Johansson, I. (2016), 'Against Fantology Again', in L. Zabert (ed.), *The Theory and Practice of Ontology*, London: Palgrave Macmillan, 25–43.

Johnson, W.E. (1924), *Logic, Part III: The Logical Foundations of Science*, Cambridge: Cambridge University Press.

Johnston, M. (2006), 'Hylomorphism', *Journal of Philosophy* 103: 652–698.

Kant, I. (1787), *Critique of Pure Reason*, N. Kemp Smith (trans.), New York: St. Martin's Press, 1965.

Kim, J. (1973), 'Causation, Nomic Subsumption, and the Concept of Event', *Journal of Philosophy* 70(8): 217–236.

Kim, J. (1976), 'Events as Property Exemplifications', in M. Brand & D. Walton (eds), *Action Theory*. Dordrecht: Reidel, 310–326.

Kistler, M. (1998), 'Reducing Causality to Transmission', *Erkenntnis* 48: 1–24.

Kistler, M. (2012), 'Powerful Properties and the Causal Basis of Dispositions', in A. Bird, B. Ellis, & H. Sankey (eds), *Properties, Powers and Structures: Issues in the Metaphysics of Realism*, London: Routledge, 119–137.

Kitcher, P. (1989), 'Explanatory Unification and the Causal Structure of the World', in P. Kitcher & W.C. Salmon (eds), *Scientific Explanation*, Minnesota Studies in the Philosophy of Science, vol. 13. Minneapolis: University of Minnesota Press, 410–506.

Koslicki, K. (2008), *The Structure of Objects*, Oxford: Oxford University Press.

Kosman, A. (2013), *The Activity of Being: An Essay on Aristotle's Ontology*, Cambridge, MA: Harvard University Press.

Kuykendall, D. (2019), 'Powerful Substances Because of Powerless Powers', *Journal of the American Philosophical Association* [online first]: https://doi.org/10.1017/apa.2019.18.

Lange, M. (2004), 'A Note on Scientific Essentialism, Laws of Nature, and Counterfactual Conditionals', *Australasian Journal of Philosophy* 82: 227–241.

Lange, M. (2009), *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*, Oxford: Oxford University Press.

Le Poidevin, R. (1988), 'The Principle of Reciprocity and a Proof of the Non-Simultaneity of Cause and Effect', *Ratio: An International Journal of Analytic Philosophy* 1(2): 152–162.

Levi, A.L. (1958), 'Substance, Process, Being', *The Journal of Philosophy* 55(18): 749–761.

Lewis, D. (1973a), 'Causation', *Journal of Philosophy* 70(17): 556–567.

Lewis, D. (1973b), *Counterfactuals*, Cambridge: Harvard University Press.

Lewis, D. (1986a), *The Plurality of Worlds*, Oxford: Blackwell.

Lewis, D. (1986b), *Philosophical Papers*, vol. 2, Oxford: Oxford University Press.

Lewis, D. (2000), 'Causation as Influence', *The Journal of Philosophy* 97(4): 182–197.

Lewis, D. (2007), 'Ramseyan Humility', in D. Braddon-Mitchell & R. Nola (eds), *Naturalism and Analysis*, Cambridge, MA: MIT Press: 203–222.

Livanios, V. (2017), *Science in Metaphysics: Exploring the Metaphysics of Properties and Laws*, Cham: Palgrave Macmillan.

Locke, J. (1689), *An Essay Concerning Human Understanding*, abridged and edited by K.P. Winkler, Indianapolis: Hackett, 1996.

Lombard, L.B. (1986), *Events*, London: Routledge & Kegan Paul.

Loux, M. (1998), *Metaphysics: A Contemporary Introduction*, New York: Routledge.

Lowe, E.J. (1998), *The Possibility of Metaphysics*, Oxford: Clarendon Press.

Lowe, E.J. (2006a), *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*, Oxford: Oxford University Press.

Lowe, E.J. (2006b), 'Non-Cartesian Substance Dualism and the Problem of Mental Causation', *Erkenntnis* 65: 5–23.

Lowe, E.J. (2008), *Personal Agency*, Oxford: Oxford University Press.

Lowe, E.J. (2011), 'A Neo-Aristotelian Substance Ontology: Neither relational Nor Constituent', in T. Tahko (ed.), *Contemporary Aristotelian Metaphysics*, Cambridge: Cambridge University Press.

Mach, E. (1896), *Principles of the Theory of Heat*, B.F. McGuinness (ed.), Dordrecht: Reidel, 1986.

Mackie, J. (1965), 'Causes and Conditions', in E. Sosa & M. Tooley (eds), *Causation*, Oxford: Oxford University Press, 33–55, 1993.

Manley, D. & Wasserman, R. (2007), 'A Gradable Approach to Dispositions', *The Philosophical Quarterly* 57: 68–75.

Manley, D. & Wasserman, R. (2008), 'On Linking Dispositions and Conditionals', *Mind* 117: 59–84.

Marek, J. (2009), 'Alexius Meinong', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: http://plato.stanford.edu/archives/sum2009/entries/meinong/.

Marmodoro, A. (2007), 'The Union of Cause and Effect in Aristotle: Physics 3.3', *Oxford Studies in Ancient Philosophy* 32: 205–232.Marmodoro, A. (2009), 'Do Powers Need Powers to Make them Powerful? From Pandispositionalism to Aristotle', *History of Philosophy Quarterly* 26(4): 337–352.

Marmodoro, A. (ed.) (2010), *The Metaphysics of Powers: Their Grounding and Their Manifestations*, London: Routledge.

Marmodoro, A. (2013), 'Aristotle's Hylomorphism Without Reconditioning', *Philosophical Inquiry* 37: 5–22.

Marmodoro, A. (2017), 'Aristotelian Powers at Work: Reciprocity Without Symmetry in Causation', in J. Jacobs (ed.), *Causal Powers*, Oxford: Oxford University Press, 57–76.

Martin, C.B. (1994), 'Dispositions and Conditionals', *The Philosophical Quarterly* 44: 1–8.

Martin, C.B. (1997), 'On the Need for Properties: Road to Pythagoreanism and Back', *Synthese* 112: 193–231.

Martin, C.B. (2000), 'On Lewis and Then Some', *Logic et Analyse* 43(169–70): 43–48.

Martin, C.B. (2008), *Mind in Nature*, Oxford: Oxford University Press.

Maudlin, T. (2004), 'Causation, Counterfactuals, and the Third Factor', in J. Collins, N. Hall, & L.A. Paul (eds), *Causation and Counterfactuals*, Cambridge, MA: MIT Press, 419–444.

Maudlin, T. (2007), *The Metaphysics Within Physics*, Oxford: Oxford University Press.

Maxwell, J.C. (1877), *Matter and Motion*, New York: Dover, rpt. 1953.

McGee, H. (2004), *On Food and Cooking: The Science and Lore of the Kitchen*, New York: Scribner.

McKitrick, J. (2018), *Dispositional Pluralism*, Oxford: Oxford University Press.

McLaughlin, B. & Bennett, K. (2018), 'Supervenience', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/spr2018/entries/supervenience.

McTaggart, J.M.E. (1915), 'The Meaning of Causality', *Mind* 24(959): 326–344.

McTaggart, J.M.E. (1921), *The Nature of Existence*, vol. 1, Cambridge: Cambridge University Press, 1988.

Meincke, A.S. (2019), 'The Disappearance of Change: Towards a Process Account of Persistence', *International Journal of Philosophical Studies* 27(1): 12–30.

Meixner, U. (2008), 'New Perspectives for a Dualistic Conception of Mental Causation', *Journal of Consciousness Studies* 15: 17–38.

Mellor, D.H. (1974), 'In Defence of Dispositions', *Philosophical Review* 83: 157–181.

Mellor, D.H. (1998), *Real Time II*, London: Routledge.

Menzel, C. (2017), 'Possible Worlds', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online*:* https://plato.stanford.edu/archives/win2017/entries/possible-worlds.

Menzies, P. (2017), 'Counterfactual Theories of Causation', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online*:* https://plato.stanford.edu/archives/win2017/entries/causation-counterfactual.

Menzies, P. & Price, H. (1993), 'Causation as a Secondary Quality', *British Journal for the Philosophy of Science* 44(2): 187–203.

Mill, J.S. (1843), *A System of Logic, Ratiocinative and Inductive*, J.M. Robson (ed.), Toronto: Routledge & Kegan Paul, 1974.

Millican, P. (2009), 'Hume, Causal Realism, and Causal Science', *Mind* 118(471): 647–712.

Mishra, H. (2013), 'Thermal field Theory', in P. Ramadevi (ed.), *Surveys in Theoretical High Energy Physics 1: Lecture Notes from SERC Schools*, New Delhi: Hindustan Book Agency.

Molander, B. (1982), *The Order There Is, and the Order We Make*, Philosophical Studies No. 35, Uppsala: Philosophical Society and the Department of Philosophy, University of Uppsala.

Molnar, G. (2003), *Powers: A Study in Metaphysics*, Oxford: Oxford University Press.

Morreau, M. (2010), 'It Simply Does Not Add Up: Trouble with Overall Similarity', *Journal of Philosophy* 107(9): 469–490.

Mozersky, J. (2015), *Time, Language, and Ontology*, Oxford: Oxford University Press.

Mumford, S. (1998), *Dispositions*, Oxford: Oxford University Press.

Mumford, S. (2004), *Laws in Nature*, Abingdon: Routledge.

Mumford, S. (2005), 'Laws and Lawlessness', *Synthese* 144(3): 397–413.

Mumford, S. & Anjum, R. (2011), *Getting Causes from Powers*, Oxford: Oxford University Press.

Mumford, S. & Anjum, R. (2015a), 'Carnap and the Anglo-Austrian Conspiracy Against Dispositions', in A.S. Spann (ed.), *The Ontological Commitments of Dispositionalism*, Berlin: De Gruyter.

Mumford, S. & Anjum, R. (2015b), 'Mutual Manifestation and Martin's Two Triangles', in J. Jacobs (ed.), *Causal Powers*, Oxford: Oxford University Press, 74–87.

Mumford, S. & Anjum, R. (2018), *What Tends to Be*, New York: Routledge.

Munsat, S. (1969), 'What Is a Process?' *American Philosophical Quarterly* 6(1): 79–83.

National Institute of Standards and Technology (2020) 'SI Units–Mass'. Available online [downloaded 2020-06-09]: https://www.nist.gov/pml/weights-and-measures/si-units-mass.

Newton, I. (1686), 'Definitions and Scholium', in *Newton's Philosophy of Nature*, H.S. Thayer (ed.), New York: Hafner Press, 1953.

Nicholson, D.J. & Dupré, J. (eds) (2018), *Everything Flows: Towards a Processual Philosophy of Biology*, Oxford: Oxford University Press.

O'Shaughnessy, B. (1971–72), 'Processes', *Proceedings of the Aristotelian Society* 72: 215–240.

Øhrstrøm, P. & Schärfe, H. (2004), 'A Priorean Approach to Time Ontologies', in K.E. Wolff et al. (eds), *International Conference on Conceptual Structures 2004: Conceptual Structures at Work*, Dordrecht: Springer, 388–401.

Paul, L.A. (1998), 'Keeping Track of the Time: Emending the Counterfactual Analysis of Causation', *Analysis* 58: 191–198.

Peppers-Bates, S. (2009), *Nicolas Malebranche: Freedom in an Occasionalist World*, London: Continuum.

Plato (1971), *Timaeus and Critias*, D. Lee (trans.), Harmondsworth: Penguin Books.

Price, H. (1996), *Time's Arrow and Archimedes Point*, Oxford: Oxford University Press.

Price, H. (2001), 'Causation in the Special Sciences: The Case for Pragmatism', in M.C. Galavotti, P. Suppes, & D. Costantini (eds), *Stochastic Causality*, Stanford: CSLI Publications, 103–121.

Psillos, S. (2006), 'What Do Powers Do When They Are Not Manifested?' *Philosophy and Phenomenological Research* 72(1): 137–156.

Psillos, S. (2009). 'Regularity Theories of Causation', in H. Beebee, C. Hitchcock, & P. Menzies (eds), *The Oxford Handbook of Causation*, Oxford: Oxford University Press, 131–157.

Psillos, S. (2019), 'Review of What Tends to Be: The Philosophy of The Dispositional Modality', *Notre Dame Philosophical Reviews*. Available online: https://ntrda.me/2CVpZTy.

Ramsey, F.P. (1929), 'General Propositions and Causality', in D.H. Mellor (ed.), *Philosophical Papers*, Cambridge: Cambridge University Press, 145–146, 1990.

Razeto-Barry, P. (2012), 'Autopoiesis 40 Years Later: A Review and a Reformulation', *Origins of Life and Evolution of Biospheres* 42(6): 543–567.

Read, R. & Richman K.A. (eds) (2007), *The New Hume Debate* (Rev. ed.), London: Routledge.

Reeder, N. (1995), 'Are Physical Properties Dispositions?' *Philosophy of Science* 62(1): 141–149.

Rescher, N. (1996), *Process Metaphysics: An Introduction to Process Philosophy*, New York: SUNY Press.

Rescher, N. (2000), *Process Philosophy: A Survey of Basic Issues*, Pittsburgh: University of Pittsburgh Press.

Rescher, N. (2012), 'Process Philosophy', in E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/sum2012/entries/process-philosophy.

Resnick, R., Halliday, D., & Krane, K. (2002), *Physics* (5th ed.), New York: John Wiley & Sons.

Robinson, H. (1982), *Matter and Sense*, Cambridge: Cambridge University Press.

Robinson, H. (2020), 'Substance', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/spr2020/entries/substance.

Russell, B. (1912), 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, New Series 13: 1–26.

Ryle, G. (1949), *The Concept of Mind*, London: Penguin.

Salmon, W. (1980), 'Causality: Production and Propagation', in E. Sosa & M. Tooley (eds), *Causation*, Oxford: Oxford University Press, 154–171, 1993.

Salmon, W. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ: Princeton University Press.

Salmon, W. (1997), 'Causality and Explanation: A Reply to Two Critiques', *Philosophy of Science* 64: 461–477.

Schmaltz, T.M. (2008), *Descartes on Causation*, New York: Oxford University Press.

Schrenk, M. (2007), 'Can Capacities Rescue Us from Ceteris Paribus Laws?' in M. Kistler & B. Gnassounou (eds), *Dispositions and Causal Powers*, Aldershot: Ashgate, 221–247.

Schrenk, M. (2009), 'The Powerlessness of Necessity', *Noûs* 44(4): 725–739.

Seibt, J. (1996), 'Existence in Time: From Substance to Process', in J. Faye (ed.), *Perspectives on Time*, Boston Studies in Philosophy of Science, Dordrecht: Kluwer, 143–182.

Seibt, J. (2000), 'The Dynamic Constitution of Things', *Poznan Studies in the Philosophy of the Sciences and the Humanities* 76: 241–278.

Seibt, J. (2018), 'Process Philosophy' in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/win2018/entries/process-philosophy.

Sellars, W. (1958), 'Counterfactuals, Dispositions and the Causal Modalities', in H. Feigl, M. Scriven, & G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science*, vol. 2, Minneapolis: University of Minnesota Press, 225–308.

Sentesy, M. (2020), *Aristotle's Ontology of Change*, Chicago: Northwestern University Press.

Shapin, S. (1996), *The Scientific Revolution*, Chicago: University of Chicago Press.

Shoemaker, S. (1980), 'Causality and Properties', in P. Van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*, Dordecht: Reidel, 109–136.

Simons, P. (2000), 'Continuants and Occurrents', *The Aristotelian Society*, Supplement 74: 59–75.

Smith, B. (2005), 'Against Fantology', in M.E. Reicher & J.C. Marek (eds), *Experience and Analysis*, Vienna: öbv & hpt Verlagsgesellschaft, 153–170. Available online: http://ontology.buffalo.edu/bfo/Against_Fantology.pdf.

Sowa, J. (2006), 'Worlds, Models, and Descriptions', *Studia Logica*, Special Issue Ways of Worlds II, 84(2): 323–360.

Sparber, G. (2009), *Unorthodox Humeanism*, Frankfurt: Ontos Verlag.

Spinoza, B. (1677), *The Ethics*, R.H.M. Elwes (ed.), New York: Dover, 1955.

Steinberg, J. & Steinberg, A. (2017), 'A Multiply Qualified Conditional Analysis of Disposition Ascription: Mapping the Conceptual Topography of Ceteris Paribus', *Erkenntnis* 82: 777–793.

Steward, H. (2012), *A Metaphysics for Freedom*, Oxford: Oxford University Press.

Strawson, G. (2008), 'The Identity of the Categorical and the Dispositional', *Analysis* 68(4): 271–282.

Suppes, P. (1970), *A Probabilistic Theory of Causality*, Amsterdam: North-Holland.

Swinburne, R. (1980), 'Properties, Causation, and Projectibility: Reply to Shoemaker,' in L.J. Cohen & M. Hesse (eds), *Applications of Inductive Logic*, Oxford: Oxford University Press, 313–320.

Taylor, H. (2018), 'Powerful Qualities and Pure Powers', *Philosophical Studies* 175: 1423–1440.

Taylor, R. (1973), *Action and Purpose*, New York: Humanities Press.

Thompson, I.J. (1988), 'Real Dispositions in the Physical World', *The British Journal for the Philosophy of Science* 39(1): 67–79.

Thompson, N. (2019), 'Questions and Answers: Metaphysical Explanation and the Structure of Reality', *Journal of the American Philosophical Association* 5(1): 98–116.

Tugby, M. (2012), 'The Metaphysics of Dispositionalism', in A. Bird, B. Ellis, & H. Sankey (eds), *Properties, Powers, and Structures*, New York: Routledge, 165–179.

Uzgalis, W. (2020), 'John Locke', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/spr2020/entries/locke.

Varzi, A. (2019), 'Mereology', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/spr2019/entries/mereology.

Vetter, B. (2015), *Potentiality: From Dispositions to Modality*, Oxford: Oxford University Press.

von Wright, G.H. (1974), *Causality and Determinism*, New York: Columbia University Press.

Wallace, W.A. (1988), 'Randall Redivivus: Galileo and the Paduan Aristotelians', *Journal of the History of Ideas* 49(1): 133–149.

Wasserman, R. (2018), 'Material Constitution', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/archives/fall2018/entries/material-constitution.

Weatherson, B. (2014), 'Humean Supervenience', in B. Loewer & J. Schaffer (eds), *A Companion to David Lewis*, London: Wiley-Blackwell, 99–105.

Weidner, R.T. & Sells, R.L. (1968), *Elementary Modern Physics* (2nd ed.), Boston: Allyn & Bacon.

Whitehead, A.N. (1929), *Process and Reality: An Essay in Cosmology*, New York: Macmillan.

Whittle, A. (2016), 'A Defense of Substance Causation', *Journal of the American Philosophical Association* 2: 1–20.

Williams, N.E. (2010), 'Puzzling Powers: The Problem of Fit', in A. Marmodoro (ed.), *The Metaphysics of Powers: Their Grounding and Their Manifestaions*, New York: Routledge, 84–105.

Williams, N.E. (2019), *The Powers Metaphysic*, Oxford: Oxford University Press.

Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.

Index