Dependencies, Connections, and Other Relations. A Theory of Mental Causation

de Muijnck, W.

Publication date:
2002

Link to publication

Citation for published version (APA):

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

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

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Dependencies, connections,
and other relations

A theory of mental causation

Afhankelijkheden, connecties,
en andere relaties

Een theorie over mentale veroorzaking

Proefschrift ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof. dr. F. A. van der Duyn Schouten, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 18 oktober 2002 om 10.15 uur door

Willem de Muijnck

gleboren op 5 november 1964 te Maarn.
Promotores:

Prof. dr. A. A. Derksen
Prof. dr. ir. A. W. M. Meijers

© 2002 W. de Muijnck
To my parents

Contents

3 A layered world ....................................................... 26
  3.1. Levels of organization ......................................... 26
  3.2. Emergence ....................................................... 27
  3.3. Upward causation ............................................... 29
  3.4. Multiple realizability .......................................... 32
  3.5. Relational properties and indirect supervenience ........... 34
  3.6. Relational properties and nonlocal supervenience .......... 36
  3.7. Against type identity ......................................... 38
  3.8. Non-identity and non-distinctness .......................... 41
  3.9. Summary ................................................................ 42

Part II. Causality ....................................................... 45

4 A duality in the concept of causality ................................ 47
  4.1. Causality: the very idea ......................................... 47
  4.2. The double face of causality .................................... 50
  4.3. Dependence and physical connection ........................... 51
  4.4. Generality and particularity ...................................... 53
  4.5. Theories of causality ............................................ 54

5 Causal dependence ................................................... 56
  5.1. INUS conditions .................................................. 56
  5.2. Noncausal conditions ............................................ 58
  5.3. Nonsalient causal conditions .................................... 59
  5.4. Counting causes and effects .................................... 60
  5.5. Indeterminism ..................................................... 61
  5.6. Collateral effects ................................................ 63
  5.7. Overdetermination ............................................... 65
  5.8. Pre-emption ....................................................... 67

6 Causal connection .................................................... 69
  6.1. From counterfactual dependence to physical connection ... 69
  6.2. The mechanistic approach ....................................... 70
  6.3. Causal processes ................................................ 72
  6.4. First criticism: the at-at theory of causal influence .......... 75
  6.5. Second criticism: causation by disconnection ................. 77
  6.6. Third criticism: causal relevance ................................ 79
  6.7. Fourth criticism: the ontology of causal processes .......... 81
  6.8. What causal connections are .................................... 82
## Contents

**Preface**

**Introduction**

1. **A brief history**
2. **The five problems of mental causation**
   1. Mental anomalous causation
   2. Normativity
   3. Phenomenal states of affairs
   4. Externalism
   5. Causal exclusion
3. **A look forward**
   1. What we need
   2. What is coming

**Part I. Ontology**

1. **Particulars, properties, and relations**
   1. Particulars
   2. Properties
   3. How properties are had
   4. States of affairs
   5. Token identity of particulars
   6. Relations
   7. Relational properties
   8. Relationism
2. **Physicalism**
   1. A negative account of 'physical'
   2. Supervenience and physical realization
   3. The free lunch principle
   4. Arguments for physicalism
Contents

7 Unifying dependence and connection ................................................. 85
  7.1. Physically linking causes and effects: Kitcher ................................... 85
  7.2. Physically linking causes and effects: Mackie ................................... 85
  7.3. Physically linking causes and effects: Dowe ................................... 87
  7.4. From relations and mechanisms to dependencies and connections .88
  7.5. The pairing problem ............................................................ 90
  7.6. The pairing principle ........................................................... 94
  7.7. Negative causal factors .................................................................. 96
  7.8. Why we think causally ............................................................... 99

8 Causation and natural law ................................................................. 101
  8.1. The first problem of mental causation ............................................ 101
  8.2. The nomological approach to causation ........................................ 101
  8.3. Against the notion of 'causal law' ............................................. 103
  8.4. Laws and patterns ....................................................................... 106
  8.5. Particularism ............................................................................... 107
  8.6. Perception of causal relations ..................................................... 108
  8.7. Causal perception as implicit knowledge ....................................... 112
  8.8. Causal perception: object and content ....................................... 114

9 The problem of causal relata ............................................................. 116
  9.1. Causal ontology and mental causation ........................................... 116
  9.2. Hume, Kim, and others ............................................................... 117
  9.3. Examples of causal statements ..................................................... 118
  9.4. Excluding concrete particulars ..................................................... 119
  9.5. Excluding negative states of affairs ............................................. 120
  9.6. Excluding regularities ................................................................. 122
  9.7. Properties of states of affairs ....................................................... 122

10 Getting events wrong ......................................................................... 126
  10.1. Event mereology .......................................................................... 126
  10.2. Events according to Davidson ..................................................... 128
  10.3. First criticism: causal relations and causal explanations ............... 130
  10.4. Second criticism: causal relevance and irrelevance of properties 132
  10.5. Third criticism: the need for states of affairs .............................. 136

11 Getting events right ........................................................................... 139
  11.1. Events according to Kim .............................................................. 139
  11.2. The narrative view: events as episodes ........................................ 140
  11.3. In favour of the narrative view ..................................................... 142
  11.4. Events, ongoing processes, and states of affairs ........................ 145
# Contents

12 Relations as causal relata .......................................................... 150  
12.1. Relational causal relata ......................................................... 150  
12.2. Causal relations as causes ..................................................... 152  
12.3. Causal relations as effects ..................................................... 154  

13 Causal efficacy ........................................................................... 159  
13.1. The need for a criterion .......................................................... 159  
13.2. Locality and intrinsicness ....................................................... 160  
13.3. The criterion of local difference ............................................. 161  
13.4. In favour of the criterion of local difference ............................ 163  
13.5. Relational properties ............................................................. 165  
13.6. The causal efficacy of relational properties .............................. 167  
13.7. The causal efficacy of the past ................................................. 172  

14 Supervenient causation ............................................................... 174  
14.1. Epiphenomenalism or downward causation: a dilemma? ........... 174  
14.2. Relationality and causal autonomy ......................................... 176  
14.3. Commensurateness and the correlation principle ................... 178  
14.4. Downward causation and causal closure .................................. 179  
14.5. Downward causation as constraint ......................................... 181  
14.6. Doing versus exploiting causal work ........................................ 184  

Part III. Mind .............................................................................. 187  

15 The concept of mind ................................................................. 189  
15.1. Introduction ........................................................................... 189  
15.2. Mind presupposes teleology .................................................... 190  
15.3. Mind presupposes representation ........................................... 192  
15.4. Mind presupposes (ir-)rationality ......................................... 195  
15.5. Mind presupposes perception and action ................................. 197  
15.6. Mind and qualia ................................................................. 201  

16 Against the computational theory of mind .................................. 205  
16.1. Introduction ........................................................................... 205  
16.2. The computational theory of mind ......................................... 206  
16.3. First criticism: epiphenomenalism ......................................... 207  
16.4. Second criticism: lack of psychological realism ....................... 208  
16.5. Further criticisms .................................................................... 209  
16.6. Propositional attitudes as mental models ............................... 212
Contents

16.7. In favour of the theory of mental models ........................................ 215
16.8. Propositional attitudes according to interpretationism .................. 217

17 Against the Theory Theory of folk psychology ................................. 218
  17.1. Introduction .................................................................................. 218
  17.2. The Theory Theory ................................................................. 220
  17.3. Against the Theory Theory ......................................................... 221
  17.4. Interpretationism ....................................................................... 223
  17.5. Simulation theory ....................................................................... 224

18 Against internalism ........................................................................... 227
  18.1. Introduction .................................................................................. 227
  18.2. Semantic externalism ................................................................... 229
  18.3. Interpretationist externalism ....................................................... 231
  18.4. Social and linguistic externalism .................................................. 231
  18.5. Etiological externalism ............................................................... 232
  18.6. Environmentalism ....................................................................... 234
  18.7. Reasons externalism .................................................................... 237

19 Against reductionism ......................................................................... 242
  19.1. Introduction .................................................................................. 242
  19.2. Nagel-reductionism ..................................................................... 243
  19.3. Functional reductionism ............................................................... 244
  19.4. Eliminativism .............................................................................. 247
  19.5. Reductionism as method ............................................................... 249

20 Against token physicalism .................................................................. 251
  20.1. Introduction .................................................................................. 251
  20.2. Mental holism and uncodifiability ............................................... 252
  20.3. The problem of arbitrariness ....................................................... 253
  20.4. Weak externalism ........................................................................ 255
  20.5. Against weak externalism ............................................................. 257
  20.6. Actions ........................................................................................ 258

Conclusion: mental causation .................................................................. 263

21 The five problems of mental causation once again ......................... 264
  21.1. Mental anomalism ....................................................................... 264
  21.2. Normativity ................................................................................. 265
  21.3. Phenomenal states of affairs ...................................................... 267
Preface

The text before you is a study of the problematic issue of mental causation: causation by minds. On hearing the expression ‘mental causation,’ you may at first think of something like bending spoons by ‘psychic’ powers. But no, I discuss something much more puzzling: doing things for reasons, i.e., what we call agency. Psychic spoon-bending would be a fairly straightforward issue. You just exert some psychic force and bend a spoon, just like you might bend it by hand, i.e., by physical force. The only trouble here is that psychic forces may not be in fact available. But now you fetch an umbrella because you expect that it will rain. How does that work? Somehow, it seems, you let an expectation move your limbs. But aren’t your limbs already moved by nerve impulses and muscle contractions? And are expectations the proper kind of items to move things around?

Mental causation is an issue that is at the heart of the mind-body problem, the problem of making it clear how minded creatures such as we are possible, and what our mindedness consists in. Unlike psychic spoon-bending, mental causation happens every day. At least, pretty much of what we take for granted about ourselves can only be right when mental causation really happens. If we came to deny the reality of mental causation, we would invalidate core notions about our existence as perceiving, thinking and acting subjects. The concept of knowledge, for instance, and related notions such as perception, rationality, science, thought and communication, seem to rely for their very intelligibility on the reality of mental causation. The same holds true of the concept of agency, with its related notions of will, freedom, autonomy, and responsibility.

This is why entire philosophical disciplines, such as epistemology, ethics, and philosophy of science, presuppose the reality of mental causation. And mental causation does not just concern philosophy: psychology, history, and a range of social sciences also presuppose mental causation. Neither is the importance of mental causation confined to the studies of academicians: any form of jurisdiction, with its use of notions like intent, responsibility, and accountability, would be impossible if the reality of mental causation were not presupposed. When advertising agencies, political parties, religious organizations and schools spend their efforts, as well as considerable sums of money, on changing our behaviour by changing our minds, they also manifest a great belief in mental causation. And so do all of us,
any time we consider it worthwhile to speak to each other, or to deliberate before we act.

Bad news, then, and a great embarassment, to learn about this important phenomenon that in contemporary analytic philosophy of mind we do not seem able to explain what it consists in, even to explain how it is possible at all. In the past three or so decades, when artificial intelligence and neuroscience became flourishing fields of research, it seemed that such explanation was within reach. Instead, however, mental causation has remained as puzzling as ever; no matter how technical and esoteric the debate on this issue has become. Given this situation, it might be helpful if we questioned some of the assumptions that have so far largely shaped that debate, and that have delimited its range of outcomes. This will draw us into daunting issues in the fields of ontology, philosophy of mind, and especially the metaphysics of causation; but the reward of such a new look at the issue may be the conviction that mental causation is real and that it can be understood.

I would like to thank the following people for helping me write this dissertation.

First and foremost, Ton Derksen and Anthonie Meijers. Anthonie Meijers, who supervised the first years of my research, has been of pivotal importance in shaping the result. Without his heavy-duty coaching and relentless criticism, I might not have gained the momentum to take several philosophical steep slopes. Ton Derksen discussed and criticized my work in the later stages. By granting me considerable elbow room in my philosophical explorations, and by leading a lively and inspiring research group in which I felt very much at home, philosophically and personally, he has made the past years of research a stimulating and enriching experience for me. The research group included Chris Buskes, Hans van Dooremalen, Monica Meijsing, Herman de Regt, Marc Slors (who read and criticized a late draft of this dissertation), Pouwel Slurink, and Arno Wouters. Thank you all for the good company and the good discussions.

Looking back, many thanks also to Jan Bransen, who coached my philosophy study in Utrecht, urged me to keep up with philosophy, and afterwards drew my attention to a very specific advertisement that I might easily have overlooked. Also thanks to Hans Koopman, who taught philosophy of art in Amersfoort/Utrecht when I studied art there. He has played an early key role in the whole process by introducing me to analytic thought.
Of all those who have in discussion helped me to shape my ideas, I would further like to thank, among many others: Filip Buekens, Terence Cuneo, Stefaan Cuypers, Carl Gillett, Irving Krakow, Rick Looijen, Anne Ruth Mackor, Fred Muller, Henk de Regt, Henk van der Velden, Sven Walter, and Ian Winchester.

Thanks to Tilburg University for enabling me to dedicate a large part of the past five years to writing this dissertation. Thanks also to the staff of the department of philosophy of Tilburg University for providing a friendly and cooperative atmosphere. I also want to thank the board of the Gomarus Scholengemeenschap at Gorinchem, where I have been continuing my former life as an art teacher for one day a week, for their well-disposed attitude and cooperation in making things go as smoothly as possible. Finally, I want to thank my circle of relatives and my close friends, especially Robert-Jan and Henk, for making the past five years worthwhile in very important non-philosophical ways.
Introduction
A brief history

The problem of mental causation in its old form originates with René Descartes' Meditations (1641). In his Sixth Meditation, Descartes distinguishes two substances, mind and matter, and claims that these interact. But Descartes is also the champion of 17\textsuperscript{th} century mechanistic thinking, and substance dualism (at least the Cartesian interactionist kind) does not sit well with mechanism. For if the motions of our bodies can be explained mechanistically, then what work is there for that strange, non-extended, invisible substance that is the mind? Substance dualism also requires at least one spot where mind and body get in touch: a Cartesian pineal gland. But as far as we know, our real pineal glands are not like that, nor is any other organ in our bodies. Finally, substance dualism implies that mind-interactions continuously perturb the laws of physics, in doing so undermining the very possibility of such laws, and making a complete mystery of the indubitable victory march of physical science ever since Newton.

Little wonder, then, that Cartesian mind-body dualism is now a marginal view and that contemporary philosophy of mind is largely physicalistic. Physicalism has put a ban on mind-stuff, immaterial forces, or the 'ghost in the machine,' as Gilbert Ryle called the Cartesian mind (1949). But banning ghosts is not yet banning minds. Minds need not be thought of as stuff- or force-like to begin with: Ryle rightly called doing so a category mistake. Very few philosophers are ready to deny that in our world of physical objects and forces there are thoughts, desires, hopes, assumptions, feelings, and so on. There may not be mental substances or energies, it is said, but certainly there are mental properties, or states of affairs. One may be a substance monist while being a property dualist, or pluralist.

Still, mental causation is waiting to be accounted for. If our actions did not have mental causes, most of our ordinary assumptions about human action would be radically false. Only by assuming that something like intentions cause actions can we distinguish, say, a vile kick from an involuntary fit, or a wink from a blink.\textsuperscript{1} Donald Davidson has plausibly pointed out

\textsuperscript{1} Example taken from Juarrero 1999.
(1963) that when we explain someone’s actions by citing her putative reasons (in terms of her beliefs, expectations, wishes, etc.) - reasons that justify, or at least motivated the actions - we give a causal explanation. A question like ‘Why did she do it?’, Davidson claims, concerns a cause, and not just a rationalization. And indeed, why should we be interested in someone’s motives, will, beliefs, etc., if they do not make a difference to that person’s behaviour?

Davidson has put forward a doctrine, called Anomalous Monism, that seemed to take mental causes into account. The ‘monism’ amounts to the claim that there are only (causally related) physical events. The ‘anomalousm’ amounts to the claim that while causation is a matter of strict physical law, there are no psychophysical and no psychological laws. Still, according to Anomalous Monism, mental events can be causes of other mental events and of actions, because they are, after all, physical events. We can subsume such events under a causal law as long as we describe them in physical terms; and we can evaluate such events as reasons as long as we describe them in mental terms. But the events described are just the same events.

Anomalous Monism split with the reductionistic mind-brain identity theories put forward in the years around 1960 by Feigl, Place, and Smart. Identity theory had already been under criticism for a time, especially since Hilary Putnam (1968) had argued that mental events like pains might well occur in brains of wildly heterogeneous physical constitutions. What Anomalous Monism did was to rescue the mind-body identity claim while rejecting mind-body, or mind-brain reductionism.

Anomalous Monism seemed to many an attractive (that is, respectably physicalist) alternative to reductionism. Like reductionism it seemed to avoid the dead end of Cartesian dualism, but in contrast to reductionism it gave the mind an autonomous status in the order of things. In particular, it recognized the fact that the domain of the mental is essentially normative, and that agents choose their actions by weighing reasons rather than blindly conforming to regularities. Anomalous Monism, then, seemed to allow one to have physicalism and mental autonomy at the same time.

Davidson’s critics, however, most notably Jaegwon Kim, have pointed out that Anomalous Monism does not in fact leave any causal autonomy for the mind. They accused it of implying that mental properties are

---

3 See, for instance, his 1984a, 1985 and 1995a.
epiphenomenal: produced by physical reality, but causally redundant themselves. For consider: if mental events are physical events, and the physical properties of those events suffice to cause other events - because, after all, these are what the laws of physics are about -, then what causal work is there left to do for the mental properties?

Davidson himself has always refused to see a problem here. He is a nominalist about properties, and a realist only about particulars (here: events). So the question whether it is 'in virtue of' physical or mental properties that one event causes another does not even arise for him. Most others, however, do believe that the question makes good sense.

We might consider Descartes and Davidson as two classic authors around whose claims the problem of mental causation has traditionally been centered; that is why I have started with them. But the current debate on mental causation is very complex and many-sided, and mental causation is considered problematic in more than one way. In the rest of this chapter, my aim is to map the problem area, to formulate the challenges that have to be met, and to give an outline of my project.

As to that project, the general idea is the following. As we will see below, there are no less than five problems of mental causation. These are distinct problems, but they are also closely related, which is why we cannot hope to solve them one at a time. Only a comprehensive theoretical framework that is able to cope with all five issues will help us out. Such a framework has to cover issues in ontology, the philosophy of causality, and the philosophy of mind, and these fields correspond to the three parts of my project.

Thus, in Part I, I develop the ontological foundations for my main arguments; in Part II, I develop a view of causality; and in Part III, I discuss how the results apply to mind and action. Because from the literature we can conclude that causality is a confusing and ill-understood phenomenon, I give particular attention to this topic. In my view, misconceptions about causality are the main obstacle to an understanding of mental causation. Yet, in received views about mental causation we encounter a theory of mind that is closely intertwined with a theory of causality, and only by undoing the entire knot can we come to see things in the right perspective. This is why I will also criticize a number of current views in the philosophy of mind.

---

4 See in particular his 1995.
Introduction

My general aim, then, is to develop a theoretical framework that can serve as an alternative to the received one, a framework in which mental causation is at home, rather than an unwelcome guest. All this may seem rather pretentious, but as will become clear, most of the work has already been done by others. The great ideas that help us understand the issue are already there; I combine and adjust them, and add a little of my own.
The five problems of mental causation

Instead of one, there are at least five distinct problems of mental causation: five distinct though closely related ways in which standard assumptions and intuitions about mind (on the one hand) and causality (on the other hand) seem to clash. They are problems, because they suggest that much of our philosophical, scientific and ordinary-life discourse is in radical error: that, contrary to what any sane person assumes, it does not matter one bit what we think. Before trying to do something about this embarrassing situation, let us have a look at each problem in turn.

II.I. MENTAL ANOMALISM

Our first problem of mental causation is the fact that, while our beliefs and desires are among the causal explanantia of our actions, in ascribing these intentional states to each other in order to explain such actions (i.e., in making sense of each other as agents) we are doing something very different from applying laws, as in physics.

Let us first take a look at causation. Causation and natural law seem to be intimately linked. Suppose that \( c \) caused \( e \). If so, it cannot be the case that \( e \) merely followed \( c \), for mere temporal succession just is not causation. It seems that in the circumstances, \( e \) would not have happened if \( c \) had not happened (barring pre-emption). That is, causation seems to be a matter of counterfactual dependence (among distinct items).\(^5\) And the only possible source of such dependence seems to be natural law. True, many empiricist philosophers mistrust counterfactuals and believe, in the wake of Hume, in regularities only. On their view, \( e \)'s being caused by \( c \) depends on the principle of 'same cause, same effect.' But note that just like the counterfactualist approach, the regularity approach treats causation as a matter of natural law: it only has a different conception of natural law! And indeed,

\(^5\) Strictly speaking, counterfactual dependence is among propositions. Below, I will speak, however, of counterfactual dependence among states of affairs, events, etc.: as Lewis says (1973, 199), 'Counterfactual dependence among events is simply counterfactual dependence among the corresponding propositions.'
something must distinguish causal from noncausal successions, the notion of natural law being the only 'something' plausibly invoked.

Now let us look at intentional states, the causes of our actions. Davidson has argued that the ascription of such states is essentially a holistic affair. For beliefs and desires can only be ascribed in large clusters that are constituted by normative connections such as 'coherence, rationality, and consistency' (1974, 231). Only insofar as specific intentional states (such as, for instance, the belief that Jones is a good sport) are embedded in such clusters, it seems, can we determine whether they are true, appropriate, or reasonable, and whether it makes sense to ascribe them in the first place.

This mental holism has been Davidson's main consideration in his rejection of psychological or psychophysical laws: the rational connections among intentional states do not seem to have a counterpart in physical theory. But apart from that, rational thought and responsible agency are not easily squared with blind, anonymous and indifferent laws of nature anyhow. Do we not have some degree of autonomy and freedom? When the telephone rings, most of us can be reliably predicted to go and answer it; yet, we could all just decide not to do so at any time. We mostly drift along with the daily course of events, but we do not seem to be at its mercy (if we come to feel that we are, we had better find help). If we cannot act according to choices that are genuinely our own, our very status as agents is compromised. But if decisions and actions are to be genuinely our own, they are not supposed to be determined by mere natural law, as it were behind our backs.

It might be thought that the way we deliberate is still highly predictable, because sane deliberation is subject to rational constraints, or rules, and as agents we all want to be rational. Might then lawfulness and acting for reasons not still coincide? Little chance, for practical wisdom is typically not codifiable. If it were, we could all use handbooks or pocket computers for practical thinking (for instance to solve ethical dilemmas), and we would not have to decide for ourselves. But that would be something for brainwashed zealots rather than for responsible agents.

But is it suggested that our minds elude the laws of nature? Does the above imply that a science of human beings impossible? No, such claims are preposterous. The relation between natural law on the one hand and thought and agency on the other hand is bound to be subtle and indirect. But it remains a fact that while our beliefs and desires are causes of our actions, they do not seem to be such in virtue of nomically necessitating them
ii. The five problems of mental causation

(or nomically raising their probability). But given that cause and law are inseparable, how can this be? This is our first problem of mental causation.

II.II. NORMATIVITY

The second problem of mental causation derives from the fact that at least some causes of our actions seem to have an essential normative aspect. Normativity comes in when, after some deliberation, we are motivated to act in a certain way. In such a case we do what we do because we have weighed the pros and cons, and have judged the pros to be decisive. The reasons we have for acting like we do carry a normative, or motivating, force.

But what is ‘normative force’? Is it something like weak nuclear force? Before dismissing this as the absurd suggestion it in fact is, we should realize that someone's being or not being motivated by a reason often makes a considerable, and thoroughly physical, difference. Think, for instance, of the flagrant injustice of a racially biased court sentence that is a reason for (i.e., motivates) someone to organize a mass demonstration.

Our second problem of mental causation, then, is this: How can motivation by reasons be a source of physical change when there is no such thing as normative force to be found in modern physical theory?

II.III. PHENOMENAL STATES OF AFFAIRS

The third problem of mental causation is this: how do phenomenal, first-person states of affairs fit within a blind causal order?

As minded creatures, we have a first-person view of the world. Apart from the objective, impersonal states of affairs about the world, there also seem to be subjective, phenomenal, first-person states of affairs. The best-known such states concern the way things feel, smell, look, taste, or sound: qualia. And they are by no means unimportant. Our behaviour would be very different from what it actually is if we did not, for instance, avoid pain.

---

6 For related issues, see e.g. Davidson 1963, Dretske 1988, and Kim 1998b.
7 For related issues, see e.g. Chalmers 1996, Jackson 1982, and Nagel 1986.
Steering our behaviour may well be the very biological point of these phenomenal states.

On the other hand, the causal order is blind. Natural law does not seem to care at all about the way things 'appear.' Our arms and legs are moved, in the end, by our muscles and nerves. The whole mechanistic shuffle of physical signals and objects, it seems, can very well go on without subjective experiences. Qualia, it seems, cannot push or pull: they are so ethereal that it is even hard to say where they are. They cannot even be perceived (or had) by more than one subject. How on earth could such entities (if they are entities in the first place) have an impact on the solid, outside, objective world of objects and forces?

Yet, we know perfectly well that they do. If qualia were really causally impotent, or epiphenomenal, it seems that we could not even so much as refer to them. Also, what we normally take for granted about experiences of pain, fear, hunger, lust, anger, and so on, would have to be utterly false: namely, that these experiences make us behave in certain definite ways, and that that is exactly why we have them in the first place, biologically speaking. This, then, is our third problem of mental causation: on physicalism qualia seem causally idle, but we just know that they are not.

II.IV. EXTERNALISM

The fourth problem of mental causation concerns the issue of externalism. It is generally held that one of the core characteristics of the mind is intentionality, or aboutness (the other being consciousness, the subjective perspective described above). Now, aboutness is a relational concept: a's being about b obviously is a relation between a and b. Of course, we can have thoughts about items that do not exist (such as Cinderella); but in such cases we find relations to, for instance, circulating fairy tales and public language.

The widespread appreciation that 'mind' is a relational concept, and that mental states require a good deal of context apart from local brain processes, is known as externalism. Now, the appeal of externalism is the source of our especially puzzling fourth problem of mental causation. Kim (1998a, 36-7) writes:

---

ii. The five problems of mental causation

That a given intentional state of an organism instantiates a certain semantic property is a relational fact, a fact that essentially involves the organism's relationship to various external environmental and historical factors. This makes semantic properties relational, or extrinsic, whereas we expect causative properties involved in behavior production to be nonrelational, or intrinsic, properties of the organism. If inner states are implicated in behavior causation, it seems that all the causal work is done by their "syntactic", or at any rate internal/intrinsic, properties, leaving their semantic properties causally otiose. The problem of mental causation generated by syntacticalism therefore is to answer the following question: How can extrinsic, relational properties be causally efficacious in behavior production?

It seems that just as the 'syntactic' operations in our desktop computers cause whatever we see on our monitors, regardless of what it all means, the 'syntactic' operations of our neurons cause whatever sounds and motions we make, again regardless of what it all means. The meanings, or contents, of our thoughts do not seem to be required to help along the local physical machinery of our nervous systems. And they could not do so anyhow, it seems, precisely because of their relational character. McGinn (1989, 133), himself embarrassed by this issue, writes:

(What happens at the causal nexus is local, proximate and intrinsic: the features of the cause that lead to the effect must be right where the causal interaction takes place. Causation is the same with brains and minds as it is with billiard balls. Their effects depend upon local properties of these entities. The causal powers of a state or property must be intrinsically grounded; they cannot depend upon relations to what lies quite elsewhere.

But this conclusion is the source of our fourth problem of mental causation, for we know very well that although metaphysical theory does not allow it, the contents of our thoughts definitely do have tangible and physical effects. Why else should we bother to think about, say, the car from the right, or the structure of the bridge that we are commissioned to build? And if intentional-relational content were really causally ineffectual, it would be a miracle how we could ever influence each other's behaviour by just talking to each other.
Introduction

II.V. CAUSAL EXCLUSION

The fifth problem of mental causation rests on a combination of three metaphysical principles. The principles, of which Kim is the most prominent defender (e.g. 1989b, 1998a), are the following:

1) *Psychophysical supervenience*. This principle says, roughly, that all mental states of affairs are determined, or necessitated, by the physical states of affairs. On the one hand, some set of physical states of affairs in a world is required for any set of mental states of affairs in that world. On the other hand, given the physical states of affairs of a world with mental states of affairs, there could not possibly fail to be the mental states of affairs of that world.

2) *Causal closure of the physical domain*. This principle says that any physical state of affairs has a complete cause that is itself physical, i.e., consists entirely of physical states of affairs.

3) *Explanatory exclusion*. This principle says that no state of affairs has more than one complete and independent causal explanation.

The fifth problem of mental causation, that of causal exclusion, is that the principles cited here seem to preclude any causal relevance for mental states of affairs. Suppose that we have an action *a* (say, my fetching of an umbrella), a bodily motion *b* (my moving my arm), an intentional state *i* (my expectation that it will rain), and a neural episode in my brain *n*. The problem is that *i*, the expectation, the reason for which I act, is excluded as a cause of the action, no matter how we look at the case.

To begin with, *n* is obviously the physical cause of *b*: it is what happens in the brain (for some time) that leads to the neural impulses and muscle contractions that make my arm move (let us for the moment ignore the fact that neural explanations are biological rather than physical). We do not seem to need *i* here, for by causal closure *n* can do all, and by explanatory exclusion, if *n* does all, then *i* does nothing. For those who think that intentional states are brain states, and actions bodily movements, this argument will already be pretty decisive.

But we may well object that *b* is mere bodily motion, not action. On most theories of action, *a* will be said to include *b* along with something like an intention, caused by *i*. Action, we might say, has an essential intentional element. Is it not the case, then, that *b* is causally explained by *n*
ii. The five problems of mental causation

while a is explained by i? Does not i have a causal job with regard to a, the fully fledged intentional action?

It seems not. For there is a physical base on which a supervenes. On externalism, this base includes environmental and historical factors besides b (because of the intention). But a broad physical base is still a physical base. Now, first, the causal closure principle says that we need only physical states of affairs as explanantia of this base. So i drops out. Secondly, a's physical base fully determines a. Again, i drops out as explanans. All has been already physically settled behind i's intentional back.

But could i not be a causal factor along with the physical states of affairs? No, the principle of explanatory exclusion precludes this. The complete physical explanation is already good enough, thank you, if slightly laborious. If i wants to be a cause of a, it seems, it will either have to disrupt the closed web of physical causes of a's supervenience base (or just: of a, for if x causes y and y determines z, then x just causes z - does it not?); which will be a violation of the principle of causal closure. Or i will have to overdetermine a, as a causal competitor of a's physical causes. But overdetermination in that sense is precluded by the principle of explanatory exclusion.

It seems, then, that on the three principles mentioned there is no possible causal role for intentional states such as expectations, beliefs, intentions, and so on. We may explain each other's actions in terms of such states, but this cannot be causal explanation. At best it will be some convenient shorthand for causal explanation in physical terms. The causal exclusion argument, as outlined above, implies that if I fetch an umbrella because I expect that it will rain, it is not the case that my expectation is a cause of my fetching the umbrella. The expectation idly rides along on the back of a physical cause that usurps all efficacy. It is epiphenomenal.

Our fifth problem of mental causation is, of course, that the implications of this conclusion are just crazy. Although we probably keep trying to change each other's minds in order to influence each other's behaviour, we now find ourselves in the company of some witch doctor who has recently learned that malaria has a biological, not a spiritual cause. He knows that his incantations rest on superstition, but he keeps using them anyhow, because they are occasionally successful (nobody knows how), and he does not know much about modern medicine. As for us: we treat each other as intentional subjects (with occasional but obscure success), but that is only because we are not good at neuroscience or particle physics. What we
think, expect, hope, want, and so on is not what in fact causes our actions. Something else does. Mental causation is just a fairy tale.

Although a position like this has been defended for a time by some philosophers (most notably P. M. Churchland, e.g., 1981), we should appreciate how extreme such a view is, and to what absurdities it commits us. So if the causal exclusion argument seems compelling, we have a considerable fifth problem of mental causation.
iii

A look forward

III.I. WHAT WE NEED

Given these five problems, what should we ask of a theory of mental causation? What kind of theory is acceptable, and what should the theory be able to accomplish? Let me start with three general demands.

First, a theory of mental causation fails if its conclusion amounts to some form of epiphenomenalism or eliminativism. There is no guarantee that these views are false, but I agree with Lynne Rudder Baker (1995) that if a piece of armchair metaphysics clashes with well-tested streetwise common sense - especially such a large and vital part of common sense as is threatened by epiphenomenalism or eliminativism - it is more likely that the metaphysics is in error than common sense.

Secondly, a good theory of mental causation ought to make the phenomenon turn out natural, and not exceptional or special. On a good theory, mental causation should not be squeezed into place, but fall into place.

Thirdly, a good theory of mental causation ought to appear plausible on independent grounds, and not just because it favours mental causation. This is why, in drawing up my analysis of causality, I will largely leave the issue of mental causation aside.

Let us now see what demands are suggested by the five problems.

First, a theory of mental causation should make it intelligible that while natural law and causation are intimately related, cause and effect need not be subsumable under any specific causal law.

Secondly, a theory of mental causation should provide an account of normatively significant states of affairs and of how such states of affairs become motivating reasons for (or of) actors, in virtue of their normative significance. It should do so, however, within plausible physicalistic, or naturalistic, constraints.

Thirdly, a theory of mental causation should clarify the ontology of qualia. Unfortunately, this is such an intricate and contested issue with so little general consensus and so many conflicting intuitions, that in the context of the present work I cannot hope to deal with this issue in an ade-
quate way. I will not, then, have much to say about the causal efficacy of qualia. A view that has my sympathy is that qualia are in fact intentional phenomena, and not some kind of essentially subjective entities. If so, there may not be a separate problem of qualia to begin with. But I cannot do more, below, than make some brief and tentative remarks about the issue.

Fourthly, a theory of mental causation ought to handle externalism. Not by limiting the damage by, for instance, distinguishing 'narrow' from 'broad' content, but by fully acknowledging the many-dimensional relationality of intentional states of affairs and by pointing out how such states of affairs can be causes of actions. It should do so, however, - and there is a catch! - without violating the plausible principle that causal influence is local and proximate.

Fifthly, a theory of mental causation should avoid violation of Kim's perfectly sound three principles of psychophysical supervenience, causal closure, and explanatory exclusion.

As a sixth point, however, it should do so while being nonreductionistic. Kim himself defends an austere reductionistic programme that winds up uncannily close to eliminativism. I do not consider such a result very satisfactory, and reduction of intentional properties to physical ones is an extremely precarious project anyhow. The sixth demand on a theory of mental causation, then, is that it explains how intentional states of affairs as such cause actions. And this without either disrupting the network of physical causes of these or overdetermining them (the fifth demand).

Such, then, are the demands on a theory of mental causation that are suggested by our five problems. Meeting these demands requires the development of a theory of causation that, while doing justice to a number of sound and familiar intuitions, puts some of the many aspects of causation in a new light. Hence, my account of the character and ontology of causation is the core part of my theory. But I will also offer a theoretical background concerning mind and action, with the necessary attacks on some standard views (most notably functionalism).

By far the most effort and space will be spent on dealing with the fourth above demand, i.e., the attempt to make it intelligible that relational properties can be causes. The reason is this. With regard to more standard views, my theory of mental causation will involve a subtle but important shift in ontological outlook, away from the common emphasis on qualities and intrinsic, or internal properties, towards what we may call relationism,
iii. A look forward

a view that stresses the importance of relationships and relational properties. This relationism will manifest itself most notably in a thoroughly externalist account of mental properties. Hence, an account of the causal efficacy of relational properties is at the very core of my project, and the solutions to the other problems of mental causation will largely depend on this account.

III.II. WHAT IS COMING

Let me now provide a brief outline of my argumentation in the chapters to come. The book is divided in three parts. In Part I, I sketch the ontological framework that is at the basis of my further arguments. I discuss the notions of particulars, properties, relations and states of affairs, and make a case for what I call a relationist view of properties (Chapter 1). I then discuss the notions of physicalism, supervenience, and physical realization (Chapter 2). Next, I provide an account of the 'layered' structure of the world that goes with physicalism, considering the notions of emergence, 'upward' causation, reduction/type-identity and token identity of events and states (Chapter 3). I argue that our articulation of physicalism does not need these notions, the first one being superfluous while the other three are downright suspect. Against type identity I propose an alternative non-reductionist viewpoint, one that is based on the relational character of higher-level phenomena.

Part II is about causality. I here propose an analysis of the notion of causality, and take up the issue of causal ontology. First, I point out that causality is a double-sided notion: it involves dependence among states of affairs as well as physical connectedness among particulars. (Chapter 4). I then discuss causal dependence (Chapter 5) and causal connection (Chapter 6) in more detail, on the basis of the work of, respectively, John Mackie and Wesley Salmon. I proceed (Chapter 7) by explicating how dependence and connection approaches can be integrated into an analysis of causality that avoids many of the perplexities afflicting traditional accounts. I am then in a position to explain how we can recognize an intimate link between natural law and causation while rejecting the notion of a 'causal law' (Chapter 8).
The following four chapters are dedicated to the issue of causal ontology. I argue that the basis of a causal ontology should be an appreciation of the fact that causes involve particulars as well as properties, and I take a critical first look at various \textit{prima facie} kinds of causal relata (Chapter 9). I then consider in more detail the category of events, which is widely believed to be central to causation. I present a number of objections to a Davidsonian event conception (Chapter 10), and argue that we can explicate what events are in Kimian ‘factualist’ terms. This is achieved by proposing a ‘narrative’ account of event individuation, instead of the more usual spatiotemporal accounts. This leads to the conclusion that causes and effects are constituted by particulars and properties, hence are states of affairs (Chapter 11). This raises the question whether relational properties, or relations, can also constitute causes and effects, and this in turn whether \textit{causal} relations can. I argue (Chapter 12) that both questions have an affirmative answer.

But saying that they can is not yet saying that they do, and the causal efficacy of relational states of affairs is controversial. I proceed to argue, on the basis of what I call the criterion of local difference, that indeed there is every reason to think of relational states of affairs as causes and effects (Chapter 13). I conclude part II by discussing the issue of ‘supervenient causation’: the question whether, and how, autonomous causal efficacy at higher ontological levels is possible given causal closure (i.e., the impossibility of genuine downward causation) and supervenience (Chapter 14).

In Part III I discuss the mind. I do not try here to develop a fully fledged philosophical theory of the mind, but I do criticize a family of received views, and suggest alternatives. I start out by sketching a general naturalistic perspective on mental properties that is based on the notions of teleology, information, and representation. I apply here work by Fred Dretske and Ruth Millikan (Chapter 15). In this chapter, I also make a few tentative suggestions about how to approach the difficult issue of subjective experience, or qualia. The remaining five chapters of Part III are all directed against specific aspects of a conception, or family of conceptions, of the mind, that has been very influential in the debate on mental causation. First (Chapter 16), I criticize what we may call the ‘computational theory of the mind,’ or ‘syntacticalism.’ In this chapter I argue that propositional attitudes, such as beliefs and desires, are not the formal entities of syntacticalism, so much as mental models. This account is based on Colin McGinn’s ‘theory of mental models.’ Secondly (Chapter 17), I criticize the
iii. A look forward

functionalist theory about ‘folk psychology,’ i.e., the way in which we ascribe mental content and explain each other’s actions. I thus take issue with the so-called ‘Theory Theory,’ denying that mental states are posits in a prescientific causal theory of the mind. As an alternative, I suggest Goldman’s version of what is known as simulationism. Thirdly (Chapter 18), I criticize internalism, i.e., the assumption that mental properties are somehow internal to the minded organism. In this chapter, I discuss six forms of externalism. This discussion is intended to substantiate my relationist approach: it suggests that mental properties are essentially environment- and past-involving, and that the assumption that mental states are brain states must be mistaken. Fourthly (Chapter 19 and 20), I revisit some metaphysical issues, this time concentrating, not on higher-level properties in general, but on mental properties in particular. In chapter 19 I criticize several theories of psychophysical type identity, along with eliminative materialism. Next (chapter 20), I criticize the version of nonreductive physicalism that is most standard nowadays, namely psychophysical token identity theory, or token event physicalism (Chapter 20).

A concluding chapter gives an overview of what has been accomplished with regard to the five problems, and briefly points out the crucial assumptions and arguments that have been put forward (Chapter 21). My conclusions with regard to the five problems will involve the following claims.

Problem 1 (mental anomalism): Although causation presupposes physical laws, it does not presuppose any laws of causal succession; nor is causal efficacy a matter of nomic subsumability. Therefore, the lack of psychological causal laws is compatible with the reality of mental causation.

Problem 2 (normativity): Perception of normatively significant states of affairs as such is possible, and it is often the perceived normative aspects of situations that make us act in specific ways.

Problem 3 (phenomenal states of affairs): On the standard conceptualization of subjective experiences, namely as non-functional ‘qualia,’ mental causation is an insoluble problem. But this conceptualization need not be taken for granted.

Problem 4 (externalism): Even though mental states are relational states of affairs, they can be causes and effects just like nonrelational states of affairs. For first, relational states of affairs can be appropriate causal re- lata. Secondly, relational states of affairs can indeed be causally nonredundant. Thirdly, relational states of affairs can meet a criterion of causal efficacy (the ‘criterion local difference’) that my analysis of causation as out-
Introduction

lined in Part II suggests is plausible. And fourthly, the causal efficacy of mental properties, being historical-relational in character, can be accounted for with an appeal to the notion of a 'structuring cause.'

Problem 5 (causal exclusion): Because physical realization is not a one-to-one affair, and because mental states of affairs are neither reducible to, or token identical with, physical states of affairs, there can be mental causes that are not physical causes. It is true that merely physical states of affairs, by physical causation and realization, collectively determine everything that is caused by mental states (such as actions), but this does not render mental states of affairs causally redundant. For due to the many-many character of physical realization, the causes of intentional states of affairs or events in particular (for instance, specific thoughts or actions), cannot be picked out physically. Cases in which mental states seem to 'downwardly' cause merely physical events can be accounted for by applying the notion of a constraint: mental states do not cause physical events, but they do constrain the physical causes of such events.
Part I. Ontology

The following three chapters outline the metaphysical framework that is at the basis of my account of mental causation. I am aware that putting forward this bit of metaphysics is hugely pretentious. But I cannot omit this difficult first part, for discussing mental causation requires talk about reality, particulars, properties, states, relations, supervenience, and physical realization; and if such terms are not sufficiently clear (as is too often the case) the discussion will lead nowhere - especially when later on we go venture out into the quagmire of causality. Conversely, if such terms are given a clear meaning, this will have substantial and interesting consequences.

In order to achieve the desired clarity and theoretical leverage, I will sometimes have to walk through a china cabinet of metaphysical subtleties and undecided issues. Some issues can be left untouched (which is fortunate), but on other issues I will have to take up a position without being able to face a number of well-argued alternatives (which is regrettable but unavoidable). Anyway, if my metaphysical framework is defective, Part I should at least make it visible where things go wrong.
1

Particulars, properties, and relations

1.1. PARTICULARS

Space-time is populated by particulars: concrete entities with a unique spatiotemporal location. Objects (stones, chairs, organisms, etc.) are good examples, but masses (heaps of sand, pools of water, etc.) also seem to make good particulars. Scattered groups (such as soccer teams), processes (such as light beams or pressure waves), wave/particles (such as electrons and photons) and fields (such as magnetic or gravitational fields) may be more controversial, but I will not have much to say about these. Events and states of affairs are generally considered particulars, but here some caution is required, as we will see later on. I have to skip issues about propositions, numbers, facts, or negative and disjunctive items; I will not treat these as particulars, but apart from noting that they do not seem to occupy specific space-time regions, I have no heavy-weight arguments concerning them on offer.

Objects are the most salient class of particulars: they have a relatively enduring and resilient structural coherence. Millikan (1984, 291-2) says that they 'retain many of their properties over time in accordance with natural conservation laws'; and even an object's very cross-time numerical identity may in this way be a causal affair. Suppose, for instance, that a particular is individuated by a property such as being an F, or having such-and-such an internal structure. Then, if that property persists in virtue of some law of physics, this will be a reason to say that the object that is an F at t and the object that is an F at t₀ are the very same object.

We may be reluctant to regard certain things, especially living organisms, as mere objects. These are better described as systems. A system may interact as a unit with things in its environment, but it processes, or transforms, causal influence rather than just propagating, or transmitting it. The energy, for instance, that we take from our environment by eating a sandwich or hearing a speech will not be propagated and translated into output in as straightforward a manner as do bouncing billiard balls with the energy they absorb.
1. Particulars, properties, and relations

By specific kinds of interaction with their environments, systems also are changed in non-coincidental ways (for instance by eating); systems can even be said to change themselves. Systems are the subject of pioneering study these days (it is even claimed that the study of complex systems is an all-new autonomous scientific discipline), but for our present purposes we can regard systems as just a special subclass of objects. We might say that objects are either aggregates or systems, depending on whether or not they have some kind of complex organization.

Let us return to particulars in general. Particulars have a spatial part-whole structure, or mereology: typically (with the exception of elementary particles - if there are any - and of the cosmos) they have smaller parts that are themselves particulars, and they are themselves part of bigger particulars. This suggests - but the suggestion can be resisted - that particulars need not be salient objects, but may also be gerrymandered or scattered groups.

Particulars are also concrete, being able, in Steward's words (1997, 31), to lead a 'secret life.' You can know them partly and discover an indefinite amount of facts about them. You can also fail or come to know that what you thought were two distinct particulars is, in fact, one and the very same one. Particulars have many aspects, or properties, of which typically only few are known.

Note that a particular's concreteness implies its having a unique spatiotemporal location. We might ask why such a location should be unique: could there not be more than one concrete particular in the same place at the same time (say, a piece of marble and a statue)? But every putative reason for supposing this to be the case is in fact a reason for thinking that one particular just has one more property.

We will see below that events and states of affairs lack a spatial mereology, are not concrete (although they do have properties), and can share one spatiotemporal location. I will therefore not call them particulars, or particular entities. Some (e.g., Kim) choose otherwise, on the ground that events and states of affairs, although they are not concrete entities, are spatiotemporally located, hence particular.

---

9 See, e.g., Waldrop 1992.
10 Note that the difference is not that systems are dynamic while mere objects are static. If object identity over time is a causal affair, mere objects are also dynamic.
1.2. PROPERTIES

Properties are abstract, rather than concrete: there can be many of them at the same spatiotemporal location.\(^\text{11}\) We can define properties as precisely the indefinitely many aspects that concrete particulars have. This is a liberal view: anything that can be the case with a particular here counts as a property. One might ask for stricter demands: Shoemaker (1980), for instance, claims that causality and properties are interrelated notions, in that anything about a particular that is of no causal relevance whatsoever is not plausibly called a property. In my view such a suggestion makes good sense, but lacking at this point a criterion of causal relevance, it will be wise to leave this issue open.

Properties as such do not seem to have much of a spatial location or mereology: only the particulars that have them do. The paper cup is on the table and has two halves. But its molecular structure or colour are on the table only insofar as the cup is, and they do not themselves seem to have two halves. No doubt, when we consider a particular under just one aspect (the cup as being a thing made of paper, or a red thing), there will be many things to know about the being-made-of-paper or the redness. This might resemble the 'secret life' of concrete particulars, but it does not make these properties concrete entities, in the sense of having a unique place and time.

Properties, it seems, presuppose items that have them. On the other hand, we sometimes speak of uninstantiated properties, such as the structure of a non-existent but possible chemical compound. This issue is related to the question whether properties are merely 'abstract particulars' (or tropes), or universals.\(^\text{12}\) Fortunately, these difficult issues can largely be left outside of the scope of my arguments. It is true that the theory of causation that I will defend later on presupposes realism about laws of nature, and it can be argued that the concept of a law and that of a universal are tightly linked. On the other hand, I will argue that by far the most properties (that is, those of other domains than basic physics) do not constitute

\(^{11}\) I here skip the various subtleties connected with the notions of abstract-concrete and particular-universal (see, eg., Macdonald 1998, 333).

\(^{12}\) The issue is about resemblance and the question whether, or how, it should be explained.
laws. And I will anyway be concerned only with properties that have actual bearers.

Properties are, as it were, ontologically promiscuous. We do not just speak of properties of particulars or states of affairs, but also of sentences, numbers, facts, kinds, probabilities, etc. The same goes for relations: it is not just particulars and states of affairs that are related, but also propositions, sums, centres, contours, types, meanings, names, regularities: anything whatsoever, no matter what its ontology. Note, however, that for such properties and relations to be part of the actual world, not mere possibilities, the existence of specific particulars is required anyway. Also, the properties and relations that will concern us are only of, or among, the basic worldly items that our states-of-affairs ontology (to be outlined below) admits: particulars and states of affairs. Not that other properties or relations are unimportant. The reason for narrowing down our view in this way is just that the brief ontological overview given here is meant as a preliminary to a number of arguments about causality.

1.3. HOW PROPERTIES ARE HAD

Let us explore the relation between particulars and their properties in some more detail. To begin with, I assume that we had best avoid both a bundle theory, on which particulars are mere bundles of properties, as well as a substratum theory, on which properties are mere attachments to the particulars that have them. For bundle theory will always have trouble explaining that, and how, properties can be bundled in the first place, while substratum theory seems to imply the possibility of featureless particulars, or at least of particulars whose identity is insensitive to what properties it does or does not have. Both kinds of theory miss precisely what I take to be the essential point, namely that the notions of ‘particular’ and ‘property’ are linked up in an important way.

The point can be brought out by noting that a particular’s having a property is not happily regarded as a relation between that particular (or ordered multiple of particulars in the case of a relation - see below), and

---

13 Armstrong (1997, 44) calls such properties ‘second-class,’ denying that they are universals.
14 For more on this issue see, for instance, Loux 1998.
15 I owe this point to Anthonie Meijers (in conversation).
another entity. We might say that properties exist, or are real, and therefore are entities. But if this makes us think that they can be related to particulars in the same way that particulars can be related to each other, it would be wise to reserve the term ‘entity’ for concrete entities. Properties are not pasted upon their particulars but are, as Armstrong suggests (1997, 30), best seen as ‘ways things are or ways things stand to each other.’

Particulars can have or lack any of their present properties: the latter are had contingently, as we might say. But this claim ought to be hedged in more than one way. First, some properties are had by particulars essentially: for instance, all particulars are identical with themselves. Secondly, particulars can have properties essentially in the sense that they would not be the kind of particular they in fact are without the property (a piece of merchandise, for instance, is somebody’s property essentially).

Thirdly and relatedly, particulars may be destroyed or created, rather than changed. And the beginning and end of a particular’s existence can only be specified by saying of what kind or type the particular is, that is, by specifying the particular under one aspect rather than another. A particular may exist as a boat, and also as a flower-box; and the life-time of the boat may not be the same as a that of the flower-box. The particular at issue does not seem to have an identity, let alone a beginning or end, independently of what kind of particular it is.

Fourthly, and most importantly, a particular cannot lose all of its properties without acquiring others. That is, the notion of a bare or featureless particular is rather suspect, if not incoherent. The reason is that we cannot pick out a particular except as an F or a G, or as the thing that is F or the thing that is G. An alleged particular that cannot be specified as anything can rightly be suspected to be nothing at all. This again refers us to the linkage between the notion of a particular and that of a property. The point is this: a particular is individuated by its properties (including its spatiotemporal relations), and by nothing else. We cannot, for instance, individuate a particular by just pointing at it.

True, from the claim that there cannot be two particulars at the same place at one moment, it follows that a particular at a fixed time could in principle be individuated just by its spatial boundaries.\(^\text{16}\) Such boundaries

\(^{16}\) This, presumably, is why it is sometimes held that particulars are capable of ‘independent existence,’ as Kim says of substances (Kim 1996, 3). Note, however, that under any of its aspects (say, as a boat, a person, a red object, or even as a square object, or an object made of copper) a particular may not be capable of independent existence at all.
1. Particulars, properties, and relations

are typically either vague or arbitrary relative to any interesting further classification (do my dead skin cells still belong to the organism that is I?) but we do seem able to individuate a concrete particular as just the-thing-that-is-here-and-here. As I will argue later on, this is an important difference of concrete particulars with states of affairs.

The view about particulars and properties outlined here will be echoed later on by my view on causal connections and causal dependencies. If we add to a particular’s having a property at a time the element of natural law and, with it, that of persistence and change (or perhaps we should say time), we have what we might call physical connection - among particulars and circumstantial counterfactual dependence - among properties. We then have introduced causation. The idea is that just as we say of mere existence that particular and property presuppose each other (but are not the same thing), we should say of causation that connection and dependence presuppose each other (but are not the same thing).

1.4. STATES OF AFFAIRS

The above suggests that we live in a world of particulars with properties. This is a fairly standard view, expressed, for instance, by one of the three mottoes that Armstrong lets precede his 1997: ‘The universe consists of objects having various qualities and standing in various relations’ (a quotation from Whitehead and Russell). This ontology may be regarded as one of a ‘world of states of affairs’ (Armstrong 1997) or, in Ludwig Wittgenstein’s original locution (1922) - which I particularly like because of its connotation of Verhältnisse, i.e., relations (see below), and that of Sachen, i.e., concrete particulars - one of a world of Sachverhalten.

Such states of affairs, situations (e.g. Menzies 1989), or events (e.g. Kim 1976) are spatiotemporally situated particular-property complexes. States of affairs are generally regarded as entities, or particulars, but let us again note that their particularity does not give them concreteness. I have not ruled it out that states of affairs lead something like a ‘secret life,’ but there can be many states of affairs at one place and time. A table, for instance, can be brown, square, and be made of wood all at one place and time.

One might wonder what place in this ontological scheme there is for states of affairs like the state of the economy, for ‘ongoings’ like the rain, or for events like the disaster or the sudden silence. Where are the particulars...
and properties that are supposed to constitute these items? At least in the case of events, some choose to regard these as a species of particular (for instance Davidson 1970b). I will, however, later on argue that, and how, such items can be understood in terms of states of affairs as defined above.

Before proceeding, let me say something about my terminology. First, we will see later on that the category of states of affairs as defined, i.e., the category of properties of/relations among (n-tuples of) particulars, does not just include ‘static’ states (e.g., Jack’s being a man), but also changes (e.g., Jack’s eating his lunch) and ongoing processes (e.g., Jack’s chewing). The term ‘states of affairs’ therefore denotes events and processes as well as states. For reasons of readability, however, I will not always speak of ‘states of affairs,’ or ‘states, events and/or processes’ but loosely talk about ‘states,’ ‘events,’ ‘states and events,’ or ‘properties,’ where maximum precision is not crucial.

Secondly, the term ‘property’ might be taken to mean something general like ‘red,’ ‘having a headache,’ etc.; but when I say ‘property,’ this will by default mean ‘property token,’ ‘property instance,’ or ‘exemplified property’ rather than ‘universal,’ ‘type’ or ‘unsaturated state of affairs.’ My preferred phrase is ‘state of affairs,’ but I will sometimes have to switch to ‘property’ in order to conform to standard usage in a debate (the reader will be warned about this).

Thirdly, I will also sometimes use the term ‘fact.’ Now, states of affairs are sometimes assimilated with facts, but the two are not quite the same. I reject the view that facts are truths, i.e., linguistic or propositional entities of some sort, but I do hold that states of affairs are ways things (i.e., particulars) are, while facts are ways the world (or reality) is. This is why we easily speak of, for instance, negative (or disjunctive, or general, or conditional) facts, such as the fact that Jones has not arrived, while talk of negative states of affairs or events tends to strike us as counterintuitive.

1.5. TOKEN IDENTITY OF PARTICULARS

We have seen that because of its concreteness a particular could at a given moment be picked out by its spatial boundaries. I have also been skipping a distinction that is sometimes made between kinds and properties/relations.\(^{17}\) For whatever the force or importance of that distinction,
insofar as a particular's being an F can be regarded as one of that particular's properties, it seems that we can ignore it here.

A token identity view of particulars now becomes plausible. At time t, for instance, there exist a structure of organic molecules, a quadruped, a cat, a pet and an idol; and we may learn that these, all being at the same place, are in fact one and the very same particular. The quadruped is the cat, the cat is the pet, and so on. Not always, and by no means necessarily, but at time t. We have at time t one particular with five properties (in fact, of course, there will be many more), rather than five distinct entities.

Such an identity view of particulars is rejected by Baker (1997). Baker argues that a statue is not identical with the piece of marble that occupies the same space, because a statue is a relational item (it can only exist in a social context) while a piece of marble is not. The relation between the piece of marble and the statue, Baker claims, is one of constitution, not identity.

It seems, however, that we can grant Baker's point that statues are not just pieces of marble, while sticking to our intuition that this particular statue definitely is this particular piece of marble, and that the piece of marble's standing before the town hall is the very same state of affairs as the statue's standing before the town hall (after all: same particulars, same relation, same time). The intuition might be strengthened by an argument to the effect that the causes and effects of both states of affairs are identical.

The token identity view of particulars outlined here seems to me the only credible form of token physicalism available. It is sometimes held\(^\text{18}\) that a (psychophysical) token identity thesis is a minimum requirement for physicalism. As it happens, standard token physicalism (the claim that mental states are token identical with physical states) will be among my main targets later on. But I do not see a problem with the claim that all concrete particulars (minded or not) are physical particulars (i.e., are of a physical kind, or have physical properties). This amounts to a rejection of substance dualism, while being compatible with pluralism, or nonreductionism, about properties.

\(^{18}\) See, for instance, Cynthia Macdonald 1990.
Part I. Ontology

1.6. Relations

In contemporary metaphysics, 'intrinsic' or 'internal' properties of particulars are generally ranked higher than relations, or relational properties. Because I will argue that relations and relational properties should be taken at least as seriously as intrinsic properties (a position that we could call 'relationism'), let us finally have a look at relations.

Relations are n-adic properties. Monadic properties (such as, for instance, the property of being made of copper, or having a spin of x) are had by just one particular at each instance. Many properties, however (such as the dyadic property of being married), are in fact relationships between two or more particulars. The bearer of an n-adic property will then be an ordered pair, or triple, or quadruple, etc., of particulars.\(^{19}\)

As we saw, relations have the kind of ontological promiscuity that also pertains to properties. But for reasons given above, the relations that will concern us are just among particulars and states of affairs. Having thus limited our scope, let us distinguish various types of relations.

In the merely physical domain we find relations of co-existence and co-occurrence, spatiotemporal and 'partitive' relations (i.e., overlap and mereology),\(^{20}\) quantitative relations, relations of resemblance, covariation and difference; distinctness; and, of course, causal relations.\(^{21}\) I will argue later on that this last category of relations splits up in causal connections among particulars and causal dependencies among states of affairs.

In contrast to the merely physical domain with its rather few types of relations, in the biological, psychological and social domains we find an enormous variety of 'new' kinds of relations. These domains are constituted by innumerable properties not referred to by physical theory: for instance,

---

\(^{19}\) In the Historisches Wörterbuch der Philosophie (598) the 18th -century philosopher Ch. A. Crucius aptly describes a relation as 'eine solche Art zu existieren zwischen zweyen oder mehrern Dingen, wodurch es möglich wird, daß man von ihnen zugleich etwas abstrahiren kann, was sich von einem alleine nicht hätte abstrahiren lassen.' 'Two or more' may not apply to reflexive relations, where something is related to itself. Note, however, that such relations are n-adic properties al the same, if we can say that the ordered pair contains one item twice.

\(^{20}\) See Armstrong 1997, 87.

\(^{21}\) This is my own classification. The reader may compare the above, in the surprisingly scant literature on this subject, with Hume's classification of 'philosophical' relations, in his 1788, 14-5, as resemblance, identity, spatiotemporal, quantitative, degree, contrariety and causation.
1. Particulars, properties, and relations

being a parasite, a symbiont, a predator, a mate, food, an eye, an alarm call, a fight, a sibling (biological domain); being a belief, a memory, a perception, a sensation, an utterance (psychological domain); being a sign, a judge, a liar, a guardian, a cheque, a hotel, an enemy, a husband, a neighbour (social domain). And it is not difficult to see that virtually all such 'higher-level' properties involve relations.

I will not attempt to classify the types of relations in the domains of the biological, psychological and social, although a perspicuous classification might be feasible. I will later on, however, argue that most, if not all, of the relations from these domains supervene on relations of the few kinds in the merely physical domain - especially causal ones.

There is another important distinction among types of relations. This distinction is important because it teaches us which relations have genuine ontological autonomy, and which do not. It is made by Armstrong (1978, 85):

(1) Two or more particulars are internally related if and only if there exist properties of the particulars which logically necessitate that the relation holds. (2) Two or more particulars are externally related if and only if there are no properties of the particulars which logically necessitate that the relation (... ) holds.

It is important to see how this distinction relates to our classification above. First, it is arguable that from the merely physical domain only the causal and spatiotemporal relations are external. Resemblance, covariation and difference are obviously internal, as are quantitative relations like being bigger, softer, more transparent, etc. -than. The co-existence and partitive relations, and identity-distinctness among particulars, are necessitated by the particulars’ spatiotemporal properties (the spatiotemporal situatedness of a particular bicycle wheel, for instance, and of a particular bicycle will ‘logically necessitate’ that these particulars co-exist, or are, respectively, part and whole), and so can be said to be internal.

Note that this latter claim rests on the assumption that two particulars cannot be in one place at the same time. Note also that if the 'properties' in Armstrong’s definition are supposed to be ‘intrinsic properties,’ it is no longer clear that the relations mentioned are internal. Note finally that the relata at issue may be concrete particulars, but also states of affairs.

---

22 A quantitative relation like going faster-than may not be internal, but arguably supervenes on spatial and temporal relations.
Part I. Ontology

The reason why, apart from spatiotemporal relations, it is only the causal relations that escape necessitation by other states of affairs is that, as we will see, causal relations depend on contingent natural law. The sun may shine and the snow below it may melt, but without the contingent laws of physics being in place we have just temporal succession, not causal relatedness. Yes, given the laws of nature causal relations are fixed by other states of affairs. But the laws are not themselves plausibly regarded as states of affairs (particular-property complexes at times, remember).

Causal relations, then, require natural law besides other states of affairs. Spatiotemporal relations are a special case: they are not necessitated by any property of, or further relation between, the relata, but are just caused. They obtain, then, in virtue of previous states of affairs in combination with natural law.

The other relations in the merely physical domain are all necessitated by further states of affairs. Of course, when a resembles b, that is a contingent state of affairs in the sense that a and b might have had other properties than they in fact have. Their causal histories might have been different. But with their actual properties fixed, there is no room left for a and b to fail to resemble. That relation does not require natural law in addition, as do causal relations.

Secondly, it is arguable that from the higher-level domains all (or in any case most) relations are external with regard to the level below it, involving as they do causal relations. Issuing a tax assessment, for instance, involves a causal relation among a person and a piece of office equipment, which in turn involves brain and muscle activity, which involves biochemical reactions. We see here types of causal relations stacked 'on top of' each other, each one being a necessary condition for the 'above' one.

This is an important point to appreciate, because arguably, external relations are an 'addition of being' (genuinely new states of affairs come to obtain when they hold) and causally efficacious, while internal ones are not. I will later on discuss the dependence of higher-level properties and relations on merely physical ones, as well as the issue of causal efficacy, in much more detail.

Note that the relation of necessitation itself does not fit the classification made above, even though unlike, e.g., the relation of entailment it holds among particular states of affairs (for instance, someone's writing

---

23 Or supervenience: see below
'Larry' necessitates her writing two times 'r' in succession; something's being red necessitates its being coloured, etc.). Now, we might just add necessitation to the list along with, for instance, causation; but I think that that would be a mistake. For we will see that if necessitation is noncausal, the item necessitating and the one necessitated are not distinct; so that, unlike the other relations (but like identity, and the non-relation of 'instantiation' among a particular and a property), necessitation is not a relation among ontologically distinct items at all.

1.7. RELATIONAL PROPERTIES

Apart from relations, we often speak of relational properties. Armstrong, for instance (1978, 78-80), makes this distinction in virtue of the fact that predicates that include reference to a particular ('revolves around the sun') are impure, i.e., not about a universal (he says that 'revolves around something' is). But if we merely hold that properties must be abstract, and not that they must be universal, it seems that impurity simply is not an issue. Also, determinable properties such as revolving around something are merely less specific than, and not really distinct from, determinate properties such as revolving around the sun. Even Armstrong himself claims that relational properties reduce to relations (78-9).

In spite of this, a good reason to maintain talk of relational properties is that there is not a one-to-one correspondence of relational properties and relations. Jan's relational property of seeing Jon and Jon's relational property of being seen by Jan both correspond - and reduce - to the one relation of seeing by the ordered pair <Jan, John>. There are also relational properties, such as being a police officer, that involve an entire network of relations. Jack's being a police officer implies that Jack is related to many people, institutions, equipment and past events in various ways. Note, however, that the relational property may still reduce to the relations: if we take the trouble to specify Jack's relations, then we are in effect saying that Jack is a police officer, even if we do not use that term. 'Is a police officer' is merely abbreviatory for: 'is related to a, b, ..., n in such-and-such ways' (at least, if this relational property does not involve intrinsic properties too).

\[24\] And, perhaps, some intrinsic properties in addition.
Part I. Ontology

I am aware that my use of the term 'relational property' is different from that of a tradition, allegedly involving Leibniz and Hegel, that Russell (e.g., 1959, 42) has called the 'doctrine of internal relations.' According to this tradition, relations among entities are in fact intrinsic properties had by the relata. Relational properties are then taken to be a variety of intrinsic properties. For reasons that fall well outside the scope of the present discussion, this view has been criticized mainly by Russell, whose Principles of Mathematics contains an argument to the effect that relations cannot be reduced to intrinsic properties of the relata. Russell's arguments were directed against a 'monistic' body of metaphysics: monistic in the sense that it seeks to establish that reality is 'one' rather than 'many.' Now, insofar as we are not monists in that sense, taking it as given that reality is 'many,' i.e., that there exist many entities among which there are relations, we will tend to regard the concept of an intrinsic relational property as unintelligible, even contradictory. I will myself assume that there is a close conceptual link between relations and relational properties, and not use the latter term in its Leibnizian-Hegelian sense.

1.8. RELATIONISM

Although it seems that most philosophers nowadays will agree that there are relations, much of present-day metaphysics is, as we may call it, intrinsicalist. By intrinsicalism I mean the assumption that intrinsic properties, rather than relations, are of prior metaphysical importance. Roughly (but only roughly), intrinsic properties are properties that particulars can be said to have in and of themselves, independently of their past or circumstances. Considerable effort has been spent on finding a good definition of intrinsicness. But I will argue that the metaphysical privilege of intrinsic properties is not to be taken for granted, and that relations are at least as real and important as intrinsic properties. Let us call the latter position 'relationism.'

Outside of metaphysics, the way we think seems in any case largely relationist: most of our predicates are, tacitly or explicitly, relational. Let us start by looking at artistic practice: one of the first things we learn as musicians, poets, dancers or visual artists is that individual notes, words, move-

---

1. Particulars, properties, and relations

ments, or streaks of paint are hardly important: it is the relationships among them that count (this is in line with mainstream theory on visual perception: what we see is contrasts and Gestalts, not sums of perceptual atoms).

Of course, an artist may work on the intrinsic properties of an object. However, such an object (say, a canvass with paint on it) counts for nothing if it is not somehow related to a public. A Renoir's intrinsic properties are required for its being the painting that it is, but these are by no means sufficient. A piece of art is an in-and-out relational entity.

Now, this is a point about works of art, and art is a social phenomenon. And social reality is all about relations. But what about merely physical kinds of particulars such as stones? Are these not what they are merely because of their internal structure and shape, their intrinsic properties? Of course, the notion of internal structure presupposes a network of relations among an object's parts. But this does not make an object's internal structure a relational property of that object.

It can be argued, however, that even a stone acquires autonomous reality mainly in virtue of its relational properties, rather than its intrinsic ones. For instance, while we see a stone lying at our feet as an autonomous particular, the countless intrinsically indistinguishable other 'stones' that are part of the rock under our feet do not very much count as individual entities to us at all. What gives a stone its ontological autonomy, it seems, are at least as much its relational properties (like its casting its own shadows, its being freely movable, etc.) as its intrinsic ones.

True, any particular does have intrinsic properties, and these may often, if not always, be required for making the particular the individual, autonomous entity that it is. But they will typically be insufficient for doing so. Also, with the possible exception of intrinsic properties of elementary particles, intrinsic properties are realized by a particular's microstructure. But this means that they are realized by, i.e., consist in, relations among the particular's parts.

On this picture it no longer looks as if a particular's internal structure is the real intrinsic base of its properties, which may or may not be described in relational terms. The picture is rather that of a particular's internal make-up (which is intrinsic with regard to the whole particular while being relational with regard to its parts) that is a requirement for further, genuinely relational properties of the particular. Now, this is not a minor issue: the relationist outlook presented here will later on suggest an alternative to
the standard accounts of intentional subjects (which are intrinsicalist), which are so inadequate in accounting for mental causation.

We might try to push relationism to its limits by saying that relationality is basic, while intrinsicness is merely derivative; or by saying that object-intrinsic states of affairs reduce to relational states of affairs. It seems, however, that dispositional properties (like fluidity, transparency, fitness, amiability, etc.) are not plausibly construed as relational. True, dispositional predicates suggest (causal) relationality, but the relations are merely possible, not actual. Also, a particular has its dispositional properties in virtue of the laws of nature, and these may well resist a relationist analysis.

An all-out relationism may also face difficulties with elementary particles, as Armstrong (1997, 85-7) warns us. Armstrong points out that if such 'basic particulars' do not have intrinsic properties, the 'metaphysical possibility' of these particulars' lacking any relational properties leads, implausibly, to the possibility of featureless particulars. On the other hand, we may well wonder whether unrelated particles, as modern physics describes them, are indeed conceivable. In any case, the point that I have tried to make above is just that we should take relations seriously as we find them, rather than declaring them ontologically inferior to intrinsic properties.
2

Physicalism

2.1. A NEGATIVE ACCOUNT OF ‘PHYSICAL’

With the ontological picture of the previous chapter in mind, I will now sketch the picture of psychophysical dependence that will be the basis of my further arguments. This picture is physicalistic, but ‘physicalism’ is a term that needs to be given a clear sense before being used. This I will try to do below.

In contemporary philosophy of mind it is generally assumed that all minded particulars, such as human beings, are physical (i.e., are of a physical kind, or have physical properties). Unfortunately, it is by no means clear what ‘physical’ is supposed to mean.26 ‘Mentioned in contemporary physical theory’ is clearly inadequate as a definition: it refers to time and culture, which makes it useless for metaphysics. ‘Constitutive of a basic law of nature’ is slightly better, but it leaves out inaccurate but intuitively physical concepts such as thick, hard, heavy, sticky, etc. It can even be held that there is not very much recognizably ‘physical’ about entities like electrons or neutrinos, or properties like the ‘flavour’ or ‘colour’ of quarks. Needless to say that a notion of ‘physical’ that depends on a Cartesian mind-matter dichotomy will not help us very much either.

In other words, there does not seem to be a consensus definition of ‘physical’ available on which we can rely; and I do not consider this the right place to mount some proposal for a definition. It can even be doubted that the project of trying to define ‘physical’ will ever succeed, for no clear criteria of success seem available.

But does this mean that a clear understanding of ‘physicalism’ is out of our reach? I think not. For the intuition behind physicalism, or materialism, seems quite clear, and does not seem to depend on any sophisticated account, let alone definition, of what it is to be physical. When it comes to it, physicalism is just the view that life, mind and society have arisen in a lifeless, mindless and nonsocial world, depend for their reality on the basic laws of that world (which are not recognizably biological, psychological or

26 See, for instance, Crane and Mellor 1990.
social) and in fact are merely special ways in which the stuff of that world behaves.

What this suggests is that by default we can call the lifeless, mindless and nonsocial the physical, or better: merely physical. The idea is that physicalism relies on a commonsense conception of 'dead matter.' This is a tacitly negative understanding of 'physical': not very much metaphysically interesting, and certainly too poor to count as a definition of 'physical.' On the other hand, insofar as we can come by an account of what is characteristic to the living, the mental and the social (not an easy affair, it seems, but more tractable than defining 'physical' from scratch), it will not be completely uninformative either.

More importantly, it seems precisely good enough to articulate the core idea of physicalism. We can imagine several parties, each with their own understanding of 'physical,' but all in agreement on one point: life, mind and society do not require an irreducibly or sui generis 'vital,' 'psychic' or 'societal' addition to the 'physical' (whatever that is). Also, a physical domain that is demarcated negatively can easily be seen to comprise, apart from properties like charge or spin, properties like weight, length, shape, transparency, treacliness, etc.; as well as chemical properties like being a catalyst or being a corrosive substance. This, it seems, articulates fairly precisely the notion of 'physical' as we find it in the discussions on the metaphysics of mind.

2.2. SUPERVENIENCE AND PHYSICAL REALIZATION

Let us now try to spell out what physicalism amounts to in a little more detail. (For convenience, let us here concentrate on physicalism about the mental, as a stand-in for the 'higher-level' in general.) We saw that physicalism is the denial of substance dualism: it holds that all particulars are physical. But many physicalists, namely the ones committed to property dualism (or better, pluralism), do not believe that all particulars are merely physical. For some particulars have mental properties.

But why should such a position be considered physicalistic? It seems that allowing nonphysical properties (though not particulars) is just as much a violation of physicalism as substance dualism. Is not physicalism just the claim that everything is physical? Why make an exception for properties? On the other hand, it is either just hollow to proclaim mental properties
2. Physicalism

physical properties or, namely on the negative account of 'physical' outlined above, it amounts to an outright denial of their reality.

We will make some progress by noting that physicalism regards merely physical properties as basic with regard to other kinds of properties, such as mental ones, in the sense that the latter depend on merely physical properties. This dependence is, in fact, the notion of supervenience that we met in the Introduction.

Let us now have a closer look at this notion. With regard to mental states of affairs we already saw that supervenience means that 1) some set of physical states of affairs in a world is required for any set of mental states of affairs in that world, and 2) given the physical states of affairs of a world with mental states of affairs, there could not possibly fail to be the mental states of affairs of that world. As Davidson (1995, 7) says:

The definition of supervenience implies that a change in mental properties is always accompanied by a change in physical properties, but it does not imply that the same physical properties change with the same mental properties.

The notion of supervenience is rather sophisticated, and Kim (1984b) has distinguished weak, strong, and global supervenience. If we understand 'covariation' as a one-way, mental-physical relation, then, roughly, weak supervenience is mere de facto but not necessary covariation of physical and mental properties; global supervenience is covariation of the undifferentiated totality of physical properties in a world with specific mental properties; and strong supervenience is necessary covariation of specific mental and specific physical properties.

Kim rightly rejects weak and global supervenience as being too weak to express psychophysical dependence. Obviously, mere de facto covariation is not dependence. And as for global supervenience, it cannot distinguish bizarre from ordinary kinds of psychophysical correlations. Intuitively, the processes in our brains make a difference to mental properties while many other physical states of affairs, such as the existence of an ammonia molecule in one of Saturn's rings, do not (see Kim 1987, 85); but on a thesis of global supervenience we cannot make such a distinction.

If this is right, a physicalistic mind-body theory needs at least the notion of strong supervenience. Kim has argued, however, that the mere assertion of mind-body supervenience is not yet a mind-body theory.27 After all, such

---

Part I. Ontology

an assertion is compatible with various mind-body theories, including Cartesian dualism, or the bizarre theory, put forward by Malebranche in the 17th century, that mind and body do not interact but are like two clocks kept synchronous by God. Such theories may well (although they need not) assert mind-body covariation in a strong sense, so that the demands of supervenience theory are satisfied. Yet, they are clearly not physicalistic.

In order to articulate physicalism, then, psychophysical supervenience should not just be asserted, but also explicated in a physicalistic way. Physicalism will only be an intelligible position if some sort of natural connection between physical and mental reality can be specified. Now, such a natural connection can perhaps be found in the notion of physical realization. As physicalists we might say that mental states do not just supervene on, but are also realized by physical states of affairs. But of course we have to explain what physical realization amounts to.

A notion of physical realization will articulate the fact that if a realizes b, then b is not something distinct or separate from, or something over and above a. Why? The answer is quite simple: anything short of this will not explain the supervenience of b on a.28

But a phrase such as 'nothing over and above' needs careful explication. A reductionistic theory such as Kim's, for instance, will take 'nothing over and above' to mean 'nothing but,' or 'consisting in.' It will tell us that physical realization of mental properties amounts to a de facto identity between any mental state and a physical one. Yet, such identities have to be established rather than presupposed, and it is doubtful, to say the least, that they can. If so, however, we should resist a 'nothing but' interpretation of 'nothing over and above,' and ask whether we can help ourselves to a nonreductionistic, non-identity reading.

Now, it seems that short of 'nothing but' or 'consisting in,' 'nothing over and above' can also be understood as 'given along with,' or 'already implied by.' Arguably, these latter expressions are weaker than 'nothing but.' For there is a difference between claiming b to be nothing over and above a (or to be given along with a), and claiming that b is identical with a. For a might be more than b, in the sense that b could either be a part or an aspect of a. A leg, for instance, is not something over and above the table that it is a part of; yet, it is not identical with that table. In the same way, the table's

28 We will see later on that notions like emergence, or even causation, are inadequate on this point.
2. Physicalism

woodenness arguably is not something over and above the table as a concrete entity (it merely is an aspect of that entity) but, again, it is not identical with it.

This difference between non-distinctness and identity is quite important. It will allow the nonreductive physicalist to avoid reductionism as well as dualism by saying that mind is an aspect of physical reality (and mental properties aspects of specific parts of physical reality): nothing in addition, but also not the very same thing. How the mind can be like this, of course, needs further spelling out.

2.3. THE FREE LUNCH PRINCIPLE

If the above is right, physicalism is the claim that biological, mental and social properties depend in a strong sense on properties that are not biological, etc. themselves. 'In a strong sense' means that biological, mental and social properties are physically realized, i.e., do not merely covary with, but to begin with are not something over and above the nonbiological, nonmental and nonsocial properties.

We should now ask, however, what justifies the 'nothing over and above' clause. If we say that our world contains physical states of affairs and biological and mental states of affairs, and especially if we are suspicious of the identity claims of reductionism, on what grounds could we still claim that life and mind are not really 'added' to the nonmental and nonbiological reality on which they supervene?

Let us first consider the claim a little more closely. Stalnaker (1996, 222) writes:

A supervenience thesis, on one way of understanding the notion, is a reductionist thesis. To say that the A-properties or facts are supervenient on the B-properties or facts is to say that the A-facts are, in a sense, redundant (…). A-facts are not facts 'over and above' the B-facts, not something 'separate.' To state an A-fact, or ascribe an A-property, is to describe the same reality in a different way, at a different level of abstraction, by carving the same world at different joints.

In the same vein, Armstrong claims (1997, 12):

What supervenes is no addition of being (…). The terminology of 'nothing over and above' seems appropriate to the supervenient. (p. 12).
The idea is that supervenient properties are real, but in an almost trivial sense: they are already part of the reality that they supervene on. Armstrong, stressing the fact that supervenient items are 'for free' for the ontologist, calls it the 'doctrine of the ontological free lunch.' Below, I will refer to this doctrine as the 'free lunch principle.' I would myself like to formulate it as follows: if \( p \) and \( q \) are states of affairs, then:

*The free lunch principle:* If \( p \) noncausally necessitates \( q \) (e.g., if \( q \) supervenes on \( p \)), then \( q \) is nothing over and above \( p \).

Note that the principle treats supervenience as a species of noncausal necessitation (if we believe that causation is nondeterministic, we can drop the 'noncausally' clause in the doctrine). And if supervenience is indeed a relation of mere necessary covariation, or counterfactual dependence, we may even identify it with noncausal necessitation. We may then, however, use the term 'supervenience' in an idiosyncratic way, so I will just treat supervenience as a species of noncausal determination, as is usual.

Now, is the free lunch principle plausible? Consider first a case in which it appears self-evident, for instance there being two apples in a bowl along with three additional apples; and there being five apples in that bowl. These states of affairs noncausally necessitate each other, and our principle will have us echo intuition in saying that they are actually the very same state of affairs.

Consider next the state of affairs of something's being a hydrogen atom, which supervenes on a number of sub-atomic states of affairs. Given these sub-atomic states of affairs (assuming, for now, that they do not include context along with 'internal' states), something's being a hydrogen atom is not only guaranteed, but it seems also the case that the sub-atomic states of affairs are all there is to that state of affairs. Nothing needs to be added, it seems, and describing the sub-atomic states of affairs is already describing something's being a hydrogen atom.

Cases like these, however, involve supervenience only because they are in fact about identities. But what about cases where supervenience base and supervenient state of affairs are not identical? Consider Jack's room being cluttered at \( t \). It supervenes on a large number of states of affairs concerning the objects in the room, but it is not identical with these (they may also subserve homeliness, for instance). Now, is the room's being cluttered something over and above the totality of the states of affairs that
2. Physicalism

It supervenes on? It seems not. Nothing has to be added, the cluttered state is ‘automatically’ given along with the rest, it is ontologically ‘for free.’ Jack could not have had his room in the same state but *minus* its being cluttered.

I admit that cases like these are at best suggestive. It is always open to the nonreductionist to argue that since there is no identity, supervenient states of affairs *are* something over and above, i.e., a genuine addition to, physical ones. There may not be an addition of matter and force, or of particulars; but one might think that ‘addition of being’ may also concern mere properties. At this point we might say that in nonreductive physicalism the distinction between ‘nothing over and above’ and ‘something over and above’ is not a clear one to begin with, so that we had best stop the discussion. As long as we are clear what supervenient states of affairs are (possibly not identical with the states of affairs they supervene on, but ontologically ‘for free’), there will be not much harm in this relaxed attitude.

I will myself, however, favour the ontologically sparing ‘nothing over and above’ over the more liberal ‘something over and above.’ The reason is that unlike causal dependence, supervenience seems to be a deterministic relation; and not just by definition. We do have *prima facie* examples of irremediably chancy causation (such as that of the terrorist who wires a bomb to a geiger counter and just one atom of a radioactive element), not of chancy supervenience. One might speak of a ‘chance’ that monetary value will supervene on some paper-ink structure, but the chanciness is easily eliminated by adding context. We might discover consciousness, unlike other supervenient properties, to have a mere ‘propensity’ to supervene, but this is a speculation that draws on our present ignorance rather than on actual evidence. Not so with chancy causation: eliminating chanciness here requires one to commit oneself to determinism, which would, in our days at least, be a daring and controversial choice.

Now, we should ask where the ‘determinism’ of supervenience comes from. How do we explain it? If we hold that supervenient properties are ‘something over and above’ their physical base we might either say that supervenience is a ‘brute’ phenomenon, or try to explain it by means of deterministic ‘laws’ of supervenience. I consider both options unattractive: the determination of a state *b* by a state *a* is not something credibly left

---

29 A diachronic account may also help: Were not, during the course of natural history, new life forms an addition of being?
Part I. Ontology

unexplained, and invoking laws introduces a questionable new piece of metaphysics (the laws would have to be very different in character from the laws of physics). All becomes much easier if we embrace the ‘nothing over and above’ interpretation: supervenience ‘determinism’ is then explained by the fact that supervenience base and supervenient state of affairs are not distinct items. The latter can then be understood as a set of physical entities’ not just being so-and-so, but also such-and-such.

2.4. ARGUMENTS FOR PHYSICALISM

Understood in the way outlined above, is psychophysical physicalism a plausible position? It is one thing to reject Cartesian dualism; it is quite another thing to claim that mental properties supervene on physical properties, if supervenience means noncausal determination, and if noncausal determination means that physical and mental reality are not even distinct. What could justify such a strong claim?

Not, it seems, systematic experiments. First, it is practically impossible to set up a test situation with duplicate working brains, or brain-environment systems. Secondly, it is unclear what would or would not count as a ‘duplicate.’ Thirdly, it would anyway be a precarious affair to determine which mental properties the duplicates do and do not share. And fourthly, of course, a sceptic might think that tests could only show de facto psychophysical covariation, but not dependence, let alone non-distinguishedness.

Short of convincing experiments, however, let me give three considerations in favour of physicalism (as here analyzed) about the mental. First, there is quite some empirical evidence of psychophysical supervenience. Lesions in the brain yield mental deficiencies that are by no means random; there is no mental capacity that could not be influenced by tampering with the brain; and none that could conceivably survive the brain’s destruction. Facts like these show that a physical base is in any case a necessary condition for mental properties. True, they do not show that there are no disembodied minds, but given the close conceptual connection between mind and action a minded being without a body is hardly conceivable anyway.

Empirical evidence that certain physical states of affairs are sufficient conditions for mental states of affairs is harder to come by, for a sceptic might maintain that, although we have never seen such a thing, tomorrow
2. Physicalism

a human being in normal conditions with an intact brain might fail to have a mind (or specific mental properties). But the burden of proof is here with the sceptic, who has only his scepticism to go by, and not with the physicalist, who has his vast body of evidence.

Secondly, physicalism is the only position that makes sense in the light of our (admittedly limited) knowledge about the origins of minded beings like us. Our mental capacities must have evolved gradually, in a world that was, as far as we can tell, not minded to begin with. And mental capacities can only have been selected for on the basis of their physically realized effects. Also, each individual minded creature has started out in the womb, it seems, as an entity having only physical and biochemical properties; and no particular point in its development stands out as the addition of that special non-physical extra. Mind seems to emerge gradually, and as an effect of physical causes. It does not come to us out of the blue. This suggests, although it does not prove, that physical reality is the arena where the whole thing takes place.

Thirdly, there is a more principled argument for physicalism. It is based on the (virtually undeniable) fact that a thing like physical science is possible in our world. Note that physicalism is of one piece with a view we have already encountered, namely Kim's doctrine of the causal closure of the physical domain. This doctrine says that any physical event has a complete cause that itself wholly consists of physical events. True, this is yet another doctrine, but it amounts to the plausible claim that physicists, in order to causally explain no matter which physical fact, need never invoke independent psychic or vital forces.

In fact, the causal closure doctrine just asserts the very possibility of physical science as we know it: an exact science with basic laws cast in mathematics (probabilistic or otherwise). But if physical science can get a grip on a world containing biological, mental and social states of affairs, such states of affairs must either be illusory, causally inert (which pretty much comes to the same thing), or physically realized. In all cases physicalism wins. A stark conclusion like this is resisted even by many scientists, but causal closure implies that believing, say, the law of conservation of energy to apply to our life- and mind-rich world, inexorably commits one to physicalism.
3
A layered world

3.1. LEVELS OF ORGANIZATION

Let us now see in more detail how the world looks from a physicalistic point of view. By far the most popular elaboration of physicalism is the layered-world view. Kim (1998a, 15-6) lucidly describes this view as

the familiar multilayered model that views the world as stratified into different “levels”, “orders”, or “tiers” organized into a hierarchical structure. The bottom level is usually thought to consist of elementary particles, or whatever our best physics is going to tell us are the basic bits of matter out of which all material things are composed. As we go up the ladder, we successively encounter atoms, molecules, cells, larger living organisms, and so on. The ordering relation that generates the hierarchical structure is the mereological (part-whole) relation (...). It is part of this layered picture that at each level there are properties, activities, and functions that make their first appearance, or “emerge,” at that level (...). Sometimes the layered model is couched in terms of concepts and languages rather than entities and their properties. Talk of levels of organization, description or languages, of analysis, of explanation, and the like is encountered everywhere - it has thoroughly permeated primary scientific literature in many fields (...) as well as philosophical writings about science.

According to this model, nature is such that apart from lawful regularities governing micro-entities (such as quarks and neutrinos), there are other regularities and patterns on different, relative macro-scales (such as that of cells or organisms). The vocabulary of ‘levels’ is metaphorical in its distinction between ‘high’ and ‘low,’ but the metaphor is harmless as long as we realize that lower-level corresponds with less complex, while higher-level with more complex (I skip the question of how we are supposed to measure complexity). Note that higher-level and lower-level is not just a matter of micro and macro, or mere scale: stars, for instance, are lower-level compared to bacteria, because they lack the complex kind of organization of the latter.

Also, the term 'levels' suggests a world built from multiple domains, one 'on top of' the other. But something can only be on top of something else when it is distinct from what it is on top of (something cannot be on top of
3. A layered world

Itself). Therefore, given the account of physicalism that avails itself of the free lunch principle, we should beware of taking the term ‘levels’ too literally.

Note also that there is no suggestion that there is always a crisp distinction between the levels (should we locate a bee colony on the level of organisms or societies?), or that they have to be causally insulated from each other. True, the doctrine of causal closure precludes ‘downward causation,’ the perturbation of causal processes on the physical and, presumably, lower levels in general, by higher-level processes. But organisms, for instance, do interact with photons, chemicals, viruses, tables and chairs and societies just as well as with other organisms. The hierarchy of the layered model, then, is by no means very strict.

There are various accounts of the relations between different levels in this layered-world picture. We have already encountered the notion of supervenience, but we will also have a look at, in turn, emergence, upward causation, and type identity. I will argue that none of these is any good in spelling out physicalism, and that ‘physical realization’ is the only appropriate term in this context. My argument against type identity theory will consist of a discussion of, in turn, the phenomenon of multiple realizability, and the physical realization of relational states of affairs.

A position that I will not discuss in this chapter is token physicalism. In general, token physicalism is just the assertion of token identity of particulars of all types with physical particulars. In that sense I have earlier declared myself a token physicalist. In the philosophy of causality and of mind, however, token physicalism is also the claim that events or states of all types are token identical with physical events or states; especially mental events and states. This claim is far less innocent than the former one, and I will later on argue that token physicalism in this sense is seriously misguided.

3.2. Emergence

As Kim (1993d) points out, the notion of emergence, developed by Broad, Morgan and Alexander in the first half of the 20th century, is similar to our contemporary notion of supervenience in various respects. The emer-
gentists held, against vitalism and Cartesianism, that life and mind were not separate factors, or entities, that were added to physical reality in the way of vital spirits or Cartesian minds, but that they somehow arose out of complex configurations of matter alone.

Yet, being nonreductionists, they also held that life and mind were more than mere physical phenomena. In their view, emergent phenomena were 'unpredictable,' or 'new' relative to the physical phenomena that gave rise to them. Also, emergent phenomena were said to bring along their own causal powers. Sometimes emergence was seen as a diachronic process (such as the emergence of consciousness in the course of evolution); but it was also seen as something synchronous, just like supervenience.

There are various problems with emergentism in this classical form. First, the fact that emergent properties are 'unpredictable' might mean that emergence is a matter of ignorance on our part, which disqualifies it as a relation with metaphysical significance (see Stephan 1992). But if 'unpredictable' means that properties do not emerge as a matter of necessity, law, or regularity, emergentism can no longer be regarded as an account of higher-lower level dependence. In both cases the concept of emergence seems of little metaphysical use.

Secondly, the doctrine of new causal powers can easily be interpreted as entailing downward causation, i.e., violation of causal closure. Finally, the term 'emergence' seems more descriptive than explanatory. This is perhaps not a criticism that we should level at the classic emergentists: these philosophers were interested in articulating a middle way between materialism and various forms of dualism, not in explaining the fact of emergence itself. But in the present context this is what we want an account of. And here we see that the concept of emergence is in the same boat with that of supervenience: it is explanandum rather than explanans.

It is true that the concept of emergence is quite popular among contemporary authors. These, however, tend to apply it in a metaphysically innocuous sense. Searle, for instance, regards liquidity as an emergent property (see below) while at the same time holding that it can be fully explained in physical terms. Indeed, most contemporary authors who use the term seem to have dropped the element of essential unpredictability. Looijen (1998) sees no conflict at all between emergence and reducibility, and an author like Holland (1998) even attempts to mathematize the concept.

Emergence is also a phenomenon actively pursued by cognitive scientists and roboticists: the concept plays an important role in debates on arti-
ficial life (see, for instance, Boden 1996). But here, emergence seems to be mainly a methodological concept, not a metaphysical one. This can be appreciated when we realize that the concept only makes sense in a framework that already takes physical realization for granted. If a property is said to be emergent, this is a matter of how one has to go about explaining it (see, for instance, the distinction between 'componential' and 'emergent' explanation in Clark 1997, Ch. 6). The discussion here concerns explanatory strategy rather than metaphysics.

If we prefer the contemporary notion of emergence to the classical and metaphysically pretentious one, we will see emergence mainly as the dependence of system properties on the dynamic relationships among the parts of that system. The properties typically called 'emergent' are dynamic system properties, from the way in which the shape of a wet-in-wet water-colour blot develops, or the behaviour of the economy, to consciousness. They are properties that result from the complex interplay of system parts, sometimes combined with the ongoing adaptive interactions of the system with its environment. Also, what is said to be emergent may not just be a momentary trait or process, such as a vortex, but a regularity, or pattern, such as weather cycles.

The term 'emergence,' then, is at home in the context of the explanation of system properties, but for metaphysical purposes it offers nothing in addition to 'physical realization.' It is on the basis of the latter metaphysical concept that the explanatory concept of emergence is significant.

3.3. UPWARD CAUSATION

The claim that the relation between higher- and lower-level phenomena is causal is rather exceptional, and probably the only philosopher who explicitly endorses it is Searle. As we have already seen, Searle claims that higher-level properties are emergent. He claims this of consciousness, but also of non-biological or -psychological properties like solidity or liquidity. But the remarkable thing about his view is that he conceives of emergence as a causal relation. He writes (1995b, 218):

Part I. Ontology

The solidity of the ice is causally explained by the behavior of the molecules (crudely, it's because the molecules make vibratory motions in lattice structures). Now consider water in a liquid state. The liquidity of the water is causally explained by the behavior of the molecules (crudely, it's because the molecules are rolling over each other in a random fashion). Neither causal relation involves a time gap nor a mechanism. In fact, the solidity and liquidity are just features of the system made up of the molecules. Liquidity and solidity are caused by the behavior of micro-elements while realized in the system made up of these micro-elements.

And so he explicitly claims that there can be causal relations 'between phenomena at different levels in the very same underlying stuff' (1983, 266), and also that 'brains cause minds' (1984, 20).

This position seems to explain how it can be that there is a natural connection between, for instance, physical and mental phenomena, that the mental phenomena do not reduce to the physical ones (effects do not reduce to their causes), and that nevertheless there is no hint of Cartesian interactionism (the causal relations are 'in the very same underlying stuff').

But is it plausible? Just to concentrate on the mind, there are obvious causal relations between brain states and mental states. Damage the brain, activate brain regions with electrodes, change brain activity with chemicals, and there will be specific mental consequences.

Another respect in which psychophysical 'emergence' (I will for the moment go along with Searle and say 'emergence' rather than 'physical realization') is like causation is the fact that it is a relation of asymmetrical dependence. Both kinds of relations can be expressed by counterfactuals, and neither kind of dependence is symmetrical. Perhaps a state of affairs can be said to supervene on itself, but to say that a state of affairs is emergent relative to itself seems as nonsensical as to say that a state of affairs has caused itself. It is true that Searle believes in downward causation (1983, Ch. 10 and 1995b); but this does not make his emergence-causation a symmetrical relation, for in his view emergence is simultaneous, while downward causation is said to work 'across time' (1995b, 219).

Yet, I will now argue that Searle's causal emergentism is at odds with a few central intuitions about causality. True, he complains against his critics (such as Kim 1995b) that they have a narrow view on this subject (1995b, 230), and one may indeed just choose to use the term 'causation' in a broad or a narrow sense. The problem is, however, that a broad conception such as Searle's obfuscates a distinction between two very different kinds of dependence relations. One might use one term to cover both, but because
the kinds of dependence are so different in character, that difference had best be taken account of in our terminology.

Calling emergence a causal relation, then, is at odds with at least two widespread and rather uncontroversial intuitions about causal relations. The intuitions are these: a) Cause and effect are related spatiotemporally in ways that are different from emergence; b) Causal asymmetry has a source that is different from the asymmetry of emergence. Let us consider both points in turn.

a) Causes and effects can very well be at the same place (an iron bar might break where it has rusted), and there does not seem to be a compelling reason for believing that immediate causation over a spatial void is inconceivable. Although it seems natural to say that cause and effect are spatiotemporally contiguous, or at least so connected indirectly, our ordinary concept of causation may be more liberal about the spatiotemporal issue than has been suggested for a long time by Humean theory.

But now consider emergence. This concept is very much committed on the spatiotemporal issue: emergence base and emergent state of affairs could neither be contiguous in space or time, nor be contingently situated relative to each other. It is flat inconceivable that emergence base and emergent state of affairs might fail to coincide spatiotemporally. What would 'emergence at a distance,' or 'contiguous emergence' mean? It is plain nonsense to say that some physical processes in my body are here (or now) while, say, the 'emergent' digestion occurs there (or later).

b) Arguably, both emergence and causation are asymmetrical relations of dependence; but when we ask where the asymmetry comes from, we find very different answers in both cases. It may be thought that the asymmetry of emergence is a corollary of the fact that emergent properties will typically be multiply realizable. But if this were so, the asymmetry would not hold good for particular emergent states of affairs, while it definitely seems that any such state of affairs does asymmetrically depend on its physical base.

Then, what does account for the asymmetry of the emergence relation? Kim (1990b, 149-60) has suggested that it is reducibility; in the same vein but less radically, we might say that it derives from the free lunch principle. We naturally say that the fluidity of this water is 'nothing over and above' the collective behaviour of its molecules (the latter noncausally necessitating the former); but we do not say that such collective behaviour is nothing over and above its fluidity (the fluidity hardly necessitates anything on the
Part I. Ontology

micro-level). And it seems less bizarre to say that Jack's mental state is nothing over and above some physical state than to say that this physical state is nothing over and above his mental state.

Where does this difference come from? The answer is that, first, lower-level properties are more general than higher-level properties, and characterize a larger range of entities. Any particular in the world has physical properties, but only a very small subclass of particulars also has mental properties. Secondly, there can exist a physical reality without mental properties, but not, arguably, a mental reality without physical properties. Thirdly, we have seen that a physical base is 'more' than the properties supervening on it, the latter being aspects of this base. These are the real sources of asymmetry in the case of physical realization, or emergence.

How are things with causal asymmetry? It is hard to say, for causal asymmetry is a notoriously difficult issue. But this much is sure: whatever causal asymmetry depends on, it depends in no way on something like reducibility, non-distinctness, generality, or being-more-than, in the way we saw was the case with emergence, or physical realization. Saying that effects are nothing over and above their causes, that causes are 'more general' than their effects, or that effects are 'less than,' i.e., 'aspects of' their causes seems nonsensical.

We might mount further objections to Searle's 'causal' conception of emergence. Causal influence but not emergence seems to depend on some kind of physical process; and causation seems to require contingent, maybe even probabilistic, laws of nature; while emergence seems to be a relation of necessitation, not contingent or probabilistic at all. But these points are more controversial than the ones discussed above and, not having said much about causation yet, I will not here rely on them.

3.4. MULTIPLE REALIZABILITY

If it is uninformative to say that higher-level states of affairs emerge from lower-level ones, and if it is false to say that higher-level states of affairs are caused by the lower-level ones they supervene on, could we not say that higher- and lower-level states of affairs are identical? After all, the free lunch principle tells us that higher-level states of affairs are nothing over and above their supervenience base. The rest of this chapter, however, can
3. A layered world

be read as an argument to the effect that identity theories are mistaken. Let us start by considering the phenomenon of multiple realizability.

In spite of the suggestion carried by the ‘nothing over and above’ of the free lunch principle, there is a class of supervenient properties that do seem to have some autonomy with regard to their supervenience bases: the so-called multiply realizable properties. Multiple realizability, especially of mental properties like being in pain or believing that p, is the classical stumbling-block for psychophysical reductionism (see, for instance, Fodor 1974 and Putnam 1968). Let us now take a closer look at this phenomenon.

Dispositional properties such as fitness, aggressiveness, toughness, fluidity, etc. are good examples of multiple realizability. Dispositional predicates, i.e., ones that say how a particular will be causally involved in specific ways under specific circumstances (given the laws of nature), are merely conditional. But dispositions, as the properties that make such predicates apply, are of course not merely conditional, for the simple reason that they are actually had by a particular. A window pane could break, but it actually is fragile. A disposition, then, has a categorical base, a determinate physical property that is actually had. Dispositions are determinable relative to such a base: we can always ask in what precise way some disposition is had. For any disposition there will then be various different possible determinate categorical bases, and that makes these properties multiply realizable. Fragility, for instance, can be realized by thin glass, but also by thin wood, ice, china, etc.

We might think that, although categorical and determinate properties can be had by indefinitely many particulars (multiply instantiable, we might say), these particulars must be very much alike to share such properties. Consider the property of being pure diamond. Physical reality seems to have only one way to realize this property, namely by a very specific configuration of carbon atoms. This does not seem to be the rule, however. Being square, having a velocity of 100 miles per hour, having a temperature of 100 degrees Celsius, having fever, being in pain, desiring that p, or being the right neighbour of Jones: all these properties can have physical supervenience bases that may diverge, a little or enormously.

But what is it that makes a property multiply realizable? This is hard to say. Indeterminateness counts, but determinate properties can be multiply realizable as well (compare sitting on something and sitting straight on the third chair to the right). Complexity also counts: while diamond, with its
simple molecular structure, is not multiply realizable, the more complex DNA is. On the other hand, being bigger than a cubic centimetre does not seem to be a complex property at all, while the variety of possible realizers is virtually unlimited.

What counts most, in my view, is relational character. Internal properties (such as being made of five planks) can be multiply realizable as well as relational properties (such as having five friends), but we may argue that the multiple realizability of internal properties generally depends on variability of internal structure, and that internal structure implies relations among parts. Perhaps, multiple realizability sometimes depends on neither relational character nor variability of internal structure. If there are elementary, i.e., structureless particulars that are different as a brute fact, one such particular may have $F$ in virtue of having $G$ and $H$, another may have $F$ in virtue of having $J$ and $K$, another in virtue of having $L$ and $M$, and so on. $F$ will then be multiply realized in virtue of neither relational character nor variability of internal structure. Because we lack any real examples, however, I do not consider this argument for 'basic' multiple realizability very impressive.

It thus may seem that relationality is the source of the multiple realizability, and therefore of the irreducibility, of higher-level properties. Whether or not this is really the case, of more interest is the following: as I will argue, relationality precludes reduction of higher-level properties to physical ones in other, probably more important ways, than by generating multiple realizability. And if that is right, antireductionist ontology is well-advised not to remain occupied with multiple realizability, but to rather exploit relationism.

### 3.5. RELATIONAL PROPERTIES AND INDIRECT SUPERVENIENCE

Let us, then, consider relationality as an obstacle to a reductionist, or type identity metaphysics, in asking how higher-level relational properties are physically realized.

As a start, let us note something peculiar about the multiple realizability of relational properties. The relational property of being a champion, for instance, can be said to be realizable in many ways (for many different people can be champions in many ways); but it can also be said to be realizable in not more than one way: namely by, say, having participated in
a contest and by having officially been recognized as the winner (call this property C). What should we say?

It is easy to see that being a champion is realized by C. The latter is an appropriate realizer (or supervenience base) of the former: if you do not have C you cannot be a champion, and if you do have C you cannot fail to be a champion. We might also say that being a champion consists in having C: in fact, one’s having C and one’s being a champion are identical states of affairs.

But from where, then, our intuition that being a champion is multiply realizable? Well, C can hardly be called a physical property (it is higher-level), and it is itself realized by yet other states of affairs; states that need not be homogeneous in different cases. In the case of Mary it is realized, for instance, by having jumped over poles, having been declared winner by Mr. Jones, having been given a silver cup, and so on (call it property D); while in the case of John it is realized by having played tennis, etc (call it E).

So, relative to C being a champion is not multiply realizable, while relative to D, E, etc. it is. But does someone’s being a champion then have more than one realizer at the same time? Not really, because C is not distinct from D, E, etc. And realization being a transitive relation, Mary’s being a champion, in being realized by C, is realized by D; and by everything that D in its turn is realized by, including some ultimate physical base (call it P).

We should note, however, that even though Mary’s being a champion can be said to be realized by P, the intermediate layers of C, D, etc. cannot be left out. Mary’s being a champion is realized by P because it is realized by C and C is realized by E and E is realized by ... (etc.) P. This fact can easily be overlooked, because if P is present, all the rest must also be present. Yet, it is in virtue of realizing E that P realizes C, and it is in virtue of realizing C that P realizes Mary’s being a champion. It is even incorrect, in a sense, to say that Mary is a champion because P obtains. Mary is a champion because she has won a contest (etc.), not because that specific configuration of atoms obtains (it might be any configuration, as long as some contest is won).

The important thing to note here is that the realization, or supervenience, of higher-level states of affairs is indirect. It is the fact that it makes no sense to say that P realizes one’s being a champion without first making it clear that P realizes D, C, etc. This is important, for on the above account, in order to find the physical realizer of a higher-level state of
Part I. Ontology

affairs, one is supposed not to skip levels, but to address all levels in turn. Disregarding the plurality of levels makes it impossible to explain why $P$, rather than some $P'$ or $P''$, should be the realizer of someone's being a champion. Acknowledging the various intermediate levels, however, has as a consequence that a realizer is seen to fan out at each level downwards, so that the ultimate physical realizer turns out wide, or nonlocal. Therefore, stressing the indirect character of physical realization is a way of arguing that the physical realization of higher-level properties is wide, not local.

3.6. RELATIONAL PROPERTIES AND NONLOCAL SUPERVENIENCE

Before seeing how the relationality of higher-level states of affairs precludes their reducibility to merely physical states of affairs, notwithstanding supervenience, let us make sure that we understand what such supervenience amounts to.

Above, when considering the layered-world view, we heard Kim say that '(t)he ordering relation that generates the hierarchical structure is the mereological (part-whole) relation.' In his hands this plausible remark becomes the basis of a principle governing the supervenience of higher-level states of affairs on lower-level ones, one that he himself endorses (1998a, 18):

Not surprisingly, supervenience theses (...) turn into claims of mereological supervenience, the doctrine that properties of wholes are fixed by the properties and relations that characterize their parts.

Now, this is certainly true for structural properties, which are intrinsic. But for relational properties, the doctrine of mereological supervenience is false: a person's being a police officer, for instance, is not at all 'fixed by the properties and relations that characterize [her] parts.' Clearly, and entirely in the spirit of intrinsicalism, the principle of mereological supervenience ignores the relations among these particulars and the rest of the world. Of course, this view can take these relations into account when it regards the 'wholes' as parts of even bigger wholes. But then what is at issue is the properties of these larger wholes, and no longer of the smaller ones. As a police officer one is, of course, part of a bigger whole (say, society). But being a police officer is a property of a person, not of a society. The thesis of
3. A layered world

mereological supervenience, then, tends to ignore such relational properties of particulars. It merely looks inside, not outside.

Relationism does look outside. It considers the relations of a particular to other particulars as just as real and important as the particulars' intrinsic, or structural-internal properties. Indeed, we will see later on that the wealth of 'new' or 'emergent' properties at each mereological level, the ones different in character from anything at a lower level, is constituted by such relations; that is, that the properties from the biological, psychological and social domain are arguably relational in character. It is, then, not a small mistake of the mereological supervenience doctrine to ignore particulars' relational properties.

Apart from an intrinsicalist orientation, Kim's motive behind defending a mereological conception of supervenience is undoubtedly the wish to keep supervenience local, non-diffuse, or narrow. We have earlier followed Kim in rejecting the notion of global supervenience, it being too weak to articulate the physicalist intuition of psychophysical dependence. But as relationists who reject mereological supervenience, we can still have more than global supervenience. For while intrinsicalism concentrates on local physical properties of objects, relationism inverts the entire picture, and lets us ask how relationships are physically realized.

The implications of both views are as follows. If a supervenience base must be admitted to be wide, intrinsicalism draws a boundary around the whole system, so that supervenience becomes global, or regional. The 'wide' property is then at least an intrinsic property of this large system. The price, however, is that supervenience becomes too weak to be interesting. But from a relationist standpoint we can see that supervenience can be as strong, and between as specific states of affairs, as one likes, without being local, i.e., confined to the boundaries of a particular. Supervenience can be strong and nonlocal at the same time, and nonlocal without being global. A simple example will illustrate this: Jake is tall 1 m 70, Jim is tall 1 m 80. So Jim is taller than Jake. Now, does the taller-than relation supervene locally? Not as far as either Jim and Jake taken separately are concerned. However, it does supervene strongly, and also nonglobally.

And when we realize that supervenience, or realization, is often indirect (see above), we can see that its point carries over to other relational properties. Consider, again, someone's being a champion: it supervenes strongly but nonlocally on that person's having participated in a contest and having officially been recognized as the winner. These relational states of affairs in
Part I. Ontology

their turn strongly but nonlocally supervene on further states of affairs (relational or otherwise); and so on, down to the domain of physical states of affairs. At no point in this cascade of levels, it seems, need we be vague about which states of affairs form the supervenience base, or realizer, of which higher-level state of affairs.\(^3\) The realizers will typically be wide and complex, but that does not mean that they are global or nonspecific. The interlevel dependencies are invariably strong, no matter how widely the realizers of many higher-level states of affairs will fan out.

3.7. AGAINST TYPE IDENTITY

If the above is right, strong supervenience of higher-level on lower-level properties can be nonlocal as well as nonglobal. This should still be acceptable to any physicalist, for supervenience of any higher-level state of affairs on physical states of affairs implies (at least on the free lunch principle) that there is nothing really ‘over and above’ physical reality. If physical reality is given, all else is given along with it.

The above will, however, be damaging to a reductionist metaphysics such as Kim’s, according to which (most notably) any mental state of affairs is identical with some physical state of affairs. This is the doctrine of type identity. A theory of type identity does not just claim that all particulars are physical particulars (which is token identity), but also that all properties, and therefore states of affairs, are physical (i.e., can be reduced to merely physical properties). It is well-known that the multiple realizability of higher-level states of affairs poses a considerable problem to such a theory.

3\(^2\) In a discussion, Marc Slors pointed out to me that vagueness and indeterminacy creep in when we ask what is the realizer of, say, the Dutch tax system (which arguably is a relatum involved in properties like being taxed). In my account, a physical realizer of a supervenient state of affairs is a sufficient condition for that state of affairs. Now, how many, and which, tax payers and officials are sufficient to realize the tax system at the moment that I am taxed? We may safely assume that a subclass of them would already make a fine tax system, and that many such subclasses can be demarcated. Which, then, is ‘the’ realizer? I would say here that we should always look at the actual, not at the minimally sufficient realizer. When I am taxed, I am related to all those involved in the tax system, not to a mere subclass of these. The property of being taxed is multiply realizable, it does not seem to require a fixed amount of relata. Note that allowing a more-than-minimally-sufficient realizer need not bring in arbitrary items (which would threaten to make the supervenience global). In being taxed I, as a Dutch tax payer, am related to all Dutch tax officials and payers, but not to Chinese ones. Nor to Saturn’s rings.
3. A layered world

Let me here, however, point out three further obstacles to this kind of reductionism, ones that do not derive from the multiple realizability, but from the relationality of higher-level properties.

The first obstacle to type identity theory is the fact that relationships among, or relational properties of, macro-entities cannot in general (depending, perhaps on the kind of relationship) be reduced to properties of, or relations among, micro-entities. For instance, a ball bounces down a staircase, two persons play tennis, two political parties form a coalition. If we descend to a micro-level of description (in terms of, respectively, molecules, organs and tissues, or individual politicians), we seem unable to describe these states of affairs. I have already suggested that this level-unicity of relations may contribute to the autonomous reality of macro-entities just as much as these entities' intrinsic properties.

The second obstacle to type identity metaphysics is the fact that relationships that presuppose, or require, further relationships may not form a hierarchy, but a self-sustaining circle. A simple example is that of employers and employees. It is impossible to specify what the relational property of being an employer consists in without referring to, among many other things, the relational property of being an employee. The reverse is equally true. It seems impossible, therefore, to describe in merely physical terms what a person's being an employee consists in.

Of course, on the view about particulars developed above, an employee (being a particular, and token identical with an organism, a bunch of molecules, etc.) can surely be described in physical terms; what cannot be specified in merely physical terms is just the fact of that particular's being an employee, or the relations that this property consists in (e.g., being on a pay-list, having duties towards an employer, etc.). Also, although no physical base of each individual property or relation of the employee/employer cluster can be established, it will surely be possible to establish the physical base of the cluster as a whole.

The third, related, and most important obstacle to the type identity doctrine is the fact that although, by the free lunch principle, higher-level, or macro-relatedness cannot ultimately consist in anything else than in physical reality (because it supervenes on it), any higher-level relation in particular fails to have its own, unique physical base. Its supervenience base cannot be 'minimal,' as Corbí and Prades (2000) would put it. Kincaid (1998, 72-4) notes:
Part I. Ontology

A property or fact \( W \) supervenes on some other property or fact \( P \) just in case fixing \( P \) also fixes \( W \). However, nothing in this relation rules out the possibility that other properties or facts may also supervene on \( P \) (...). There might be a many-many relation between the supervening and the base predicates or facts.

The physical base of a higher-level state of affairs, then, will often be so wide as to be the base of countless other higher-level states of affairs. These higher-level states typically do not supervene one by one, but collectively, en masse. Suppose that we want to specify the physical base of a purchase.\(^{33}\) The purchase is a four-place relation between a customer, a piece of merchandise, money, and a trader. Now, we can physically describe customer, merchandise, money and trader. But that will not do to physically describe the purchase, because a purchase not only requires the four relata mentioned, but also a context. Especially being a piece of merchandise, being money, and being a trader will consist in a number of relational states of affairs involving the national bank, currencies, licenses, taxes, laws, and so on.

Of course, of all these relational states it will be possible to specify what relations they consist in. And what those consist in. And so on. We can in this way patiently peel our way through layers and layers of relationality, until we reach a wide, but very specific set of physical states of affairs on which all supervenes. We then have, in fact, specified the physical base of the purchasing relation. We have done so in a non-arbitrary way, and although the physical base will be wide, it will by no means be the entire world.

We then know by what set of physical states of affairs the purchase is realized (be it indirectly). I am aware that many will say that the purchase is just a local event that takes place within a context, but can itself be specified physically. I will later on argue, however, that the concept of an ‘event’ that this account relies on is extremely problematic, in that any

\(^{33}\) Kincaid himself takes electrical conductivity and translucency as examples, properties that both supervene on the molecular structure of water. They seem to be causally distinct: if the water is involved in a short circuit, it seems to be its conductivity and not its translucency that caused it; and it also seems to be impossible to individuate these two dispositions in terms of molecular structure. I am myself not impressed by this example, because the free lunch principle tells us that at least disposition tokens are nothing over and above their determinate and categorical base (given the laws of physics). In my view, a short circuit is as much caused by a quantity of water’s conductivity as sleep is caused by a pill’s dormitivity.
physical specification of ‘it’ could only be an arbitrary affair. The purchase is physically realized, but by a wide, nonlocal set of physical states of affairs. Yet, being ‘nothing over and above’ this set, is it perhaps not identical with it? No. The purchase is but one aspect of that part of physical reality. The latter is the physical base of a great many higher-level states of affairs, for insofar as a purchase requires a context, the physical base must also realize the states of affairs making up that context in order to realize the purchase. And it seems impossible to abstract the purchase from that wide context using only physical terms. We might physically describe the bodily movements and shuffling around of physical particulars as they occurred in the shop in such terms, but that would not be the realizer of the purchase: like physical events might take place, say, on a stage without realizing a purchase.

3.8. NON-IDENTITY AND NON-DISTINCTNESS

Relationism implies, then, that type identity theory is mistaken. On the other hand, physicalism implies that higher-level states of affairs are ‘nothing over and above’ their physical basis. But then, is the combination of relationism and physicalism not unstable? I have above sometimes used the phrase ‘consists in’: we may, for instance, say that being a champion ‘consists in’ having participated in a contest and having officially been recognized as the winner. And no doubt these latter states of affairs will themselves ‘consist in’ other, lower-level states of affairs. But then, does not someone’s being a champion ‘consist in’ some wide physical state of affairs, and does that not suggest reduction?

Note, however, that there is a difference between ‘consists in’ and ‘is identical with.’ Like ‘is identical with,’ ‘consists in’ is a transitive relation. But it is not symmetrical. It does not seem to be the case, for instance, that someone’s having participated in a contest and having officially been recognized as the winner ‘consists in’ her being a champion. If so, ‘consists in’ may neatly articulate the notion of ‘being nothing over and above but also not identical with.’ True, it will then be a small step to the reductionistic ‘is nothing but,’ which also seems asymmetrical. But while ‘consists in’ describes what there is, ‘is nothing but’ seems to point in the direction of elimination.
Part I. Ontology

Unfortunately, expressions like 'is nothing over and above,' 'consists in,' 'is no ontological addition with regard to,' 'is different without being distinct,' etc. will probably remain sensitive to subtle misinterpretations, no matter how carefully we explicate them. So I want to formulate my point in less ambiguous terms. Briefly: higher-level states of affairs are instances of properties of wide parts of physical reality. While being relational properties of 'narrow' particulars, we may say that higher-level properties are intrinsic properties of the sets of particulars constituting their wide physical realizer. And they are so along with, but not reducible to other (i.e., merely physical) properties of these, although any higher-level state of affairs will no doubt be necessitated by a set of merely physical states of affairs. Higher-level properties are just additional properties of (sets of) physical particulars. The properties are indeed 'additional' with respect to merely physical properties, but not with respect to their concrete bearers. For we have seen that a property is a 'way to be' for the thing that has it: not identical with that thing, nor a second entity. This, then, is what my nonreductive physicalism amounts to, and this is what should be kept in mind in case terms like 'is nothing over and above' seem confusing.

The combination of non-identity and non-distinctness is surely puzzling: it is partly this combination that fuels the ongoing debate about particulars and properties, and gives rise to the question whether something's having a property is a relation or not. But the fact that it is hard to articulate in a perspicuous, non-ambiguous way should not tempt us to settle for something simpler and easier like identity (reductionism), or relatedness (dualism). Like a particular and its properties, physical and higher-level reality are not identical; and they are also not distinct items that are related. If this is difficult to grasp, and to articulate, so be it.

3.9. SUMMARY

So far, I have argued that if we want to give an account of the way in which in our world multiple domains, or levels of organization, hang together, we should do so in terms of physical realization. 'Supervenience' and 'emergence' are unexplanatory notions, while the theories of upward causation and type identity are false.

Physical realization of higher-level states of affairs, as I have explicated it, implies that higher-level phenomena are nothing over and above their
physical realizers (the free lunch principle). But because of the relational
ccharacter and wide realization of higher-level states of affairs (rather than
because of their multiple realizability), the thesis of physical realization
does not imply that higher-level states of affairs are identical with any spe-
cific lower-level state of affairs, or set of states of affairs.

There is also, of course, token identity theory, the claim that all particu-
lars are in fact physical particulars. I have earlier pointed out that I support
this view, but it should not be confused with what is perhaps better known
as token physicalism. Such token physicalism assumes that, because we
may regard events and states as particulars, we may also regard mental
events and states as identical with physical events and states; just as we
may do this with physical and minded objects. It will take considerable ef-
fort, in later chapters, to dislodge psychophysical token physicalism in this
less innocent sense.

This concludes my brief overview of the general ontological issues that
are relevant to the discussion to come. Let us now turn to the issue of cau-
sality.
Part II. Causality

In the following chapters I will attempt to give an analysis of the concept of causality, or causation, and also an account of the ontology of causes and effects. This will probably tax the reader's patience, but our understanding of mental causation will only improve if we understand causation, at least to some extent. We have to go through it.

It might help when I show my larger agenda here. I am planning to bring to the surface what I consider an easy-to-grasp but neglected duality in the concept of causation that is extremely important. This duality, which is one of counterfactual dependence and physical connection, provides a fresh view of what we call causal efficacy, or causal power. It is often thought that genuine causal efficacy pertains exclusively to basic and intrinsic physical properties, while higher-level and relational properties have explanatory relevance at best. We will see, however, that on the one hand mere propagation of physical quantities (force, energy, or whatever) in and among particulars is only half the story about causation, or 'efficacy'; while on the other hand we will see that such propagation is an indispensable ingredient of all causation, be it basic-physical or higher-level.

To develop this duality, I will in turn consider what we may call the 'dependence approach' to causality and, later on, the 'mechanistic approach.' The dependence approach is that of traditional nomological and counterfactual theories of causation (such as Mackie's); the mechanistic approach (such as Salmon's) concentrates on physical infrastructure. My aim is to show that none of these two approaches can on its own clarify what causation is, and that we need a way to unite them. If, however, we do have such a unified account (which I will try to provide), then interesting consequences follow.

For instead of a much-encountered dichotomy between on the one hand basic physical, lawful, and therefore genuine causal efficacy and, on the other hand, mere causal-explanatory relevance of properties at higher levels, we find that causal relations at all levels involve both physical connections as well as counterfactual dependencies, as two sides of one coin. This analysis will, then, tell us that genuine causal efficacy is found at all levels of nature, and that it is not mere physical push-and-pull. It will also enable us to see how particular causal relations can involve laws of nature
without themselves being covered by such laws: an important requirement, remember, for solving the problem of mental anomalism.

On the basis of the duality, I will also argue for a ‘factualist’ causal ontology, in the spirit of a state of affairs-ontology (see above) rather than in the spirit of those, most notably Davidson, who conceive of causes and effects as concrete particulars. The connection-dependence duality comes in here when we realize that connections are among concrete particulars, while dependencies are among states of affairs. This suggests that regarding causes and effects as concrete event-particulars conflates connections with dependencies. The alternative is a fine-grained conception of causal relata, according to which properties are constitutive of, rather than merely had by, causes and effects. If such a conception can be made to work, we can appreciate how mental properties constitute causes of their own, rather than being mere (epiphenomenal) aspects of physical event-particulars.
4

A duality in the concept of causality

4.1. CAUSALITY: THE VERY IDEA

We all know what a cause is. A cause is what causes something else. But what is causing? Yes, it is bringing about, producing, or effecting, having as a consequence, making something happen, or making it the case that; influencing, affecting, doing something to, or making a difference to something. All very obvious, but only to those who already understand what causing is. But now try to explain what causality, or causation, is in terms that do not appeal to such understanding! Anscombe (1971, 93) claims that

the word 'cause' can be added to a language in which are already represented many causal concepts. A small selection: scrape, push, wet, carry, eat, burn, knock over, keep off, squash, make (e.g. noises, paper boats), hurt. But if we care to imagine languages in which no special causal concepts are represented, then no description of the use of a word in such languages will be able to present it as meaning cause.

And Armstrong (1997, 210-1) tells us:

All we have in our concept of singular causation (...) is that something (...) makes something happen (...). At the conceptual level this making something happen, this singular causation, is a primitive. We can walk round it, but we cannot analyze it conceptually.

'Causality' and 'causation' belong to the same cluster of concepts. Let us restrict our use of the term 'causation' to referring to singular causal relations that exist between particular items, and use the term 'causality' as a broader term that covers singular causation as well as causal influence, causal powers, etc. Now, is causality, or causation, really a primitive notion, as the above quotations suggest? Many philosophers think not. With the pretention of noncircularity, causation has been claimed to be:

- Regular or lawful succession ((neo-)Humean views)
- Counterfactual dependence (Lewis, Mackie)
Part II. Causality

- Energy transfer (Fair, Harré and Madden)
- Trope persistence (Ehring)
- Persistence of a conserved quantity (Dowe, Salmon)
- 'That theoretical relation that determines the direction of the logical transmission of probabilities' (Tooley 1987, 251) and
- Spatiotemporally contiguous change (Ducasse).

Does not such an impressive list show that a conceptual analysis is possible after all? Perhaps, but the sheer diversity of accounts is unsettling. There is no consensus at all here. The only thing that can be agreed on is that all proposed analyses are beset by trouble, and that there is no undisputed champion among the above accounts.

Yet, an eliminativist or deflationary stance towards causation does not seem to be a good option. Russell (1912) has once claimed that the concept of 'cause' is obsolete; and Hart and Honoré (1959) have regarded 'causation' as a mere umbrella term. But causal eliminativism utterly disrupts our normal worldview, while deflationism just does not clarify anything.

The problem with eliminativism is that every scientist, engineer, insurance agent or judge knows that whether or not there is a causal relation between co-occurrent phenomena is often crucially important. What would it mean to a judge or engineer to hear that the concept of 'cause' is obsolete? Would it mean that nobody ever does anything, or that nothing ever makes a difference to something else? Causality is a key concept in our thinking about the world. If we were to delete the causal concepts from our reasoning, hardly anything would be left. Perhaps we would still be able to do some logic or mathematics, but we would no longer be capable of having such simple thoughts as that it is raining, or that we should have a cup of tea. David Fair (1979, 221) writes:

'(C)ause' functions as a kind of "dimension-word" (Austin) for a large class of transitive verbs. These verbs, call them causatives, presuppose causal notions in the sense that sentences in which they occur can easily be transformed while roughly preserving meaning into sentences in which 'cause' explicitly appears. For example, 'John knocked the glass onto the floor' means roughly, 'John caused the glass to fall onto the floor.' Once we realize that words like 'go,' 'lie down,' 'perceive,' 'speak,' 'think,' i.e., words for anything we do, may well be regarded as causatives (obviously or less obviously), we will appreciate how the idea of causality pervades our thinking. But it does not stop with verbs: 'footprint,' 'text,' and sunglass' are just a few of the
vast amounts of terms that presuppose causal relations for their meaningful application.

Remarks like these suggest that we should not adopt eliminativism until we are really desperate.

The problem with a deflationary view of causality is that it makes a mystery of the fact that we attach such a great importance to the class of dependencies among phenomena that ordinary usage would call causal, and much less, or only a derived importance, to noncausal dependencies such as supervenience. The latter concept is a philosophers’ invention, it is not part of our everyday speech. If ‘causation’ is just an umbrella term, why do we not just call all dependencies causal? Baker (1995, 12 and note) writes:

Maybe ‘causation’ does not denote a single phenomenon; maybe ‘causation’ is just the word that we use when we think that we have found an explanation of a certain sort. (...) Schematically: A causal explanation purports to explain one event B in terms of another event A, where event A precedes event B, the occurrence of event A does not entail the occurrence of event B, and in the context, event B would not have occurred if event A hadn’t occurred but given event A in the context, B was guaranteed to occur. For different domains, different kinds of counterfactuals are relevant in the last two clauses.

The characteristics that Baker mentions belong to dependencies in the world that are vitally important to our survival and well-being. Of course, this alone does not unify them into a kind. But should we not expect there to be something that underlies, or explains the characteristics, rather than these to be just brute facts? Why should this mishmash of constraints on causal explanation be assumed to constitute causal character, rather than to be its corollary?

I suggest, then, that we should not take it easy with causality. Because eliminativism is unacceptable and deflationism unclarifying, and because causality does not seem to me as primitive a notion as the above quotations from Anscombe and Armstrong suggest, I will seek to analyze it. What I will not do is to cover all aspects related to causation. I will not, for instance, try to explain why there should be causation in the first place, rather than pervasive natural caprice. I will also not address the difficult methodological issues related to scientifically establishing causal relations. Nor will I deal with the many puzzling legal issues about causation. I will leave alone the numerous difficult issues about causal and temporal asym-
Part II. Causality

And so on. My account of causation will have to be limited. For just explaining what, in general, we are talking about when saying 'cause' will be difficult enough.

4.2. THE DOUBLE FACE OF CAUSALITY

It would be best if we could say in just a few words what causality is; if we could find an elegant analysis out of one piece. Such a monolithic account would take the following form: 'causality (or causation) as X.' It would provide one sophisticated concept that articulates what the phenomenon essentially is.

I will now argue that a monolithic account of causality is out of the question, because the concept has a dual nature. The duality concerns the causal relation itself as well as the causes and effects, as causal relata. Let us see what it consists in.

Causation is a relation of dependence as well as of physical connection. This can be appreciated when we remind ourselves of two facts about causation that are generally acknowledged. On the one hand, effects depend on their causes, in the sense that causes are necessary conditions for their effects. This is why counterfactuals often indicate a causal relation: if c had not happened, e would not have happened either. This is not true if c is backed up by c' that would have caused e if c had been absent, but even in

---

34 Causation seems to be necessarily asymmetrical. If causation were symmetrical, causation backward in time would also be causation forward in time. This would enable self-causation and pervasive overdetermination of events, but accepting these notions seems to undermine any intelligible notion of causality. But causal asymmetry is a notoriously difficult issue. Consider the following list of alleged sources of asymmetry: direction of time (Hume, among many others); 'fixity' of causes, not effects (Mackie 1974); direction of irreversible processes (Reichenbach 1956, Popper 1956); direction of 'open causal forks' (Salmon 1984); 'logical transmission of probabilities' (Tooley 1987); overdetermination of causes by effects (Lewis 1979); mutual independence of conditions, not of effects (Ehring 1997, Papineau 1985). The latter account, it seems, best fits a counterfactualist theory such as I will endorse. One may object that it does not point to a connection with temporal asymmetry, but we may question the assumption that there should be such a connection. Causal asymmetry falls largely outside of the scope of my arguments. It does seem, however, that the notion of a 'structuring cause,' that later on plays an important role in my argument, introduces asymmetry in the network of causal dependencies, in suggesting that causal relations are caused without, in general, being able to cause anything themselves.
such a case, it would be strange to say that e did not depend on either c or c'.

On the other hand, causes and effects are physically connected: if there are no physical influences connecting alleged cause and effect, we would sooner suspect coincidence than 'action at a distance.' The relative spatio-temporal locations of a cause and its effects just cannot be arbitrary; otherwise, there would not be any ground for declaring two items causally related. Anything could then be causally related to anything. But let me stress, here and below, that physical connectedness means more than mere contiguity: it means influence, a physical signal. If, as seems the case, there is no 'action at a distance' in our world, contiguity is required for connectedness. Still, connectedness is more than mere contiguity.

We can approach the duality from a slightly different angle, one that has to do with the ontological structure of causes and effects. Kim (1968, 71), discussing causal relata, writes:

(W)e must have a way of relating the talk of necessity and sufficiency to individual events that are spatio-temporally localized. Thus, we need entities that possess both an element of generality and an element of particularity; the former is necessary for making sense of the relations of necessity and sufficiency, and the latter for making sense of singular causal judgements.

We will see later on in detail how this aspect of the duality can be taken account of in a theory of causal relata, namely by conceiving of these as states of affairs (see Part I), i.e., structures of particulars and properties at times.

4.3. DEPENDENCE AND PHYSICAL CONNECTION

Let us examine the duality more closely. Effects depend on their causes; that is a central and widespread intuition about the causal relation. If this is put into question (as may happen in the context of Humean regularity theory), the discussion does not seem to be about the character of the causal relation so much as about the question whether there is causation.

But should we hold that causes and effects are physically connected also? May causality not just consist in 'regular co-occurrence,' as Hume has claimed? Perhaps. But as is well-known Hume, even though he relinquished any explanatory resources such as forces and influences, had to
hold that causes and effects are spatiotemporally contiguous. Even if his scepticism about causal influences had been warranted - I will have more to say about this below -, he could not possibly have denied spatiotemporal contiguity as a characteristic of causes and effects.

For on such a denial, everything can be causally related to anything else, and the very concept of causation becomes hollow. If you just claim that effects must follow their causes, you might try to look for regularly co-occurrent phenomena (at any spatial distance). If you just claim that effects must be next to their causes, you might pick out a number of locations and look (through time) for local patterns of co-occurrence. But if you let go both the temporal and spatial constraint, any intelligible notion of dependence, or even regularity, is lost. You will be left with an enormous bag of events (past and present, here and there), among which any semblance of pattern is as real as any other.

Now, the requirement of spatiotemporal contiguity may just be accepted for these reasons. But this leaves it unexplained why salient regularly co-occurrent phenomena are indeed typically contiguous. To explain this, we need the notion of influence, i.e., physical connectedness. As a Humean, one may reject such an explanans on the ground that we cannot perceive it. But that leaves it mysterious why something should ever be the cause of something else. Whether or not as empirically responsible philosophers we can actually afford it, a full account of causation does require some notion of physical connectedness.

Now, physical connectedness is quite a different relation from dependence. My desk and my PC are connected (i.e., by pressure), but are not very much dependent. Being owned by somebody and having monetary value are dependent, but not very much physically connected. Considerations like these strongly suggest that dependence is a relation among states of affairs, while physical connection is a relation among objects. Yet, in causation dependence and connectedness come together. Causation is a mixture of dependence and connection. There is no way to have causality without dependence: there would be nothing to distinguish a causal relation from a merely spatiotemporal relation. Also, there is no way to have causality without physical connection: there would be nothing to account for the dependence of effect on cause.35

---

35 For the moment, I omit the complication that causation may also involve disconnection. More about this important point later on.
4. A duality in the concept of causality

4.4. GENERALITY AND PARTICULARITY

Let us approach the duality from the angle of the ontological structure of causal relata. As Kim has noted (see above), causes possess an element of particularity as well as of generality. On the one hand, causes and effects seem to be particular items, situated at particular times and places, that are causally related to other particular items. Causes and effects are singular. If they were not, they could never be physically connected, or satisfy the Humean criterion of spatiotemporal contiguity (see above). Davidson’s claim (1967a) that causation is an ‘extensional’ relation among particulars articulates this aspect of the notion.

It is also particular occurrences that we are interested in when we ask for a causal explanation. When a plane crashes, what we do not want to know is, why in general planes crash (although we may very much want to know this), but why this one did, in this particular case. Knowledge of why in general planes crash may help us in finding the cause of particular crashes; but what we are trying to explain is a particular occurrence, not just a general law or pattern.

Relatedly, if causes and effects are not regarded as particular occurrences, it becomes impossible to account for such facts as that a bomb explosion at a particular time and place caused a plane crash at a particular time and place. Causes and effects must be paired: not anybody’s smoking causes somebody’s cancer.

On the other hand, causation involves an element of generality: properties. A first reason for holding this is that we typically give causal explanations (‘e is causally explained by c’). In making causal statements (‘c caused e’), we typically will not be satisfied with relating two concrete particulars. We characterize them (e.g., one was a car, the other a pedestrian), and we tell something about them that specifies as well as explains, to a certain extent, the causal relation. The car was moving on the wrong side of the road near a corner, and the pedestrian got wounded. In order to make an informative causal statement - that is, any causal statement at all - we must ascribe causally relevant properties to particular entities. This is not obvious in elliptical statements like ‘The storm caused the damage’; but as we will see later on, terms like ‘storm’ and ‘damage’ are abbreviatory, and not the names of causally related concrete particulars.

Secondly, if causality were a mere particular affair, it would be mysterious how causal generalizations like ‘Dynamite causes explosions’ or ‘AIDS
is caused by HIV infection' are possible. It is true that particularistic theories of causality have been put forward (see, for instance, Ducasse 1926). But the problem with particularism is that it seems unable to make sense of the fact that effects depend on their causes (see above). For dependence must be spelled out in terms of counterfactuals, and counterfactuals must, it seems, be backed by laws. Even the concept of 'interaction' cannot be understood purely particularistically: if item a interacts with item b, what happens to b must, at least in part, depend on what a does.

These implications may not seem immediately obvious. It is true, for instance, that Lewis (1986a, 160-1) distinguishes his counterfactual analysis of causation from regularity analyses. In a Lewisian possible world certain laws may or may not hold, and Lewis explicitly distinguishes counterfactual dependence from nomic dependence (167-9). Mackie (1974, 77-8) also claims, that 'a singular causal statement need not even imply even the vaguest generalization.' Moreover, I will myself defend something along these lines later on. Even so, without there being any laws, the counterfactuals that Lewis' and Mackie's analyses depend on would be without ground. Neither analysis could be particularistic: counterfactual dependencies can only obtain in a world that is governed by natural law.

Thus, causality turns out to be a dual-aspected notion, and only an analysis that accounts for the fact that such divergent aspects as physical connection and dependence complement each other in constituting what superficially seems a unitary notion, can be successful. One may hold that causality is really two things, dependence (involving generality) on the one hand and physical connection (involving particularity) on the other; in the same way that - to cite an hackneyed example - jade is really two things. However, I will below explain in detail why we can only achieve a viable analysis of causality by combining both elements.

4.5. THEORIES OF CAUSALITY

If the concept of causality is indeed dual in character, we should expect existing theories about causality to sustain this duality. If causation involves physical connection as well as counterfactual dependence, we should expect that existing theories explicate causation in terms of concrete particulars as well as properties, and that no account is purely generalist or purely particularist. Is this indeed the case?
4. A duality in the concept of causality

It does seem so. Even Hume, who is the father of the regularist approach (which, almost by definition, centres on the general features of causes), supplements regularity with something as 'particularistic' as spatio-temporal contiguity. And although some propose to reduce causation to transactions between objects (e.g. Fair 1979), it is in such views something about the objects (energy, in Fair's case) that makes them causally related. It is true that Salmon advocates a conception of 'causality without counterfactuals' (1994), but from the fact alone that his account appeals to 'conserved quantities' (quantities explained by conservation laws) we learn that this account is not particularistic. The same goes for a radically anti-Humean theory like Harré and Madden's (1975), meant to revive the idea of causal powers in objects.

Acknowledging the duality one can choose, it seems, between two kinds of approaches to causality: a nomological or counterfactualist approach that emphasizes that causation is dependence among states of affairs; and a mechanistic approach that emphasizes that causation is physical connection among, and in, particulars. Any credible approach will have to borrow from the other side, but there are significant differences. In the following two chapters I will discuss a dependence and a mechanistic approach in turn. We will see that neither a dependence nor a mechanistic approach can on its own adequately analyze the notion of causality: the idea is that we should seek an analysis that integrates both.
5
Causal dependence

5.1. INUS CONDITIONS

There are various theories of causality along the lines of the dependence approach, and it is impossible to discuss them all. But although the present chapter is just about Mackie's theory (and not, for instance, Lewis'), it is intended to point out what are the virtues of and the troubles with dependence theories in general, and in what ways, if any, they stand in need of supplementation with a mechanistic account.

Mackie's principal claim (1974, notably Ch. 2) is that a cause explains its effect by being a necessary condition in the circumstances for its occurrence. This claim is the basis of Mackie's famous doctrine that states that a cause is an INUS condition for its effect (see Mackie 1965 and Ch. 3 of Mackie 1974).36

Let us see what INUS conditions are. Mill (as quoted by Mackie 1974, 63) claimed that 'the cause (...) philosophically speaking, is the sum total of the conditions positive and negative.' (Mill, J. S. System of Logic, III, 5, 3.) One may use the term 'cause' in this way, but such a 'total,' or 'full cause' is not what in ordinary or scientific practice is called a cause. It is too broad: causes, it seems, should be individuated more finely. Mackie (1967, 33-4) writes:

Suppose that a fire has broken out in a certain house (...). Experts investigate the cause of the fire, and they conclude that it was caused by an electrical short-circuit at a certain place. What is the exact force of their statement that this short-circuit caused this fire? Clearly the experts are not saying that the short-circuit was a necessary condition for this house's catching fire at this time (...). Equally, they are not saying that the short-circuit was a sufficient condition for this house's catching fire (...). At least part of the answer is that there is a set of conditions (of which some are positive and some are negative), including the presence of inflammable material, the absence of a suitably placed sprinkler, and no doubt quite a number of others, which combined with the short-circuit constituted a complex condition that was

---

36 Mackie's account has been heavily criticized by Kim (1968) for being ontologically unintelligible. I will here postpone the issue of ontology, and later on defend the Kimian view.
5. Causal dependence

sufficient for the house’s catching fire - sufficient, but not necessary, for the fire could have started in other ways. Also, of this complex condition, the short-circuit was an indispensible part (...). In this case, then, the so-called cause is, and is known to be, an insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result. (...) Let us call such a condition an INUS condition.

We can call the set of INUS conditions that is a complex sufficient condition for a certain state of affairs the ‘total cause,’ if we want; but this total cause should not be regarded as something over and above the sum of the separate INUS conditions composing it.

Mackie’s is a counterfactualist account of causation: first, something’s being a necessary or a sufficient condition presupposes the truth of a counterfactual, and secondly, it could not be spelled out in any other than counterfactual way. Mackie (1974) writes: ‘Necessary’ means: ‘X occurred and Y occurred and Y would not have occurred if X had not’ (31); ‘sufficient’ means: ‘given the circumstances, if Y had not been going to occur, X would not have occurred.’ (39)

It is crucial to the above account that INUS conditions are allowed to be negative, such as ‘the absence of a suitably placed sprinkler.’ Such negative conditions are a necessary ingredient in the account, because there can always be counteracting causes that will defeat a presumed counterfactual dependency. A negative condition is the absence of a particular kind of counteracting cause. This need for negative conditions (and, also, negative effects) may raise an urgent ontological issue: does not the counterfactualist account require us to embrace a dubious ontology of negative entities? I will postpone this issue for the moment, but I will take it up when discussing causal relata.

Connected to the counterfactualist account there are, however, many more complications and subtleties. In the rest of this chapter I will discuss several of these. My aim of this discussion is, first, to give an idea of the complexities involved in causal dependence and, secondly, to point out how well, on the whole, an account like Mackie’s can deal with these complexities and is up to the challenges they represent (given, that is, some unexplained assumptions). In my view, the counterfactualist approach to causation is eminently plausible on a number of points. It is a very powerful approach indeed.

I will thirdly, however, point at a notorious stumbling-block for the counterfactualist approach in its pure form: the possibility of one state of
affairs backing up, or pre-empting, another state of affairs as a cause. This is a stumbling block, for it seems to straightforwardly invalidate the core claim of the counterfactualist approach, namely that causes are necessary conditions. Here, the lesson will not be that a counterfactualist theory of causation is doomed; it will be that it is very good but not complete. This will then set the stage for my subsequent discussion of the mechanistic approach.

5.2. NONCAUSAL CONDITIONS

Unconditionally identifying causes with INUS conditions would yield counterintuitive results. A driver’s exceeding the speed limit may be an INUS condition for his breaking the law; someone’s death may be an INUS condition for another person’s becoming a widow; and the destruction of my car may be an INUS condition for your car’s becoming the fastest car in the world; all without causing the dependent fact.

In answer to examples like these, Mackie claims that ‘what is required is that the cause and the effect should be, as Hume says, distinct existences’ (1974, 32). Note that this nicely supplements the free lunch principle (see above): Mackie claims that if a dependency is causal, it is between distinct items; the free lunch principle tells us that if a dependency is not causal, it is not between distinct items. The reader is invited to check whether or not the above noncausal dependencies involve, or seem to involve, distinct items. Is, for instance, the destruction of my car a distinct state of affairs from your car’s becoming - because of it - the fastest? I would say no: there is just a change in relations among cars (one faster-than relation, supervening on, and therefore not something over and above the intrinsic characteristics of the cars, disappears). Note how here relationism supports the free lunch principle.

Note also that the noncausal dependencies mentioned all seem to have a symmetrical character, while causation is typically asymmetrical. Finally, note that on the counterfactualist approach the ‘distinct existences’ requirement itself remains unexplained. Why should it only be distinct states of affairs that can be causally related? What is so special about causal relations that supervenience-dependencies do not qualify? On a mere counterfactualist approach we must draw a blank here, but we will see that mechanistic theory answers this question for us.
5. Causal dependence

5.3. NONSALIENT CAUSAL CONDITIONS

On Mackie's account a state of affairs can have many INUS conditions and, therefore, many causes. Above we saw that states of affairs can have conditions that do not cause them; but of the conditions that are causal many are, so to speak, nonsalient. These would not be mentioned in any informative, interesting causal explanation, just because we take them for granted. The fact that the house was not flooded (or that it had been built in the first place, or that it contained oxygen) was an INUS condition for its burning down. These conditions could in principle be called causes, but many people would be reluctant to do so. Mackie (1974, 34) remarks that 'we tend to be a bit selective, to be more ready to call some kinds of factors causes than others. There are no firm rules governing this selection (...).'

The point is that what we are willing to regard as 'the' cause depends heavily on the explanatory context. It seems that on Mackie's account the distinction between causes and mere background conditions is just pragmatic. We can see this when we look at what kinds of conditions normally are accorded a privileged status. Honderich (1988, 19-20), for instance, writes:

In many cases we speak of a given condition as cause and it is the one action or piece of behaviour involved, something to which responsibility attaches. (...) In a second class of cases we refer to a condition which is unusual or abnormal. (...) In a third class of cases we may take the latest condition, the one closest in time to the effect, as the cause. Fourthly, we may single out what is ordinarily called an event or change, partly because of its brief duration, and thereby distinguished from what is called a standing condition (...). Fifthly (...) we commonly talk of causes in situations where we are attempting to explain some event and have not yet done so.

Mackie himself (1974, 34) notes the following 'privileged' kinds of conditions:

Even among events [as distinguished from standing conditions, WdM] ones which are seen as intrusive are picked as causes in preference to ones which occur within some going concern (...). (W)hat is normal, right, and proper is not so readily called a cause as is something abnormal or wrong.

There is hardly any reason to think that with the above we have a complete list that enables us to distinguish 'real' causes from mere conditions. But then, there does not seem to be any problem in admitting that such a
Part II. Causality

distinction can only be made in a pragmatic way. Mackie himself distinguishes conditions that belong to a 'causal field' (1974, 35) from those that do not: 'What is said to be caused, then, is not just an event, but an event-in-a-certain-field' (1974, 35). A causal field consists of all the conditions that are considered normal, that is, taken for granted, in a causal explanation. Mackie claims that what is part of the causal field does not deserve the name of cause, a cause being precisely a difference relative to the field. And in Mackie's view causal explanations have a contrastive, rather-than character.

'Causal field,' however, is obviously an explanatory, not a metaphysical notion. The weakness of Don's rope is a cause of his fall as well as his decision to undertake a climb, regardless of whether the explainer is the person who sold Don the rope or the femme fatale who wanted an Edelweiss from Don. Causal fields will be different for both, but it seems that Don's fall was caused the way it was.37

5.4. COUNTING CAUSES AND EFFECTS

It might seem that in contrast to 'the' cause, there really seems to be such a thing as 'the' effect. In causal explanation, we typically pick one explanandum state of affairs, which can then be said to have a multitude of necessary conditions, some being mere background conditions, belonging to the causal field, some salient and interesting. The relation conditions-effect can then be seen to be many-to-one.

This is an important point that can probably be used to clarify causal asymmetry (see Ehring 1997, Papineau 1985). But it has to be qualified. First, although an effect has many INUS conditions, it is also true that one state of affairs typically is an INUS condition for many effects at the same time. The difference between cause and effect in this regard, then, is not just that the causes are many while there is just one effect.

Secondly, even if we pick out one state of affairs as an explanandum, it will typically contain different explananda. Consider: Sebastian strolled though the streets of Bologna at midnight, and various explainers may want to know why. Yet, they might be interested in different things: one

37 It is true that Mackie makes a difference between 'producing' and 'explanatory' causes (1974, 262). He does, however, give priority to the latter, and there is no suggestion in his account that the former have a metaphysical status the latter lack.
explainer may want to know why it was Sebastian (rather than someone else) who strolled then and there; another why it was strolling (rather than running) what Sebastian did then and there; yet another why it was through the streets of Bologna (rather than the surroundings of Bologna, or the streets of Rome) that Sebastian strolled then; and again another why it was at midnight (rather than at noon) that Sebastian strolled through the streets of Bologna. There may also be someone who wants to know why Sebastian strolled then and there after having had a good meal, or did so using a walking-stick, etc. Explainers typically do not want to know all these things at once; and they will take diverging states of affairs to be causal explanantia, dependent on which constituent of the explanandum is emphasized.

It seems, then, that even a single state of affairs contains a ‘field’ and an interesting difference, and can be approached from different angles to yield different explananda. As we have seen earlier, states of affairs have particulars, properties and times as constituents, and these may be emphasized or suppressed according to explanatory interests. As a result, different strands of the state’s causal history will be mentioned as its cause. Of course, an explanandum remains one state of affairs: there is no point in explaining loose constituents such as ‘Sebastian,’ ‘strolled,’ etc.

### 5.5. INDETERMINISM

Although Mackie calls causes INUS conditions, he acknowledges (1974, Ch. 2) that according to our ordinary conception of ‘cause,’ causes need not be *sufficient* conditions, not even along with the circumstances, for their effects. When a gambler wins money, his getting money from the gambling device is caused by his act of gambling, although he is not guaranteed to win money when he gambles. His act of gambling does not have to necessitate, i.e. be a sufficient condition for, his getting money from the gambling device in order to cause it. Even the acts of an omniscient being may be gambles: perhaps some events cannot occur with a probability of 1.

Note that causes *do* need to be necessary conditions (given the circumstances). Suppose that in some rare cases a gambling device just dispenses money, whether or not someone takes the trouble of gambling. In such cases, an act of gambling is as causally superfluous as a magic spell of a con-
Part II. Causality

jurer. With an unreliable gambling device, we just cannot know whether a gambler really causes the money to come out when it does come out. But if gambling is to be a cause, it is supposed to be necessary in the case at hand.

Causes in our ordinary conception, then, seem to be INUS conditions minus 'S.' In some cases they may also be INUS conditions minus 'U': some effects might be producible in just one way. Mackie does not, however, embrace a probabilistic conception of causality. For him, causes being INUS conditions remains the rule, counterexamples invoking quantum mechanics being odd cases (62). He also maintains that in a sense causes always necessitate their effects (see 48-50), even if the 'laws of working' that underlie this necessitation are of a statistical kind. For sufficiency can be taken in an ex post facto sense: if something actually happens, this shows that its causal antecedents must have been sufficient.

But we may well ask why we should continue to speak of causes as INUS conditions in the face of the probabilistic worldview dominating modern physics. As has been pointed out by Salmon (1984, 184-90) and others, we might live in a fundamentally indeterministic world. Now if so, why do we not rather analyze causality in terms of statistical, or probabilistic, correlations? Various authors have indeed made that choice, in general claiming that if c causes e, it must be the case that c raises the probability of e's occurring.

These probabilistic theories are very sophisticated, and I cannot discuss them here in an appropriate way. But the basic trouble with such analyses is that mere correlation just is not causation. If it were, there would not be the daunting epistemological and methodological issues facing those (social scientists, for instance) who look for the causal connections underlying, and explaining, their statistical data. It is true that when, under specific conditions, some state of affairs raises the probability of another, this often indicates a causal relation. But it will only be recognized as a cause when it can be seen to be a necessary condition. Moreover, it seems that although causes do often raise the probability of their effects, this is not always the case. Causes may sometimes even lower the probability of their effects. For instance, I may take measures to prevent a fire, and in doing so, perhaps by incredible coincidence, set the house on fire. Now, if such probability-lowering causes are possible (as Dowell claims they are), statistical correlation does not seem to be the right concept for clarifying causation.

38 E.g. Suppes, Mellor and, before he developed his 'process' account, Salmon.
39 See, for instance, McKim and Turner 1997.
5. Causal dependence

Also, indeterminism does not seem to be a real problem for a counterfactualist theory of causation to begin with. First, as we have seen, on the counterfactualist account causes need not necessitate their effects (not even given all circumstances). Secondly, insofar as events can occur by chance (and so have no necessary conditions), we will say that they are uncaused rather than say that some unnecessary factor was the cause.

Nevertheless, the above does teach us that the ‘S’ in INUS is not really needed in a counterfactualist account. This is a remarkable thing to establish, given the popularity of the view that causes ‘necessitate’ their effects. But causes need not necessitate (or even probabilify) their effects at all. We just do not need a clause referring to sufficient conditions in our definition of the causal relation. Causes may in fact often necessitate their effects, but being a cause just is something different from being a sufficient condition.

And neither do we need the ‘U’ in INUS. Again, causes may in fact often be just one of many possible ways to produce an outcome; but if a state of affairs happens to be the only possible cause of some effect, that does not make it less of a cause. True, the ‘U’ might seem to be required for dealing with the phenomenon of pre-emption, which we will see below is deeply problematic for the counterfactualist approach. But I will suggest a different treatment of pre-emption. For this reason I will not myself adopt the term ‘INUS condition,’ but just say ‘circumstantially necessary condition.’

Finally, a warning. In the context of the mental causation debate, it might seem tempting to expect something from indeterminism. We might think that it enables us to get rid of the ‘strict law’ requirement (Davidson 1970a) for the causing of our actions. Or that it creates room for a determining agent along with merely probabilifying events (see Cuypers 1998). But let us not count on indeterminism. An indeterministic world is as much governed by natural law as a deterministic one, and the very chanceiness of the course of events implied by indeterminism seems, if anything, to hinder free agency rather than to enable it. In my view, the issue of (in)determinism is irrelevant to the problem of mental causation.

5.6. COLLATERAL EFFECTS

We have collateral effects when two states of affairs (for example, two symptoms of a disease) are counterfactually interdependent and satisfy the
distinct existences requirement, while not being cause and effect. Such cases are due to the fact that one state of affairs can have multiple effects. I would now like to argue that a dependence approach can deal with collateral effects on its own resources.

The problem here is not so much to distinguish, in terms of necessary and sufficient conditions, between two kinds of dependencies, one causal and the other noncausal; it is, rather, to determine, on particular occasions, which of several co-occurrent and counterfactually interdependent states of affairs are causally related to which other ones. How can we, on the counterfactualist approach, distinguish a causal relation between two states of affairs from mere counterfactual interdependence due to a third event?

It seems that asymmetry considerations can help us out here. Suppose that a disease (say, influenza) has fever and headache as its symptoms. In that case, two things are possible. Either the fever and the headache have a symmetrical counterfactual dependence. But in that case, none of the two can be a cause of the other, for causation is asymmetric. Or fever and headache could have occurred independently. But in that case too, none of the two could be a cause of the other, for causes are necessary conditions.\footnote{Except in cases of pre-emption (see below).}

However, imagine the following scenario: influenza guarantees headache and fever; influenza is the only way to get a headache; and fever can be caused by a number of illnesses. In such a situation, the person's fever is a circumstantially necessary condition for his headache (you cannot have a headache without fever), but not \textit{vice versa} (you can have fever without a headache). So there is asymmetrical counterfactual dependence.

Now, is the counterfactualist account fooled into assuming a fever-headache causal relationship? Probably not. The fever may be required for the headache, but it also seems redundant in a sense. We can say that the fever's being a circumstantially necessary condition (of the headache) is merely parasitic on the influenza's being such a circumstantially necessary condition. Absence of the fever would mean absence of the headache, but only \textit{because} it would mean absence of the influenza. Thus, the structure of counterfactual dependencies involved does not seem 'right' for the fever to cause the headache.

Of course, substantiating this claim requires a more detailed account of possible lay-outs of counterfactual dependencies involved in causal rela-
5. Causal dependence

tions than I have given. Such an account would probably have a bearing on causal asymmetry, such as in Ehring (1997) or Papineau (1985). In my view, there is no reason to think that such an account is beyond a counterfactualist account of causation. If so, it seems that the counterfactualist account can handle collateral effects without external help.\textsuperscript{41}

5.7. OVERDETERMINATION

Overdetermination is generally considered a problematic issue, at least for counterfactualist accounts. Metaphysically speaking, no such thing as overdetermination seems possible; this is a consequence of the thesis of causal closure and explanatory exclusion. But there are certainly cases of overdetermination in a non-strict sense. Let me now argue, however, that the counterfactualist approach is able to handle such cases. Consider the following example by Mackie (1974, 44):

A man is shot dead by a firing squad, at least two bullets entering his heart at once, either of which would have been immediately fatal.

In this example, none of the alleged causes are circumstantially necessary conditions: there is too much causal efficacy, and half of it can be missed. It seems that only an (inclusive) disjunction of causes is required to bring about the effect. Normally, the causal efficacy of disjunctions is arguably pre-empted by that of one of the disjuncts. But here it is not obvious that one of the disjuncts (say, either one of two shots) is the cause.

On the other hand, disjunctive causes are quite hopeless: when a cause is said to be disjunctive, each of the disjuncts on its own must screen off the causal relevance of the disjunction as a whole. For only one of the disjuncts is actual, and merely possible items cannot cause anything. And anyway, it is the very point of the example that the double-shooting event cannot be divided up into causally significant and causally insignificant constituents (if it could, we would have a case of backing-up, or pre-emption, which I will discuss below). Insofar as it is a cause, this double shooting must, then, be considered as a unit.

\textsuperscript{41} In dealing with collateral effects, I will not consider Lewis' strategy: working with possible world constructions. For it is not obvious to me what can and cannot be shown with, or what should and should not be assumed about, possible worlds. But see Lewis 1986, 159-72.
Still, it is a form of overdetermination, and overdetermination threatens counterfactual dependence. But we can rescue the counterfactualist account by following Mackie's advice (45) to tell 'a more detailed causal story.' What does such a story look like? Let us consider another example by Mackie (43):

The hammer-blow caused the chestnut's becoming flatter. But the whole of the blow was not necessary for this result, though it was more than sufficient: a somewhat lighter blow would have sufficed.

This is a case of quantitative overdetermination. My idea is that in the firing squad example either causally effective and causally redundant factors can be distinguished, in which case we no longer have overdetermination but backing-up (see below); or that the overdetermining cause can only be treated as a unit, in which case we have quantitative overdetermination. Mackie, indeed (43-4), says that the hammer-blow can be treated as a unit.

The hammer-blow was, in a sense, too much: it was much heavier than was needed for the chestnut to flatten. Now, the counterfactualist account does not have to stop at rough-hewn claims such as that the hammer blow-unit, or the shooting-unit, were circumstantially necessary conditions; it can do better. Mackie says that '(i)t is possible to go on to a functional dependence view of causation, which would relate exact quantities on the cause side with exact quantities on the effect side' (43). In order to follow this advice, we would have to refine our specification of the cause and effect event in such a way that the exact force of the hammer blow and the exact amount of flattening were described (and, perhaps, such things as the transfer of energy to the table or floor). 42

The same kind of solution will do in cases like that of the firing squad. Specify the damage to the man's heart and the exact way in which his death came about, and you will see that one shot can no longer be said to have been enough to produce that effect.

42 Note that the free lunch principle ensures that, because we go from the determinable to the determinate, the former supervening on the former, we do not change the subject, i.e., talk about a different set of causes and effects, in applying this strategy.
5. Causal dependence

5.8. PRE-EMPTION

So far, the counterfactualist account of causes seems to be fairly well up to a number of complications and challenges. For all we have seen, we might even think that a counterfactual analysis of causation is all we need. There is one kind of case, however, that a counterfactualist account cannot handle on its own resources. And here it is that the need for a mechanistic viewpoint (see below) makes itself felt.

Mackie (1974, 44) describes the following case:

A man sets out on a trip across the desert. He has two enemies. One of them puts a deadly poison in his reserve can of drinking water. The other (not knowing this) makes a hole in the bottom of the can. The poisoned water all leaks out before the traveller needs to resort to his reserve can; the traveller dies of thirst.

Typical to situations like these is, that what intuitively is the cause of a certain state of affairs (the latter being a death, in both cases) is not a circumstantially necessary condition for it, for if it were absent another state of affairs would produce the effect anyway. We can say that in the example the poisoning of the water backs up the piercing of the can. If the piercing somehow does not work, the poisoning will do the killing instead. We might doubt, of course, whether the effects would be the same. Is not dying-of-poisoning something different from dying-of-thirst? But it does not seem difficult to come up with examples in which 'competing' causes would have indistinguishable effects, such as one involving, say, two generals intending to push a button that unleashes mass destruction (if one gets killed, the other will push the button anyway, and the effect will be none the different).

The problem for the counterfactualist account is as follows: Neither the backing-up nor the backed-up state of affairs is a circumstantially necessary condition for the traveller's death. Nor is their conjunction. Yes, their (as it happens, exclusive) disjunction is. But apart from the ontological hurdles surrounding disjunctive entities, it would just be counterintuitive to say that in such a case the death would have a disjunctive cause. We know all too well that the poisoning of the water had nothing to do with the traveller's death, while the piercing did have something to do with it. True, as it was it saved the traveller from death by poisoning, but only by causing his death by thirst.
Part II. Causality

Note that merely specifying cause and effect in a finer grain, the counterfactualist account's way out with overdetermination, will not help here. It is true that by specifying more details of the traveller's death, we could rule out the poisoning as a circumstantially necessary condition for such a death; but we saw above that more obstinate examples (such as that of the two generals) are available that will defy such a strategy.

Now, how can we account for the fact that it was the piercing, rather than the poisoning, that was causally involved here, if it was not a circumstantially necessary condition for the traveller's death? It would have been such a condition if there had been no poisoning. The poisoning would likewise have been such a condition if there had been no piercing. But that is not the actual situation. Then, what is?

To answer this question, we have to step out of the counterfactualist framework. Note that when discussing examples like the one under consideration, we have very clear intuitions about which accounts are acceptable and which are not. For instance, hardly anybody will accept that the poisoning caused the traveller's death along with the piercing. The very point of the example is that we have a clear picture of what is going on, and clear intuitions about the causal lay-out of the situation. But we do not thank these intuitions to the counterfactualist account, for that account tells us that that either the death has no cause, or has a conjunctive cause, or has a disjunctive cause; and all of this is at odds with the intuitions.

It is, however, quite obvious where our intuitions about causation and pre-emption in the example come from. What shapes them is the structure of physical interactions among puncturing instrument, can, water, poison and man, or precisely the absence of such interactions where they would have had salient consequences. The question to be asked is now: If such physical connections shape our causal intuitions, why are they absent from the counterfactualist account? Could an analysis of causality that ignores such an intuitively important aspect of it be right, or complete?
6

Causal connection

6.1. From counterfactual dependence to physical connection

Especially the phenomenon of causal pre-emption teaches us that counterfactualist approaches do not tell us all about causation. Our intuitions about pre-emption do not seem to depend on a merely counterfactualist understanding of causation; nor, for that matter, do our intuitions about noncausal dependence and collateral effects, the near-consensus that immediate action at a distance is impossible, or the assumption that cause and effect must be 'distinct existences.' But what is the missing element?

I will argue below that it is the fact that any dependency among cause and effect itself depends on an infrastructure of physical connections among the particulars involved. This notion of a dependent dependency is crucial to my account, so let me already sketch the basic idea. Remember my discussion of particulars, properties, and relations, where I argued that 'particular' and 'property' are interdependent notions.43 Remember also the concept of a state of affairs as a particular's having a property at a time. Now, such states of affairs, as they are regarded as time-slices (e.g., a's being F at t) can only be mere abstractions. Particulars persist, interact, and change, and properties appear and disappear; all, it seems, by virtue of physical law. Add, then, lawful persistence and change to mere states of affairs (i.e., mere existence at times); and you have introduced causation.

If such mere existence is to be explicated in terms of particulars and properties, it is reasonable to expect that full-blooded temporal and lawful (i.e., causal) reality inherits this duality. After all, it is the reality of particulars and properties, only with time and law in addition. Now, we have so far concentrated on counterfactual dependencies among states of affairs, i.e., instances of properties. So we should ask how things are on the side of

43 As far, of course, as properties of particulars are concerned. We saw earlier that properties are ontologically 'promiscuous', and a property of, say, a number does not seem to presuppose a particular. However, any property presupposes something that has it.
the particulars. This is where the notion of physical connection comes in. Just as mere existence-at-a-time is a matter of both properties and particulars, fully fledged causation is a matter of both counterfactual dependencies and physical connections. Just as we cannot do with only properties or only particulars, we also cannot do with only dependencies or only connections. It seems that we need both.

Now, note that any instance of a property (i.e., state of affairs) depends counterfactually on the existence of a specific particular. For instance, if this square table had not existed, this instance of squareness had not obtained. But in just the same way, any counterfactual dependency among property instances (states of affairs) depends counterfactually on the persistence of, or interaction among, specific particulars. For instance, if the bullet and the victim had not persisted and interacted, the momentum of the bullet would not have caused the death of the victim.

That is the core of my analysis of causation. It demands that we understand counterfactual dependence, physical connection, and the relation between them. I now have discussed counterfactual dependence, and I will discuss physical connection below. After that, I will discuss the relation between counterfactual dependence and physical connection.

6.2. THE MECHANISTIC APPROACH

Having seen what a counterfactualist approach to causality amounts to, let us now see how we can approach causality from a different angle, which we may call mechanistic. The mechanistic approach concentrates on physical connections. Precisely the banning of these from our ontology may seem to have been the point of the influential neo-Humean theories of causation. But for anyone who accepts the well-known spatiotemporal contiguity requirement for causal relations, such a rejection can be half-hearted at best. According to neo-Humean theories, what connects causes and effects is regularity, natural law, or counterfactual dependence, and not some sort of unobservable influence, or force. But we have seen that in spite of this, no theory can afford to ignore restrictions on relative spatiotemporal location of causes and effects - whether or not these restrictions take the form of a contiguity requirement. And it seems that without an explanation of such restrictions in terms of physical connections, these restrictions must strike us as arbitrary. Humean or neo-Humean theories
6. Causal connection

may for respectable reasons (i.e., unobservability) refuse to tell us where
the restrictions come from, but this does leave us with a flagrant theoretical
gap.

It is no surprise, then, that the notions of causal influence and physical
connection have made their comeback in the philosophy of causality, in a
family of mechanistic approaches. Fair (1979), for instance, has proposed
to reduce the concept of causation to that of energy transference; Papineau
(1985) advocates an 'evidential' approach to causal explanations, accord-
ing to which causal explanations provide evidence for causal connections;
Harré and Madden (1975) have revived the idea of 'causal powers' of par-
ticulars; and Salmon (1984) has defended a theory of 'mark transmission'
before he decided to follow Dowe (1992) in claiming causality to be con-
served quantity persistence. But even Mackie, whose theory may be con-
sidered Humean in essence, by no means ignores underlying physical proc-
esses. He writes (1974, 221-4):

(...) I argued for extensions of the concept of causing to include both the persistence
of objects and the persistence of self-maintaining processes, saying that in both of
these we could regard an earlier phase as a cause and a later phase as an effect. I am
now suggesting that this sort of causing plays a larger part, underlying processes that
at the perceptual level are cases of unrelieved change, of a cause being followed by
an utterly different effect. A match is struck on a matchbox and a flame appears: on
the face of it this effect has nothing in common with its cause. But if we were to re-
place the macroscopic picture with a detailed description of the molecular and
atomic movements with which the perceived processes are identified by an adequate
physico-chemical theory, we should find far more continuity and persistence. (...) What is called a causal mechanism is a process which underlies a regular sequence
and each phase in which exhibits qualitative as well as spatio-temporal continuity.
(...) On the other hand, there can be no question of reducing a process of change wholly
to some persistence (...). Qualitative or structural continuity of process, then,
may well be something in the objects, over and above complex regularity, which
provides some backing for the conditional and especially counterfactual statements
that emerge in the analysis of our ordinary causal concept.

The mechanistic approach, as I will understand it, concentrates on the
physical infrastructure or processes underlying causal dependencies, rather
than on these dependencies themselves. Below, I will consider a leading
theory that follows this approach, namely Salmon’s. On Salmon’s theory,
the units of causation are continuous causal processes, rather than events
or states.
Part II. Causality

Having started with discussing Salmon's theory of causal processes, I will raise various objections to it and make alternative suggestions. The reader may come to wonder why I am discussing Salmon's theory if I disagree with his account on so many and such vital points. But Salmon's is no doubt today's most influential and sophisticated mechanistic theory, and I have found that contrasting my own views with it is the best way to bring out what they are, and what is at issue when we discuss physical connection.

My first criticism will concern Salmon's conception of 'causality without counterfactuals' and the related 'at-at' theory of causal influence. The lesson here will be that causality is a counterfactual notion to its very core, no matter whether we concentrate on causal dependencies or causal connections. This is underscored by an instructive point that I will discuss next, raised by Schaffer (2000): the fact that there is causation by physical disconnection as well as by physical connection. I will further criticize Salmon's account on the issue of causal relevance: an exchange with Hitchcock (1995) brings out that Salmon's theory of causal processes cannot on its own explain the causal relevance of any properties other than conserved physical quantities. I will then contest the ontology of Salmon's causal processes: we will see that although the latter are not causal chains, neither are they, as Salmon claims, interacting entities. I will finish by summing up what the above criticisms teach us about causal connections. That will then be my account of what the dependence approach needs as its supplement. How the supplementation is supposed to work will be the subject of a subsequent chapter.

6.3. CAUSAL PROCESSES

The idea of a causal process was reintroduced into the contemporary debate by Salmon. Reintroduced, because Salmon exploits notions such as 'mark transmission' and 'causal line' that were developed earlier by, respectively, Reichenbach and Russell. Several decades ago, Salmon introduced a Statistical-Relevance model of scientific explanation (1971), a model that was intended as an alternative to the influential Deductive-Nomological model by Hempel and Oppenheim (1948). The latter model was often regarded as a model of causal explanation, but as such it had turned out to have serious shortcomings (see, for instance, Salmon 1984).
For a number of reasons, however (see, again, 1984), Salmon concluded that neither nomic subsumability nor statistical relevance constituted causal relatedness. They were at best evidence for such relatedness, Salmon reasoned, but they had themselves to be explained by underlying causal processes.

As a result, Salmon radically reconceptualized causal explanation, as well as causality itself. His new theory involved an important distinction between causal connections and relations of statistical-explanatory relevance. Empiricist tradition has it that talk of causal connections and causal powers is unscientific, because these things are supposedly unobservable. But Salmon found an answer to this Humean challenge: the theory of special relativity (not available, of course, to Hume) tells us, among other things, that signals cannot travel faster through space than 300,000 km/sec. This means that there is an upper limit to the speed at which causal influence, understood as a physical signal, can be transmitted; and it suggests that paths of causal influence can in principle be traced and measured, if not directly observed. Transmission of signals in and among physical entities, then, can be argued to be the causal influence that gives the world its causal structure (see Salmon 1984 and 1998a).

A trajectory of a transmitted property Salmon calls a causal process. Such a causal process is not an event, nor a chain of events, but a persisting physical entity that, in persisting, transmits a characteristic. Not only objects and quantities of stuff are typical processes, but also air waves or travelling amounts of energy, like photons. Salmon's theory says that every particular physical entity is, or contains, a causal process, as long as it has the capacity to propagate a change in a nonrelational (i.e., measurable) property, and to interact with (i.e. produce a change in such a property of) other causal processes.

After Reichenbach, he called such a change a mark. For some time, Salmon considered capacity of mark transmission to be the defining characteristic of a causal process. But in 1994 (253) he wrote:

It has always been clear that a process is causal if it is capable of transmitting a mark, whether or not it is actually transmitting one. The fact that it has the capacity to transmit a mark is merely a symptom of the fact that it is actually transmitting something else. That other something I described as information, structure, and causal influence (...).
And he followed Dowe (1992) in claiming that what is actually transmitted by a causal process is a conserved quantity: a nonrelational physical property, such as momentum, for which there is a conservation law. Such a property could, in principle, be measured at any space-time point of the process by some detector (see Salmon 1994, 301). Salmon's point in switching from: 'having the capacity to transmit a mark' to: 'actually transmitting a conserved quantity,' as the defining characteristic of a causal process is, that in this way counterfactuals no longer seem to enter into the definition of causal influence.

Propagation of causal influence according to Salmon, then, consists in a process's having, when left alone, a property that remains constant over a space-time interval. The property must be present at any point of that interval, the interval being a continuum. So, a process is not a causal chain, but a causal line. When there is an interaction with another process, the property will change; and there will also be change in the other process. Those changes will be permanent, at least until there is another interaction. If they are not permanent in this way, there has not been a genuine causal interaction in the first place, but a mere intersection of processes. An interaction, we might say, is not just any intersection of processes, but one with lasting consequences.

Causal processes such as light rays can intersect without interacting. But in most cases, when two processes merely intersect, at least one of these is what Salmon calls a 'pseudoprocess.' Pseudoprocesses manifest constancy of characteristics, but they do not transmit these. A shadow, for instance, is a pseudoprocess: if you paint it (if such a thing is possible), it will lose its colour when it moves, even if nothing interferes with it. It does not transmit 'its' colour. We cannot tell from observation that the shadow is a pseudoprocess if it does not move; yet, when it would have moved - and would have been left alone -, it would not have kept its colour. In contrast, cars are real causal processes: if you paint them, the colour will stick until it is removed by further interactions, such as sandpapering or weathering.

Note that we can use counterfactuals, such as the one in the preceding paragraph, to clarify the distinction between a process and a pseudoprocess; but that on Salmon's recent view such counterfactuals are inadmissible, at least as defining characteristics of processes or pseudoprocesses. The defining characteristic is the continuous presence, or absence, of a conserved quantity.
6. **Causal connection**

If, apart from a counterfactualist account of causal relations, we also need an account of causal connections, is the above what we need? I will now argue that, although we definitely need something like Salmon’s causal processes to account for the physical infrastructure that underlies causal dependencies, the account as it stands can be criticized on at least four points.

### 6.4. **First Criticism: The At-at Theory of Causal Influence**

Although Salmon’s ‘New Look at Causality’ (1998b) seems very different from Humean approaches, it is not really anti-Humean: Salmon avoids a commitment to imperceptible causal powers. ‘Transmission’ is of course a causal term, and it has been pointed out by Kitcher (1989) that it involves a counterfactual aspect. In reaction, however, Salmon has claimed that the transmission of causal influence just consists in a property’s being observably, or measurably present at every stage of a continuous space-time path. This is his ‘at-at’ theory of causal influence, according to which causal influence is just the continuous presence of a property along a space-time trajectory.

Salmon holds, in effect (1997, p. 462), that

(a) process transmits a conserved quantity between A and B \((A \neq B)\) if and only 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.

The ‘if and only if’ clause expresses Salmon’s at-at theory of causal influence.\(^{\text{44}}\) Now, I want to claim that the at-at theory is misleading: it seems that precisely what it is for a connection among two items to be causal, is not captured by ‘at-at-ness’ (continuity). For the latter is a merely spatio-temporal notion. Continuity of a property along a space-time path is in itself merely de facto, where what a causal connection requires is nomic, or counterfactual force. This is why Dowe (2000, 98-101) points out that ‘timewise gerrymanders’ cannot be excluded as causal processes on Salmon’s theory. As far as continuity is concerned, pseudoprocesses seem just as much causal influence as genuine processes.

\(^{\text{44}}\) See also his 1977.
Of course, Salmon's at-at theory does distinguish pseudo- from causal processes, for it counts only continuity of conserved quantities, and by definition, for such quantities a conservation law is at work. Other space-time continuities do not qualify. If so, however, it is not continuity that confers causal character (i.e., transmission), but the fact that the persisting properties at issue are conserved quantities. It is the nomic character of a property, and not its presence along a continuous space-time path that makes it causal influence.

For perhaps a quantity can even be conserved along a discontinuous space-time path. Kitcher (1989, 472) writes:

Time-travel can be defended as logically possible (...). Imagine a time-traveller whose worldline terminates at t, just after he has eaten a peanut butter and jelly sandwich. The time-traveller reappears at a much later time t' with the peanut-butter and jelly sandwich in his stomach. (...) There is no continuous process that links the two parts of his worldline, but the pertinent counterfactual holds: if he had not eaten the sandwich, then it would not be in his stomach.

This is of course just a thought experiment: for an account of actual, this-worldly causal influence, the mere logical possibility of the above situation is irrelevant. But as Salmon himself points out (e.g. 1998d, 276-8) the at-at theory runs into trouble with a phenomenon that does seem to be actual and this-worldly: quantum entanglement. In certain experimental setups it seems that a measurement conducted on a photon at one place has immediate implications for the state of a photon quite somewhere else. There is no agreement on which conclusions this does and does not warrant, but it seems at least awkward for the at-at theory, with the requirement of spatiotemporal continuity at its core.

Now, if we just drop this requirement we are out of trouble. We can just say that what constitutes a causal process is the conservation of a quantity according to a conservation law, no matter whether or not spatiotemporal gaps are allowed. True, we have seen that there are spatiotemporal constraints on causal relations, and causal influence does in fact seem a local phenomenon. But even if action-at-a-distance, faster-than-light travel, or time travel, were possible, there would still be causal processes. Their identification would be a tricky affair, but probably not impossible, at least as long as not everything went nonlocal. The simple reason is that if physical law allows phenomenon X, then phenomenon X is part of the causal order, local or otherwise.
Salmon intends his at-at theory to pave the way to a conception of 'causality without counterfactuals' (the title of his 1994 paper). The idea is that de facto spatiotemporal continuity can substitute for counterfactual dependence. But if the above is correct, and causal character is a matter of nomicity, and not of spatiotemporal continuity, then the attempt to get rid of counterfactuals will fail. For one either holds that laws imply the truth of counterfactuals, and then a conserved-quantity account of causality is not free of counterfactuals. Or one holds that laws just describe regularities in what has actually happened and is happening; but then it seems hard to sustain a distinction between genuine causal processes and pseudoprocess.

I conclude, then, that the at-at theory of causal influence fails, and that, contrary to Salmon's claims, counterfactual dependence is at the heart of causality. I would myself recommend to just accept this. It is true that counterfactual dependence cannot be observed, involving as it does the actual as well as the merely possible; and this fact no doubt raises difficult epistemological and methodological issues. But it does seem that if we throw out counterfactuals, we throw out causality: the at-at theory will not save it.

6.5. SECOND CRITICISM: CAUSATION BY DISCONNECTION

A second criticism of Salmon's account will show us that mere physical processes are not sufficient to ground a theory of causality. The criticism has been put forward by Schaffer (2000), who argues that causation can work by the interruption of physical processes as well as by their continuation. He writes:

There are many ways to wire a causal mechanism. One way is to have the cause connect to the effect, but another is to have the cause disconnect what was blocking the effect. (283)

Schaffer argues that accounts of causation in terms of causal connections, or processes, cannot capture such causation by disconnection, ubiquitous as it is. A heart-piercing causes death by interrupting oxygen supply to the brain. A trigger-pulling causes a gun-firing by taking away the obstacle that keeps a spring from uncoiling. Nerve signals cause muscle contraction by making a troponin-calcium complex stop tropomyosin from blocking myosin-actin binding. And so on: Mackie's example of the traveller with the
Part II. Causality

pierced can (see above) is also an example of causation by disconnection. In cases like these, a causal connection is not created, but cut off. Such cases, however, do seem to be typical causal relations, as Schaffer rightly stresses (285-9).

Schaffer's argument raises a problem for all those accounts that explicate causation wholly in terms of persistencies, or paths of causal influence; that is, in terms of actual connections. Not that in cases of causation by disconnection there is no network of actual connections! It is precisely new connections that undo the old ones. The idea of causation by disconnection is only possible in virtue of the very idea of causal connections. The point is, rather, that there is no actual connection from the cause to the effect. It is precisely here that we find disconnection rather than connection.

Schaffer's argument is instructive: I can see as much as three lessons to be drawn from it. First, along with the point about causal relevance (see below), the argument underscores my contention that the concept of causation cannot be explicated in terms of physical infrastructure alone. For as long as we limit our attention to paths of causal influence, we cannot appreciate the causal significance of either the continuation or the disturbance of such paths.

Secondly, it teaches us once again that causation is a counterfactual notion. Just as we saw that negative circumstantially necessary conditions count as much as positive such conditions, we now see that causal disconnections count as much as causal connections. The point is precisely not that negative conditions or negative connections should be regarded as some kind of entity. No: just as a negative condition is the mere absence of a positive condition (one that significantly might have been present), a disconnection is the mere absence of a connection where one might significantly have been present. Speaking of the underlying physical infrastructure of a causal dependency turns out to involve a tacit comparison with possible alternative situations, just as speaking of a causal dependency tacitly involves a comparison with alternative situations. One is always into counterfactuals, even with the simplest and most straightforward causal claims.

Thirdly, Schaffer's argument brings it out that a picture of causes and effects as 'events' and causal connections as what lies 'between' these (so that one can easily swap causal process-talk for event-talk, or that both are ways of talking about the same thing) is too simplistic. Yes, any causal de-
6. Causal connection

dependency requires a physical infrastructure. But causally related events, or states of affairs, are not always like two telephones connected by a cable. A physical infrastructure underlying a counterfactual dependency is not like one connection so much as an intricate network of connections and disconnections.

6.6. THIRD CRITICISM: CAUSAL RELEVANCE

Salmon has claimed that his mechanistic approach to causality obliterates the need for accounts like Mackie's, ones that conceive of causality as a matter of dependencies among states of affairs, events, or facts. He suggests that as far as causality is concerned, talk of persisting and interacting processes is enough, and talk of causal dependencies among events or facts at best accomplishes the same thing. In his 1984 (183) he says:

"I have laid particular stress upon the role of causal processes, and I have even suggested the abandonment of the so-called event ontology. It might be asked whether it would not be possible to carry through the same analysis, within the framework of an event ontology, by considering processes as a continuous series of events. I see no reason for supposing that this program could not be carried through, but I would be inclined to ask why we should bother to do so."

But can we indeed have a theory of causality entirely in terms of causal processes? I will now argue that we cannot: that talk of causal connections is not a better way to accomplish what talk of causal dependencies tries to accomplish, but that connections-talk and dependencies-talk are about different things and that a theory of causality ought to consist of both.

The limitations of Salmon's account come to the surface when we realize that there is a variety of levels at which causal explanations can be given. Salmon himself suggests that the issue of explanatory levels is merely pragmatic (1994, 258):

"As an anonymous referee remarked, "You'd want to say that the speeding bullet transmits energy-momentum from the gun to the victim, but what about its incessant, negligible interactions with ambient air and radiation?" (...) Pragmatic considerations determine whether a given "process" is to be regarded as a single process or a complex network of processes and interactions. In the case of the "speeding bullet" we are not concerned with the interactions among the atoms that make up the bullet. (...) It all depends upon the domain of science and the nature of the question under investigation."
Part II. Causality

But the problem is that relations between explanatory levels are typically far more complicated than is suggested here. The bullet may have causally relevant properties other than energy-momentum: it may be poisonous, for instance. And poisonousness does not seem to be a conserved quantity. Now, there are innumerable such properties. Being red, being solid, having sharp teeth, being a notary, believing that it is five o'clock, etc. are all properties that in ordinary causal discourse have unquestioned causal relevance, but that obviously are not conserved quantities transmitted in causal interactions. Salmon's account does not, then, account for the 'efficacy,' or explanatory relevance, of properties that are not physical quantities. Hitchcock (1995, 319) writes:

In the new mechanical philosophy, as in the old, the explanatory store contains nought but geometrical properties.

One might try to argue that relevance in causal explanation is not the same thing as causal efficacy, and that no matter which causal relevance pertains to higher-level properties, it is just the physical quantities that have genuine efficacy. But such a move requires a viable notion of efficacy, and therefore of causation. For 'efficacy' and 'causation' are perfect synonyms. Without such a notion, any attribution or denial of efficacy must remain hollow. But whether or not a purely mechanistic account can provide a viable notion of causation is precisely what is at issue. In any case, Salmon himself has not tried to make the move just suggested, but has instead conceded Hitchcock's point (1997, 469-76):

There is nothing wrong with this sort of geometrical approach as long as its limitations are recognized. It furnishes something like a model of a telephone network that exhibits the lines of communication and the connections. (...) It does not, however, reveal anything about the messages that are sent. (...) As a result of Hitchcock's analysis, I would now say (1) that statistical relevance relations, in the absence of connecting causal processes, lack explanatory import and (2) that connecting causal processes, in the absence of statistical relevance relations, also lack explanatory import.
6. Causal connection

A theory of causality merely in terms of physical connections, then, is onesided and needs supplementing.\(^45\) Note that Salmon's statistical relevance relations are different from Mackie's counterfactual dependencies. The former are mere regularities (and therefore not explanatory), the latter are dependencies among particular states of affairs (and explanatory in their own right). Yet, an instance of a statistical relevance relation that has an underlying physical connection may well be in fact a counterfactual dependency.

6.7. FOURTH CRITICISM: THE ONTOLOGY OF CAUSAL PROCESSES

My final point is about ontology. In his 1984 (140) Salmon claims that material objects are specimens of causal processes. And indeed, he seems to regard processes as genuine entities. But a process as he defines it in his 1994 is a mere space-time path of a physical property, and although we might, depending on how liberally we apply the term, call such a path an ‘entity,’ or a ‘particular,’ it is not concrete. A concrete particular will typically have several persisting properties at a time. In Armstrongian terms, we will have to say that a causal process is a state of affairs (or, less attractive, a continuous series of states of affairs),\(^46\) rather than a material object.

But then, can we not regard concrete particulars as complex bundles of causal processes? We might see it this way: my desk, realized as it is by buzzing subatomic activity, is sustained and kept in shape by gravitational, electromagnetic and nuclear forces, and these may all, in fact, be causal processes. Also, if we describe a specific set of such causal processes, we have described my desk; although, of course, the desk will have many properties in addition to its physical ones (for instance, its being a desk).

The problem with this suggestion, however, is that it amounts to the claim that particulars are mere bundles of properties. When we are saying that objects are causal processes, or bundles of them, we are in effect saying that processes are the basic entities. But these would have to be property

\(^45\) See also Kitcher 1989 (470-1) for a criticism along roughly the same lines as the one presented here.

\(^46\) Depending on whether we can reduce properties constituting ongoing processes to properties constituting brief events or states. Although such a reduction may succeed sometimes (e.g., when we reduce ongoing hammerings to blows), treating continuous processes as successions of events, or of snapshot-like states, may well be unwarranted.
instances (for instance, instances of charge) without a particular that has the properties. This view can perhaps be supported by a bundle theory of properties, or tropes; but adopting such an ontology of properties is at least controversial.

I must admit that the ontology of causal processes is a hazy issue anyhow. We may insist that properties have concrete bearers, but what are these in the case of fields or light beams? It is unclear whether here a particular-property distinction makes sense to begin with. However, once we start to speak of persisting properties, we do seem to presuppose such a distinction. If so, causal processes as Salmon defines them seem to have the ontological character of states of affairs rather than of concrete particulars. But an interaction between two processes (a ball smashing a window, etc.) is clearly a relation between concrete particulars. For note how strange it sounds when we say that, for instance, a thing's mass interacts with another thing's momentum (or something like that). True, an interaction will certainly take place in virtue of some, and not other properties of the particulars. But it is the objects that interact, not their properties.

6.8. WHAT CAUSAL CONNECTIONS ARE

The great achievement of Salmon's theory of causal processes has been to revive a way of thinking about causality that appreciates that underlying the regular-irregular jumble of states of affairs in our world there is a nomadic physical infrastructure in and among objects that makes things go the way they do; an infrastructure that was not given due credit by Humean theories, to their own disadvantage (mainly, the problem of pre-emption). We have also, however, found reasons to complain: the physical connections underlying the counterfactual dependencies that we call 'causal' are not exactly what Salmon tells us that they are. Let us now see what a modified account might look like.

First, although causal connections may in fact be spatiotemporally continuous, that is not what makes them causal. Causal connections are (gappy or non-gappy) paths of physical properties through space-time with a very definite location - namely the location of the particulars that have the properties -, but what makes the connections causal is the lawfulness of the properties. Assuming that Salmon and Dowe are right on this, we can say that it is conservation laws that make up causal connections. For due
6. Causal connection

to these laws some properties can be said to be propagated or transmitted. Causal influence is then, in effect, (physical-) property conservation.

Secondly, laws and counterfactual dependence are closely related. The above has as a consequence that causal connections are constituted by counterfactual dependencies, in the sense that an earlier stage of a causal connection is a circumstantially necessary condition for the later stages. We had better not, however, think of causal processes as causal chains of states of affairs: there does not seem to be a basis for delimiting separate stages of processes. I endorse Salmon's claim that processes are causal lines, and not chains. Still, when we say that a physical infrastructure 'underlies' some counterfactual dependency, we should realize that the infrastructure in its turn consists of further counterfactual dependencies (more on this in the next chapter).

Thirdly, what underlies a causal dependency may be a disconnection as well as a connection. This shows us that underlying causal dependencies there typically is a network of physical connections and disconnections, rather than one straightforward connection. It also shows us once again that causation is an essentially counterfactual notion: any causal claim is a tacit comparison between what actually does happen and what might have happened, and in this respect positive and negative conditions are on the same footing. Not only do we routinely call negative circumstantially necessary conditions causes; as Schaffer's argument stresses, we also routinely invoke disconnections.

Fourthly, a causal connection is not an entity, and neither does it interact or transmit causal influence. Concrete particulars do these latter things. A causal connection is not a transmitter of causal influence, but the causal influence transmitted; and it is not a thing interacting with other things, but a state of affairs in virtue of which interaction takes place. A causal connection is a path of causal influence (or causal influence along a path); not a concrete entity.

The above characterization of causal connections/disconnections, of which I claim that these 'underlie' the counterfactual dependencies that we call 'causal,' implies that these connections are themselves constituted by counterfactual dependencies. But if all is a matter of such dependencies, why the talk of 'physical infrastructure,' 'connections' and 'underlying'? And why my claim that these are fundamentally different aspects of the causal relation? I will address these questions in the next chapter, where I
Part II. Causality

will propose a way to unify the counterfactualist and the mechanistic approach to causality.
7

Unifying dependence and connection

7.1. Physically linking causes and effects: Kitcher

In the previous chapters we have seen that neither a counterfactualist nor a mechanistic account of causality is adequate on its own resources. The two kinds of account really seem to clarify important aspects of the phenomenon, but if we want a full analysis, the two approaches should supplement each other. Kitcher's conclusion (1989, 472) is different:

I suggest that we can have causation without linking causal processes (...). What is crucial to the causal claims seems to be the truth of the counterfactuals, not the existence of the processes and the interactions. If this is correct then it is not just that Salmon's account of the causal structure of the world needs supplementing through the introduction of more counterfactuals. The counterfactuals are at the heart of the theory, while the claims about the existence of processes and interactions are, in principle, dispensable. Perhaps these notions might prove useful in protecting a basically counterfactual theory of causation against certain familiar forms of difficulty (problems of pre-emption, overdetermination, epiphenomena, and so forth).

Now, the view that I defend is that indeed a theory of causality has to be based on counterfactual dependence, but that the linking processes are not, as Kitcher suggests, dispensable or at best helpful, but on the contrary are as essential to causal relations as are the counterfactual dependencies. Circumstantial counterfactual dependencies and physical connections, I will argue below, are two sides of one coin.

7.2. Physically linking causes and effects: Mackie

We have already seen that Mackie's view also tends in that direction. Mackie claims that causation is a relation of counterfactual dependence among facts, but he also claims that causal necessitation is a form of lawful persistence (1974, Ch. 8). He writes (216-8):

The suggestion, then, is that we advance beyond a view of causation as mere regular succession when we conjecture that there really is some causal mechanism underly-
Part II. Causality

ing the succession and explaining it. (...) The cause-event is identical with something unobservable that has a certain mathematically describable structure. The latter develops, necessarily, into something else which is also unobservable and also has some such structure. The latter something is identical with the effect-event. (...) Mathematically there will be some form of description applicable to both the earlier and the later phases of the developing structure; the two particular descriptions of this form will have identical values for some of the quantities (...) while the values for other quantities in the latter phase will be derivable from the earlier values with the help of the integration over time of some differential equation, that is, some law of the functional dependence variety.

The attraction of this view is that it shows us how causal dependence can be a matter of natural, i.e., physical law without committing us to the problematic assumption that there must be specific causal laws covering any type of cause and effect (I will have to say much more on causal laws in the next chapter). Unfortunately, I cannot straight-away accept Mackie’s account as it stands. Not because I consider it mistaken (I consider it nearly right), but because it leaves a crucial issue unclarified, and also because it may involve a wrong suggestion. The unclarified issue concerns the ontology of causes and effects; the wrong suggestion concerns the nature of the causal relation.

Let me explain. Mackie speaks of INUS conditions and their effects as facts, but he also considers causes as events (see quotation), and speaks of ‘causation in the objects’ (223). Now, if we ask how these are related we get an answer that is somewhat elusive. Mackie illustrates the way in which a physical process ‘underlies’ a causal dependence ‘at the perceptual level’ (221) by the example of a match that is struck and ignites (see the quotation earlier on). This example suggests that the causal dependencies that we refer to in ordinary causal discourse and the physical processes underlying them are related in a macro-micro way. On such a picture, discovering underlying physical processes will then be a matter of ‘micro-reduction’ (222). The example reminds one of Salmon’s speeding bullet (see above). Examples like these work fine for straightforward causal interactions among macro-entities, but they leave us in the dark about cases where the relation between dependency and connection is more opaque (for instance, cases of mental, or social, causation).

Now, I believe that this is not a minor issue. If we do not push the matter any further, examples like those of Mackie and Salmon will leave us with the impression that what we ordinarily call causal relations are (counterfactual or statistical) dependencies among facts, states of affairs, or
7. Unifying dependence and connection

events, while the real, scientifically and metaphysically respectable causation takes place at the level of the physical processes (in Mackie's case, a development of a cause-event into an effect-event; in Salmon's case a space-time trajectory of a conserved quantity). But that impression is misleading. It is precisely the picture of causation that has made mental causation look so problematic, and that I aim to dislodge.

7.3. PHYSICALLY LINKING CAUSES AND EFFECTS: DOWE

A more recent attempt to deal with these matters is Dowe (2000). Remember that Dowe is the philosopher who convinced Salmon that a 'conserved quantities' theory of causal processes is right. Dowe explicitly addresses the question how what we normally call causes and effects are physically linked, and he does so endorsing, like me, Armstrong's state of affairs ontology, regarding causes and effects as particulars' having properties at times. And because he raises the issue of causation at the 'manifest' level, one expects him to deal with the issue of 'causal relevance,' as discussed above.

Dowe even starts out by rejecting what he calls the 'naïve process theory,' according to which 'two token events are connected in a causal relation if and only if a continuous line of causal processes and interactions can be traced between them' (146); for he correctly notes that on such a view causes and effects can be 'misconnected' (147), a telling problem that I will myself below address as the 'pairing problem.'

Dowe deals with these issues by assuming that causally related facts 'at base involve' possession of, or change in, a conserved quantity (171): which in his account amounts to the assumption that possessions of, or changes in, a conserved quantity are the only thing of causal relevance. That assumption yields the following account of the way in which causes and effects are physically linked (171-2):

Causal connection: there is a causal connection (or thread) between a fact q(a) and a fact q'(b) if and only if there is a set of causal processes and interactions between q(a) and q'(b) such that: (1) any change of object from a to b and any change of conserved quantity from q to q' occurs at a causal interaction involving the following changes: Δq(a), Δq(b), Δq'(a), and Δq'(a); and (2) for any exchange in (1) involving more than one conserved quantity, the changes in quantities are governed by a single law of nature. (...) (C)ausation can occur at the manifest level. If c_m, e_m
Part II. Causality

are manifest facts, and \( c_p, e_p \) are physical facts, then \( c_m \) causes \( e_m \) only if \( c_p \) causes \( e_p \) and \( c_m \) supervenes on \( c_p \) and \( e_m \) supervenes on \( e_p \).

With our criticisms of Salmon in mind, we can easily guess why this is not satisfactory. First, the issue of causal relevance at the 'manifest' level is not really addressed. The most that Dowe's account can offer us here is the supervenient causation of token (event or state) physicalism, which implies that there is no genuine causal work left for higher-level properties. Secondly, this account is as vulnerable as Salmon's to Schaffer's problem of 'causation by disconnection' (see above). It is true that Dowe does address the related issue of causation by 'omission' (Ch 6), even in a way that I find quite plausible (I will myself deal with alleged 'negative' causes and effects later on). But causation by omission is not causation by disconnection; nor, it seems, can we construe the latter as causation by omission. If, for instance, I pull a trigger and cause someone to die, that is causation by disconnection, but not by omission.

7.4. FROM RELATIONS AND MECHANISMS TO DEPENDENCIES AND CONNECTIONS

Considering the failure of the above accounts, my alternative suggestion is this: from a picture that distinguishes causal mechanisms from causal relations, we have to move to a picture that distinguishes causal dependencies from causal connections. At first sight there may not seem to be much of a difference, but on a closer look the pictures turn out crucially different.

Here is how we can make the move. We start by considering two seemingly contrary kinds of views on causal relations: counterfactualist and mechanistic ones. Then we note that in their pure forms both kinds of views face what may well be insurmountable difficulties. At least, this is what I think that the previous chapters teach us. We then come to suspect that causal relations involve counterfactual dependence as well as physical connection. Not just higher-level causal relations, but all causal relations, at all levels, including the physical.

We then realize that this goes for what we normally call causal 'mechanisms' as well. Two particles collide in a particle accelerator. A plant grows. A rat pushes a lever. Two lovers quarrel. A president is elected. These are all causal mechanisms, no matter whether they concern physical, biologi-
7. Unifying dependence and connection

cal, intentional or social causes. They all involve dependencies among states of affairs as well as connections among concrete particulars. It is not as if scientifically respectable physical mechanisms underlie, or realize, merely 'manifest' causal relations. That picture of causation is mistaken. We have causal relations, or mechanisms, at various ontological levels: distinct, and all equally real - at least wherever reduction fails. And the mechanisms consist of 'own-level' dependencies as well as of an underlying physical infrastructure: all of them, not just the higher-level ones. Thus, 'mechanistic' theories of causality are not occupied with 'mechanisms' in this sense, but with physical connections that are only one aspect of genuine mechanisms. This is why I have taken care not to call physical connections 'mechanisms.' (Even so, the terminology may be confusing.)

To triangulate, let us shift our perspective for a while, and see how an ontological point of view yields the very same picture. Remember that states of affairs, the items related by causal dependence, are particulars having properties at times, the particulars being the relata of physical connections (persistencies and interactions) - or disconnections. Remember also our discussion of particulars and properties: we saw that there is a conceptual relation between particulars and properties, in the sense that the notion of a featureless particular and that of a property without a bearer makes no clear sense.

Now, a state of affairs conceived of as a particular's having a property at $t$ is a mere snapshot, as it were, of a particular under an aspect. It is an abstraction, for real-world states of affairs are not just at times, but have duration. We should therefore take time into account. If we do so, we get a world of particulars with properties and relations in which persistence, development and sudden change take place. Particulars come into being or are destroyed, lose or acquire properties, and change relations. And all this persistence, development and change takes place according to physical law (this at least we can assume).

We now have in fact added causation to an ontological picture that initially contained only particulars and properties. What brought causation into the picture was law and (if it is a separate factor) time. On the present picture, instances of properties and relations (states of affairs) are circumstantially necessary conditions for other states of affairs, while the particulars involved are physically connected to other particulars (when they interact), and to themselves at earlier and later times (when they persist). The idea is that just as we say of bare existence that particulars and proper-
ties presuppose each other (but are not the same thing), we should say of causation that physical connection and counterfactual dependence presuppose each other (but are not the same thing). The conceptual pair of particular and property is closely related to that of connection and (causal) dependence; it is the same one when you subtract duration and law.

Note that the picture is not such that there are physical connections at some basic micro-level and counterfactual dependencies at higher levels. No: if we have a physical connection, constituted by, say, the law of conservation of energy (or whatever), a situation at an earlier stage (or a later one if we believe in backward causation) is a circumstantially necessary condition for situations at later stages - at least, if Salmon's at-at theory fails this may be the only viable alternative.

This may look like a reduction of physical connections to causal chains, and I admit that the two come extremely close. Yet, physical connections do not, as such, seem to have a direction, while the causal relations making up the alleged chain do have one. A trajectory of conserved energy, for instance, may just be there in space-time without there being a difference in whether it runs from A to B or vice versa. Even so, the circumstantial counterfactual dependency among stage e and c of that pathway will have a direction. I have not put forward an account of causal asymmetry, so I cannot explain where the direction comes from, but various accounts of causal asymmetry are available.47

It is true that only a very limited range of properties will be relevant to physical connections, while a virtually unlimited range of properties will be relevant to counterfactual dependencies. This should not tempt us, however, to claim that only physical properties are causally relevant, or 'efficient.' 'Efficacy' is a term that presupposes full-blooded causation, and as we have seen, there is more to causation than mere physical connection.

7.5. THE PAIRING PROBLEM

Causation, then, is circumstantial counterfactual dependence that has an underlying infrastructure of physical connections and disconnections. But what is 'underlying'? Consider the following causal statement: The passing of the impopular law caused the downfall of the politician (i.e., state of affairs c is a circumstantially necessary condition for state of affairs e).

47 See my earlier note on causal asymmetry.
Hardly anyone, I trust, will doubt that this dependency among states of affairs will have an underlying physical infrastructure. But how do we pick out that infrastructure? And how are cause and effect related to it? We saw that both Mackie and Dowe have an account of the physical link among such a cause and effect, but if we reject the token identities that these accounts presuppose, we need an alternative.

And we had better provide a clear and satisfactory one: for so far we have just the claim that causes are circumstantially necessary conditions and that everything is physically realized (so it all happens in virtue of the laws of physics). That is still standard counterfactualism. If we want to improve on this in claiming that any causal relation is in fact a counterfactual dependency-plus-physical connection, we ought to be able to pair dependencies and connections. Which connections underlie which dependencies? What principle enables us to tell? That is, if we have indeed an adequate analysis of causal dependence, then we ought to be able to fully specify what particular causal relations consist in. Can we do this?

Let us start with an example of a straightforwardly physical causal relation: A flying ball breaks a vase. Let us suppose that internal make-up, momentum and spatial trajectory of the ball are all circumstantially necessary conditions for the vase's going to pieces. Which causal processes underlie this counterfactual dependency? Eberhardt (1999) tells us that the breaking of brittle materials is due to the fact that any stress on such mate-

---

48 A theory of causation that attempts to come to terms with causal relations as we know them as well as with continuities in nature is Ehring's (see Ehring 1997). Ehring elegantly defines causation as trope persistence. Could we not perhaps adopt Ehring's crisp formula and dispense with the dependency-connection duality? I think not, for Ehring's theory does not seem able to satisfactorily deal with change. For trope persistence seems to be precisely the absence of change. Ehring presents a solution to this problem in introducing the notions of trope fission, fusion, creation and elimination. He writes (1997, 124-5):

'Consider an everyday example. Jones cuts a string with a pair of scissors. The effects are the two separate-length tropes of the cut string pieces, and a cause is the movement and sharpness of the scissors. The causal process corresponds to unstable fusion followed by fission. The string is characterized by a structural trope - its length, which consists in two string half-lengths being in a certain relation to each other, the relation of 'being attached.' The scissors are characterized by a movement/sharpness trope. (...) The 'attached' relation is eventually eliminated.'

My general objection to this account of change is, that it is just too complicated and far-fetched to be credible. More specifically, the tropes 'persisting,' 'fusing,' 'fissioning,' being 'created' and 'eliminated' seem properties (i.e., tropes) themselves. If so, an infinite regress threatens when these higher-order tropes also persist, fuse, etc.
Part II. Causality

Materials concentrates in the tips of microscopic cracks. The molecular bonds at these locations will then be disrupted, which further widens the crack, as a result of which further molecular bonds are disrupted (in non-brittle materials the stress is relieved by atoms' being able to slide a little past one another, and in this way to blunt the crack).

This explains the macro-breaking by referring to micro-breakings. But Eberhardt proceeds by explaining the micro-breakings themselves in terms of variations of charge densities. The breaking of molecular bonds does not consist in the disruption of bonds among even smaller particles. The Encyclopaedia Britannica tells us that 'if the total energy of a group of atoms is lower than the sum of the energies of the component atoms, then they bond together and the energy lowering is the bonding energy.' In other words, the breaking of molecular bonds is a matter of spatial rearrangement of electrons and nuclei in such a way that the lowest possible energy state is achieved. The breaking of the vase, therefore, can be seen to be at base a matter of energy flow and spatial rearrangement.

So far so good. But what about the case of the unpopular law and the politician? It is obvious that in such a case an account like the above is out of the question. The breaking of a vase is a local and merely physical event, the interacting particulars can be easily identified, and the counterfactual dependencies can be linked quite straightforwardly to the rule of physical laws. But the passing of a law is a highly relational affair, as is the downfall of a politician. These are social states of affairs with presumably a very wide physical base. According to what principle can we relate counterfactual dependencies among such higher-level states of affairs to physical connections among particulars? Causal dependencies and causal connections obviously are not neatly paired one to one: there will typically be many connections and disconnections underlying one dependency. Also, as we have heard Salmon concede, causal connections will often tell us very little about which dependencies they underlie.

It is not difficult to establish that the causal infrastructure connecting $c$ and $e$ must connect at least one pair of particulars. For if it is claimed that two states of affairs are causally related (say, a magician's pointing his staff and a vase's going to pieces) and no physical infrastructure whatsoever links the particulars involved in these states of affairs, there is no reason to believe that the states of affairs' co-occurrence is more than a coinci-
7. Unifying dependence and connection

dence. But the connection among the particulars may be indirect rather than direct (someone seeing the magician pointing his staff might shoot the vase in reaction); we might have, in fact, a disconnection (the magician might by pointing his staff remove an obstacle to a vase-shattering process) and there will typically also be irrelevant connections (such as gusts of air).

Discussion of merely physical examples like the one with the vase above, then, can be seriously misleading if it is suggested that something general can be learned from them. What we want is a general principle of relating a physical infrastructure to a causal dependency, and an example like the one above is not going to provide such a principle; not, at least, one that can be carried over to causal relations with a wide physical base.

49 Discovering that nevertheless the magician can and does 'shatter' vases at will would be a source of puzzlement, comparable with the phenomenon of quantum entanglement, described above.

50 On our search for such a general pairing principle, we should also realize that dependency-connection relations can be rather diverse. Consider two causally related states of affairs, c and e. It seems that c and e may be either simple states of affairs involving just one particular each (e.g., the rolling downhill of a boulder), or complex states of affairs involving multiple particulars, properties and relations (e.g. an avalanche). If so, we seem to have four possibilities: 1) c and e are both simple; 2) c and e are both complex; 3) c is simple and e is complex; and 4) c is complex and e is simple.

Now consider the particulars, or sets of particulars, a and b, involved in the causal relation c-e. It seems that a and b can be 1) fully distinct (as in the case of bouncing billiard balls); 2) identical over time (as in the case of a green banana turning yellow); 3) overlapping (as in the case of the boulder's causing an avalanche); 4) related as part-whole (as in the case of a soccer player making the goal that renders his team world champion.

One might think that there is an option 5), in which the particulars are related as whole-part (as in the putative case of a rolling boulder causing one of its surface molecules to follow a certain trajectory); but 5) is not causation. The reason is this: we can enumerate all the circumstantially necessary conditions for, say, the movement of the molecule (for instance, the movements of its neighbouring molecules); and if we do so, the rolling of the boulder will (by the free lunch principle) certainly turn out not to be an additional causal factor. And this is not surprising: the example would violate both causal closure and the distinct existences requirement. Yes, the boulder's rolling certainly determines, constrains, steers, programs, guides, or even explains the movement of the molecule; but it would be a mistake to call that causation.

The picture that we get is one of causal dependencies among complex and simple states of affairs indifferently, and of causal connections or disconnections among micro- and macro-particulars indifferently (with the exception of whole-part causation). This picture suggests that, in different singular cases of causation, the relation between counterfactual dependence and physical connection may be simple and straightforward, but will more typically be very indirect, opaque, and gerrymandered.
7.6. THE PAIRING PRINCIPLE

I order to come by the desired principle, let us consider an example. Suppose that you look at a radioactive tomato. The tomato emits X-rays as well as it reflects light. It affects you in two ways: the X-rays cause cataract of your eye lens, and the light causes a visual impression. There are, then, two causal dependencies: your eyes’ facing the tomato causing cataract, and your looking at the tomato causing a visual impression. There are also two distinct causal connections: the light and the X-rays. Yet, there is only one pair of particulars: you and the tomato. How, then, are the connections and the dependencies paired in the right way?

This case is instructive. For the only appropriate kind of answer to our question seems to be one in terms of counterfactual dependence, and this immediately suggests a general solution to the pairing problem. We knew all along that the X-ray connection belongs to the radioactivity-cataract dependency, while the light connection belongs to the redness-visual impression dependency. How? No doubt by appreciating that without an X-ray connection not even the most radioactive tomato will harm the eyes of an onlooker (you may look at the tomato with a telescope); and that without light not even the reddest tomato will give one a visual impression. The general principle of pairing causal connections and dependencies will then be the following:

Pairing principle: A physical connection, or disconnection (or set of these) $p$ among particulars (or sets of particulars) $c$ and $e$ underlies a counterfactual dependency $c$-$e$ among state of affairs (circumstantially necessary condition) $c$ (involving $c$) and state of affairs (effect) $e$ (involving $e$) iff $p$ is a circumstantially necessary condition for $c$-$e$.

It can be imagined, of course, that only radioactive tomatoes are visible, or that only visible tomatoes are radioactive. In such hypothetical cases the pairing principle does not work. In reaction, we can imagine experiments to pry the causal links apart. About cases in which this is said to be impossible, we can then claim that it is just indeterminate which connection underlies which dependency (if speaking of two different connections makes sense in the first place).

Let us now apply the pairing principle. Can we determine which physical connections belong to the counterfactual dependency described above
7. Unifying dependence and connection

(of the impopular law, call it c, and the downfall of the politician, call it e)? Yes, in principle we can do so. We have already seen how the physical realizers of even such higher-level states of affairs can be demarcated. What we can do now is to enumerate the particulars involved in those realizers (the presses, the people, the televisions, the politician, and so on), and also the physical connections and disconnections among them. After that we have to figure out, for this particular case, which minimal set of those connections and disconnections was a necessary condition for the counterfactual dependency c-e to obtain.

So we ask: Given the situation, which minimal set of (interruptions of) sound waves, electrical currents, object persistencies, trajectories of momentum, light beams, etc. was required for it being the case that without the passing of the impopular law there would not have been the politician's downfall? The question is tricky: we are not asking for a cause, but for a necessary condition for a relation of counterfactual dependence.

In spite of some complications that I will discuss below, the pairing principle in itself is quite simple: it just tells us that what makes specific counterfactual dependencies and pieces of physical infrastructure belong together (in together making up a full-blooded causal relation) is (circumstantial) counterfactual dependence. Just that. But the principle is not trivial, in that it uncovers one more way in which the notion of counterfactual dependence is indispensible for understanding causation. Counterfactual dependence is really at the heart of our notion of causality: causal dependencies are counterfactual dependencies; the underlying physical infrastructure consists of space-time trajectories determined by counterfactual-sustaining conservation laws; and the pairing principle tells us that the relation between the two is also counterfactual dependence.

Admittedly, nested counterfactual dependencies such as I have proposed them are a bit awkward. In the possible worlds that they presuppose, the states of affairs that in the actual world are cause and effect do obtain (c and e respectively), but e is not dependent on c. For a pairing of a piece of physical infrastructure with a counterfactual dependency c-e, the claim is, in Lewisian terms, that a possible world in which the infrastructure is lacking and in which e is not dependent on c is closer to the actual world than one in which the infrastructure is lacking and in which e is nevertheless dependent on c. We have here two tiers of possible worlds to reckon with, and establishing relative closeness will not be simple. But in my view, this does not invalidate the pairing principle.
Part II. Causality

7.7. NEGATIVE CAUSAL FACTORS

The above summarizes my account of causal relations. One may, however, suspect that on a closer look this account is invalidated by complications that derive from ‘negative’ causation: causation by omission or prevention. Moreover, the issue of physical infrastructure is complicated by the fact that causation by pre-emption and disconnection are possible. So let me indicate how, in general, such cases should be approached on my analysis.

First, note that my account of causation does not straightforwardly represent cases of causation by omission or prevention. A doctor’s absence may cause a death, and a doctor’s presence may prevent a death; but in these cases, there is no concrete particular that constitutes cause (omission) or effect (prevention), let alone a physical connection or disconnection.

Does this block my attempts to define causation? In my view, such ‘negative’ causation teaches us a more modest lesson. I will later on, when discussing causal relata, argue that causation by omission is not in fact a relation among states of affairs, and explain our ordinarily taking it for a relation by pointing out that causal discourse is always counterfactual, comparing actual to merely possible situations. This implies that causation by omission or prevention is not what it seems: it is not a causal relation of which one relatum is negative in character. Instead, it is causation by, or of, something else, where it is significant that (barring pre-emption) if the absent item had been present, things would have gone otherwise. I consider this a satisfactory solution that is very much in line with the counterfactualist approach that I endorse; although I have to admit that to a certain extent it violates ordinary usage.\(^{51}\)

Secondly, pre-emption and causation by disconnection raise questions about physical infrastructure. At first, it seems that cases of pre-emption are easily handled by a modified pairing principle:

**Pairing principle for pre-emption:** A physical connection, or disconnection (or set of these) \(p\) among particulars (or sets of particulars) \(c\) and \(e\) under-
lies a counterfactual dependency \((c \lor d) - e\) among disjunctive state of affairs (circumstantially necessary condition) \(c \lor d\) (involving \(c\) and \(d\)) and state of affairs (effect) \(e\) (involving \(e\)) iff \(p\) is a circumstantially necessary condition for \((c \lor d) - e\).

This distinguishes the pre-emptor from the pre-empted, because \(p\) is among the particulars \(c\) (rather than \(d\)) and \(e\). So, never mind that circumstantially necessary conditions can be disjunctive: there will be a physical link from the pre-emptor to the effect, and not from the pre-empted to the effect.

But the worry may be that, because an underlying physical infrastructure may be more than just one simple straightforward link, this infrastructure will sometimes involve many particulars, among which, perhaps, the particular constituting the pre-empted factor. If so, talking about physical infrastructure will not help us to distinguish pre-emptor from pre-empted.

As an example, consider the following case. Suppose that a missile will be launched after a button is pushed and a key is inserted. The button can be pushed only once. The button is accessible to several people, among whom the president, but only the latter has the key. One day, both a general and the president try to push the button. The general is faster and pushes the button, thereby pre-empting the president’s efforts. Yet, the president has to insert the key in order for the general’s action to cause the launching of the missile. If so, it seems that the infrastructure underlying the dependency among button-pushing and missile-launching includes the president as well as the general and the missile. Thus, picking out the infrastructure seems to fail to distinguish pre-emptor from pre-empted.

Causation by disconnection seems to complicate the demarcation of the underlying physical infrastructure. For causation by disconnection, implies an absent physical link. Suppose that I kill the president by shooting the guard who would have shot the assassin who shoots the president.\(^{52}\) How to demarcate the physical infrastructure underlying this killing? Must the gap between guard and assassin, the crucial absence of a connection, be

---

\(^{52}\) I owe this example to an anonymous referee. We may ask whether it is not, rather, a case of prevention. In a case of prevention one has intermediate links in a causal chain, while in a case of causation by disconnection speaking of such intermediate links would be a confusion of levels. In the example I can, perhaps, be said to kill the president; but it may be more plausible to say that I prevent the guard from saving the president. I am accessory, of course, but the person who kills the president is the assassin, not I.
counted? And the guard’s bullet? How much of the assassin’s past must be included?

In order to deal with these complications, let me suggest three criteria for narrowing down what we should consider a physical infrastructure. First, I want to claim that a suggested piece of infrastructure underlying a counterfactual dependency ought not to belong to the circumstances that must be assumed fixed for the counterfactual dependency to obtain. For instance, we may argue that in the example above, the president’s inserting the key belongs to the circumstances that must be assumed fixed for the button-pushing to cause the missile-launching. A good reason for claiming this is, that the president’s inserting the key is a cause of the missile-launching in its own right, with its own underlying piece of physical infrastructure involving president and missile. If we say that this very same piece of physical infrastructure is a part of the infrastructure underlying the button-pushing/missile-launching dependency, we double-count. If this first criterion is sound, pre-emption will be less of a problem than it may have seemed.

Secondly, I want to claim that a piece of physical infrastructure can involve either a connection or a cut in a connection, but not a dead end, such as the guard’s bullet. After all, causation is either by connection or by disconnection. Also, the piece of physical infrastructure must originate, not anywhere, but with the particular constituting the cause. This suggests that the past of the assassin is not relevant; only at the point where the guard’s shooting would have made a difference to the assassin (i.e., from a possible but fatefully absent interaction onward) does he enter the infrastructure.

Thirdly, I want to claim that a piece of physical infrastructure may involve counterfactually significant absences of either connections or cuts, such as the absence of a bullet trajectory between the guard and the assassin, as well as actual connections and cuts. An absence here is a merely possible piece of physical infrastructure that would have underlied a counterfactual dependency that is precluded by the actual one. For instance, the bullet-trajectory between guard and assassin is a merely possible piece of physical infrastructure that would have underlied the counterfactual dependency shooting-by-guard/dying-of-assassin, a dependency that is precluded by the actual one (dying-of guard/shooting-by-assassin).

---

53 The absent, or merely possible, connections involved in causation by disconnection create spatiotemporal gaps between putative causes and effects, so that cases of causation by disconnection seem to violate the traditional spatiotemporal contiguity requirement for
7. Unifying dependence and connection

In this way, the notion of an absent connection or disconnection raises the same ontological problem as the negative causes and effects of cases of omission and prevention; but it can be handled in the same way. Non-actual connections and disconnections, and non-actual states of affairs are part and parcel of any causal relation, even those not involving disconnection, omission or prevention; not in the sense that they are somehow negative entities, but in the sense that causal relations are essentially counterfactual, always involving a contrast between the actual and the merely possible.

7.8. WHY WE THINK CAUSALLY

If the previous arguments have been right, causation is circumstantial counterfactual dependence with an underlying physical-nomic infrastructure. Now, the above has been rather laborious. Yet we, perceiving and acting creatures, experience causation in a very direct way, and have an intuitive grasp of it. Presumably, our ancestors have known what 'efficient' causation is long before philosophers began to theorize about it. Why, then, should it not allow for a more unified analysis?

I think, however, that there are some facts about our condition as living creatures that suggest that my attempt at an analysis is in the right direction. First, we are creatures with many behavioural options. And the more such options a creature has, the more it will have to anticipate phenomena that are important, in a positive as well as a negative sense, to its survival, well-being and reproduction. So a successful complex species can be expected to have a good appreciation of the ways in which such phenomena hang together with other phenomena; especially a thinking species (it is the very thing that thinking is for).

But why should our thinking be biased towards dependencies with underlying physical processes? Supervenience is also a form of dependence, but it is not causal. What is so special about causal dependencies that ordinary language is pervaded by (explicit or implicit) mention of them, while a kind of dependence such as supervenience is merely the subject matter of academic philosophical discourse?

causation. However, because those gaps presuppose possible connections bridging them, to allow causation by disconnection is not to reject the significance of the spatiotemporal contiguity requirement.
Part II. Causality

The answer seems obvious. The physical processes that underlie causal dependencies provide the in-principle opportunity to interfere with the course of events. You may want to prevent some event that depends on the occurrence of another event - one you are presently observing (a predator approaching, a flame held near combustible stuff, etc.); but you can only do so if the event that you want to prevent (being eaten, the hut going up in flames) is somewhat later than the event that informs you that something unwelcome is about to happen. And of course, physical processes also allow you to grab opportunities: when you see that certain conditions obtain, you are well off when you realize that you can yourself add maybe just one condition in order to make something wonderful happen. But you need time, if ever so little. Because, then, we need to appreciate how states of affairs depend one on the other, and because time to interfere is so important, it is causal dependence that occupies center stage in our thinking, rather than any old kind of dependence.
8

Causation and natural law

8.1. THE FIRST PROBLEM OF MENTAL CAUSATION

With the above analysis of the causal relation in hand (i.e., causation as a circumstantial counterfactual dependence with a specific physical infrastructure as its circumstantially necessary condition) we have reached a point where we can address the first of the five problems of mental causation, described in the Introduction: the problem of mental anomalism. The problem was that while our beliefs and desires are causes of our actions, they do not seem to be such in virtue of nomically necessitating them (or nomically raising their probability). But given that cause and law are inseparable, how can this be?

I will now, on the basis of the above account of causal dependence and connection, argue that although the concept of causality and natural law are closely connected (so much so that in a sense causality and natural law are virtually the same thing) the notion of a ‘causal law’ should be abandoned. The idea is that, no matter how ‘nomological’ causality is, a’s having F at t1 may well be the cause of b’s having G at t2 without there being anything like a nomic connection between the properties F and G. If so, the non-nomicity of mental properties will in no way be detrimental to their causal relevance, and the problem of mental anomalism dissolves.

I will in this chapter also point out how the counterfactualist-mechanistic analysis of causality that I endorse is compatible with the fact that we have an intuitive grasp of the notion of ‘cause’ and can also directly experience causal relations as such.

8.2. THE NOMOLOGICAL APPROACH TO CAUSATION

As Kim (1996, 139-43) points out, counterfactuals can only be true in virtue of specific laws of nature that support them. In a lawless world, no counterfactual is ever justified. The counterfactual ‘if you were to decrease the volume of this amount of gas, you would (ceteris paribus) increase its pressure’ is true in virtue of, if anything, a law that relates gas pressure and
Part II. Causality

volume. In its turn, the truth of this counterfactual is what makes the inverse correlation of gas volume and pressure lawful rather than coincidental. The relation between counterfactual dependence and lawfulness is, then, very tight.

But it is difficult to say what laws of nature are. The laws of physics are so-called 'functional' laws: correlations among specific determinables such as velocity, mass, charge, and so on. Such laws are expressed by mathematical equations. It seems that such laws describe constraints on possible physical situations or processes: only a limited range of values of physical quantities, given others, is possible.

Often, however, philosophers endorse quite a different conception of laws of nature, namely as laws of causal succession. Such laws are supposed to link subsequent state-of-affairs types, or properties that are involved in causal dependencies. Fales (1990, 52-3) writes:

(3)he view that causal relations relate properties is suggested by the independently developed views of Michael Tooley (1977), David Armstrong (1978b), and Fred Dretske (1977). At the same time, none of these three philosophers would wish to deny that events can be causally related. (...) The two these can be shown to be consistent if we make the distinction between properties, understood as universals, and their instances. (...) If we make this distinction, then we may distinguish between the causal relation understood as a second-order relation between first-order universals, and (derivatively) its instances which obtain between the instances of these universals. (...) Causal relations are, therefore, taken to be second-order relations. This view has the attraction, for those who believe that causal relations entail universal laws, of explaining immediately why this should be so.

The idea here is that if c causes e, and c is an instantiation of the universal F, while e is an instantiation of the universal G, there is a causal relation between the universals F and G. This causal relation is supposed to be a universal (over and above the universals F and G) that, like F and G, may be multiply instantiated. The attraction of this idea is supposed to be that it explains why F-instances are always followed by G-instances and, indeed, must be so followed (under specific circumstances, that is). For because the relation between F and G is itself a universal, it (i.e., numerically the same item) will be present on all occasions when Fs cause Gs. The fact that it is numerically the same item that is present each time is then supposed to explain why Fs have to be followed by Gs: this one universal cannot be
different from one occasion to another, because otherwise it would not be a
universal.54 This view, indeed, links up causation and natural law.

If it is correct, for all causal dependencies each there must be a law of
nature that covers them. It is this approach that has rendered mental anom-
alism a problem: our first problem of mental causation. For if any causal
relation requires a corresponding relation among the properties (taken as universals) that are instantiated, then if there is mental causation, there
must be psychological, or psychophysical laws. If, for instance, I undertake
an action of type G (say, fetch an umbrella) because I have a mental state
of type F (say, believe that it will rain), on the nomological approach there
must be a law of nature that links Fs to Gs, i.e., rain-beliefs to umbrella-
fetchings. But as we saw in the Introduction, there are reasons for denying
that there are such laws.

True, it is not uncommon to hold that there are inaccurate, ceteris par-
ibus clause-ridden, laws of 'folk psychology.' But we will see later on that the
so called 'Theory Theory' of psychological explanation on which this belief
is based is misguided. So let us assume that the domain of intentional phe-
nomena is anomalous. Not in the sense that it is independent of the laws of
nature, but in the sense that mental properties are not subsumable under
any laws of nature. Then, if we want to solve the first problem of mental
causation, we must either render mental anomalism compatible with the
nomological approach to causation, or reject that approach. Davidson's
Anomalous Monism is an attempt to do the former (although his nominal-
ist conception of causal laws is, of course, different from the realist concep-
tion just described); I will myself try to do the latter.

8.3. AGAINST THE NOTION OF 'CAUSAL LAW'

So let me mount an argument against the nomological approach. Earlier
on, we have seen that the physical infrastructure underlying any causal
dependency consists of lawful persistencies in, and interactions among,
particulars; and the laws involved in this are all the laws needed to fuel
causal dependencies. Take the world at any moment; add the basic
physical laws and time (that is, the persistence and change in that world);
and, at least on physicalism, all causation is automatically given along. And
if you leave out the laws, there will be no causal dependency whatsoever.

54 For an account along these lines see also Armstrong 1997, Ch. 15.
Part II. Causality

We might say that causality is lawful persistence and change in a world. But because all of it supervenes on physical reality, physical law underlies all causation.

In the spirit of Russell's attack on the principle of 'same cause, same effect' (1912), let us now see why the prospects for further conceptions of natural law, ones that are supposed to subsume causal relations, are not good and why, therefore, a nomological approach like the one described is misguided.

The source of trouble is the fact that any alleged causal law has to be ceteris paribus. The counterfactual: 'If you decrease the volume of this amount of gas, you will (ceteris paribus) increase its pressure' already contains a ceteris paribus clause, even if it is very straightforwardly linked to a law that relates gas volume and pressure. When we turn the counterfactual into a 'causal law' by substituting 'any gas' for 'this gas,' we see that the ceteris paribus clause is still needed: if, while decreasing the volume of an amount of gas, we also cool it down enough, pressure does not increase at all, so that our causal law is violated.

The reason why ceteris paribus clauses cannot be missed in any causal generalization is that causation is circumstantial. No state of affairs can be the cause it is if the circumstances do not allow it. Causes as such are insufficient conditions. For some causes it might be difficult to find possible counteracting circumstances (it seems that, whatever the circumstances, thin glass always breaks when forcefully hit by a hammer); but if such causal relations are really immune to no matter which circumstance (which we should not believe too soon), they are most plausibly seen as borderline cases of the much more common circumstantiality of other causal relations.

Now, the general problem with ceteris paribus clauses in laws is that they are prone to carry too much weight. We might say, for instance, that there is a ceteris paribus law that relates rook moves to checkmates, meteor impacts to extinctions, assassinations to world wars, and so on. The problem with this is that laws become ubiquitous. Any kind of event, no matter whether it has occurred only once, or only now and then, and in virtue of no matter which exceptional circumstances, will be said to be the manifestation of a very specific law of nature (one that might manifest itself only once in world history). For only in very specific circumstances will we trigger a world war when we assassinate someone, checkmate a person when moving a rook, and so on.
8. Causation and natural law

But a law that is appropriate to world wars will typically specify one circumstance condition out of thousands, leaving the rest to the ceteris paribus clause. Hence my claim that such clauses are prone to carry too much weight. Now, speaking of a causal law while letting the ceteris paribus clause do almost all the work may not be false, but it is in any case vacuous. For stating the law will be uninformative. Note that such laws cannot be interpreted as being probabilistic instead of ceteris paribus. A law that says that smoking makes lung cancer more probable, or that assassinations raise the chances of a world war (if ever so slightly) still needs a ceteris paribus clause, for if such a clause is omitted the probability could be defeated, say by some form of medication in the case of smoking and international supervision in the case of assassinations.

True, there is a way of avoiding ceteris paribus handwaving: making the law unwieldily complex, perhaps applicable to one situation in world history only. But a law that applies to one situation can hardly be called a law. Laws are supposed to be general, or even universal. The point is not that laws can only be derived from a large number of cases. A physicist may, with the help of rare and expensive equipment, succeed in creating an unusually heavy atom only once, or only a few times; and this will by no means keep her from making inductions and formulating laws on the basis of the experiment.

The point is, rather, that the law, if interesting, will not be a causal generalization. The reason is that any attempt to formulate such a law will either result in near vacuity or in endless complexity. In being more explicit about circumstances we will diminish the vacuity of our law, but at the same time limit its scope. And no matter how much complexity we allow, it seems that there is no end to possible disturbing factors that only a ceteris paribus clause could deal with. What a ceteris paribus law says, in effect, is that if a total situation like the one under consideration were to recur, an effect like the present one would recur, too. Claiming this is not really vacuous: it amounts to claiming that nature is lawful, and that is a claim with great metaphysical weight. But as a claim about specific causes and effects it is worthless.
Part II. Causality

8.4. LAWS AND PATTERNS

My conclusion is that there are no causal laws: there are no laws of causal succession, the principle of ‘same cause, same effect’ is false or empty, and causal relations cannot be instantiations of relations between universals: for there do not seem to be such causation-universals. The reason is simply this: causation is irremediably circumstantial.

It is interesting to consider in this light Cartwright’s claim (1980) that the laws of physics do no ‘state the facts.’ Cartwright notes that no known law of physics accurately describes the behaviour of ‘real concrete things’ (55). Her point is that the laws, as stated by fundamental equations, always act in combination, never alone. The law of gravitation as we know it, for instance, ignores the workings of Coulomb’s law, which describes the force of electricity. This, according to Cartwright, gives it explanatory power, but at the cost of not being true to the facts.

If she is right, it seems that precisely the examples of laws of nature that are least controversial cannot be causal laws in the sense outlined above. Cartwright herself regards this as a reason to reject realism about laws of nature; I myself would rather say that laws of nature are traits of physical reality, where physical reality just does not have one such trait in isolation. Alternatively, one might say that the laws of physics are constraints. The law of conservation of energy, for instance, constrains what states a system may be in at different times, without, however, determining these states. Other laws will then be further constraints, narrowing down the range of alternatives of that system (perhaps to one, perhaps not to one). On this view physicists, in describing the nature of nature, so to speak, do state the facts: not facts about inner-worldly entities, but facts about nature as such. These facts do not concern connections among event types, but functional dependencies among physical quantities.

Note that my denial of the existence of causal laws does not at all preclude stable causal regularities and patterns. These are ubiquitous: if you let go a brick it will fall, if you turn the knob, water will come out of the tap, if you pull a dog’s tail it will get angry; the presence of gene X leads to disease Y; if a person desires that \( \neg p \) and believes that \( p \), she will try to bring it about that \( \neg p \); low inflation leads to high unemployment. Such patterns will no doubt obtain in virtue of the laws of nature, and discovering them can also be of great scientific and social importance. And we bet our lives on them every day. The point is just that it is a mistake to
8. Causation and natural law

call such patterns laws of nature. Causal relations do require the laws of nature, but not at all regular patterns.

8.5. PARTICULARISM

Although my account rejects the notion of a causal law, it is not particularistic. It is sometimes held, by those who are sceptical about nomological theory, that causal relations do not presuppose any laws whatsoever. Ducasse (1926, 129), for instance, writes:

(T)he supposition of recurrence is thus wholly irrelevant to the meaning of cause; that supposition is relevant only to the meaning of law. And recurrence becomes related at all to causation only when a law is considered which happens to be a generalization of facts themselves individually causal to begin with.

Something like this also characterizes Cartwright’s position (1983), and it has motivated Anscombe’s attack (1971) on Davidson’s nomological conception of causation. But also Fair (1979, 225), who proposes to reduce causation to energy transference, claims:

Watching a baseball shatter a window, we would all agree that the baseball was (at least apparently) the cause of the shattering. I maintain we would conclude such even if we falsely believed that glass was unbreakable. (...) We recognize unique occurrences as causally connected even if they defy regularities in our experience. (...) Knowledge of causes would seem to be epistemologically prior to knowledge of regularities.

And Mackie, who holds that ‘laws of working’ underlie causal dependencies (see 1974, Ch. 8) nevertheless claims (1974, 22) that no general knowledge is required for appreciating some or other counterfactual dependency.

In my view, all these claims are false. Not only does causation presuppose laws of nature, causal knowledge also presupposes an implicit grasp of laws, or at least - in less straightforward cases - general patterns. This is even the case in exceptional cases: pace Fair we should say that we do have general knowledge of the world that enables us to appreciate that the ball shatters the glass; it is the same knowledge that would be the source of surprise and doubt when we see a butterfly breaking a window. At least, a
Part II. Causality

glass-breaking incident would teach us the general fact that glass is breakable.

The alternative is to suppose that our capacity to recognize causation is primitive and unanalyzable. If so, however, it will be a complete mystery what it is that we recognize and how we could do a thing like making causal judgements. This is why, in my view, particularism is an obscure doctrine. On the other hand, if causal knowledge involves a grasp of physical law, how can this be? Where did we learn the physics?

The answer is that our grasp of the notion of causality is like the ability to move our limbs: not explicit and theoretical knowledge, but know-how. Just as we do not calculate our motions but just carry them out, we do not theorize about causal relations but just find them in the world. The fact that it would take a supercomputer to figure out how to make even a simple grabbing movement does not mean that we need equal computing power and mathematical skill to just make it.

This explains how even small children are well able of causal discourse (they know what breaking, eating, talking, etc. is), even though they do not know a thing about physical quantities. It is one thing to have mastery of causal vocabulary, and another to explain what that mastery consists in; just as we should not expect that fishes know that what they are intimately familiar with is H2O.

Our conclusion here can be that a counterfactualist and (in a sense) nomological analysis of causality like the one that I have put forward is quite compatible with the fact that we have an unreflective appreciation of causal relations and an unreflective grasp of causal notions. In order to account for this fact we need not resort to the obscure doctrine of particularism; we should just realize that causal knowledge is know-how, and not explicit, or theoretical knowledge. There is, however, a related aspect of causal knowledge that may shed doubt on a counterfactualist analysis in another way: the perception of causation.

8.6. PERCEPTION OF CAUSAL RELATIONS

My analysis of causality might seem to violate our intuition that we can directly observe causal relations as such. The idea is this: the counterfactual dependence that is so fundamental to my analysis of causal relations involves not just an actual state of affairs, but also a possible one
8. Causation and natural law

with which the actual one is compared. For instance, 'If that car had stayed on the right side of the road, this terrible accident would not have taken place' is a comparison between an actual and a possible course of events. And there can only be counterfactual dependence between the car's getting on the wrong side and the accident in virtue of a possible course of events in which the car stayed on the right side and there was no accident.

Also, if I have been right in arguing against Salmon's at-at theory of causal influence, even the underlying physical connections have this aspect of counterfactual dependence. Moreover, the relation between these connections and the causal relation that they underlie is also counterfactual. The problem with all this is that we can talk about it, and infer it, but we cannot observe it. For we can only observe what is actually the case, and not what is merely possible; we cannot, therefore observe any relation between something actual and something merely possible. This is what has made Hume's denial that we can see one event necessitating another event so compelling.55

That counterfactuals make singular causation a relation that cannot be perceived has also been argued by Menzies (1993, 202-2) and criticized by Armstrong (1997, 215-6). I am definitely with Menzies here, and do not support Armstrong's claim that '(t)he counterfactual is not part of the central essence of the causal situation (...).' (216) But this seems at odds with ordinary experience: I can see a hammer flattening a chestnut, feel a person pushing me aside, or hear one person insulting another; and most directly of all, I can perceive my own acting, or will-exertion. Many causal relations, that is, do seem to be directly observable. If we take this intuition seriously, how then can my thoroughly counterfactual analysis of causation be right?

It is even sometimes thought that the concept of causality derives from that of action and exertion of will, these latter notions being the basic ones rather than natural law and counterfactual dependence. Adherents of such an agent-centered view are, for instance, Gasking (1955), O'Connor (1995), and, in a sense, Von Wright (1973). These authors claim that we

55 The actual background of this denial is, of course, Hume's psychologistic claim that causality is a matter of association of ideas, motivated by regularities in our sense impressions, rather than matters of fact in the world of objects. But Hume was right, in my view, in claiming that the idea of a relation's being causal could not get into one's mind via the senses.
regard the interactions among things in the outside world by analogy with what we directly experience when we act, or are acted on.\textsuperscript{56}

In my view, such an anthropocentric view - implausibly implying that there could not be causality in a world without subjective experience - gets it backwards: any intelligible notion of action, will and sensitivity relies itself on the notion of cause. But the intuition that motivates it is not so easily dismissed: it definitely seems that we are aware of even our most nonreflective acting on or being affected by external reality as action or perception (which means: as causal).

I will below argue that a counterfactualist picture like mine can in fact accommodate this intuition. In doing so, I will not be content with merely pointing out that we perceive causal relations. Fales compares the relation of causing with the relation of being between (1990, 36):

There is nothing like an isolable impression here either - nothing strictly distinct from our impressions of the relata of the relation. Yet when we are aware of the relational fact, we are certainly aware of something over and above the existence of those relata.

This suggests that the fact that it is only the related particulars that interact with our senses (rather than the relation itself which is not, after all, a particular capable of interaction) does not seem to preclude our observing relations. We might then happily admit that we perceive shatterings, softenings, robbings, cheerings-up, and so on, even though we hold that it is only particulars that exert forces, reflect light rays, or emit sound waves capable of affecting our senses.

But the worry will be that while admitting this, a counterfactualist might still deny that we perceive these relations as causal. For on this point there is a difference between spatial relations like being between and causal relations: the former involve only actual states of affairs, while the latter also involve merely possible ones that cannot be perceived. In a Humean vein we might claim that we perceive causal relations as spatiotemporal relations and instances of regularities, and that our mental habits, or inferential capacities, add the (unperceived) causal component.

We then accommodate the fact that we perceive causal relations while, however, still violating a powerful intuition, namely, that we do perceive

\textsuperscript{56} A more sceptical version of this view would be the claim that the notion of causality is a naive projection of our own will, action and sensitivity on other items in the world.
8. Causation and natural law

causal relations as causal. Sometimes we have indeed to go through theory to establish a causal relation between two perceived states of affairs; although we can now be reasonably sure that smoking is unhealthy, we have never been able to directly observe a cigarette's causing a case of lung cancer, in the way that we see a hammer flattening a chestnut. But ordinary life is full of cases in which no theory or inference is required to make a causal judgement. We unreflectively feel pushes and pulls, just see balls break windows and dogs scare children, and just exert our will. Harré and Madden (1975, 49) rhetorically ask,

Can anyone seriously deny that we sometimes veridically perceive the waves eating away the shore, the ax splitting the wood, and the avalanche destroying the countryside, where 'eating away,' 'splitting,' and 'destroying' are clearly causal concepts?

And Anscombe (1971, 93) says that Hume

confidently challenges us to 'produce some instance, wherein the efficacy is plainly discoverable to the mind, and its operations obvious to our consciousness or sensation.' Nothing easier: is cutting, is drinking, is purring not 'efficacy'?

It might at first seem that such phenomenological considerations miss the point, and that our experience could be defeated by philosophical argument. Consider Tooley's suggestion (1990, 191) that

there could be worlds - call them Berkeleian worlds - where the contents of one's experiences would be as they are now, but where the events that one observed did not stand in causal relations to one another.

But I suggest that here we should be firmly on the side of naive perception. For it may be true that we do not learn that our world is not Berkeleian from perception; but we do not learn it from science either. The above thought experiment is only worrisome insofar as any form of radical scepticism is. So the question remains: how do we explain the direct perception of causal relations as causal when we analyze causation in counterfactual terms and reject particularism?
8.7. CAUSAL PERCEPTION AS IMPLICIT KNOWLEDGE

I do think that we should stick to our anti-Humean intuition that we perceive causal relations as causal, rather than as spatiotemporal and fitting into a pattern. But if we can directly perceive causation as such, how is it that we are taken in by the pseudo-causation in the virtual reality of a computer interface or a video game? In a word processor, we seem to perceive ourselves to be pushing buttons on a task bar; but how can we? For no button-pushing is in fact taking place.

It seems that in contrived set-ups our perceptual apparatus cannot distinguish real from pseudo-causation. When looking at the screen, even if we do our best we see buttons rather than arrays of pixels or areas of colour. When looking at two objects moving at exactly the same speed and making coordinated movements, we see them as attached. If that turns out not to be so, we are surprised and we cannot get rid of our initial impression. Now, if we straightforwardly do perceive causal relations as such, then why do we not note their absence in cases like these?

This situation is not exclusive to causal perception. It resembles that of perceiving figure-ground relationships on flat paper or filling out incomplete forms (e.g. the indentations in four circles to a square). In a way, we see causings just as we see physical objects: we see a house when we see its fore-front, a bicycle when looking at a part of its wheel just above the surface of a dirty pond, etc; and we can very well be fooled. The 'house' might be a mere façade, and what seems to be kicking, stabbing, insulting, etc. might just be a stage act. Again, how could we be taken in here if we can directly perceive causation as such?

The answer is that perception of causal relations as such is not something like the perception of causal connections, as if these were a kind of visible threads. Note that we have rejected Salmon's at-at theory of causal influence on which such perception is possible in principle! The key to understanding causal perception is to note that perception is never passive, an external impingement on a purely receptive subject. It is right indeed to reject particularism: in perceiving causation as such we bring to bear our implicit knowledge of physical law, as a matter of know-how. Just as we manifest a knowledge of the laws of physics by walking, jumping and grabbing, we also manifest such knowledge when making causal judgements and, indeed, perceiving causal relations as such.
Such causal knowledge comes naturally to us. We just know that solid objects do not pass through each other, or that bouncing objects behave in certain ways rather than others: even babies manifest this knowledge in showing surprise when confronted with situations in which these regularities are violated. The surprise does not seem to be in the fact that they perceive causal relations they never saw before: we may assume that they (and we ourselves) do so all the time without being surprised. Rather, what elicits the surprise is the violation of a body of general causal knowledge. This is not explicit, (proto-) theoretical or declarative knowledge, and if it is learned at all it is acquired very rapidly and early in life. But it definitely is an understanding of the way physical reality is. It is an intuitive grasp of physical law.

This grasp can only be inaccurate and fallible. We occasionally find our ingrained causal beliefs to be wrong, e.g., when we find that metal containers with hundreds of people can stay in the air; not to mention the surprises that modern physics had in store for us. Also, many beliefs about causal relations (such as, say, our belief that a push at the button at 8.00 p.m. will bring us the daily news, or superstitious beliefs) will be the result of Humean association, not substantially different from conditioned responses in laboratory animals, rather than of a natural grasp of physical laws. But that our grasp of causality is in general a matter of tacit background beliefs about physical laws is shown by the fact that superstitions (e.g., about illnesses, or lightning) can often be dispelled by new general, often physical knowledge.

The question when and how we have acquired such practical knowledge has an obvious answer: phylo- and ontogeny have hard-wired it into us. It is not at all a mystery that we and many other species as well are such capable intuitive physicists: physical knowledge is an obvious prerequisite for survival and reproduction, and it will always have been relentlessly selected for.

The reader may of course balk at my use of ‘knowledge’ in this context. For what I am claiming is just that we are adapted to a world with specific physical laws by means of dispositions to behave in ways that are appropriate given these laws. Some authors indeed suggest that biological

---

57 See, for instance, Plotkin (1997, 186-200), who refers to experimental work of Baillageon, Leslie, and Spelke.
adaptation and knowledge are, in the end, the very same thing, but in the context at hand we may well stay uncommitted on this issue.

For note that on the present account, and whether or not we call our capacity for making causal judgements and applying causal concepts ‘knowledge,’ we can explain both why there should be direct, unmediated, unreflective perception of causation as such, and why this should be the perception of something’s being subject to counterfactuality and law, rather than just of something entirely and particularistically here-and-now. Against particularism we can now hold that each time we perceive a causal relation the laws of nature are at play and there is counterfactual dependence; and against Humean regularism and associationism we can hold that we can immediately perceive this because we have a grasp of these laws wired into us by evolution and individual development.

8.8. CAUSAL PERCEPTION: OBJECT AND CONTENT

To round out this account of causal perception let me explain why I say that indeed we perceive causal relations as such, rather than perceive some objects (and, perhaps, a spatiotemporal relation) and understand, or appreciate, or grasp, that it is causal. For there is an important lesson to be learned from Searle’s distinction between the object of a perception (c.q. an experience) and its content (1983, 123-4):

When, for example, I raise my arm, part of the content of my experience is what makes my arm go up, and when I see a flower, part of the content of the experience is that this experience is caused by the fact that there is a flower there. In such cases we directly experience the causal relation (...). On my account the Humeans were looking in the wrong place. They sought causation (force, power, efficacy, etc.) as the objects of perceptual experience and failed to find it. I am suggesting that it was there all along as part of the content of both perceptual experiences and experiences of acting. (...) Neither flower nor movement is part of the content of the experience, rather each is an object of the relevant experience. But in each case causation is part of the content of the experience of that object.

The object of an experience, we might say, is just out there in the world (it might be a concrete particular or a state of affairs); but its content, i.e., what this object is to us, or what information it carries, is constrained by who we are: our perceptual capacities, interests, expectations, and ways of categorizing. The objects of our perceptions, intentional states and actions
are mere particulars; but what contents pertain to them is a thoroughly relational affair in which especially our past plays an important role.

In the case of a perception of causation, we might say that the particulars interacting with each other and with our senses are the objects of the perception, while it is (among other things) our instinctive grasp of physical law, constraining the information that we get from the perceived episode, that provides the content. This explains why in specific circumstances (for instance in twilight, before a movie screen or at the circus) our perceptual apparatus is fooled\(^\text{59}\) about what is really going on: the piece of rope under the table is not a dangerous snake, we do not really see two people kissing, and we do not really see a woman being sawed in two.

\(^{59}\) We will see that this normative notion is appropriate: our senses are not there to create illusions, but to make us aware of what is really the case. But biological economy does not care for infallibility.
9

The problem of causal relata

9.1. CAUSAL ONTOLOGY AND MENTAL CAUSATION

In the previous chapters I have argued that causation is a matter of counterfactual dependence among states of affairs (particulars'-having-properties-at-times) as well as of physical connection in and among particulars (mostly objects). On that account, mental causation can be expected to be a matter of counterfactual dependence involving mental states as well as of physical connection involving a minded actor. But physical connections are indeed just that: physical. As far as causal connections are concerned, the only relevant properties will be physical-nomic, not mental ones. I came across this passage about Gödel, the mathematician:

(H)e remarked to his close friend, economist Oskar Morgenstern, that in the future it would be deemed a great oddity that 20th-century scientists had discovered the elementary physical particles but had failed even to consider the possibility of elementary psychic factors.' (Dawson Jr., 1999, 68)

But physicists, if ever they come across what seem to be autonomous 'elementary psychic factors,' would certainly classify these as physical, not as 'psychic.' This would not be out of materialist dogmatism; there would just be no point in calling some of the factors interacting with one's measuring devices 'physical,' and others 'psychic'!

Mental causes and effects can thus be expected to have the ontological profile of states of affairs, standing in relations of dependence; rather than of particulars, standing in relations of connection. However, causal ontology is a difficult and much-disputed subject, and the 'factualist' conception of causal relata is not a consensus view. Neither is causal ontology a minor issue, for our conception of causes and effects will crucially affect our understanding of mental causation. It is often held, for instance, that causes and effects are a kind of local particular (mostly called 'event'); and it is also widely held that relations, or relational states of affairs, cannot be
9. The problem of causal relata

causes or effects. These views are among the major obstacles to a satisfactory account of mental causation, and we have to carefully dismantle them.

9.2. HUME, KIM, AND OTHERS

Saying what causal relata are requires a sophisticated conceptual toolkit. Hume, in his famous discussion on causality, worked with relatively primitive equipment. Although in some places he spoke of events, he thought of the relata of a causal relation as 'constantly conjoined' 'like objects in like relations of succession and contiguity,' the 'ideas' of which objects were, by a 'habit' or 'custom' 'determining' the mind to join them, always in each other's company, and so, by the supposition of a 'uniform' 'course of nature,' thought to be 'necessarily connected' with each other (1740, I, III).

Kim (1973a) has elegantly summarized the ways in which this account, as it stands, is ontologically incoherent. First, 'constant conjunction' only makes sense as a relation between generic, or type items, while spatial and temporal contiguity can only be between particular, or token items. Secondly, the temporal component of these items ('succession') best suits events or states of affairs, while the spatial component is more fit to physical objects. Third, 'necessary connection' typically holds between linguistic items such as propositions, while 'constant conjunction' applies more naturally to things in the natural world.

Of course, the charge of incoherence ought to be hedged by an awareness of the fact that for Hume, causation was a psychological phenomenon. He did not aim to give an account of which kinds of items in the natural world are causally related to each other, because in his view causation was not a relation in the natural world to begin with. The 'necessity' of causation, according to Hume, merely consists in habits of the mind. Yet, the first two of Kim's criticisms do hold good in any case.

But then, what are causal relata? Even today there is not much of a general consensus. Davidson acknowledges only events (e.g. Davidson 1967a); his numerous critics recognize states of affairs, whether or not under that name, or whether or not including negative ones (see Armstrong 1997, Fales 1990, Kim 1976, Lewis 1986c, Mackie 1974, Mellor 1995, and

60 Where Hume spoke of 'ideas' and 'habits,' his modern followers who adopt Hempel's Deductive-Nomological model of explanation speak of 'logic' and 'propositions.'
Part II. Causality

Menzies 1989). Kim and Lewis happily include states of affairs in the category of 'events,' Fales opts for the converse, Mackie speaks of facts alongside with events, Mellor of facts and facta, and Menzies of 'situations.' The latter two explicitly acknowledge negative conditions as causes. Instances of event properties are recognized by Dretske 1977 and Sanford 1985; relations and instances of relational properties are, implicitly or explicitly, denied causal relatum status by most (but Armstrong 1978, Baker 1995a and Fodor 1991 think otherwise); and causal relations seem to be acknowledged by Mellor 1995 (106) and Dretske 1988 (in the form of 'structuring causes').

9.3. EXAMPLES OF CAUSAL STATEMENTS

Let us start our discussion of causal relata by seeing what kinds of causes and effects are allowed by ordinary causal discourse. Its rich diversity jumps to the eye right from the start. Consider:

1. The ball broke the window; The sand caused the machine to jam; Bill makes a boat.
2. The explosion caused the fire; The tax measure caused the capital flow.
3. The water is 78° after all this time because the container is insulated (example from Steward 1997, 174); The room is dark because the curtains are closed.
4. John's shouting caused irritation among his colleagues; The incessant downpour caused the gutters to overflow.
5. Don does not die because he does not fall (example from Mellor 1995, 132); John died because there was no doctor around.
6. John's saying 'Hello' caused Fred to greet John in return; John's saying 'Hello' loudly did not. John's state of tension caused his saying 'Hello' loudly, rather than his saying 'Hello.' (example from Lewis 1986c, 255); The collapse was caused by the fact that the bolt gave way suddenly (example from Davidson 1967a, 161); The operation's being a failure caused John's sickness to aggravate.
7. Don's fall is the first because his rope is the weakest (example from Mellor 1995, 117); There is more wine than whiskey left because most guests have been drinking whiskey.
8. I saw the explosion set the ship on fire; The placental species' having been subject to heavy environmental tests caused them to supersede the marsupial species (example from Gould 1980, 240-5).

9. PC-keyboards come with a QWERTY-layout because such a layout has been the standard one for typewriters; Zebras have stripes because natural selection has been favouring that trait in the past.

These are all causal statements, in the wide sense that they tell us that one item causally depends on some other item. The reader may check for herself whether all the dependencies are genuinely causal; I think they are. Now, 'items' is a vague catch-all term. And in the examples there is a great variety in prima facie causal relata:

- In 1, concrete particulars: objects, an amount of stuff, and a person;
- In 2, events and an action;
- In 3, states of affairs;
- In 4, ongoing processes;
- In 5, non-occurrences, or negative events and states of affairs;
- In 6, event and state-of-affairs properties;
- In 7, relations, or relational states of affairs;
- In 8, causal relations;
- In 9, general facts or regularities.

Items not listed above are, for instance, numbers, triangles, fictional characters, or general properties like redness. The reason why these are not even prima facie causal relata, at least not on the analysis of causation given earlier, is that they do not involve particulars, and therefore cannot be physically connected to anything else. But from the items that are listed, not all are plausible causal relata either.

9.4. EXCLUDING CONCRETE PARTICULARS

First, we should note that although concrete particulars are relata of physical connections, they cannot be relata of counterfactual dependencies. Certainly, example 1 gives us good causal statements, but these are best seen as abbreviatory. The causes and effects are not the objects in all their concreteness, so much as these objects' having
properties like moving, breaking, or obstructing. In the case of Bill's making a boat, we should not just say that Bill is the cause of the boat. No, Bill's acting as he did is the cause of the boat's coming into existence, or of an aggregate of planks' becoming a boat (the boat did not exist before Bill started hammering).

But although concrete particulars cannot be said to be causally related in the sense that one is a circumstantially necessary condition for the other they are, of course, causally connected, being, after all, the items that persist and interact. Insofar as these are ways of being causally related, concrete particulars are of course causal relata. But they are not relata of counterfactual dependencies.

9.5. EXCLUDING NEGATIVE STATES OF AFFAIRS

Apart from 1, the examples of 5 ought to be met with some suspicion. The causal statements of 5 suggest that causes and effects can be non-events, or non-states of affairs. It seems, indeed, that there is such a thing as causation by omission, and that prevention is also a species of causation. But what could negative events or states of affairs be? It seems that their very negative character precludes their being part of actual reality! For Mellor (1995), examples like those of 5 are a reason not to call causation a relation at all. Yet, causation seems very much to be a paradigm of a relation (it is a relation of dependence), and 5 does seem to cite routine causal statements. In ordinary speech, something negative like Don's failure to fall, or the absence of a doctor, is definitely treated as a cause.

It seems, then, that we should affirm negative causes, but deny negative states of affairs (not, though, negative facts). If so, we should deny that the alleged causes in the examples are actual events or states of affairs. And indeed, the very point of the causal statements from these examples is that something that (saliently and significantly) might have been the case is in fact not the case. If the phrase 'the cause' suggests that we have to do with an actual state of affairs, this suggestion should be resisted.

We might object that if the examples of 5 are examples of causation, there must be a relation. I want to resist that objection, for it would commit us to holding that mere absences can be terms of a relation. So let me argue that what the examples of 5 refer to are not actual relations. First, remember that causation is a thoroughly counterfactual affair. What we do
9. The problem of causal relata

when making causal statements, straightforward ones like 'the fire caused the explosion' as well as less straightforward ones like 'John died because there was no doctor around,' is contrasting an actual and a merely possible situation. Negative as well as positive conditions are implicit in any causal claim. The great frequency of negative conditions in causal discourse underscores this central role of counterfactuality in our concept of causality.

Secondly, note that a statement like 'Don does not die because he does not fall' implicitly compares the actual situation with a causal relation that might have obtained, but in fact did not. It sounds strange to speak of an alternative, negative causal relation, as if Don’s not falling causes Don’s not dying. Also, negative causal statements without negative causal relations are easy to explain: such statements give us as much causal information as positive ones.

Thirdly (and most contestably, I assume), I would like to claim that a statement like ‘John died because there was no doctor around’ also implicitly compares the actual situation with a causal relation that might have taken place, but in fact did not. In my view, it compares the actual course of events with the merely possible presence of a doctor preventing John’s death. It does not refer to a relation between a non-presence and a death. Nor does the causal statement: 'The absence of a doctor caused John’s death.' It is an informative causal statement: as any other causal statement, it compares the actual with the merely possible. But it does not refer to a relation that actually obtains.

My claim is, then, that negative conditions are not, as negative states of affairs, terms of causal relations. They certainly are causal explanantia and explananda, but just in virtue of the counterfactual, contrastive nature of causal statements. Note, however, that this leaves intact a notion that we encountered earlier: causation by disconnection. If the pulling of a trigger causes a shot, this is a straightforward dependency between two positive events, no matter whether the underlying physical infrastructure involves connections or blockings of connections. This teaches us once again that a causal dependency and its underlying physical infrastructure are not the same thing.

I am aware that my denial that causal explanations involving negative conditions refer to actual causal relations may be found unsatisfactory. As an alternative we might hold that there are causal relations involving negative conditions, and that causes and effects are facts rather than states of affairs. I do not favour this alternative myself: it introduces a wealth of
Part II. Causality

complications (an instructive example is Mellor 1995), and it violates our intuition that causes and effects have a spatiotemporal location.

9.6. EXCLUDING REGULARITIES

Example 9 raises questions as well: it suggests that general facts and regularities can be causes or effects. But taking them as such violates our intuition that causes and effects are somehow particular occurrences. If they were not particular, remember, facts could not be paired as cause and effect. On the other hand, example 9 seems to show that there can be particular events ('frozen accidents') that have as a result that henceforth some regularity obtains.

What to say? We should note that general facts can perhaps be effects but not causes: their causal efficacy will always be screened off by their particular instances. If effects, they terminate causal chains. This result violates intuitions about causal transitivity (according to which effects will be causes of further effects), although it might also be regarded as an interesting source of causal asymmetry. Yet, we can draw a parallel with the screening-off on the effect side, and hold that it is just the instances of the regularity that are caused, not the regularity itself. It thus seems that we do not need to recognize regularities as either causes or effects.

9.7. PROPERTIES OF STATES OF AFFAIRS

Example 6 introduces what we may call iterativity in our causal ontology: there is a particular's having a property (John's saying 'Hello') that is a cause, but this state of affairs seems itself to have a property that constitutes a further cause (John's saying 'Hello' loudly). A saying-'Hello,' it seems, can be loud, sudden, cheerful, take place in the corridor, etc. Typically, we encounter such cases as the referents of adverbial modifications in causal sentences: 'suddenly,' 'on the beach at noon,' 'patiently,' etc. But reference to them can also be more explicit:

The slowness with which the controls were applied caused the rapidity with which the inflation developed. (Davidson 1967a, 161)
There is something about Sebastian’s stroll that makes me uneasy, and something about it that makes you uneasy. (...) Whatever caused Sebastian to stroll at two in the morning can be different from what caused him to stroll slowly and different again from what caused him simply to stroll. (Sanford 1980, 293)

The authors quoted here endorse an ontology of events as concrete particulars. To state of affairs-ontologists, these examples suggest that even if events and states of affairs are themselves property-havings, they can have properties in turn. Perhaps there can even be several orders of nesting in this way.

Note, however, that a state of affairs’ status as itself a bearer of properties is not sufficient to make it a concrete particular. Perhaps states of affairs can have a ‘secret life’ of sorts (so that there are things to discover about, say, someone’s being human, as well as about concrete human beings). But we saw that concrete particulars have a spatial position and spatial parts, are token identical with a physical particular, and can be physically connected to other such particulars. Now, no matter whether states of affairs can have properties themselves, they do not have these characteristics.

Do properties of states of affairs, or events, constitute causal relata, i.e., further causally effective states of affairs or events? The answer is that while some do not, some do. As to those that do not: states of affairs or events can be short- or long-lasting, early, delayed, at midnight, etc. But such temporal characteristics are constitutive of, rather than had by, states of affairs, at least on Armstrong’s account that I adopted earlier on. A time individuates a state of affairs rather than that it characterizes it.

Also, there are properties that merely supervene on a number of states of affairs simpliciter, i.e., on a state of affairs together with the circumstances. Something may take place on the beach, in the palace, close to a fire, etc.; but such properties do not constitute further causal relata because their causal efficacy is screened off by that of the items on which they supervene (properties of particulars among which, significantly, spatial relations).

Yet, many cases are different. A stride may be stealthy or clumsy, a saying-‘Hello’ can be loud, submissive, cheerful, etc. (with effects that vary accordingly); a stroll can have various aspects with various causal roles; a government decision can be unjust; and so on. Now, consider Menzies (1989, 76):
Part II. Causality

(A) n adverbial phrase like "with a coarse towel" can play two roles in a nominal like 'Flora's drying herself with a coarse towel.' It can play an identifying role in which it serves merely to identify the referent situation. (...) An adverbial phrase can, however, have a modifying role in which it modifies the verb of the nominal in an essential way. Used in this way, an adverbial phrase is not redundant and indeed plays a crucial part in determining the constitutive property of the situation referred to by the nominal. (...) (T)he adverb or adjective plays an essential, indispensable role in determining the constitutive property of the situation denoted by the nominal. If this explanation is correct, there is no need to see nominals like 'the quickness of the climber's fall' and 'the climber's falling quickly' as referring to event aspects.

Such nominals refer to events and states of affairs in much the way as ordinary nominals with modifying adjectival or adverbial phrases (....).

If Menzies is right here, when we speak of a property of a state of affairs, we do not characterize an item that might or might not have that property. Rather, we describe in more detail what property we had in mind. We will then say that John's saying 'Hello' is, in fact, nothing over and above John's saying 'Hello' loudly (remember that the free lunch principle implies that less specific states of affairs, or determinables, are nothing over and above more specific ones, or determinates).

Note, however, that the two may be causally distinct. John's state of tension is not among the causes of his saying 'Hello' even if the latter is nothing over and above a saying-'Hello'-loudly (of which the state of tension is a cause). For John would also have said 'Hello' (though not so loudly) if he had been more relaxed. This suggests that loudness can be a property of the state of affairs that is John's saying 'Hello' (and a causally efficacious property at that), rather than that we have just one state of affairs (John's saying 'Hello' loudly) that may be described more or less precisely.

The case for properties of states of affairs as causal relata is strengthened by considering causal, meaning, and valuational properties. With causal properties I mean here causal-relational properties of states of affairs, such as being devastating, loud, harmless, etc. I will later on argue in detail that causal relations can themselves be causes of mental states, namely when they are perceived. And as we will see, they can also be effects. If this is indeed the case, it follows that causal-relational properties of causes and effects can themselves be further causes and effects.

Secondly, there are meaning properties. The position of a thumb, for instance, might represent approval or disapproval, a pattern of pixels on a screen might represent a tree, a firing pattern in the brain might represent
the presence of a possible mate, the burying of an axe might represent the end of a conflict, etc. Again, such properties of states of affairs can be perceived, in which case they are causes. They will also, of course, be effects, for instance of an agreement.

Thirdly, there are valuational properties. A situation can be dangerous, an action can be admirable, an event dramatic, and so on. Again, it is arguable that states of affairs like these can be causes (namely when they are perceived) and effects.

I am aware that on considering the latter three cases many will doubt whether we can indeed perceive the causal relatedness of two events, an action’s meaning, or a situation’s goodness or badness. My discussion of the perception of causal relations earlier on suggests, however, that we can perceive these properties. The idea was that we distinguish the intentional content of perception from its object. I will return to these issues in Part III.

Although our list of candidate causal relata has shrunk, we still have events, actions, states of affairs, ongoing processes, properties of events and states of affairs, and relations (among them causal ones) as causal relata. Because all of these kinds of items can play the role of circumstantially necessary condition (of no matter which other kinds of items on the list), and are enabled as such by underlying physical connections, there is no reason to single out one of them as the primary kind of cause or effect. If so, our conception of causal relata ought to be liberal.

This result is at odds with the widely held belief that causation is primarily between events. Indeed, because events are traditionally regarded as the units of causation, the ontology of events has been very much discussed. For the problem of mental causation it is significant that thoughts and actions, whatever else they are, in any case seem to be events. Let us, then, see what events are.
10 
**Getting events wrong**

### 10.1. Event Mereology

Events are widely supposed to be the natural, worldly entities related by causation. In many a philosopher's view, that is precisely what they are good for, ontologically speaking.\(^{\text{61}}\) Most philosophers go along with common sense in holding that events are spatiotemporally situated items that happen, or occur.\(^{\text{62}}\) But beyond that minimal assumption there is hardly any consensus to be found in the literature on this subject. I would like to start my discussion of events, in the present chapter, by rejecting two event conceptions. The first one is that of events as having a spatiotemporal mereology, a bit like objects. The second, to which I will dedicate the most effort, is Davidson's conception of events as concrete entities. In the next chapter I will outline what seems to me the correct view.

Let us first consider the view that events are some kind of building blocks of reality. This is the view that Salmon has in mind when writing (1981, 286):

> The main difference between events and processes is that events are relatively localized in space and time, while the processes have much greater temporal duration, and, in many cases, much greater spatial extent. In spacetime diagrams, events are represented by points, while processes are represented by lines.

Now, if this is the way in which we think about events and hold that events are causal relata (which Salmon himself, of course, does not), we must explain how one event gets its causal message across to another event. Should we not say here, risking endless regress, that intermediate events are required as connectors? Ducasse (1926, 128) has pointed out that, whereas this supposes that there are gaps between events, events are better regarded as separated by cuts. Brand (1980) tackles the problem by claiming that events partly overlap, that the overlaps are to be seen as event

---

\(^{\text{61}}\) E.g. Lewis (1986c, 241): 'Events are not much of a topic in their own right.'

\(^{\text{62}}\) the only exception that I know of being Chisholm (1970) who claims that events can occur more than once.
parts that are themselves events, and that there is simultaneous causation between the parts of the cause-event and that of the effect-event where these parts overlap.

Issues like these become spurious, however, once we simply exclude events in the above sense from our ontology. It seems that we can dismiss them without any harm: such events are metaphysical constructions rather than entities that we perceive, or explanantia of what we perceive. Salmon writes (1984, 183):

I have laid particular stress upon the role of causal processes, and I have even suggested the abandonment of the so-called event ontology. It might be asked whether it would not be possible to carry through the same analysis, within the framework of an event ontology, by considering processes as a continuous series of events. I see no reason for supposing that this program could not be carried through, but I would be inclined to ask why we should bother to do so.

On this view, clarifying the nature of the physical connections underlying the causal dependencies that we normally speak of is a much more useful job than explaining how endlessly divisible ficticious entities are linked up. This stance towards events will raise suspicion, of course, when it is claimed that the particulars and events that we do know and talk about are the mereological wholes of sets of point-events. Lewis, for instance, claims (1986c, 258-60):

Events have a spatiotemporal mereology (...). Each of Sebastian's steps is a spatiotemporal part of his stroll; so is the entire half-stroll performed by the left half of him. (...) Events overlap iff they have some event as a common part; an atomic event is one that has no events except itself as parts (...).

It is suggested here that 'events' defined in space-time terms, are the same kind of thing (related by mereology) as 'events' in ordinary speech (the feast, the capsizing, the stroll, etc). The corollary of such a view for the debate on mental causation is quite straightforward: all you find in the world are merely physical events and their mereological aggregates. These are all the causal relata. Then, where are the decisions, the thoughts, the perceptions and the actions? They must be identical with a mereological aggregate of merely physical events. But the behaviours of such aggregates are entirely accounted for in physical terms, so that only the physical, and not the mental properties of these events are causally efficacious. The unfortunate result is epiphenomenalism.
Part II. Causality

Considering the troubles with the above mereological conception of events, I want to suggest that we had better dismiss it. There do not seem to be any metaphysical costs in doing so (particulars and properties serve us equally well), and we avoid a number of spurious issues. We may well continue to talk about points in space-time diagrams as 'events,' but we can regard such points as locations (rather like contours or centres), rather than entities.

10.2. EVENTS ACCORDING TO DAVIDSON

Davidson approaches events from a somewhat different angle. The fact that we have in our language nominalized verbs, or even proper names (hurricane Mitch, Doris' capsizing the canoe, the French Revolution, etc.) is Davidson's reason to regard events as 'true particulars.' He says that

> without events it does not seem possible to give a natural and acceptable account of the logical form of certain sentences of the most common sorts (...). I propose to legitimize our intuition that events are true particulars by recognizing explicit reference to them, or quantification over them, in much of our ordinary talk. (1969, 166)

Davidson also claims that such events are not like states of affairs, because they cannot be uniquely picked out by a sentence (except, of course, a sentence in which the event is just named). When we describe a state of affairs we pick out a particular (say, Jack), and ascribe it a property (say, is a tall man). But, reviving a point made earlier by Ramsey, Davidson claims that we cannot do this with an event (1969, 169). A sentence such as 'Doris capsized the canoe yesterday,' he claims, does not refer to the unique event that is Doris' capsizing the canoe, but merely states the fact that there occurred at least one such event.63

---

63 Davidson (1967b) also has an argument, put forward earlier by Frege and Quine, that is intended to show that if events are referred to by true sentences there can only be one event, because all true sentences have the same denotation: the True. This 'Slingshot argument,' as it is called (see Searle 1995, 113), is not, in fact, an argument against a factualist causal ontology, but against a correspondence theory of truth. But it is clear that if correspondence theory is false, the prospects for fact-like causal relata will not be good. I will not try to dismantle that argument here, but see Menzies (1989, 78-82), Searle 1995, 221-6 and Mellor 1995, 113-9.
What are events in Davidson’s view? He claims that ‘(E)vents are non-abstract particulars, and (...) causal relations are extensional relations between such events.’ (1995, 6) An event as characterized by Davidson (e.g. 1969) is an ‘unrepeatable particular’ or ‘concrete individual,’ describable in an indefinite number of ways. For instance, ‘He signalled’ can be a description of the very same particular event as is ‘He raised his arm’; and ‘the catastrophe in the village’ can refer to the same event as ‘the avalanche’ (see 1969, 164-5).

According to Davidson, then, events are particulars with a ‘secret life’ (Steward 1997, 31), that may be token identical with each other, in spite of diverging descriptions. For what gives events their identity is their unique place in ‘the framework of causal relations’ (1969, 179), and ‘(i)f causality is a relation between events, it holds between them no matter how they are described.’ (1995, 6) Davidson (1963, 12) acknowledges that ‘states, dispositions, and conditions are frequently named as the causes of events,’ but claims that ‘(m)ention of a causal condition for a cause gives a cause only on the assumption that there was also a preceding event.’

Davidson allows events to be very complex. He says that the predicates that go into the description of an event may have an indefinite number of places (see 1969, 166). This means, that descriptions of an event may be quite elaborate, without thereby posing a threat to the event’s identity as an individual. So, World War II can be regarded quite unproblematically, in Davidson’s view, as one event. ‘(E)vents have parts that are events, and the parts may be discontinuous temporally or spatially (think of a chess tournament, an argument, a war).’ (1970b, 183)

Below, I will mount three points of criticism against Davidson’s causal event ontology. The first point is that it severs the close link between causal relations and causal explanation. The second point is that it either renders unintelligible any distinction between causally relevant and causally irrelevant properties, or implies a thoroughgoing epiphenomenalism concerning all except physical properties. The third point is, that in spite of Davidson’s claim that only events can be causes and effects, causal-explanatory practice requires states of affairs to be causal relata.

I think that these points together show that a Davidsonian view of causal relata is rather problematic. They do not, of course, show that there are no events, but they do show that they are not the particulars Davidson claims them to be.
10.3. First Criticism: Causal Relations and Causal Explanations

It is one thing to claim that causation is a relation among Davidsonian events such as explosions, fires, capsizings and deaths; it is another thing to accommodate this picture to examples of causal statements such as: 'The fact that the dam did not hold caused the flood,' or: 'The collapse was caused, not by the fact that the bolt gave way, but by the fact that it gave way so suddenly and unexpectedly.' Davidson realizes this, but he claims of these examples (1967a, 161-2):

(1)n addition to, or in place of, giving what Mill calls the 'producing cause,' such sentences tell, or suggest, a causal story. They are, in other words, rudimentary causal explanations. Explanations typically relate statements, not events. I suggest therefore that the 'caused' of the sample sentences in this paragraph is not the 'caused' of straightforward singular causal statements, but is best expressed by the words 'causally explains.'

But Mellor (1995, 131) notes that

Davidson needs more than a distinction between causation and causal explanation: he needs a dichotomy. (...) This dichotomy is very hard to accept. It is (...) a strong connotation of causation that causes explain their effects (...). Why else, after all, should we use 'because' as an explanatory connective? Why else should we take 'C causes E' to entail 'C explains E' (...)?

And Menzies (1989, 64-5) says:

If there were a genuine equivocation in our use of the verb 'caused' in the two sorts of sentence, it would make illegitimate many of our ordinary ways of talking and reasoning about causation. For example, if you say 'The shortcircuit caused the fire' and I say 'So too did the failure of the sprinkler system,' my remark would be, strictly speaking, out of place, since I would be using the word 'caused' in a different way from you.

I myself will now argue that informative causal statements and causal explanations are the very same thing, so that the distinction that Davidson needs (one between straightforward causal statements mentioning causally related events on the one hand, and 'rudimentary causal explanations' on the other) cannot be made at all.
Lewis (1986b, 226-7, 237-8) notes that explanation is something that we can have more or less of, and that can be more or less adequate to the demands of an inquirer (it is assumed, of course, that the explanations are by and large true). This is plausible: if only full explanation were explanation, causally explaining even the most trifling fact would be a humanly impossible task. Also, an explanation that does not sift out irrelevancies is hardly an explanation. If Lewis is right, a causal statement may be brief and relatively uninformative ('The illness is caused by some virus'), while still being an explanation. Completely uninformative causal statements, such as: 'The illness is caused by something that happened,' or: 'The illness is caused by its cause' are not causal explanations. True, these examples might show that not all causal statements are causal explanations, but then, we might also refuse to call them genuine causal statements.

All informative causal statements, then, can be regarded as 'rudimentary causal explanations.' They can be improved on by characterizing objects and situations, applying laws (if available), and mentioning circumstances; until we reach what Mackie calls a 'minimally complete causal account' (1974, 260). Such improvements, however, are not required for causal statements to be explanatory, for we typically have quite some background knowledge already.

The connection between causal statements and causal explanations also holds in the other direction: causal explanations generally involve causal statements. The reason is that otherwise they will not be recognizable as causal explanations to begin with. The exception is, of course, causal explanation by negative conditions, for we have seen that these do not yield causal relata. On the other hand, even cases of negative explanations confirm the intimate link between causal explanations and causal statements, for we have seen that negative causal explanations are best analyzed as tacit counterfactual claims. Claiming that Don does not die because he does not fall is, in effect, claiming that it is salient that if Don had fallen he would have died and that actually he did not fall. And there is no way of making such a counterfactual claim without making hypothetical causal statements.

If, as the above implies, causal explanations and causal statements come to much the same thing, it ought to be possible to question Davidson's distinction between extensional causal relations between events and intensional explanatory relations between propositions.
Is it? We can say, for instance, (an example from Mackie 1974, 249) that the speech caused the crisis, (extensional causal statement) and also that there was an international crisis because President De Gaulle made a speech (intensional causal explanation). The point is that in the first case we can substitute co-referring terms while preserving truth value, for instance by saying, ‘The public talk caused the international emergency situation.’ In the second case, however, truth may be lost by such a substitution, as is shown by: ‘There was a crisis because the man with the biggest nose in France made a speech.’

But while certain forms of causal explanation do indeed introduce an intensional context, it is also true that causal explanations can very well be given without such a context; so that the point about in- and extensionality is largely irrelevant. It is easy to turn an intensional causal explanation like: ‘There was a crisis because President De Gaulle made a speech’ into an extensional causal statement like: ‘President De Gaulle’s making a speech caused a crisis.’ And if we delete the intensionality in such a way, no causal information need be lost. There may be a difference between extensional connectives like ‘caused’ or ‘is causally responsible for’ on the one hand, and intensional ones like ‘because’ or ‘causally explains that’ on the other; but for the goals of causal explanation this difference is hardly relevant.

Causal statements and causal explanations, then, are not different in any interesting way. If so, Davidson’s way of dealing with the examples of the dam and the collapse is inadequate, so that such examples remain a threat to his view of causal relata.

10.4. SECOND CRITICISM: CAUSAL RELEVANCE AND IRRELEVANCE OF PROPERTIES

Our next issue was raised in the context of the debate on mental causation, so that I will for the moment abstain from my habit of postponing discussion of specifically mind- and action-related subjects. But

64 I say ‘may be,’ not ‘is’: I am not at all sure whether someone who knows that the man with the biggest nose in France is President De Gaulle would consider this sentence false. Also, someone who does not not know that President De Gaulle is such an influential person (but does perhaps know that the man with the biggest nose is influential) needs not at all accept ‘There was a crisis because President De Gaulle held a speech.’ Intensionality seems less important than background knowledge on the part of the inquirer.
the points made here against Davidson's view could easily be generalized to apply to non-psychological examples. The reader is cautioned that I will sometimes have to abandon my preferred vocabulary of 'states of affairs' and say 'properties' instead. For this is the way in which the authors involved in the debate discuss the present issue, and mixing up vocabularies would be confusing here.

We have already encountered Davidson's doctrine of 'Anomalous Monism,' on which, first, all events are physical events but can be given mentalistic descriptions; secondly, causation is a matter of causal laws; and thirdly, there are no psychological or psychophysical, but only physical laws. We also saw that Anomalous Monism has been criticized for implying epiphenomenalism, for instance in Heil and Mele 1995 by Kim, McLaughlin, and Sosa. The idea is that because Anomalous Monism holds that all events are physical events and that such events are related by strict physical laws, it must also hold that all events are causally related in virtue of just their physical properties. This would mean that it is irrelevant to the causal efficacy of mental events that they are mental events.

Davidson opposes this objection by pointing out that the supervenience of mental properties (or 'predicates,' as he prefers to say) guarantees that mental properties do make a difference. If two events are mentally different, they must also be physically different, and that means that they must be causally different:

\[(P) \text{roperties are causally efficacious if they make a difference to what } \text{individual events cause, and supervenience ensures that mental properties do make a difference. (1995, 15)}\]

I think that in itself this is a good reply: if a higher-level property is a circumstantially necessary condition for change on a lower level, why not say that the higher-level property is causally efficacious? Unfortunately, this good point is spoilt by Davidson's token physicalism, which has it that any mental event is token identical with a physical event. Now, if a certain mental event is the very same event as a physical event, it seems that any alleged causal efficacy of the mental event properties is relentlessly usurped by that of the physical ones.

Davidson's reaction to this objection amounts to downplaying the distinction between an event's causally relevant and causally irrelevant properties. He says (1995, 6) that
given my concept of events and of causality, it makes no sense to speak of an event being a cause 'as' anything at all. (...) If causality is a relation between events, it holds between them no matter how they are described.

Expressions like ‘(M)entality is causally effective only if events are causes in virtue of [or: 'because of'] their mental properties'; ‘(I)t is only under its physical description that a mental event can be seen to enter into a causal relation with a physical event (or any other event) by being subsumed under a causal law'; ‘(T)he mental does not cause anything qua mental; the mental is not efficacious as such.' (1995, 13) are claimed not to make any sense. Davidson explicitly denies, for instance, (1995, 17) that when somebody is killed by a loud shot (where the shot could also have been silent because of a silencer), one could show the loudness to be causally inefficacious by reminding that a silenced shot would have been deadly, too.

What is at stake in this discussion is whether such a refusal to distinguish between causally relevant and causally irrelevant properties is a credible move. The main trouble with it is that it implies that events are causally related while there is nothing that the causal character of the relation could possibly consist in. True, Davidson holds that causal character consists in nomadic relatedness; but that is exactly what lends the charge of epiphenomenalism its force. As soon as causes fall under laws, they become explanatory - as they must; but then, properties have entered the picture, which is exactly what Davidson needs to avoid in order to pre-empt the charge of epiphenomenalism. This is why McLaughlin notes (1995, 34):

Critics hold that Davidson himself is committed to the view that events cause other events in virtue of certain of their properties by his principle of the nomological character of causality.

So the dilemma for Anomalous Monism is that either it must stop treating causes as explanatory (which would make it impossible to distinguish causal relations from noncausal ones), or it must admit the causal role of properties (which makes it epiphenomenalist). Kim (1995a, 22) claims, that

it makes sense to ask questions of the form 'What is it about events c and e that makes it the case that c is a cause of e?' and be able to answer them, intelligibly and informatively, by saying something like 'Because c is an event of kind F and e is one of kind G (...).’ How could anyone refuse to acknowledge this - unless, that is, he
believed that causal relations were brute facts about events, having nothing to do with the kind of events they are?

McLaughlin (1995, 32-3) thinks that Davidson does indeed hold this last belief:

He [i.e. Davidson] appears to claim that when an event causes another, there is nothing about the events in virtue of which this is so. But that claim seems quite implausible. (...) (T)hat two events count as causally related seems an implausible candidate for a brute fact that admits of no explanation.

Davidson, of course, does not claim this at all, because we have already seen that he allows the possibility of causal explanation (1967a, 161). And there is no alternative: it may not just be ‘implausible,’ but outright contradictory to hold that causal relations are ‘brute.’ But Davidson does not claim that they are brute; he claims that they are nomic. On nomological theories of causation, however, laws connect properties, and there must be properties for the causal laws to connect. It is hard to see how this could be otherwise, i.e., how laws could be insensitive to properties. But laws cannot connect all properties of a pair of causally related events (a position that is called ‘supergen’ by Macdonald and Macdonald 1995b, 69). So there must be a distinction between causally relevant and causally irrelevant properties. Sosa elegantly captures what is wrong with a refusal to admit such a distinction (1995, 42):

Is it assumed that every event - such as death d - has its every property essentially? This would explain why d’ must be distinct from d, since it would be caused by a silent shot, unlike d, which is caused by a loud one. (...) But if this is the way the mental is efficacious, then the mental seems no more efficacious than a speck of dust on the butt of a murder gun.

What we are seeing in this debate is, that Davidson’s attempt to keep the ‘in virtue of’ criticism at bay derives its initial plausibility from his distinction between straightforward causal statements and causal explanations. But the distinction does not give him a stable position. It seems that insisting on such a distinction makes him face the following dilemma: either regard the causal relations referred to by causal statements as brute, nonexplanatory, and inexplicable (which would make it impossible to tell why causes are causes); or do allow ‘in virtue’ talk, but that amounts to undoing
Part II. Causality

the distinction and (on the rest of Davidson’s assumptions) to embracing epiphenomenalism.

The significance of this issue for Davidson’s event ontology is, of course, that as long as events (i.e., causal relata) are regarded as concrete particulars, causation must implausibly be claimed to be a brute, unexplained and unexplanatory relation.

10.5. THIRD CRITICISM: THE NEED FOR STATES OF AFFAIRS

My third objection to Davidson’s view of events as concrete particulars derives from the fact that unchanging states of affairs are causal relata in good standing along with changes; and being states of affairs, I will claim, they cannot be concrete particulars.

For Davidson, events are changes. He says: ‘States and dispositions are not events, but the onslaught of a state or disposition is.’ (1963, 12) On Davidson’s view of events, then, such things as the presence of gas in a room or the weight of an iron bar cannot be counted as events themselves, although whatever changes in the world that started these situations can.

Now, states of affairs (not the ‘onslaughts,’ but the states of affairs themselves) are, in ordinary causal discourse, often counted as genuine causes or effects. When someone strikes a match and there is an explosion, for instance, we will readily say that the cause of the explosion (given that people often strike matches without there being explosions) is the presence of gas in the room. It will not do to mention the event of the gas’s leaking into the room (the ‘onslaught’ of the state) as the cause, because what is important is that the gas is still there. If we cited the leaking-event as the cause, we would need some negative still-not-having-drifted-away event as an extra cause, but that would not really be very much of an event.

One might protest that we are always free to claim that any time a state of affairs is mentioned as a cause, what is really meant is the ‘onslaught’ of that state of affairs. After all, we have seen earlier that when causes are regarded as circumstantially necessary conditions, there are various ways in which causes can be distinguished from a ‘causal field.’ There even seems to be some plausibility in emphasizing changes: intuitively, causation is primarily about things happening, not about things staying the same. We may well believe Mackie, Salmon or Tooley to be right in claiming that
10. Getting events wrong

object persistence is a form of causation, while admitting that this does not sound quite as obvious as the claim that causation is between events.

But standing conditions are undeniably of crucial causal importance, and indispensable as causal relata. Suppose that I work at a military base, and my job is to wipe clean the nuclear missiles. One day my bosses launch the freshly-cleaned missiles, with horrible consequences. If only events can be causal relata, I might be suspected to be a mass murderer, for it seems that the first event featuring the missiles after I have dusted them should count as the effect of my dusting; at least when we want to avoid the claim that the dusting was epiphenomenal. Of course, my dusting only caused the missiles to be clean, but that sensible conclusion is out of reach for a Davidsonian view.

Yet, can we not admit the causal indispensability of states of affairs while construing such ‘unchanges’ as concrete particulars, just like events? It does not seem so. Suppose that someone started to talk to me about the state of my shirt by saying things like, ‘I do not like that state,’ or ‘It is the same as last year.’ My first reaction would be to ask: What state? Two things could be the case: my interlocutor might mean my shirt with all its properties. But in that case he just means the concrete particular, and the ‘state’-talk is empty. Or he is talking about something that is the case with my shirt: its being crumpled, striped, yellowish, made of cotton, wide, etc. In that case the state talk makes sense, but it is about a property instance instead of a concrete particular.

The difference between concrete particulars and states can also be brought out by noting that we can point at a concrete particular (or lift it, throw it in the air, sell it, etc.), and say, I do not know what this thing is. Is it an F or is it a G? If we do so, there need not be any doubt about what particular we have in mind: in specifying its spatial boundaries we can make it unambiguous which particular ‘this thing’ is supposed to be. There can exist, then, unspecified concrete particulars. 65

This is not the case with states of affairs (or, for that matter, events). If you say, There is this thing the case, but I do not know what it is, you do not pick out anything at all, as you would in the case of a particular. There

---

65 Hence the intelligibility of the characterization UFO. Surely, making sense of unidentified events, let alone of unidentified states of affairs, is much harder. Note that I am not claiming here that there can be featureless particulars. I am also not claiming that particulars can be picked out without characterizing them at all; just that they can be picked out by their mere spatial boundaries.
Part II. Causality

is no such thing as an unspecified state or event. Even pointing to, for instance, a falling rock, and saying 'that event over there, beginning...now and ending...now' does not make it unambiguous that 'that event' is, for instance, the rock's fall or its rotation. If this is right, we could in principle count concrete particulars before we had characterized them, but we could not do so with states or events.

These remarks depend in part, of course, on the groundwork in Part I, but one need not buy all my claims there in order to appreciate that treating a state of affairs as if it were a concrete particular is a strange thing to do. If, then, states of affairs cannot be understood as concrete particulars, and if ignoring them leads to causal gaps, Davidson's ontology of causal relata can hardly be adequate.
11

Getting events right

11.1. Events according to Kim

If the previous chapter has taught us how not to think about events, what is the right way? What we should be looking for is an event ontology that avoids the troubles of the Davidsonian view, and that squares with the factualist metaphysics, or states of affairs ontology, that was outlined in Part I. Such an event ontology is Kim’s. Kim (1976, 34) claims that

an event (or state) is a structure consisting of a substance (an n-tuple of substances), a property (an n-adic relational attribute), and a time.

Kim allows that his events can have properties themselves, but he warns us that (1976, 43) that

the properties an event exemplifies must be sharply distinguished from its constitutive property (which is exemplified, not by the event, but by the constitutive substance of the event). (...) Thus, events can be redescribed by the use of different predicates expressing the properties of (exemplified by) them; what cannot be done is to redescribe them by tampering with their constitutive properties.

On this view, ‘Brutus’ stabbing Caesar has the property of occurring in Rome, it was intentional, it led to the death of Caesar and caused Calpurnia to grieve, and so on.’ (43) According to Kim we can omit these properties and add others in our description, without thereby picking out a different event than before. But substituting ‘killing’ for ‘stabbing’ would change what event was described: unlike Davidson, Kim holds that the killing was not the stabbing, no matter how closely related these events are.

Note that in Part I I have adopted the basic structure of Kim’s ontology (his ‘substances’ are my ‘particulars’), but that my analysis of causal relations is not the same as Kim’s. It is Kim’s view that the property making up an event (its ‘constitutive property’) is a nomic type (a ‘generic event’):

“(R)eal changes” or “real events” seem to be just those that make a causal difference, and generic events seem to be just those properties whose possession by an ob-
Part II. Causality

ject bestows upon it a causal power or potency, or whose possession by an object indicates its being subjected to such powers. (...) The basic generic events may be best picked out relative to a scientific theory (...). (37)

However, we can accept Kim's event ontology without supporting his nomological theory of causation and properties.

11.2. THE NARRATIVE VIEW: EVENTS AS EPISODES

Kim's theory of events, though in my view largely adequate, as it stands does not account for all kinds of events. I have in mind events like wars, parties, vigils and avalanches: ones that cannot directly be construed as property instances, and that seem to best fit a Davidsonian framework. I would now like to supplement the Kimian view by giving an alternative, non-Davidsonian account of such events. The account is inspired by Steward's claim that we 'carve out events from space and time' according to 'narrative significance' (1997, 71); it is a radicalization of that claim.

The idea is the following: when talking about events like wars, parties, vigils and avalanches (say, World War Two) we are recapitulating a history, but in a condensed form. It seem to be a fact that when we talk about the world, we do not just describe loose states of affairs, but related ones: they form, more or less, a coherent whole. We can always choose to tell such a history, or a part of it, in detail: 'In 1939, the Germans invaded Poland. And then... (so that..., and..., while at the same time..., which led to...) ...and then the Japanese capitulated.' But instead of doing so (and on the assumption that we could do so), it is often more convenient to just say 'World War Two.' World War Two is just that stretch of history, or causal episode, as described by the detailed narrative. It is not something more than that, not something in addition; so that if there was World War Two, it must have been that episode.

Events are, then, in fact causal episodes, and nouns or proper names that refer to events are in fact abbreviations of histories, or narratives. This

---

66 To help the reader keep track of the terminology: a stretch of history is any set of states of affairs involving causal sequences among those states of affairs; an episode is any part of a stretch of history. Episodes may be stretches of histories themselves, and vice versa. A history, or a narrative, is our description of a stretch of history, or episode. Along the same lines, episodes as occurring can be distinguished from episodes as being told; but here I will assume that it is clear from the context what is meant.
11. Getting events right

claim has as a consequence that events as concrete particulars vanish from the scene. For when telling a detailed history, we are describing states of affairs: we are describing the world in a fine grain, not picking out Davidsonian events so much as Kimian property instances. Only by lumping the latter together in causal episodes do we obtain events that have a seeming resemblance to particulars, in that we can name and count them.

If the above conjecture is right, we should expect that when we speak of events as the x or the y, we cannot do so unless we have constructed a history in which each stage has its unique place. In that case, there can be no referring to events outside of a narrative context. And this, indeed, is what we find. It does not make sense to talk about Doris' capsizing the canoe (this, not that capsizing) without a narrative context. Compare this with an event such as a woman called Doris capsizing a canoe: we can talk about it outside of a specific narrative while remaining intelligible. Not so with that woman's (note the 's'!) capsizing of that canoe. If there is no history as a salient part of which that capsizing can figure, then speaking of the capsizing does not make sense.\(^{67}\)

Events, then are just causal episodes, or parts of these, mostly described in an abbreviatory way, not in detail. They are stretches of history with narrative significance; a significance that justifies referring to them as a unit. When, for instance, we are referring to my writing of an article, we are telling a part of the story of my life (or the story of a magazine, or of a university), namely that part in which I wrote an article, no matter how long it took me and no matter how many breaks there were. This scattered part of the story of my life can, however, be treated as a unit (probably with vague, or indeterminate boundaries), because it is concerned with that article we, as narrators, are interested in.

Any stretch of history will have parts of which there is no reason for anybody to pick it out. We may or may not call such episodes events, just as we may or may not call spatiotemporal gerrymanders objects. We may demand additional criteria to distinguish real, natural-kind events from any stretches of history; but I suspect that such attempts will be futile. Consider, again, the event that is my writing of an article. Is it a real event? I would say yes, but it may well be temporally scattered, and different people may include different states of affairs in it, depending on their interests. For

\(^{67}\) Nor, I think, does it make sense to talk about Doris and the canoe outside of narrative context.
me, my coffee-drinking may belong to it; others might count only my hours in front of the computer.

As we saw, Steward, from which I borrow the idea of 'narrative significance,' herself claims that we 'carve out events from space and time' according to 'narrative significance' (71). But in my view, events are not carved out from space and time at all, even if, of course, they do occur somewhere and at some time. For instance, when speaking about my decision, we are not identifying some specific stretch of space-time, for in order to intelligibly mention the decision, we need not have the slightest idea when or where I took it. But we must have an idea of what stretch of history it is a part of. At best we can say that it, as an event, was where I was when I decided. If events are 'carved out' at all, it is from a narrative or an episode. What matters to an event's individuality and essence is not the spatiotemporal region where it took place, but the position and role of that event relative to other events. In saying this I am advocating a relationist perspective on events, just as I have advocated such a perspective on particulars.

On my theory, then, events are not concrete entities. For as they are made up of states of affairs, they are constituted by properties (the ones that constitute the states of affairs). But note that all the same, the concept of an event is being taken quite seriously. Events are not eliminated, or reduced to states of affairs. For insofar as episodes are declared real, events are. Also, it is not denied that events have properties of their own. As the episodes stand in their own relations, have their own structure, and have narrative saliency as wholes, they do not reduce to the states of affairs making them up. Furthermore, the narrative view allows events without narrators and narratives; it just says that narratives are required to pick events out, not that narratives are required for the events to occur. But the narrative view does deny that events are concrete entities.

11.3. IN FAVOUR OF THE NARRATIVE VIEW

Let me now point out what advantages the view of events just proposed has. One important advantage is, of course, that in adopting it we avoid the numerous problems with the Davidsonian view: it can account for the role of properties in causal explanations. It will say that, because events are wholly constituted by states of affairs, it is the properties constituting these
states of affairs that do the causation. Admittedly, it faces something like the converse of the problem for Davidson, who had to downplay property causation in favour of event causation. For ordinary causal speech trades in event causation, as in: 'The fire caused the explosion'. On the narrative view, this sentence must be abbreviatory of a narrative, one to the effect that some essential facts from the fire episode caused some essential facts from the explosion episode. I will leave it to the reader to decide whether this is enough to do justice to ordinary causal speech.

Note also that the narrative view can easily accommodate the point made by Ramsey and Davidson (mentioned above) that referring to events is different from describing them with sentences. An episode, after all, can be narrated in sentences as well as it can be referred to by a noun or a proper name. Furthermore, when we take a closer look at event discourse as we find it in our ordinary speech, the narrative account can deal with five further issues that are a source of trouble for other accounts of events.

First, we are enabled to deal with questions such as: 'Was Jack's eating his sandwich a part of the conference or was it not?' Such questions are an embarrassment when we try to individuate events in spatiotemporal terms, but on the present theory we can just answer that it depends on whether we would include Jack's eating in a full report of the conference.

Secondly, as Steward (1997, 69-70) points out, ordinary usage has room for 'changeless events' such as vigils, or standings-still (or, we might add, failures). This is not easily explained by a Davidsonian view, but it fits easily within an account of events as (parts of) episodes. If someone is keeping still, not much happens, so that it might seem a mystery why ordinary speech would call this an event. The mystery evaporates, however, when we realize that the keeping still might fulfill a key role in some history of danger and heroism, and that it is this role that lends it event status.

Thirdly, my narrative account makes less mysterious our frequent mentioning of what Steward calls 'subjectless' events, such as the rate of inflation's rising in January (59-60), or the fire. Such events are not a problem for the Davidsonian account so much as for the Kimian one as we met it above. One might ask: If events are substance-property-time complexes, then, in the case of an inflation's rising or a fire, where is the substance and what is its property? On the present account we see that we may just tell about all the particulars, their relations and properties, and the causal dependencies among these, that are involved in such events (there always are), and then say: 'That is the inflation's rising,' or: 'That is the fire'; even
Part II. Causality

if ‘the fire’ does not refer to the same item as ‘the burning of the house.’ The point is just that the latter event may be part of the episode that we refer to as the fire.

Fourthly, the view of events as episodes can accommodate claims such as that my wave (i.e. the very same wave) might have been a right-handed as well as a left-handed one (example from Bennett 1987). Kim also notes that some events do not seem to be essentially what their specification says they are, which worries him because he has proposed as an identity condition for events, say, a and b, that a’s and b’s respective constitutive substance, property and time be identical (1976, 35).

And indeed, the identity of many events (waves, declarations, speeches, birthday parties, etc) does not seem to depend on the actual form they take on. This, in fact, is what makes Davidson’s claim (1995, 17) that a death that was caused by a loud shot would have been a different death when caused by a silent shot sound so strange. For people, it seems, have just one death, and it may take one form or another.

Now, if we regard events as particulars, or as mere property instances, what could be meant by claims to the effect that ‘the’ wave might have been right-handed, or that ‘the’ stroll might have taken place somewhere else, in a different manner, or even by someone else? Why do we not say that if someone had waved with his left hand instead of with his right one, it would have been a different event?

This is not easy to explain on the basis of a Kimian or Davidsonian event ontology, but my narrative account does not have trouble with it. What makes the ‘loud’ death the same one as the possible ‘silent’ death, or the left-handed wave the same one as the possible right-handed one is, that we are talking about a part of an episode worth telling that has a unique place in that episode. The episode-part may take on different forms, but its identity does not depend on that form. It depends on its playing the appropriate role in the wider context (again, note the relationality).

Finally, and relatedly, my account of events makes it understandable how events such as weddings, deaths, collapses, parties, etc. can be hastened or delayed.68 This last aspect of events is sometimes alluded to in discussions about event identity and event essence, such as the following criticism of Kim-events as ‘fragile’ (1986c) by Lewis. Lewis (1986c, 249-50) writes:

---

68 See also Bennett 1987.
You might hope that an essential specification of an event could easily be extracted from the sort of nominalisation whereby we standardly denote it. Suppose we denote an event by a nominalisation: “the F-ing of A at T.” (…) The trouble is that an event with such a rich essence is a fragile thing. (…) The causes and effects whereof we ordinarily speak are more robust than that. (…) Who would dare be a doctor, if the hypothesis under consideration were right? You might manage to keep your patient alive until 4:12, when otherwise he would have died at 4:08. You would then have caused his death. For his death was, in fact, his death at 4:12. (1986, 250)

The narrative account of events tells us how such cases are to be handled. The patient’s death just is not the same item as the patient’s dying at 4.12. The former is a salient stretch (i.e., the end) of an episode: the patient’s life. The latter is one state of affairs from this stretch. Remember in this context that causal explanantia and explananda tend to split up: causal inquirers generally do not want to know just why the patient died at 4.12; they want to know why the patient died at 4.12, why the patient died at 4.12, or why the patient died at 4.12. Lewis seems to assume that causal inquirers explain events wholesale, but often that is not the case at all. In this case, we should say that you did not cause the patient’s death, so much as his time of dying.

We have now seen that events are not very well regarded as particulars, and that our event discourse can very well do without such Davidsonian events. Events are real enough, but they are not concrete, and their ontology can very well be explicated in terms of an ontology of states of affairs.

11.4. EVENTS, ONGOING PROCESSES, AND STATES OF AFFAIRS

To conclude this chapter, let me now argue that even if we adopt the Kimian view of events outlined above, it would be a mistake to say that events are the primary kinds of causal relata. My point will be that although events are constituted by property instances, by no means all property instances are plausibly called events. I will argue below that events, states of affairs and ongoing processes are significantly different, which should be taken as a final argument against the causal ontology that we might call ‘event causation.’

This dismantling of event causation will be useful in understanding intentional action (it might help put an end to the habit, as popular as it is awkward, of describing beliefs, desires, intentions and the like as ‘mental
events'); but my main agenda behind it is to set the stage for a further argument, in the next chapter, to the effect that in contrast to what is often held, relations (or relational states of affairs), though very unlike events, are causal relata in good standing.

There is a significant difference between events, ongoing processes and states of affairs, and it is not in the particulars or times involved in them. Any particular can be involved in an event, a process, and a state of affairs indifferently, even during one and the same time. For instance, my computer may crash, hum, and be made of plastic and metal all at the same time. And no matter whether some item is an event, ongoing process or state, it may either take a long time or just a moment. For instance, the cooling down of a planet (an event) may take aeons, Peter's looking for fun (a process) may go on for just two seconds as well as for all his life, and the car's being at two kilometres from a gas station (a state) may either last for years or a split second. Events happen, processes go on, and states obtain or endure; but states and processes need not last any longer than events occur.

The difference, then, lies in the properties involved. Now, what is it about a property instance that makes it an event, process or state? Decisive is what Steward (1997) calls 'temporal shape':

(...I should like to (...) introduce here the concept of temporal shape. It is often observed that in merely giving the temporal dimensions of an existent thing - in specifying the beginning- and end-points of its existence - one does not thereby determine its temporal character. For vastly more than these temporal reference points, in determining the ontological category of any item, is the way in which that item fills the relevant period of time - whether it persists through the time, or occurs during the time, or obtains throughout the time, etc. (72-3)

I think (with a slight modification to the above) that it is plausible to claim that an event occurs or happens during a time, that a process goes on or continues through time, and that a state obtains or persists throughout a time. Of course, a state has a 'duration,' as does an event, but 'during' as mentioned by Steward above does not mean something neutral as 'taking time,' but more something like 'needing time for completion.' Events, proc-

---

69 'Process' will below mean something different from Salmon's 'causal process', which I usually refer to as 'physical connection.'

70 Also, of course, I would not say 'existent thing' in the case of events, processes and states.
1. Getting events right

cesses and states have, then, different temporal shapes. Let us explore these differences a little.

Note first that events and processes consist of temporal parts, or stages. In this they differ from concrete particulars as well as from states of affairs. To begin with the former, these are wholly present at any moment of their existence and have spatial parts. True, I have suggested earlier that the identity of a particular through time is a causal matter; also, particulars (for instance persons) may change and develop. But at any moment of your life you are a whole person (even though it is your past that has made you that person), and what has stages is that life (or just, in the case of mere objects, existence). Remember also that a particular can unambiguously be picked out by its spatial boundaries, but that there are no unambiguous temporal boundaries of a particular (the only way in which a particular has temporal boundaries is as an F or a G.)

States of affairs also lack temporal stages. The room's being in a mess at $t_1$ can be called a different 'stage' from this state at $t_2$; but this makes little sense, for the room is wholly messy at any moment of its messiness (just as an object is wholly present at any moment of its existence). Note that in the case of states there is little to the notion of persistence through time. Is the messiness at $t_2$ numerically the same state as the messiness at $t_1$, just like the chair in the corner at $t_2$ is numerically the same chair as the one in the corner at $t_1$? This question does not seem to make clear sense. Suppose that I bend an eraser, after which it springs back in its old shape. Is the latter numerically the same state of the eraser as the one before I bent it or is it another state? As far as I can see this question has no intelligible answer.

In fact, states do not have much of a temporal shape at all compared with events. Events as I have characterized them are causal episodes, whether (at their simplest) they are just changes in particulars, or whether they are complex stretches of world history. In any case they unfold, or develop, and they take time to complete. When an event is over, things are different from what they were like when the event started. An event has distinguishable stages and a non-arbitrary start and end.

Ongoing processes also have stages, because they, too, involve change (if nothing changes, it seems, nothing can be said to 'go on'). The difference with events, however, is that processes start and end at arbitrary moments relative to the other stages, while events do start and end at mo-

---

71 Note that this commits us to holding that existence through time (as distinguished from existence at a time) is a process, or event (in the sense of causal episode).
ments that are non-arbitrarily related. You may continue or stop pushing against a surface of suspended foil, but you may not continue or stop breaking it. You may continue or stop whipping cream (and continue or stop growing satisfied), but you may not continue or stop taking a decision and jumping out of bed.

Note that with processes, and not with events, there is a sense in which things stay the same, namely as long as the process goes on ("There is no change, they are still going on with that horrible drilling"). There also seems to be some difference in temporal variability between processes and events: while a lightning flash (an event) cannot last a year as well as two seconds, an ongoing flashing of a lamp can. Yet, the distinction is not very sharp or reliable: some events (such as a dressing-for-dinner) could be brief as well as near-endless, and some processes (such as the burning of a candle) cannot go on indefinitely. But then, it is in the first place not very clear that a dressing-for-dinner is an event and not a process, and the burning of a candle a process and not an event.

We can also distinguish processes from events by saying that processes always contain events, but that events need not contain processes. One could even regard certain processes, such as someone's pounding on the door, or the radiation of a piece of uranium, as series of events. Yet, we should resist trying to reduce ongoing processes to events in this way: chopping up continuous processes such as the moon's circling Earth into events is an arbitrary thing to do, and it leads to the spurious question what connects those events. Also, we need ongoing processes as causal relata as much as events and states: Juarrero (1999, 34), for instance, points out that an event model of causation is inadequate to account for intentions that guide and sustain actions.

Steward (1997, Ch. 3), on the basis of ideas by Kenny (1963), Mourelatos (1978) and Vendler (1953), takes a linguistic approach to clarifying the differences among events, processes and states. Such an approach points us, for instance, to a distinction (Vendler's) between verbs of accomplishment (such as 'running a mile'), of achievement ("recognizing") of activity ('running') and stative verbs ('knowing'). Verbs of these different kinds allow and disallow different kinds of adverb (for instance, saying that you are running unceasingly makes sense, but saying that you are recognizing someone, or running a mile, 'unceasingly,' does not); and they allow and disallow different kinds of questions about them ('For how long...'; 'When exactly...'; 'Did you...'; 'Were you....-ing?'; etc.).
11. Getting events right

An approach like this has led Mourelatos to a 'typology of predication' (see Steward 1997, 83), distinguishing states, processes, developments and punctual occurrences. It has also brought to light that we can refer to events only with count terms, and to processes only with mass terms: we can say, for instance, that there was a bang on the door, but saying that there was a (continuous) pounding on the door does not seem correct: there just was (some) pounding.

I think that a linguistic approach such as Steward's is very useful (even if it is mainly centered on human action rather than on events, processes and states in general), and that it is to this approach that we owe the important distinction in 'temporal shape' between events, processes, and states. On the other hand, language does not always wear its ontological commitments on its sleeve, and it can also mislead. For instance, Mourelatos' division of events into two categories, namely developments (such as sunrises) and punctual occurrences (such as the snapping of a rope) may tell us something about event discourse, but it does not seem to point us to a genuine distinction in reality, where the difference between split seconds and hours seems to be only one of degree.72

---

72 Another intriguing, and quite recent example of a linguistic approach, not just to ontology but to causation in particular, is Lakoff and Johnson 1999, Ch. 11. Lakoff and Johnson argue that the 'skeletal concept' of causation (causation as a 'determining factor for a situation' (177)) is so weak that in the face of the wealth of different causal metaphors in our language the concept of causation does not allow of a uniform and general analysis. I disagree, of course. In my view, Lakoff and Johnson provide a fine analysis of our rich and complex causal speech and causal thinking, but not much of a metaphysical study of causal reality itself.
12

Relations as causal relata

12.1. RELATIONAL CAUSAL RELATA

We have seen that causes and effects can be events (among which we find actions), changeless states, and ongoing processes. All these are states of affairs, or sets of these. Events are no exception: if my narrative account is right, these are episodes, i.e., sets of (mostly causally related) states of affairs that are individuated according to narrative significance. If so, we can say that causal relata are essentially property instances.

Now, it is often held that only intrinsic, or internal properties of particulars can have causal efficacy, or can be causal powers. Yet, we have seen that many states of affairs are relational. We should try to determine, then, whether relational properties, or relations, can constitute causes and effects, just like intrinsic properties. In the present chapter I will argue that relational states of affairs, and especially causal relations, do at least qualify as prima facie causal relata.

In ordinary causal discourse, relational causes or effects are nothing special. Causal statements like 'The privatization of the Dutch Railways will cause the shutdown of railway lines in thinly populated areas,' 'Because he thought that his neighbours were thieves, he purchased a safety lock for his front door,' 'Greenhouse warming will cause the spreading of tropical diseases,' or 'The high tide caused the dikes to burst' are brimming with relational concepts (no disease without a relationship among victims and micro-organisms, no tide without a relationship among moon, sea and shore, etc.) as much as they are routine. And there is nothing in our use of expressions like these that suggests that they are not literally true.

Yet, of some kinds of relations it can be argued that they must be causally inefficacious. We have seen of resemblance, covariation, difference, co-existence, sameness-distinctness, identity-distinctness and quantitative and partitive relations that they are necessitated by other properties or re-
12. Relations as causal relata

relations, and so are not external. If so, it is plausible to hold that any alleged causal efficacy of such relations is screened off by the properties that they supervene on.

For instance, ‘The object in Johnny's hand is a gun’ describes a relation of identity (at least if the form of token physicalism that I have defended is right), which is not external, and therefore does not constitute a cause over and above states of affairs like Johnny’s fetching the object, or his aiming the gun. Also, relations such as being faster than, being as big as, etc. might seem to make a causal difference but do not really do so. True, we say that the animal lost the fight because it was the smaller one; that the administration had problems because the computers were not all of the same type; that the oil was on top because it was lighter than the water, etc. But such relations are nothing over and above other states of affairs, and all the causal work in the situation is done by the latter. The reader might find this claim pedantic, but it is a corollary of the free lunch principle.

Note that what I am not claiming here is that in general any alleged supervenient causal efficacy pertains to the physical base. Higher-level relational states of affairs, I will argue later on, do have their own causal autonomy insofar as they are not identical with their physical base. But in the case of non-external relations, no causal-explanatory information is lost if mention of the relation is omitted. For instance, ‘Jones was 40 kg heavier than Smith’ does not add anything to ‘Jones was 100 kg and Smith was 60 kg.’ For reasons of explanatory transparency an inquirer may well mention the relation instead of the subvenient states of affairs, but this is a point about explanatory strategy, not about causal efficacy.

Non-external relations, then, lack causal efficacy. We have seen, however, that spatial, temporal and causal relations are external. Such relations do not seem to supervene on other states of affairs, and therefore are a genuine ‘addition of being.’ If so, relations like these are at least good prima facie causal relata, items that may not have their causal efficacy screened off by other states of affairs.

Spatial relations are the only external relations that hold between concrete particulars. Their causal efficacy, in my view, is rather obvious: the burglary is successful because you are far from home; because the bus and the truck are only one metre apart, you cannot pass them with your car;

73 Calling them ‘internal’ will be confusing insofar as some of these relations supervene, not just on intrinsic properties of the relata, but also on spatiotemporal relations among these.
Part II. Causality

etc. Note that mention of no amount of either intrinsic or nonspatial-relational states of affairs can convey the same causal-explanatory information as these statements.

With temporal relations (that, like causal relations, hold between events, processes or states) things are a little less obvious. They do seem to make causal differences: timing one's actions well is the key to their success, as any soccer player or stockbroker knows. But we have seen that temporal event properties individuate rather than characterize events and states; and insofar as we do not believe in some notion of absolute time, an expression like ‘at/during t1’ seems to be implicitly relational. If, then, we just mention the temporally-related events, we do not also have to describe their temporal relation. That relation, I would say, does have its causal relevance, but it is already given with the mere occurrence of the events, processes and states themselves; at least when their time is indeed constitutive of them.

12.2. CAUSAL RELATIONS AS CAUSES

Now, how about the causal efficacy of causal relations? Mellor, who holds that causation is between facts, argues that indeed causal relations can be causes over again. He writes (1995, 106):

One advantage of the view that causes and effects are facts is that it lets causation be iterated: it lets the fact that C causes E have its own causes and effects. (...) On this view a true ‘E because C’ just states another fact, no different in that respect from C and E. In particular, this fact can also have causes and effects, i.e. there can be, and are, true instances of: (1a) ‘(E because C) because B,’ e.g. ‘Don dies because he falls, because his bones are brittle’; (1b) ‘F because (E because C),’ e.g. ‘The climb is halted because Don dies because he falls’; and so on, to any degree of complexity.

But I do not find this very compelling. Does Don die because he falls because his bones are brittle? It rather seems that Don’s bones’ being brittle and Don’s falling are just two cooperating circumstantially necessary conditions. And as to the second example, it seems plausible to say that the climb is halted because Don dies, or because he falls; separately, these events already sufficiently explain the halting. And the climb would also have been halted if Don had died by some other cause. The causal efficacy
of the causal relation between his dying and his fall, then, is not obvious, be it as cause or as effect.

In contrast to Mellor, Kim denies that causings can be causes all over again. He writes (1976, 50-1):

(M)any transitive verbs are implicitly causal; e.g., 'pull,' 'push,' 'break,' 'shatter.' (...) Are we to accept these causal relations themselves, i.e., one event's causing another, as events? (...) His answer is negative: causings at least are no 'basic events,' even though they may be called 'facts.' But he suggests that we do not need causings anyway:

(A)ll causes of the killing are among the causes of the action that caused the death or of the death, and all effects of the killing, too, are among the effects of the action or of the death.

And he is echoed by Fales (1990, 66):

Can causal relations themselves be among the constituents of events which are related by still other causal relations? (...) The answer (...) is 'no.' The correct ontological parsing of 'a caused b to cause c' is 'a caused b, which in turn caused c.'

What should we say? In my view, the authors just quoted ignore a plausible kind of case of causation by a causal relation: the kind of case in which a causal relation is mentally represented. For instance, it is not inconceivable that Don’s companions have decided to halt the climb only in case they see Don die by a fall; and this would vindicate 'The climb is halted because Don dies because he falls.' Indeed, Kim's and Fales's claims are at odds with the fact that causal relations, such as killings, can be observed. For it seems that if we do so, the killing (not just the stabbing, not just the dying, not just both, but exactly their being causally related) causes a mental state such as a belief, panic, or indignation.

True, we may 'see' causal relations where there are none (as on stage) and we may fail to see such relations where they are present (as in the case of an unknown allergy). Also, we have seen that the causal character of a relation is not perceived in the way that one perceives a tree or a house: the causal relation is not an object that interacts with one's senses. However, we do perceive causal relations as causal, and even if we occasionally fail, our perceptual apparatus is such that causal judgements quite reliably
track actual causal relations of a number of kinds. It is, after all, designed to do so. In the case of normal perceivers and normal circumstances, then, it is causal relations that cause causal judgements, never mind the occasional misshot.

One may well ask how Kim's denial of causal relatedness of breakings, shatterings, etc. squares with the bracketed phrases 'an n-tuple of substances' and 'an n-adic relational attribute' in his definition of events (see above), for according to these, 'the ball breaks the window' obviously is an event. In the text quoted above Kim has in mind causal relations among events, but a ball's breaking a window arguably involves a causal interaction among particulars as well as a causal relation among events. This suggests that Kim's denial of the causal relatedness of causal relations cannot be right on his own definition of events.

Note that the only clear examples that we can give of causal relations as causes concern cases in which causal relations are perceived. In other (i.e., nonintentional) kinds of cases, it does not seem possible to argue that it is the causal relation c-e, rather than merely event e, that causes a further event f.\[^{74}\] This means that before there were intentional subjects in our world there may well not have been causation by causal relations. Below, we will find that causal relations can definitely be effects without intentional states. This means that at least in nonintentional reality, something strange is going on: where normally, everything that is an effect can also be a cause, this seems not to be the case with causal relations!

We might regard this strange asymmetry as a reason for mistrusting the claim that causal relations can themselves be causal relata. But another suggestion is also possible: perhaps we have here a source of causal asymmetry.

12.3. CAUSAL RELATIONS AS EFFECTS

So far, I have argued that causal relations can be causes. But can they also be effects? An affirmative answer will enable us to exploit Dretske's impor-

\[^{74}\] Millikan (1984, 26) argues that a present property A of an entity may result from earlier causings, because a series of such causings may create a positive correlation, and such a correlation may figure in an explanation of A's presence now. But I cannot use this as an argument for the causal efficacy of causal relations, it seems, for I have above ruled out regularities (the 'positive correlations') as candidate causal relata.
12. Relations as causal relata

tant notion of a 'structuring cause' (see Dretske 1988, 42-4). A structuring cause (as distinguished from a 'triggering' cause) is a past event, or episode, that has made it the case that in appropriate circumstances instances of property C cause instances of property M.

Dretske distinguishes among a (triggering) cause of an event e and a (structuring) cause of an event c's causing event e, because in this way he can explain why intentional properties, say, beliefs, are relevant to causal explanation. He regards beliefs as 'physical structures' (1988, 80): brain states that are the internal causes of an actor's bodily motions. But he also holds that the contents of an actor's beliefs are what they are in virtue of the actor's learning history. So on Dretske's account, content is a relational property of an intentional state.

Because Dretske does not believe in causal efficacy of relational properties, he faces the problem of still doing justice to our intuition that belief content is causally efficacious. His ingenious solution consists in claiming that belief content does not explain mere motor output, but rather explains behaviour, which is the entire process consisting of the causing of motor output (M) by the 'internal cause' (C), which is the belief. While brain states are triggering causes of motor output, intentional characterizations of these brain states relate them to the structuring causes that made the brain states the triggers they now are. These structuring causes caused a present situation in which Cs cause Ms in appropriate circumstances. This account leaves intentional content causally inefficacious, but at least it explains why such content is of explanatory value.

I cannot endorse this account: it relies on a computational theory of mind, intrinsicalism, and event token physicalism, which are all views that I oppose (as will become more clear later on). It also leaves the mind, in the end, epiphenomenal. But I do think that the notion of a structuring cause is of key importance in accounting for the causal efficacy of intentional properties (which, I will claim, along with Dretske, are historically-relational). What I would like to do is, then, to adopt this idea of a structuring cause, but to give it a slightly different twist. In that way it will enable us to make sense of fully fledged causal efficacy for a number of relational properties, among which (we will see later on) intentional ones.

The notion of a structuring cause requires that causal relations be effects, not just causes. We might argue, of course, that because of causal transitivity, if something can be a cause it can also be an effect. But our case would be strengthened considerably if we could make sense of causal
relations as effects in an independent way. I would now like to suggest that appealing to the circumstantial nature of causation enables us to do so.

The idea is this. Consider physical event $e$ that is caused by physical event $c$. Event $c$, of course, has its cause $b$: another physical event. But there is no reason to hold that $b$ is the cause of $c$'s causing $e$: for all $b$ is worth, its effect $c$ might have caused quite something else than $e$. Then, what is the cause of $c$'s causing $e$, i.e., the structuring cause? Remember here that causes have the effects they have in virtue of the circumstances being as they are. It is not the case that a push against a chair will in each and every circumstance cause the chair to move. The chair may be nailed to the floor, there may be an obstacle, etc. A specific causal relation can obtain only in appropriate circumstances. Now, circumstances have their own causes. There must have been events that have shaped them. But in having caused the circumstances required for $c$'s causing $e$, they are circumstantially necessary conditions for that causal relation. What shapes the circumstances, then, can be regarded as the cause of any causal relation that obtains in virtue of these circumstances.

Note that on this account, every event or state is a structuring cause, or part of one: any event is a cause of another event as well as a cause of a causal relation, presumably even countlessly many. Events do double causal work, in a sense: they trigger and structure at the same time. On the recipient side, however, only causal relations have structuring (but not triggering) causes, while only mere events or states can have triggering (but not structuring) ones. Note also that because the required circumstances of a causal relation typically are many, complete structuring causes typically are complex episodes rather than simple events and states.

Here is yet another way to reach the conclusion that causal relations can be effects. Suppose that we defend a counterfactualist theory of causation (as I do), and also that we believe that causation is transitive (the plausible idea that if $c$ causes $e$ and $e$ causes $f$, $c$ causes $f$). Now, here is a case that threatens to drive a wedge between counterfactual dependence and causal transitivity:

\footnotesize

75 In order for them to be causal conditions, it will have to be shown that the particulars constituting them and those constituting $c$ and $e$ are physically connected in such a way that this connection is a necessary condition for the structuring causal relation. This does seem to be the case: after all, the structuring causal relations are also triggering causes.
12. Relations as causal relata

While skiing, Suzy falls and breaks her wrist. Her injury is such that she can type but not write. The next day, she types a philosophy paper and sends it off to a journal. The journal subsequently publishes it. Is Suzy's fall a cause of the publication of the paper? Of course not. Yet many theories of event causation tell us otherwise.\(^{76}\)

This example exploits the fact that causes can be backed up, in combination with the fact that on a counterfactualist theory, c may cause e, not in a direct way, but just by cutting off, or disabling, any alternative(s) of e.

For clarity, compare this example with a parallel case: someone wants to cause a scaffolding to collapse by removing pole 1. But pole 2, idle until then, backs up pole 1 and nothing happens. I think that we are not tempted to say that the removing of pole 1 caused the scaffolding to remain standing. But why not? The case is exactly parallel to that of Suzy's article. The analogy would be: the removing of pole 1 (c) caused pole 2 to catch the weight of the scaffolding (e). Pole 2's catching that weight caused the scaffolding to remain standing (f). And because of transitivity, it seems that on a counterfactualist theory of causality, you can cause a scaffolding to keep standing by removing one of its supports!

The mistake made here is the assumption that there is a causal chain c-e-f. But why not say that the removing of pole 1 caused pole 2 to keep the scaffolding standing? We would then have, instead of c-e-f, c-(e-f). In the case of Suzy's article, we would say that Suzy's fall (together, of course, with other states of affairs) causes her typing the article to cause the publication. The idea is that the fall disables one causal chain, or history, and puts another one into operation.

It seems that we have a choice here between relinquishing either a) a counterfactualist account of causation; b) causal transitivity; or c) Kim's

ban on causal relations as causal relata. In my view, relinquishing c) is by far the most credible option.\textsuperscript{77}

\textsuperscript{77} In conversation, Marc Slors has pointed out to me that my claim that causal relations can be caused is ambiguous: what can be caused is a relation (which is in fact a causal relation) - rather than some other causal relation - but what cannot be caused, it seems, is that relation's being causal. This ambiguity, however, does not seem to invalidate the points that my claim is intended to support.
13
Causal efficacy

13.1. THE NEED FOR A CRITERION

In the present chapter I will argue that relations, or relational states of affairs, are not just *prima facie*, but indeed actual causal relata. Relations, I will claim, are not just the right ontological category for being causes and effects, they can also actually be causally efficacious.

Before starting my argument, I want to point at a terminological issue that we encountered before. Because what is at stake in the debate is clearly the causal efficacy of relational states of affairs, not the causal efficacy of relational properties in abstraction from their particular instances, I would myself prefer to talk about the causal efficacy of relational states of affairs, not that of relational properties. But this is not the way in which the issue is generally raised in the literature: causal efficacy is standardly attributed, or denied, to properties. Below, I conform to this more common usage. Let me stress that this issue is merely terminological.

Are relations, or relational properties, causally efficacious? We can only answer this question if we can establish causal efficacy. But how to do this? Sometimes it is quite clear which properties deserve causal credit and which do not. When playing pool, the colours of the billiard balls are obviously causally irrelevant to the way they put other balls into motion, but obviously causally relevant to our decisions as to which ball to hit, and how. The problem of mental causation, however, shows that at other times things are less obvious.

The trouble is that properties can, in some sense, make a difference without making a causal difference. Suppose that Bill throws a biased die that has cost one dollar. The die lands at six. We could say that the die’s price makes a difference to Bill’s gambling: after all, Bill might have played with a die that has cost two dollars instead of one. But it would be strange to call this a *causal* difference. Bill’s throwing a six, however, does seem to be a causal difference. Intuition says that the die’s being biased is causally efficacious here, while its price is not.
Part II. Causality

But what is it that shapes our intuition? Many would say that it is the fact that being biased is (or at least implies) an intrinsic property of the die, while having cost one dollar is merely relational. But then, what about the cases in which the die’s price does seem to make a causal difference, for instance to the amount of money in Bill’s wallet? It is clear that if we want to settle issues about causal efficacy we need a criterion.

13.2. LOCALITY AND INTRINSICNESS

To find one, let us start by having another look at McGinn’s remarks (1989, 133) to the effect that

what happens at the causal nexus is local, proximate and intrinsic: the features of the cause that lead to the effect must be right where the causal interaction takes place. (...) The causal powers of a state or property must be intrinsically grounded; they cannot depend upon relations to what lies quite elsewhere.

McGinn starts out with what we may call the locality assumption about causation: the assumption that causal interactions are local, and that there can be no action at a distance. But almost immediately he switches to what we may call the intrinsicness assumption: because causal influence is local, McGinn claims, only intrinsic properties can be causal powers.

It is important to note that apart from the locality assumption there is no independent support for the intrinsicness assumption. If we held that causal influence is nonlocal, the claim that only intrinsic properties can have causal efficacy would have no plausibility whatsoever. In such a case, causal influence could come from anywhere without taking time; and there would be no ground for insisting that causal powers must nevertheless reside ‘in’ a locally present particular.

It should, then, be possible to show the intrinsicness assumption to be unjustified just by showing that the locality assumption does not actually support it. And if it could be argued that the locality assumption even works against the intrinsicness assumption, the latter would be in a bad position indeed, as maintaining the intrinsicness assumption without the support of the locality assumption seems a precarious thing to do.

The locality assumption is a very basic intuition about causality, and if there is any common ground between relationism and intrinsicalism, we are likely to find it in this assumption. I suggest, then, that a non-question-
begging criterion of causal efficacy had best be based on the locality assumption.

13.3. The criterion of local difference

The locality assumption says that causal influence (i.e., a physical connection) is always at, and always goes to, very specific locations. Note that this could be maintained even if paths of causal influence were discovered to be gappy. It is just assumed that if a physical quantity goes somewhere, it does not go somewhere else.

On the analysis of causation that I have given, there can be no causal relation, i.e., no counterfactual dependency among states of affairs of the causal kind, without some underlying physical infrastructure (consisting of connections and/or disconnections). If so, we can safely assume that on the one hand causal efficacy implies local differences (or, if one wants, either intrinsic differences to particulars or differences in their spatial relations). Differences that, for instance, can be observed or measured at very specific points (in the sense that actual measurement at a specific location differs in outcome from hypothetical measurement at that location in a different situation). And, on the other hand, we can assume that such local differences guarantee that there is causation going on.

If so, a property can only be causally efficacious if its presence makes a local difference. Not just any local difference, but one of the right kind. To see what this means, let me suggest the following criterion of causal efficacy:

The criterion of local difference: Property c is causally efficacious with regard to a given distinct property e if and only if (in the circumstances and barring pre-emption) c is a necessary condition for e and for a local state of affairs l involving a particular constitutive of e.

The criterion of local difference is, in effect, a way in which my conception of causation can be formulated, but cast in the form of a criterion of causal efficacy: ‘causally efficacious with regard to’ might just be paraphrased as ‘is a cause of.’
The criterion says that $e$ has to be distinct (i.e., somewhere else in space-time) from $c$, for on even the weakest forms of physicalism, any property makes a local difference to itself. But that, of course, is not a causal difference: it is just necessitation without a physical connection. Things are already different, however, when a little temporal delay is introduced: in the case of persistence, or internal change, of a particular, $c$ is distinct from $e$.

Let us demand here that $c$ be a circumstantially necessary condition for the right, not of any old local state of affairs concerning $e$. If, for instance, my speaking to you is to be causally efficacious with regard to your changing your mind, it will have to be a circumstantially necessary condition for specific local changes to your brain, rather than, say, your left foot. Now, what makes local state of affairs $l$ the ‘right’ local state of affairs? The answer is that the pairing principle (see above) will pair physical infrastructure and counterfactual dependency in such a way that the infrastructure runs through your brain, and not your left foot. The local difference that the above criterion of causal efficacy appeals to is thus determined with the help of the pairing principle.

For a proper understanding of the criterion of local difference, two subtle but important points demand attention. First, although $c$ must be a circumstantially necessary condition for $l$, it would be a mistake to say that $c$ must be a cause of $l$. Yes, $c$ makes a local difference (which is why we call it causally efficacious), but it does not cause that local difference. It causes $e$, not $l$, even if it is a circumstantially necessary condition for the latter.

For $l$ is caused by $c$'s physical realizer rather than by $c$ itself. On the doctrine of causal closure, $c$ could never causally interfere at the level of physical changes where $l$ is to be found. My trying to convince you (call it $c$), for instance, will not cause the relevant changes in your brain. All the same, $c$ is a circumstantially necessary condition for $l$. For given the circumstances, $c$ (in causing $e$) is required for the causal infrastructure terminating in $l$ to obtain. For instance, if I had not spoken, the sound waves connecting you and me would have been different, or there would not have been any. We might say that $c$ in some sense ‘selects’ or ‘activates’ $l$'s physical causes. But not that $c$ itself causes $l$. I will return to this subtle issue when discussing so-called ‘downward causation.’

Secondly, we should not, as many intrinsicalists and token physicalists are used to doing, confuse $e$ (i.e., the causal difference made) with $l$ (the local difference made). Here is an example to bring this out. Suppose that a
vandal breaks off the nose of a statue. This, obviously, makes a local difference to the statue (the lower part has remained as it was, but in the upper part something has changed). But another, not so local difference to the statue is made as well: it is damaged. This is a difference to the whole statue, not just to its upper part: think, for instance, of the decrease in monetary and aesthetic value, a decrease that will not be proportional to the amount of physical damage.

13.4. IN FAVOUR OF THE CRITERION OF LOCAL DIFFERENCE

Why should we adopt the criterion of local difference? It does fit my own analysis of the causal relation. But are there not perhaps any better criteria? Let us consider some in turn. First, we have seen that intrinsicness might be taken as a criterion for causal efficacy. Yet, this option has its appeal only insofar as it is supported by the locality assumption. The locality assumption is more basic, we might say, than the intrinsicness assumption.

Now, note that the very concept of local causal influence makes sense only in virtue of a tacitly presupposed criterion of local difference. For local influence without local differences seems inconceivable. This suggests that anybody who takes a property’s intrinsicness to be a criterion of its causal efficacy must already be committed to the criterion of local difference. The intrinsicness assumption, apart from presupposing my criterion above, puts an additional constraint on causal efficacy. But insofar as it derives its initial plausibility from the locality assumption, nothing seems to justify that additional constraint.

Another criterion of causal efficacy is provided by Fodor’s (1991) ‘no-conceptual-connection test.’ Roughly, Fodor’s criterion says that when it is a conceptual truth that a difference in cause properties results in a difference in effect properties, the property that makes the difference cannot be a causal power. Applied to the example of the biased die that has cost one dollar, Fodor’s criterion correctly rules out the dollar’s price as a causal factor. For it is a conceptual truth that when the die has cost one dollar, and

---

78 Fodor, who admits that relational properties can be causal-explanatory, intends his test to show that broad contents of intentional states cannot be causal powers. Baker (1995, 47-51) has mounted an intricate argument to prove that broad contents, contrary to what Fodor himself claims, do pass his test. Although the agenda that motivates my present argument is to contest some of the alleged reasons for doubting the causal efficacy of broad mental properties, I will not here try to decide whether or not Baker’s argument succeeds.
Bill plays with that die, he plays with a die that has cost one dollar; while it is not a conceptual truth that when the die is biased, Bill wins money.

The problem with Fodor's no-conceptual-connection test is, however, that it is rather limited in its application. Fodor has developed the test for the specific purpose of exploring the connection between propositional attitudes and actions, and it does not seem to be a criterion for causal efficacy in general. It only works with pairs of properties and descriptions, and there is no guarantee that what passes his test is causally efficacious. For all the test is worth, my shirt's being grue, or my saying 'abracadabra,' might be causally relevant to the Dow Jones index.

A third candidate criterion that I want to consider is that of nonredundancy, or causal-explanatory indispensability. It says that the causal efficacy of a property can only be causal efficacy over and above the causal efficacy of some other property or properties. Jackson and Pettit (1990, 108) write:

A causally efficacious property with regard to an effect is a property in virtue of whose instantiation, at least in part, the effect occurs; the instance of the property helps to produce the effect and does so because it is an instance of that property. (...) A property F is not causally efficacious in the production of an effect e if these three conditions are fulfilled together. (i) there is a distinct property G such that F is efficacious in the production of e only if G is efficacious in its production; (ii) the F-instance does not help to produce the G-instance in the sense in which the G-instance, if G is efficacious, helps to produce e; they are not sequential causal factors; (iii) the F-instance does not combine with the G-instance, directly or via further effects, to help in the same sense to produce e (nor of course vice versa): they are not coordinate causal factors.

Causally efficacious properties, then, are not supposed to be redundant. Surely, this is a perfectly sound restriction on causal efficacy. After all, a redundant property is not a necessary condition. But my worry with the above is that in some cases a property will slip by condition (i) while not being causally relevant. This will be the case if the dependency relation at issue is not of a causal character to begin with. For instance, my raising my hand might be thought to 'produce' my greeting, for (i) it does not ride on the back of another property: my greeting may well counterfactually depend on my raising my hand.

It may be objected that Jackson and Pettit have designed (i), (ii) and (iii) exclusively for relations that are already known to be causal. In that case, however, their conditions do not shed any light on what makes a rela-
tion causal; and in my view, that is not something to recommend them as a criterion of causal efficacy. We should note here that ‘causal efficacy’ is in fact a pleonasm: any efficacy is causal, anything causal is efficacious. If so, however, we should expect a criterion to tell us what efficacy is, and if the Jackson/Pettit criterion does not distinguish causal from noncausal dependencies (as I have suggested above), this is precisely what it fails to do.

This point carries over to the other criteria discussed, and also to ones that I have to skip here, such as Ludwig’s ‘nomic sufficiency’ (1994). To my knowledge, the criterion of local difference is the only criterion of causal efficacy that reliably tells causal dependencies apart from noncausal ones, such as generation (Goldman 1970), physical realization and, of course, conceptual dependence.

The reader is invited to check for herself how the various criteria handle questions like the following: Did your clicking on the ‘send’ button cause you to greet your friend by email? Do the spatial relations among its molecules cause the ball’s roundness? Does your brain activity at \( t \) cause your consciousness at \( t \)? Does the death of your parents cause you to become an orphan? The criterion of local difference echoes common sense in each time saying ‘No’; the other criteria, I suspect, will run into trouble with at least some of the questions.

I hope that the above considerations have done something to pardon me for having brought along my own criterion of causal efficacy in order to vindicate my relationism. If so, let us see what relational states of affairs can, or cannot, bring about.

### 13.5. RELATIONAL PROPERTIES

Earlier on, we have found reasons to hold that internal relations lack causal efficacy, being screened off from causal relevance by the properties (and, sometimes, relations) that these relations immediately supervene on. But with spatiotemporal and, most importantly, causal relations (these being external relations), things are different; and if so, things may also be

---

79 Ludwig provides three ‘necessary conditions on causal relevance’ (334): a nomic sufficiency condition, a logical independence condition, and a screening-off condition. The reason why I cannot here discuss the first of these conditions is that it would require me to take up issues about ‘event types’ (335) and causal laws. But my discussion does capture Ludwig’s two further conditions.
Part II. Causality

different with all those higher-level relational properties that, more or less indirectly, supervene on (among other things) spatiotemporal and causal relations.

Relational properties are not held in high esteem, and they are sometimes derided as 'mere Cambridge' properties.\(^{80}\) Now, that is fine in the case of relational properties such as being admired by someone: determinable properties that supervene on determinate properties like being admired by Sue, which in turn are, in fact, relations. Indeed, Armstrong (1978, 78-80) distinguishes relations (and the corresponding 'impure' relational properties) from relational properties in this way.

Yet, relational properties in that sense need hardly interest us. At least, I have not found much reason to follow Armstrong in his dismissal of impure properties (ones containing a particular), depending as it does on a conception of properties as universals, and of corresponding causal laws. On the more liberal view of properties that I endorse, we can take seriously not only relations, but also the 'impure' properties that, exclusively or not, and directly or indirectly, supervene on relations. The corresponding determinable 'purified' properties add nothing to the latter, and can safely be ignored.

Armstrong (see, e.g. 1978, 86), although in general on the intrinsicalist side, is one among the not so many analytic philosophers to explicitly acknowledge that relations do have causal efficacy. I will argue in a moment that the criterion of local difference does indeed give us ample reason to think that relational properties are every bit as causally efficacious as intrinsic ones.

We have seen earlier that higher-level relational properties (for instance, biological, intentional, historical, or social) typically supervene on (mostly past) causal relations. Of course, they supervene on very much else, but the causal component stands out. Here are some examples: only an animal with a specific ancestry can be a whale; only an object with a specific kind of production history can be a portrait; only officially certified objects can be money; an area of soil with a corpse in it can only be a grave if it was intended as such; and so on.

These necessary conditions are also sufficient ones: if they are met the item at issue is guaranteed to be a whale, portrait, and so on. For instance, even a very malformed whale will be a whale if it has the right causal his-

---

\(^{80}\) See, for instance, Kim 1974, 29.
13. Causal efficacy

tory. If so, the relation between the conditions and the properties is one of 
supervenience, and a particular's being a whale, a portrait, etc., can even 
be said to consist in its having the appropriate kind of past, one that 
includes numerous environmental factors, and that reaches right up to the 
present.

Now, such a conceptual connection between history and item-kind may 
be plausible in the case of animal species, or artifacts like money or Rem-
brandts, but it may well be denied of natural items such as hearts and liv-
ers, or believers-that-p. For why not say that something is a liver in virtue 
of its present causal role rather than its history; in the same way that a 
plank is a bridge because of the way I now use it, rather than because of its 
history?

We enter here a complex debate about functional properties that I can-
not here begin to explore, let alone settle (but see Wouters 1999). I will 
argue in Part III that in any case for intentional properties the minded sub-
ject's relation to her past is essential, i.e., constitutive. We should note that 
terms like 'whale,' 'PC,' etc. do in fact exclusively refer to items with a spe-
cific kind of history (we can safely assume that there are no whale- or PC-
like items with a seriously wayward genesis); and also that it is arguably the 
function of these terms to refer to such items. But I admit that this does 
not settle the debate on whether in general causal history enters the mean-
ing of functional terms, as it obviously does in 'needlework,' 'vomit,' or 
'memory.'

In any case, the class of properties that supervene on (among other 
things) past causal relations is quite important, especially if intentional 
properties belong to that class. Let me now, then, assess the causal signifi-
cance of such properties.

13.6. THE CAUSAL EFFICACY OF RELATIONAL PROPERTIES

To do so, I will discuss an example (borrowed from Dretske 1998) of a rela-
tional property that is largely constituted by past causal relations: monetary 
value. This example will provide an occasion, not just to apply the criterion 
of local difference, but also to compare the lesson I draw from it with Dret-
ske's own, intrinsicalist version of what is going on.

The example at issue is that of a vending machine that delivers a Coke 
when the appropriate coins are thrown in. To appreciate the significance of
Part II. Causality

this example, note that the way in which coins cause a vending machine to deliver a Coke is meant by Dretske as an analogy to the way in which intentional states cause behaviour. Dretske's point is that beliefs, desires, and the like can only steer, or guide, or cause, an organism's behaviour when they are current and internal. Just as coins, such states are taken to have relational specifications, but it is their intrinsic and current characteristics that are supposed to do the causal work, just as, it is thought, in the case of coins.

The vending machine, then, delivers a Coke when specific coins are inserted. Not all kinds of coins, but, it seems, only those that really count as money, that have the relational property of having a specific monetary value. Dretske writes:

The monetary value of an object is a relational property of that object. It has to do with its history - was it produced in a mint or in someone's basement? - and the economic practices of the community in which it exists - are such objects generally accepted as a medium of exchange in the community? (158)

The 'economic practices' are mentioned as a separate factor, along with the object's history; but obviously they are closely related to that history. The disposition of the members of the community to accept an object as a medium of exchange exists in virtue of the (known) kind of history the object has; and there must have taken place in the past an official event as a result of which objects with such histories were henceforth treated as a medium of exchange. The object's monetary value, then, is mainly a historical matter.

Common sense has it that monetary value is a causally efficacious property. Otherwise, why should we have real money in our wallets rather than chewing gum leaflets and loose buttons - or counterfeit money? But, it seems, we should know better. Counterfeit money is just morally objectionable and hard to fabricate; and occasionally people do successfully use it. Cashiers, after all, do not have a record of the history of the objects we give them. With vending machines, things seem even more obvious:

---

81 It seems false to say that as soon as an inserted object makes the vending machine respond (i.e., comes to be treated as money by both parties), it thereby becomes money. No monetary system could function that way, i.e., without a distinction between genuine and fake.
13. Causal efficacy

Though we predict the behavior of vending machines by mentioning the value of the money we put in them ("You have to deposit $.75 to get a coke.") we all know that it isn't the value of the money that explains the result. It is the shape, size, weight, and (for machines that take bills) visible marks of the objects we put in them that explain why machines behave the way they do. (...) We ignore the intrinsic properties that are causally relevant. We ignore them because, often enough, we don't even know what they are. (Dretske 1998, 160)

The criterion of local difference, however, suggests that the above is not correct. For the coins' monetary value definitely makes a local difference. Look what happens: we insert some coins with the right value, and as a result the vending machine dispenses a Coke. The 'right' local difference (to vending machine and Coke, two particulars involved in the effect) is very obvious: the machine is set in motion, and the Coke comes rolling down. All this would not have happened with coins that did not have the right value. If I throw a coin at a glass and it breaks, its value does not make a local difference. But with vending machines, things are different. Monetary value is here causally efficacious.

But may the inserted objects not have been counterfeit? Yes. But note that a reasonably well-designed vending machine will not respond to an object that is neither a real coin nor a counterfeit coin (and I do not discuss ill-designed vending machines, in the case of which monetary value may indeed fail to make a difference). So what we throw in must either have real monetary value or be counterfeit in order to make the machine deliver a Coke.

Now, monetary value is a relational property, but being counterfeit is a relational property as well. Being counterfeit is parasitic on real value. So being counterfeit depends on economic practices, just like real monetary value. It is also a historical property in a straightforward sense: an object's being counterfeit implies that it has a wayward, illegal production history. So whether the machine responds to a real or a counterfeit coin, it is relational properties that make it respond, for without these nothing happens.

But what about objects that have all the intrinsic properties of real coins, but purely by coincidence? Well, in those cases in which such objects are available, monetary value or counterfeitness will indeed be causally redundant. But only in those cases! And vending machines are built in such a way that such cases are rare. So with well-built vending machines, monetary value or counterfeitness will be causally efficacious in most cases. One might try to defeat this point by demanding that causally efficacious
properties must be so under all circumstances. But that will leave no causal efficacy whatsoever. On such a demand, even wearing a bullet-proof vest will be causally inefficacious, even if it saves your life on a regular basis.

Finally, is it not the case that when we change the value of the coins without changing the physical properties of coins or machine, the Coke will come out anyway? Coke-dispersion, it seems, tracks the size and shape of the coin and not its monetary value. Now surely, the coins' value may fluctuate a little without directly affecting the vending machine's physical properties. Such fluctuations, I admit, will be causally inefficacious (at least as far as the vending machine is concerned). But this does not show us that monetary value is in general causally inefficacious. For the suggestion that one can vary the coins' value while keeping fixed the vending machine's physical properties makes sense only for minor fluctuations in value. In other cases the suggestion is mistaken. If the value of the coins changes substantially (say, by inflation) and the Coca Cola Company does not allow their product to become ever more cheaper, the physical properties of the vending machine will vary along with the monetary value of the coins. Coke-dispension, then, does track monetary value.

I am aware that the argument just offered may not be able to dislodge the intuition on the part of many readers that in spite of the above it is really the coins' physical properties that do all causal work. Note, however, that the example of coins and a vending machine, although instructive, is also a little treacherous, in that tricking a vending machine that works with coins is not too difficult. The simpler the system under consideration is, the stronger will be the suggestion that all efficacy is merely local and physical, and that relational properties are irrelevant. I want to ask the unconvinced reader, however, to also check her intuitions against a vending machine that works with chipcards. It seems quite unlikely that an object that is not either a properly loaded chipcard or a clever forgery will ever make such a vending machine respond.

We now have seen that the criterion of local difference vindicates the causal efficacy of relational properties, at least the kind that supervenes on past causal relations; but there is still room for scepticism. Does the alleged causal efficacy of relational properties like monetary value not ride piggyback on that of intrinsic properties? It seems that the relational properties can do their causal work only when the right intrinsic properties are also in place. A lottery ticket or a bar of chocolate may have the same value as the coins, but it will not make the machine respond. If so, relational properties
like monetary value run afoul of Jackson and Pettit's conditions for causal efficacy (see above): they are redundant as causal factors.

But a coin's physical properties and its monetary value cause different things. There are different ontological levels to be reckoned with. It is true that the physical properties are required for the value to be efficacious, but I wish to argue that the former could, on their own, never cause what the latter causes. Dretske, endorsing event token physicalism, loosely talks about vending machines yielding their contents, or giving us a Coke, under the assumption that such events are token identical with the shifting and clanking inside the machine. But that assumption is highly problematic.

For in contrast to the physical events inside the machine, a vending machine's giving a Coke is an event that is highly (social-) relational in character. In order for there to be things that are a vending machine and a Coke, an entire economy must be in place, and there can be no literal 'giving' without rules of ownership. Therefore, if you explain to someone the physical events inside the machine by mentioning the coins' shape, size and weight, you do not explain the Coke transaction. You just bother your interlocutor with physical details without telling her what causes a vending machine to make someone the owner of a Coke, that is, what explains the change in ownership of a piece of merchandise.

Note that Dretske himself does believe that the coin's monetary value explains (although it does not cause) something of its own (169):

The value doesn't explain why the cokes come out, but it does explain why coins - objects of that size and shape - cause cokes to come out.

But if 'explain' is meant to be causal explanation, then this is surely not what the value explains. Dretske obviously means here a local, physical process. This process will have a structuring cause that explains it, but if we take seriously the doctrine of causal closure, this can only be a past physical episode, not something involving a higher-level property such as monetary value. Coins' monetary value does explain the dispensing by vending machines of Cokes, just as common sense tells us.

82 Perhaps we should write 'Cokes' (capital C) when we mean branded products, and 'cokes' (small c) when we mean merely physical particulars. Dretske writes 'cokes,' and he seems to mean the latter; but he does not distinguish between the two possible readings.
Part II. Causality

13.7. THE CAUSAL EFFICACY OF THE PAST

Relational properties, then, especially higher-level historical-relational ones, can have causal efficacy over and above the causal efficacy of physical-intrinsic properties. They are 'coordinate causal factors,' along with physical properties, in causing states of affairs at ontological levels that cannot, for reasons that we have explored before, be understood in terms of basic physics alone. Of course, higher-level states of affairs need not always be caused by same-level states of affairs: the dinosaurs allegedly went extinct as a result of a meteor strike. But they often are.

Now, a significant amount (I suspect all of them, but I will not insist on this) of the higher-level relational properties supervene on past events and causal relations. But how could historical causes ever enter present causal relations in a world in which all causal influence is here-and-now?

It is here that Dretske's notion of a structuring cause becomes important. Let us remember once more that all causation is circumstantial: my pounding on the keyboard would not have produced a visible text if the computer had not been switched on, my pushing would not have opened the door if it had been locked, etc. And the causally relevant circumstances typically are many. This means that every causal relation requires the world to be structured in a certain way in order to take place. Seen in this light, McGinn was quite wrong: causal powers more often than not do 'depend upon relations to what lies quite elsewhere.'

And here is the rather obvious but crucial point: the present state of the world is what it is only in virtue of past events. These past events can be seen as the structuring causes of present causal relations. Circumstances have to be brought about, the world has to be structured, the causal lay-out of things has to be configured so that events such as c will have events such as e as their effect. But something must do the structuring, and we do not seem to live in a world in which miracles do it. Therefore, past events are indispensable for present causal relations to obtain.

But being predisposed by some earlier event or episode to cause e, or being endowed by it with its present causal powers, is a property that c (or better, the particular constituting it) must now have. It cannot do its present causal job without that historical-relational property. Its intrinsic properties are not enough, because the combination of these with the circumstances is crucial to what will and will not be caused.
13. Causal efficacy

True, if this whole combination of factors is presently in place, that will be sufficient for bringing about e. But in a world without miracles, such a situation must be shaped by history. An appropriate history is, then, causally required. The design procedure of money and the publicity that went along with it, for instance, have caused present states of affairs involving particular coins, but have also caused states of affairs involving the people using them (they know that coins looking like that are quarters). The design procedure, publicity and manufacture are not just parts of the coins' history; their causal impact has fanned out to the wider circumstances in which their presence has its specific effects (vending machines respond to them, as do cashiers). A particular's historical-relational properties, then, cause events in virtue of the fact that these events would not have occurred without that particular's special history, barring miracles.

One might object that it is the history, a past episode, that has been causally effective, rather than the current historical-relational properties of an entity. But the free lunch principle tells us that there is no difference between these. The past occurrence of an episode involving some present item and that item's possessing its historical-relational properties at present are just one and the very same state of affairs. For the latter directly supervenes on, and therefore merely consists in, the former. Compare: 'this coin has monetary value,' and: 'there has taken place an episode involving the development of economic practices, official events, publicity, and exclusive manufacture of legal tender - among which this coin.' These two sentences say the same thing. A coin's being worth a quarter now consists in some history's having taken place.

And the causal efficacy of the coin's monetary value consists in the fact that without such a history circumstances would be such that the coin would not now make vending machines, cashiers, etc. respond appropriately. That is, historical-relational properties are doing their causal work in the present. They are presently had and bring about future events. And these properties do their causal work along with and in no way disrupting the causal efficacy of nonhistorical properties. They cause different things.
14

Supervenient causation

14.1. Epiphénomenalism or downward causation: a dilemma?

I have argued above that the relational character of higher-level properties such as monetary value need not at all be an obstacle to their causal efficacy. It can be claimed, however, that the relationality of such properties is not the real issue anyway: what threatens their causal efficacy may rather be their supervenient, higher-level character.

For remember the doctrine of causal closure: it tells us that no merely physical state of affairs has any other than merely physical causes. We might think for a moment that this leaves room for higher-level causes of higher-level states of affairs, but when we realize that such states of affairs supervene on, and therefore are not even distinct from merely physical ones, we realize that in causing specific lower-level states of affairs one also causes anything supervening on it. If we want to cause a higher-level state of affairs (say, a Coke-transaction) it will do to cause its physical realizer: the transaction is, we might say, already present in this realizer. But the realizer has only merely physical states of affairs as causes, and all further alleged causal efficacy could only be redundant.

The upshot seems to be that lower-level, merely physical states of affairs usurp all causal efficacy. All causal work is at the lowest level, and no matter how useful or transparent causal explanations invoking higher-level states of affairs can be, they do not, strictly speaking, refer to any causally efficacious factors. Properties like energy, charge and spin do the real work; all the rest (hungriness, monetary value, believing that p, etc.) idly rides along.

Or, to make the same point in a different way: no level except the physical one is causally closed. Downward causation is problematic, but upward causation (not in Searle's sense, but through time) happens all the time. Sunlight and chemical reactions cause plant growth; sunsets cause perceptions; chemicals cause hallucinations and states of mystical elation;
radiation causes illnesses; actions of individuals change societies; meteor impacts cause new evolutionary processes; and so on.\textsuperscript{83}

If so, however, upward causation will not just be possible, but pervasive. For because the physical realizers of all higher-level states of affairs have complete physical causes, all higher-level states of affairs can be said to be 'upwardly' caused by physical states of affairs (namely the causes of their realizers). The latter, then, seem to take care of all causal efficacy, at all levels at once. What causal addition could a higher-level state of affairs make, given that ontologically speaking it is not even distinct from its physical realizer?

And for reasons outlined earlier, a reduction of higher-level domains such as the intentional or the social to physical reality is out of the question. A coin's having monetary value cannot be construed as a physical state of affairs, because it involves all kinds of relations (such as ownership, meaning, etc.) that are absent in the domain of merely physical states of affairs. Reductionism, then, is no way out. Nor is any thesis of event token identity: even if there were a way to non-arbitrarily identify an economic event with a physical one, this would still lead to the causal privilege of the physical over any higher-level states of affairs.

So, again: we physicalists who accept causal closure seem compelled to deny higher-level causal efficacy. Alternatively, noting that this epiphenomenalism is counterintuitive, we might simply insist on the causal efficacy of higher-level states of affairs. But in that case it seems that the principle of causal closure has to go. For if causal efficacy for higher-level states of affairs implies causal nonredundance, they must have effects that are not already fully accounted for by merely physical states of affairs. But there could not be such effects, it seems, for every state of affairs has a physical realizer that is entirely physically caused. Unless, of course, there were 'downward causation,' i.e., disruption of causal closure.

This would require scientists to give up the idea of a complete physical theory, or even the very idea of laws of physics. For on such a picture, every time there are higher-level factors at work one can no longer trust the laws that govern the merely physical domain. The higher-level states of affairs will take over the causal steering wheel, and this can only happen if the

\textsuperscript{83}This is the reason why, in the context of the problem of mental causation, the focus is on causation by intentional states (namely action) and much less on causation of intentional states (namely perception). The latter just does not seem to raise metaphysical issues, although it does raise epistemological ones.
merely physical factors are pushed aside for a moment. So if physicists want to have reliable laws, it now seems that they should be ready to quantify not just over physical, but also over biological, intentional and social factors. Now, given that modern physics does not seem to have anything of the kind on its agenda, it seems that we should be sceptical about its results. Never mind how well-tested the Standard Model of quantum mechanics, because of the causal efficacy of phenomena that it does not cover (such as our digestion, our thoughts, or our money) it just cannot be right!

Of course, this outcome is at least as preposterous as epiphenomenalism. But could we not avoid both? In the present chapter I will explain why the options that I have just sketched present a false dilemma, and how there can be genuine higher-level causal efficacy without disruption of causal closure.

14.2. RELATIONALITY AND CAUSAL AUTONOMY

We have seen earlier that in spite of supervenience on physical states of affairs, higher-level states of affairs do have their ontological autonomy. This is so because first, macro-entities can be related in ways micro-entities cannot, and because, secondly, higher-level relations are mostly between items that are themselves relational in yet other ways (i.e., supervenience is very much indirect). There are, then, entire domains of (kinds of) relations over and above the merely physical states of affairs. It is not that these domains are not somehow present in, or implicit in, or given along with, the totality of merely physical states of affairs; it is just that they are extra aspects of that (one) reality. Reality is all physical, but it has many levels of relationality and, hence, many different kinds of properties.

We will now see that the same relationality (and the holism that goes with it) that gives higher-level states of affairs ontological autonomy gives them also causal autonomy. Consider again the coins’ monetary value’s causing a Coke-dispension by a vending machine. Let us start by calling the physical realizer of the vending machine’s dispensing a Coke $P^*$. Because this dispensing is a highly relational event, as we have seen, $P^*$ will be complex, wide, and diffuse (spatially as well as temporally). It does not just consist in some shiftings-about in a piece of machinery, but it involves peo-
ple, banks, factories, etc. The coins’ being worth so-and-so much also has such a physical realizer: let us call it P.

Here is a first reason for suspicion of the causal exclusion argument presented above. In order for one to be able to claim that the physical realizers usurp all causal efficacy, one has to claim that P causes P*. But does it? No. Parts of P and P* lie in the past (occupying, we can be sure, diverging space-time regions), and it is hard to see how such temporally scattered physical items could be each other’s causes or effects. The most that we can say is that perhaps the here-and-now part of P causes the here-and-now part of P*, and that in virtue of this, P* is completed, as it were. But P’s causing P*, as far as such a thing does happen, does not seem to be the same thing at all as the value’s causing the Coke-dispension.

It remains a fact, however, that causal closure implies that P* has a complete physical cause (the real P, whatever it is). There is a P, then, that causes P*, and P* realizes the Coke-dispension. So P, whether or not it realizes some coins’ monetary value, ‘upwardly’ causes the Coke-dispension, and so usurps the value’s causal efficacy, does it not?

In a sense, yes, but in the sense that matters: no. P (by definition) causes P*, and P* (by definition) realizes the Coke-dispensation. But because P* is wide and diffuse, it realizes far more than just that. For instance, it realizes the vending machine’s colour, the roundness of the Coke can, events concerning people partaking in the Coca Cola business, etc. The totality of states of affairs realized by P* can surely be said to be caused by P. But not the states of affairs separately, at least not the ones involving relational properties.

For that is the whole point of having a wide supervenience base: many relational properties (e.g., economic ones) form a tangled web, and they arguably do not have physically isolable realizers. What, for instance, would be the exclusive physical realizer of Coca Cola’s being a trade mark? There is no such exclusive realizer; there is, rather, a wide base that realizes it along with a host of other states of affairs. The realization relation between physical and economic states of affairs is not one-to-one, or many-to-one, but many-to-many.

The result is that P cannot be the cause of the vending machine’s dispensing a Coke in particular, for P* cannot be the realizer of that event in particular. Along with the Coke-dispensation it realizes a complete package of relational states of affairs. So in a sense P causes the dispensing, but
Part II. Causality

merely in the sense of causing a cluster of states of affairs of which the dispensing is one.

14.3. COMMENSURATENESS AND THE CORRELATION PRINCIPLE

We may ask, however, whether that is not perfectly legitimate causation. But a good reason for suspicion is provided by Yablo’s plausible demand that causes be ‘commensurate’ to their effects (1992, 404):

(N)ething causes an effect that leaves out too many relevant factors, or brings in too many irrelevant ones.

For instance, if I were asked what caused my headache, and I answered: ‘The situation of my body and environment of the past two weeks,’ I would bring in too many factors and so give an answer that was empty. That answer just would not describe the cause of my headache. In this example it would no doubt be possible for me to be more precise: it is always possible to be more precise about causes.

But now consider the question: ‘What was the cause of the Coke-dispension?’ We have seen that this question has the straightforward answer: ‘I inserted so-and-so-much’s worth of coins.’ But why not just say: ‘P’? Well, that brings in too many irrelevant factors. Saying it, in fact, amounts to saying: ‘The entire physical situation underlying all that Coke business.’

So as before, we will have to be more precise. But can we? Not in a merely physical vocabulary, if the supervenience base of a Coke-dispension is indeed wide. So in physical terms, we could only give an answer full of irrelevant information. The relevant and irrelevant factors would be inseparable at the physical level. Now, Yablo’s ‘commensurateness’ requirement on causes suggests that if we cannot do better than that, we just are not able to specify a cause.

Let me try to make this same point in a slightly different way. Consider the following principle:

Correlation principle: for every possible cause, a causal explanation is in principle (although, perhaps, not in fact) available that identifies that cause.
This principle articulates a basic intuition about causality. Allowing it to be violated will render the very notions of 'causal explanation' or 'causal statement' unintelligible. A cause that could not in principle be described, or referred to, seems an absurd, perhaps even a contradictory notion.

Those who claim that all causal efficacy is usurped by physical states of affairs should say that \( P \), rather than the coins' value, causes (in causing \( P^* \)) the Coke-dispension. But \( P^* \) is a wide situation, and the Coke-dispension is merely an aspect of it. \( P \), then, does not cause the Coke-dispension but, rather, the large and tangled cluster of relational states of affairs supervening on \( P^* \).

In order to honour the correlation principle, one would have to divide up \( P \) into 1) the cause of the Coke-dispension and 2) the cause of all the other aspects and parts of \( P^* \). But this cannot be done, for the Coke-dispersion does not have a physical realizer of its own. If so, it does also not have a physical cause of its own. It definitely does have a cause of its own, but that cause is same-level. If it is nevertheless maintained that the Coke-dispersion has a physical cause, although that cause could not possibly be specified, the correlation principle is violated.

My conclusion here is that because of their relational and holistic character, higher-level states of affairs do have their causal autonomy: they cause things that no local phenomena can cause. Physical realizers of higher-level states of affairs may, because of their broad character, not be able to cause other such states of affairs in particular. Identifying a higher-level cause or effect in terms of its supervenience base is about the same thing as pointing at an individual in an ant-hill with the help of a tree trunk.

### 14.4. DOWNWARD CAUSATION AND CAUSAL CLOSURE

Higher-level states of affairs, then, cause same- or higher-level states of affairs. And insofar as such states of affairs are not identical with merely physical states of affairs, higher-level states of affairs do their causal job without disrupting or overdetermining the course of events at the merely physical level. On different grounds, such a dual-explanandum account is also advocated by Corbí and Prades 2000, among others. They write (200-1):

---

179
Part II. Causality

(N)onbasic causation does not necessarily entail overdetermination because, strictly speaking, the nonbasic and the basic cause do not explain the same effect, since they are individuating the effect in contrast with different sets of other possible outcomes that might have taken place if the cause at stake would not have occurred.

It may seem, however, that this is not sufficient to account for even the most routine cases of supervenient causation: intentional action. For does not intentional action often have plain physical effects? Consider the following example: I take a walk on my own. One moment I kick a stone because I want to. The stone lands into a pond and it causes ripples. The ripples are clearly caused by my desire and action (something higher-level), but are themselves merely physical. How, given causal closure, can my action, or my desire to kick, be in the ripples’ causal etiology? There is just one effect: ripples in the pond. And that effect, being merely physical, can be fully causally explained in merely physical terms.

It seems that if we are to honour causal closure, we should admit that although the desire to kick was really a circumstantially necessary condition for the ripples (as it was, had I not wanted to kick that stone, the pond would have remained undisturbed), the causal efficacy of this desire, as a higher-level state of affairs, must be trumped by that of its supervenience base.

True, a physical explanation of the ripples must involve a description the (presumably wide) physical base of the desire. Otherwise, mentioning the latter will still provide causal information lacking in a merely physical account. This will make a full physical explanation of the ripples extremely complicated, opaque and and uneconomical. By following the ripples’ causal history we will surely notice that neat causal threads quickly unravel. The ripples have a broad set of immediate circumstantially necessary conditions, and these will themselves have many such conditions, and so on: we will face a genuine combinatorial explosion of circumstantially necessary conditions, especially when we have to include the wide and historically based supervenience base of a desire. But all this does not change the fact that in principle a full causal explanation of the ripples can be given in merely physical terms.

But then, are we not committed to saying that the desire (or at least the ensuing action) is epiphenomenal? Not necessarily. It would be epiphenomenal if it had no further effects. But it may have further effects: it may cause further mental states of affairs on the part of the actor, such as satisfaction or frustration, knowledge of his own action, etc. I admit that there
14. Supervenient causation

need not be further effects: a desire may be very brief, it may be very soon overridden by other desires, it may be completely forgotten, etc. It may also be one’s very last mental state. It does not seem possible to exclude higher-level epiphenomena. We should not even rule out the possibility of an argument to the effect that most of our mental states are without effect. But an all-out epiphenomenalism about higher-level states of affairs is in any case unwarranted.

14.5. DOWNWARD CAUSATION AS CONSTRAINT

Saying that causation can only be same-level or lower-higher-level, but never downward seems rather austere. And there does seem to be downward causation in some sense. No doubt, many will feel that the causal closure doctrine violates common sense in cases such as the above: the desire causes the action, which causes the stone’s trajectory, which causes the ripples. By transitivity: the desire causes the ripples. Mental states of affairs (or, in general, higher-level ones), it may be insisted, do cause states of affairs that are merely physical. Also, being uncomfortable with epiphenomena, would we not expect that our last silent thought should at least cause a little brain activity?

Note also that earlier on, in the context of the criterion of local difference, I argued that the phrase ‘c is a necessary condition for e and for local state(s) of affairs l involving the particular(s) involved in e’ should not be taken to mean: c causes the local state(s) of affairs l. But suppose that by talking to you I change your mind. Do I then not also just change the physical state of your brain?84

Or suppose that I do not kick a stone into the pond, but jump in myself and drown. Now, to simplify a little, the only plausible effect of my desire-

---

84 Or consider the claim that same-level intentional causation is typically mediated by merely physical events. For instance, if I write a book that will years later change your mind, I seem to be creating a physical structure of paper and ink that will later on be a cause of your new thoughts. But this claim is mistaken: if my writing is a circumstantially necessary condition for your change of mind and there is an underlying physical infrastructure, the conditions for causal relatedness have been satisfied, and worries about intermediate physical events are beside the point. We should not conflate levels. My writing does not cause the physical structure of paper and ink (that is taken care of by my bodily movements), but the meaningful message embodied by that structure.
action is the ripples (I might never be found). Would this not be downward causation?

Or consider the motion of a ball that is plausibly said to cause the trajectory of one of its surface molecules; a mass hysteria that causes individuals to shout slogans; the structure of organisms, or other complex systems, that cause its chemical parts to behave in very special ways. Are these not definitely cases of downward causation?

Juarrero (1999) indeed suggests that in order for us to understand complex systems (as well as human action) it is essential that we appreciate the latter form of downward causation. It is, she claims, the very nature of complex systems to bring micro-elements under a macro-regime that at the same time constrains their behaviour as well as it makes them behave in new, coordinated ways. In the same spirit Van Gulick (1995, 252), noting that physical systems (from rivers to living creatures) often exhibit stable patterns of macro-behaviour (from vortices to maintaining homeostasis) writes:

Such patterns can affect which causal powers of their constituents are activated or likely to be activated. (...) Thus the whole is not any simple function of its parts, since the whole at least partially determines what contributions are made by its parts. (...) (H)igher-order patterns can have a degree of independence from their underlying physical realizations and can exert what might be called downward causal influences without requiring any objectionable form of emergentism by which higher-order properties would alter the underlying laws of physics. Higher-order properties act by the selective activation of physical powers not by their alteration.

And Holland (1998, 239-40) notes that

when a projected change in interest rate causes a sell-off in the stock market, we have an unambiguous example of aggregate quantities affecting the action of individual agents.

Let us now see whether we can do justice to such allegedly downward causation without violating causal closure. It will be appreciated that none of the latter cases involves a disruption of causal closure. If anything, a causally closed physical domain makes such cases possible rather than precluding them. What is to be causally explained (the ‘activation’ of some ‘physical power,’ or the behaviour of an individual investor) can be fully causally explained on the relative lower level, in the sense that there is no mystery at all in what is going on when the macro-states of affairs are
14. Supervenient causation

ignored. And the totality of micro-states of affairs will be seen to causally screen off the macro-states of affairs.

The kind of explanation appropriate to the above examples, although at first sight it seems causal is, in the words of Jackson and Pettit (1990), program instead of process explanation. In explaining the trajectory of a molecule by saying that it is at the surface of a rolling ball, for instance, we do not just trace its causal antecedents (in which case a host of micro-states of affairs causally screens off the rolling of the ball as an autonomous, or extra causal factor); we say that a trajectory like that typically occurs in cases of rolling spherical objects.

The idea here is that macro-states of affairs constrain the behaviour of the micro-elements, rather than straightforwardly cause it. For constraints are not causes. Your being named Larry constrains the amount of consonants in your name; the surface area of my garden constrains how many plants I can grow there; the law of conservation of energy constrains what can happen next; and so on. True, causes also constrain what can happen afterwards, but that is not what their being a cause consists in. Being a circumstantially necessary condition of something is different from being a constraint on it, although a constraint can be a circumstantially necessary condition. We might say that to cause is to enable possibilities, while to constrain is to exclude them.

Now, by constraining, or shaping, or selectively activating, or coordinating, or controlling, the micro-behaviour of its parts, a system also controls what that behaviour causes. And that is an important fact indeed: we will see later on that human action is all about control, or lack of it. Yet, pointing out the existence of such downward causation to the epiphenomenalist who invokes causal closure misses the latter's point. For it still seems to be micro-behaviours that do all the causal work.

What to say? Can we not say that the state of a system s, by programming for micro-behaviour of type C (which implies an event c), causes e? Indeed, is not saying that it was not the kicking but rather its physical realizer that caused the ripples just pedantic?

I would say that it certainly is, and that of course we may very well speak of some form of 'downward causation' in the examples discussed. The problem, however, is that this will not put our worries with causal closure to rest. Even if the 'downward' causes typically will be circumstantially necessary conditions for the alleged effect, they will not be such conditions in addition to, or over and above, the lower-level states of affairs realizing
them. So they are redundant. Which means that they are not genuine causes.

14.6. DOING VERSUS EXPLOITING CAUSAL WORK

Our discussion of supervenient causation, then, teaches us two things. First, that there is autonomous higher-level causation. It teaches us that we can safely speak of intentional states of affairs (including beliefs, desires, and actions) causing other intentional states of affairs. Secondly, however, our discussion has taught us that the constraining, or control, exerted by a macro-system on the behaviour of its micro-parts, is not causation. This implies that it is not the case that intentional states can cause mere bodily movements, or the effects of these. It implies that an intention can cause an arm-moving, but not an arm-motion. It also implies that the merely physical effects of our actions cannot be the effects of any intentional states on our part.

This conclusion was foreseeable: it is a fairly straightforward consequence of the causal closure principle. But is it acceptable? Does it not leave us with a very meagre notion of mental causation? In my view, such a conclusion can be accepted when we realize that we are embodied selves, that is, physical systems. The causal work of our micro-parts may not be causal work that we do, but it is at least work that we organize, orchestrate, coordinate, control, or put to use. We exploit it, and in that sense it is definitely ours.

We saw, however, that the control cannot be externally imposed by us on these micro-parts: if genuine downward causation is impossible, there cannot be a self that is distinct from these parts doing the imposing. The only possible imposer left is the micro-parts themselves, collectively. We do not have here, we might say, control by causation (as is the control exerted by a CEO on a firm, or by a processor on the data flux in a PC), but control by structure.

The view that there must be a controller, coordinator, etc. distinct from what is controlled, coordinated, etc. will remind us of Cartesian dualism. Now, it is true that in contemporary philosophy of mind this view is largely abandoned. Yet, the view that controller and controlled system are distinct entities is still very much alive. Today, the putative controller is no longer the Cartesian ego; it is the brain. Many philosophers assume that the mind
(or the self) is the brain, or is in the brain, or consists of brain states. Pinker (1997, 21) crisply expresses this view in saying that 'the mind is what the brain does.' Indeed, is it not obvious that the brain controls the rest of the body? Does it not receive the perceptual inputs, process the information so gained, and issue the motor outputs? Where else should we look for the mind, or the self?

But now suppose that we take seriously the suggestion that we are self-steering and self-organizing physical systems (in which, of course, the brain plays a significant role). Why then should the mind, or self, not be in the organism as a whole, that is, in its dynamic and adaptive organization? Indeed, given that we do our self-steering and self-organizing in response to the world around us, why should a self, or mind, not be in the entire ongoing process of mutual development involving an organism and the environment in which it is embedded? It is time that we should ask what having a mind consists in.
Part III. Mind

The previous discussion of causality is the core part of my account of mental causation. The results so far will already have suggested in which directions solutions to the problems of mental anomalism, externalism and causal exclusion are to be sought; and they will constrain my account of the problems of normativity and phenomenal states of affairs. Yet, in order not to prejudge issues any more than necessary, I have tried to keep my account general, largely abstracting away from mind-related issues.

It is now time, however, to explicitly discuss the mind. For although in my view the problem of mental causation is mainly a matter of misconceptions about causality, the debate about mental causation also suffers from a number of problematic assumptions about mental properties. In the following chapters my main targets will be syntacticalism, functionalism, internalism, reductionism, and token (event) physicalism. Together, these views yield a picture of the mind that is congenial to the metaphysical view that I have earlier called intrinsicalism. For on all these views, mental states are most plausibly construed as brain states.

Surely, these views can square a good deal of known facts about mind and action with an intrinsicalist metaphysical picture. But they also make the problem of mental causation loom large, render the phenomenon of subjective experience an intractable mystery, and do not at all sit well with the ever more prevalent scientific wisdom that nothing like what one might call a boundary separates us embodied and embedded selves from our natural and social environments. That is why we need a change of perspective, away from intrinsicalism and towards relationism.

The reader should be aware, however, that Part III will be different in character from Part I and II. In Part I, I have tried to say something original about physical realization; in Part II, about causation and causal relata. For Part III, however, no such thing is on the agenda: Part III will be critical rather than original. In this part, I will target the views just mentioned, and only tentatively suggest available alternatives. This will leave the reader with a host of suggestions and undecided issues, rather than with a rounded-out theory of mind. In my view, however, it will in any case be enough to achieve the goal of my project: to put the five problems of mental causation in their proper perspective.
I will proceed as follows. In the first chapter of Part III, I will attempt to give a general naturalistic sketch of what being a minded creature amounts to. This introductory chapter is more exploratory than argumentative in character. In the subsequent chapters, I will criticize the following problematic positions: the computational theory of mind, or syntacticalism; the so-called Theory Theory of folk psychology, and functionalism; internalism (including a related view called 'weak externalism'); and finally, a variety of psychophysical identity theories, in particular Kim’s functional model of reduction, and token physicalism about mental states.

Although I will indicate the significance for the five problems of mental causation of the issues discussed in these critical chapters, I will there abstain from drawing final conclusions. Many implications with regard to mental causation will become clear enough during these discussions, but making up the final balance is best postponed until all the relevant materials are in. In a concluding chapter, outside of Part III, I will summarize my arguments, and then state my verdict on what has and has not been achieved with regard to the five problems of mental causation.
15
The concept of mind

15.1. INTRODUCTION

Let us assume, as is common, that a general account of what it is to have a mind should be stated in terms of intentionality and consciousness. Such an account may then sound like the following:

*Having a mind amounts to having a subjective, first-person perspective on the world. A perspective that is one's own, and that guides one's behaviour, often but not always by way of practical reasoning. Having a mind amounts to representing aspects of the world to oneself, having preferences and aversions about these and, normally, to make practical decisions that are (respectively) informed and motivated by these. Typically, there will be a feel to all this, a subjective phenomenal quality. These features, intentionality and consciousness, uniquely characterize the mind.*

Consciousness is one's having a first-person, subjective, phenomenal view at reality; intentionality is one's world-directedness, or the 'aboutness'-character of some of one's properties. Although consciousness and intentionality are closely related, it seems that genuine intentionality is possible without any consciousness (much is unconscious in our own minds), while genuine consciousness is impossible without intentionality (arguably, you cannot be conscious of nothing whatsoever). This seems to make intentionality the basic feature of minded creatures.

Let us in the sections below see how we can make naturalistic sense of the above. Assuming that intentionality is a biological phenomenon, or at least arises from biological processes, and that biology implies goal-directedness, I will first outline how we may understand such teleology. On the basis of this notion, I will proceed to explain intentionality by discussing representation. I will then highlight, first, the fact that intentionality is a thoroughly normative notion, and secondly, the fact that it involves perception, thought and action as a tight conceptual cluster.
Consciousness, of course, deserves separate treatment. In the last section I will say a few things about consciousness, especially about its mysterious 'phenomenal' character. This phenomenon, remember, gave rise to our third problem of mental causation: how can phenomenal qualities, or qualia, have any causal relevance? Let me already admit that I will be unable to solve this problem as it is standardly raised. If we first abstract the 'phenomenal' from anything 'functional' and then ask how the phenomenal can still have causal efficacy, epiphenomenalism seems the inevitable outcome. But then, this very problem may well rest upon faulty assumptions. And this is what I would like to suggest below. But turning this suggestion into a solid argument would require an extensive discussion of consciousness, which would go far beyond the scope of my project. This is why I will limit myself to the few remarks below.

15.2. MIND PRESUPPOSES TELEOLOGY

From a naturalistic standpoint, mindedness is best regarded a biological phenomenon. It is the only way of accounting for the fact that there are minds at all. Having a mind is a way of making a living, functioning, or exploiting an ecological niche typical of a variety of living species. If so, we should begin by explicating what being alive amounts to. Of course, living creatures are causally receptive to and active in their environments. But that is not a difference with mere physical objects. The difference is that being alive implies that the causal processes involved serve a goal, and can therefore be appropriate or inappropriate, or normal and abnormal.

Although there can be plenty of scepticism about the reality of goal-directedness, or teleology, in our 'world of causes,' there do seem ways of making good naturalistic sense of it. There is a large body of literature on this subject that I cannot begin to properly address, but let me pick out two complementary naturalistic approaches to teleology.

First, there is the dynamic systems approach explored, among others, by Juarrero (1999). Juarrero points out that when we see an actor, in the spirit of systems theory, as following a trajectory through a many-dimensional 'state space,' we will expect the system that is the actor to move towards attractors in that space, if there are any. 'Attractor' is a mathematical concept that has proven a powerful tool in describing well-documented cases.

85 Quoted from the subtitle of Dretske 1988.
of resilient behaviours of dynamic systems, from simple pendulums to complete ecosystems. Very roughly, an attractor is a type of state in which a system is likely to end up (that is, until it is disturbed) when it is in a range of specific ‘nearby’ states.

Juarrero argues that the manifold of attractors that can be expected to be present in an actor system (and also the occasional major shifts in such an attractor regime as a result of environmental perturbations) very well accounts for the phenomenology of action, such as the uncodifiability of practical wisdom, weakness of the will, or pathological phenomena like compulsory behaviours.

For our present purposes it is important to note that regarding an organism as a multi-dimensional attractor regime may well account for behaviour that intuitively strikes us as goal-directed (such as searching or hunting behaviour) without introducing something like Aristotelian final causes. The dynamic systems approach can be regarded as a replacement, or perhaps further development, of the older ‘cybernetic’ model of goal-directed behaviour, according to which feedback mechanisms were at the core of such behaviour.

Secondly, there is the etiological approach to teleology, advocated by, most notably, Millikan 1984. On this approach, which centres on the notion of proper function (i.e., what properties of living beings are for), we can say that a persistent structural or behavioural trait of a system or subsystem (an entire community, an organism, or some part of an organism) has a function if its presence is not a coincidence. The latter must be the case in a very specific sense. A trait (such as the chemical composition of a protein, a pelt’s colour or thickness, an animal’s inclination to sweat or fight under specific conditions, and so on) must owe its presence to previous instances of it, be it in the present system or in earlier systems connected to it by a process of copying (That is, it must be a member of a ‘reproductively established family,’ see Millikan 1984, 23-5). The previous trait must then have helped in bringing about, or sustaining, the presence of the present trait. But if this is the case, this entitles us to say that the trait is there for some goal. It will then also justify talk of malfunction or abnormality.

Function so understood becomes a historical notion: when an organism’s present traits can have a function only if their presence is non-
Part III. Mind

coincidental in the way described, then having a function implies having a specific kind of past. Only if a trait is the product of a very specific kind of causal process can it be a function.

There are of course numerous issues that complicate such an account of proper function. Traits can be backed up, change function, be recruited for multiple purposes, and so on. There will no doubt be indeterminate cases in which a clear ascription of function can only be arbitrary. Functions can also belong, not to a trait of an organ or organism, but to items in the environment, such as artifacts. And arguably some functions, such as linguistic meaning, can be instantiated only when entire communities are present.

Millikan 1984, however, has developed a sophisticated conceptual apparatus (featuring 'higher-order reproductively established families,' 'adapted devices,' 'proper functions,' 'relational' and 'adapted,' 'derived,' 'stabilizing,' 'serial' and 'focused' proper functions,) that may be up to a good deal of the many tangles.

15.3. MIND PRESUPPOSES REPRESENTATION

In order to understand what the mind is we have now to move beyond mere teleology. For mindedness is a way of being alive, but it is quite a special such way: it involves representation. Special to mindedness, or intentionality, is that the minded creature represents parts or aspects of the world; including, significantly, the creature itself.

More precisely, and importantly, a minded creature represents the world to itself. Books, flags, and television shows do represent aspects of the world, but not to themselves. They do not have intrinsic or original, but merely derived intentionality. Representations are always there for the purposes of users (or 'consumers' as Millikan calls them), and books, flags, etc. cannot be users, at least insofar as they do not have purposes of their own. It is typical of a minded creature that it, as a consumer of representations, has a producer (mostly the brain) as a part.

Now, what is representation? How does it come about? As with function, this is a vast field of issues with its own vast body of literature. But let me just briefly sketch, roughly in the spirit of Dretske's 'informational semantics' (1981, 1988), how representation can come about in a world that is merely causal and teleological.
15. The concept of mind

Let us first look at what we could call a proto-intentional concept: indication. Indication, or 'natural meaning' (Dretske 1988, 54)\(^8\) involves the stable, non-coincidental co-occurrence of at least two phenomena: the indicator and the thing indicated. Well-known examples are fire and smoke, and walking-by and footprints. Barring pre-established harmony, such co-occurrence will have to be a matter of natural law. It will be defeasible, but given a number of background conditions it will also be quite reliable.

Jacob (1997, 49) suggests that the thing indicated will, in fact, be the cause of the indicator (for instance, a state of a 'receptor system'):

\[(T)\text{he indication relation between state } r \text{ of receptor system } S \text{ and what it indicates,}
\]
\[i.e., \text{the fact that object } s \text{ is } F \text{ (the source) is the converse of some sort of causal relation between the fact that object } s \text{ is } F \text{ and the fact that state } r \text{ has property } G.\]

Indication so understood is a merely causal phenomenon. As Dretske (1988, 56) writes, 'there can be no misindication, only misrepresentation.' One might think that 'indicates' is a three-place predicate (e indicating c to s), but this is not indication in the technical sense outlined above. The idea is precisely that e carries information about c, whether or not someone is present, or able, to make use of it.

We must now proceed to explain indication-to. Consider the following example by Dretske (1988, 63):

Some marine bacteria have internal magnets, magnetosomes, that function like compass needles, aligning themselves (and, as a result, the bacterium) parallel to the Earth's magnetic field (Blakemore and Frankel 1981). Since the magnetic lines incline downward (toward geomagnetic north) in the northern hemisphere, bacteria in the northern hemisphere, oriented by their internal magnetosomes, propel themselves toward geomagnetic north. Since these organisms are capable of living only in the absence of oxygen, and since movement toward geomagnetic north will take northern bacteria away from the oxygen-rich and therefore toxic surface water and toward the comparatively oxygen-free sediment at the bottom, it is not unreasonable to speculate, as Blakemore and Frankel do, that the function of this primitive sensory system is to indicate the whereabouts of benign (i.e., anaerobic) environments.

It is suggested here that the bacteria are consumers of information, where the Earth's magnetic field embodies the information, and where the magnetosomes indicate geomagnetic north and, in doing so, indicate oxygen-

\(^8\) The originator of this concept is Grice 1957.
free surroundings. Because it seems in virtue of the latter that the bacteria have their magnetosomes, we should say that although the magnetosomes indicate geomagnetic north, what they indicate to the bacteria is oxygen-free surroundings.

If an account like this works, we have the beginnings of an account of intentionality, one that is based on merely causal concepts. However, we must proceed from mere indication to fully fledged representation and intentionality, and how to do that is a muddied issue. The kind of indication at issue, although it functions, and is really used, is still a mere trigger of motion; not the subject of reflection, as in cases of more sophisticated intentionality. And the bacteria described above do not have the ability to learn, i.e., to adaptively readjust their information-processing infrastructure during the course of their lifetime.

Suppose that they could learn. They might still do so in a passive way, as a mere effect of external influences (as in classical conditioning), not actively, as a result of internal currents of information. From a minded creature we can at least expect that its information processing is iterative (where information processing provides input for further processing), flexible (where depending on the circumstances, c may indicate an F, or a G), fine-grained (c may come to indicate, not e, but the fact that e is F, or G, or H) and multi-layered (simultaneous multi-modal channels of information cooperating).

We may ask whether these things would also be sufficient for mindedness. But how liberal or restrictive we should be exactly in ascribing mindedness will only be a good question insofar as the concept of mind is so precise as to give us clear criteria. But it is not that precise. At some point we will find reasons to speak of genuine representation, intentionality, or belief, rather than of mere indication. But what that point is supposed to be is an issue that, I think, can at best be explored tentatively. Here are three quotations from Dretske (1988) that illustrate the difficulty of the issue:

When a rat acquires a disposition to avoid a kind of foodstuff that is poisonous and has caused it sickness in the past, we might, to explain its behaviour, assign it a belief that the food is poisonous. But this belief, which I am calling an implicit belief, is quite different from our belief that the food is poisonous. Our belief that the food is poisonous has achieved explicit status and is, therefore, available for a variety of jobs and applications. (120)

As beliefs become integrated into more tightly structured cognitive systems, their indicator functions become more interdependent. Not having to do as much, they
are free to become more specialized. As a result of this increasing specialization, they begin to exhibit a finer-grained intentionality. (150)

... (I)nternal elements, by becoming more interdependent (...) affect each other's indicator function, and, hence, each other's meaning. (151n)

It seems plainly false to say that only human beings have minds, but the above quotations suggest that there is no clear and unique way of sorting out exactly which animals do and which do not have genuine minds. It does seem, however, that the prospects for mindedness on the part of mere artifacts are dim: if intentionality requires teleology, PCs and household robots do not seem to qualify, at least not on the account of teleology given above. Interestingly, the reason why artifacts do not have genuine beliefs seems to be very different from the reason why, say, snails do not. Snails may just not have a sophisticated enough representational apparatus to count as 'true believers.' PCs and household robots are not true believers insofar as they do not have representational functions that are for the benefit of any survival and reproduction on their part.

With the notion of representation, we have also helped ourselves to the notion of intentional content. This notion is metaphorical, and it may be misleading if taken literally: it is not a payload of ethereal stuff that is carried along by a freight train of causal influence. But we can say that intentional content is information that is processed for the goals of an organism and that is available to that organism (i.e., on which it can act). I will return to the notion of intentional content when discussing propositional attitudes.

15. The concept of mind

15.4. MIND PRESUPPOSES (IR-)RATIONALITY

The concept of mind is inextricably linked up with that of rationality and irrationality. In analytic philosophy of mind, mental states are traditionally thought to be (mainly) either beliefs or desires, depending on their 'direction of fit' (see Searle 1983, Ch. 1). Rationality requires that beliefs fit the world (i.e., be true or veridical) and desires are fitted by the world (i.e., become satisfied). Beliefs are also supposed to be consistent with one another and with what is perceived, and desires are supposed to be such that

88 Expression borrowed from the title of an essay in Dennett 1989.
their satisfaction would be a good thing. And actions are supposed to be
guided by beliefs, motivated by desires, and justified by both.

Insofar as having a mind implies being a perceiver/actor, a normatively
neutral minded being seems a conceptual impossibility. For the very notion
of perception seems to involve distinguishing the significant from the insig-
nificant, and the very notion of action seems to involve aims, the further-
ing of goals or goods. If so, speaking of intentionality without the possibility
of rationality and irrationality, of proper and improper thinking or behav-
iour, of conforming to and violating rules, will be self-contradictory. There
seems to be no such thing as a purely cognitive, normatively uncommitted
mind. The more so because, as we have seen, the very notion of repre-
sentation constitutive of intentionality itself implies the normative notion of
proper functioning.

Note that rationality is typically mental, while mere normativity is not.
The reason is that proper functioning is already typical of biological phe-
nomena. Consider the stomach: while it does not represent food, it does
distinguish between good and bad stuff. It does not have desires, and no
notion of any stuff as good or bad; but that is in virtue of its inability to rep-
resent anything, not in virtue of any neutrality about good or bad. Al-
though your stomach performs well, it is not rational; neither is it irrational
if it goes out of order. You can only curse it, not blame it. Even if it does
not do what it supposed to do, it does not violate any rule.

The normative aspect of mindedness, then, is already captured in the
notion of teleology, and does not require the additional aspect of representa-
tion. 'Bare' teleology provides a basis for a minded creature's preferences
and aversions, for these will be rooted in the way the creature, as a member
of its species, normally exploits its ecological niche. Also, it makes it intel-
ligible that the creature actively seeks or avoids certain situations: its at-
tractor landscape, in the terminology of systems theory, will have been
shaped in a way corresponding to its normally exploiting its niche.

We can even go so far as to say that mere teleology provides an objec-
tive notion of good and bad: good is what helps a creature flourish (i.e., do
its optimal proper functioning), bad is what hinders it. If so, right and
wrong have a reality that is to some extent independent from what particu-
lar minded beings happen to think about it, and there is room for norma-
tive states of affairs. Needless to say, goods so understood are diverse and
need not be neatly coordinated; no matter whether they concern just one
or more organisms.
15. The concept of mind

Although I have distinguished mere teleology from intentionality, it is sometimes claimed that any adaptive trait of a living species, such as the polar bear’s white fur, does embody information, or even knowledge (Plotkin 1994, Buskes 1998). In my view, this evolutionary approach rightly stresses the continuity between sophisticated human knowledge and biological adaptation. Nevertheless, ‘knowledge’ as that embodied by a polar bear’s white fur is not full-blooded representation. We may credit the species, or its genome, with knowledge about, even wisdom concerning polar conditions. We may even regard the genome, over the generations, as a selfish perceiver, learner and agent. But if representation implies the use of indicative relations by a system, it is not obvious that we find such a thing in a genome.

15.5. MIND PRESUPPOSES PERCEPTION AND ACTION

The concept of mind is inextricably bound up, not just with the concept of thought, but also with the concepts of perception and action. This idea, at least in its behaviourist guise, has motivated the ‘interpretationist’ accounts of intentional states, from Quine to Davidson and Dennett. There can be radical and less radical versions of interpretationism (see Child 1994, Ch. 1), but they all embody the idea that it does not seem possible to make sense of mental properties without action on a specific perceived environment that can in principle be observed by an external observer.

One reason why explicitly linking up mind with perception and action is important is, of course, that this is the only way to ‘naturalize’ the mind. My more specific aim in bringing up this issue, however, is to make it clear that perception and action are just as much intentional phenomena as thought is. It is not as if perception and action are mere in- and output of intentional processes. They are themselves intentional processes, fully continuous with thought, and with each other. There is at least wide consensus that perception is a process that terminates in an intentional state, and that action is a process that starts with an intentional state. Start and end of such a process may also both be intentional states, but then, mere reflective thinking is also something we do.

---

89 ‘Thought’ is not a simple notion, and I use it in a more liberal way than many other philosophers of mind, for instance, Pettit 1993, Ch. 2. Yet, ‘thought’ and ‘mind’ are taken to be closely connected notions by everyone.
Part III. Mind

More controversial, perhaps, is the claim that mental states cannot be distinguished from perception and action in being somehow inside the organism or the brain. But first, we will later find reasons to criticize the view that mental states must be internal states. And secondly, if such states are taken to be internal, it must be admitted that action and perception may also be entirely internal: think of flexing one's muscles, keeping one's breath, repressing a thought, feeling one's stomach, or introspecting a wish. We may say that some of these examples are not perceptions or actions, but thoughts; but seeking definite points in the nervous system (or seeking such points introspectively) where perceptions end and thoughts begin, or where the latter end and actions begin seems a hopeless endeavour anyhow.

Perceptions and actions, then, should be regarded as themselves intentional events, along with thoughts. This can be brought out by making a distinction between the perceptions and actions of a subject on the one hand, and the things that merely happen to her on the other hand. Suppose that somebody pushes you and you fall. Then obviously, the toppling over of your body is not perception, and your falling is not an action (although you can perceive the pushing, and, perhaps, actively interfere in your falling). It seems that mere causal in- and output, without intentionality, cannot constitute perception or action.

There is also causal in- and output that, although it has more to do with you than being pushed and falling, intuitively does not constitute perception and action: breathing is not perceiving, the processing of gaseous substances in your lungs is not thinking; and exhaling is not, as such, action (although you can perhaps voluntarily control some of it). Now, it seems that it is precisely an aspect of intentionality that makes the difference between perception-action and mere causal, or mere biological input and output.

Thus, perception, thought and action form a tight conceptual cluster. And although they can be distinguished, none of them will normally occur without the other. Genuine perception and action require intentional states; and in the process of acting one gets continuous feedback from perception, while in the process of perceiving one has to continuously steer by acting, e.g., moving one's head. Blind action, purely contemplative

---

90 What Hurley 1998, 10 calls 'instrumental dependence of perceptual content on output.' There is also noninstrumental such dependence, but the 'output' will not be plausibly called 'action.'
thought, and purely passive perception may be possible for a while, but only, it seems, in virtue of the normal continuous perception-thought-action loops.

We can say, then, that intentionality is constitutive of perception, thought and action, along with some form of causal influence (perception, thought or action that does not involve causation seems inconceivable). In the spirit of our foregoing analysis of causal relations, this suggests that perception, thought and action are all ‘mental causation,’ in the sense that perceptions, thoughts and actions are all causal counterfactual dependencies that involve intentional states (be it as effect or cause), the underlying physical infrastructure of which involves the particular organism that is the actor.\(^{91}\) Perception will then be the causation of an intentional state by worldly and bodily events (and, via these, also by actions); thought will be the causation of intentional states among themselves; and action will be the causation of worldly or bodily events by intentional states (and, via perceptions, of further intentional states).\(^{92}\)

Yet, going from a belief-desire complex to an intention, from there to an action, and from there to the worldly consequences, it seems pedantic and fruitless to seek a precise parsing of the events and states involved. Terms like ‘belief’ and ‘intention’ seem descriptions of aspects of the intricate intentional reality culminating in a subject’s interaction with the world, rather than terms for steps in a causal chain.

---

91 It might seem that on the above definition phenomena that we call ‘psychosomatic,’ such as the alleged occasional regression of tumours caused by certain kinds of experiences (see Schilder 1996), or the wide range of psychosomatic illnesses, seem also to be actions; which conclusion should obviously be resisted. Examples like these suggest that the causal dependencies involved should be of an appropriate kind. What makes a causal dependency appropriate here may perhaps be explicable with an appeal to the notion of proper function. Another interesting suggestion is Juarrero’s (1999, Ch. 6), who argues that there must be an ‘uninterrupted flow of information from intention to behavior’ (82). I will not here try to give a watertight criterion of distinguishing perception, thought and action from ‘wayward’ forms of mental causation.

92 All on the assumption that ‘wayward causal chains’ can be barred by saying more about the precise character of the causal relations involved. Note that my discussion of causal relata earlier on may help to explicate various different forms that action may take: omissions, for instance, can be understood along the lines of my account of ‘negative’ causes; mere bodily actions, such as raising one’s arm (which do not seem to cause much of a ‘worldly event’), can be understood as the causation of a change in spatial relatedness of parts of the body; and guidance, instead of triggering, of actions by intentional states becomes intelligible once we realize that ongoing processes are causal relata just as well as events.
Part III. Mind

Perception, thought and action, then, have no sharp boundaries; and neither seems there to be a sharp boundary between an action and its mere worldly consequences. A salient aspect of action is the so-called 'accordion effect': we do things by doing other things. For instance, we greet a friend by writing a letter; we write a letter by handling a pen; we handle a pen by moving our hand; and we move our hand by contracting muscles. Here we trace the causal sequence inward, and there will not be a clear answer to the question at which point a train of intentional states ends and action takes over.

The accordion effect also goes outward. I might, for instance, tip off a band of dangerous criminals by greeting my friend; and when tracking the consequences a little further, I might turn out to have killed ten people by writing my letter. But is it really me who did the killing? Where is the boundary between what we do and the side-effects of our actions? This is a difficult issue in the philosophy of action that I will not try to settle here. But it does seem that the notion of action is linked to that of control. For the more we are able to control the situation that we are in, the more we are held responsible. Now, if control is not all-or-nothing, neither seems agency. If so, the 'outward' indeterminacy of the line between action and consequence mirrors the 'inward' indeterminacy of the line between thought and action.

---

93 See, e.g., Davidson 1971.
94 This remark is based on Lewis's helpful notion of 'insensitive causation': 'Maybe there is a time after which every death that occurs is one that would not have occurred but for my act. (...) And still I deny that I have ever killed. (...) So killing must be a special kind of causing to die. But what distinguishes this special kind of causation? Not that there must be one single step of causal dependence, as opposed to an intransitive chain. (...) Not that the chance of the effect must be high. (...) Not that the causal chain must be short. (...) Not that the chain must be simple. (...) Not that the chain must be foreseeable. (...) Not that the chain must pass through no later human actions. (...) Perhaps a cluster of these conditions, inadequate if taken one by one, would work to distinguish the kind of causing that can be killing. I think not. (...) I suggest a different way to distinguish the right kind of causing: by its insensitivity to circumstances. When an effect depends counterfactually on a cause, in general it will depend on much else as well. (...) Sensitivity is a matter of degree, however. (...) Jonathan Bennett restates my suggestion in this way: killing requires "that the causal chain run through a stable and durable structure rather than depending on intervening coincidental events." (...) If a chain is insensitive enough that you can predict it, then it is insensitive enough that you can kill by it. (...) What if you are much better than I am at predicting chains that are somewhat sensitive? I am inclined to say that if so, then indeed you can kill in ways that I cannot.' (Lewis 1986a, 184-7)
15. The concept of mind

15.6. MIND AND QUALIA

In the above sections I have given an outline of the concept of mind as I will apply it in my discussion of mental causation. I have claimed, roughly, that mindedness amounts to underived intentionality, such intentionality to representation, and representation to consuming information (in the sense of ‘natural meaning’) in a sufficiently sophisticated way (‘sufficiently’ remaining undetermined). It may now seem that the most important feature of the mind, consciousness, or subjective experience, has been left out. So I would now like to say something about consciousness, and its alleged lack of causal efficacy.

The debate on this issue, however, is intricate and, above all, very much undecided. And I do not have the resources at hand to present something like a solid case: all I can do here is make a few remarks. But I do not want to skip this subject entirely. Subjective experience is a core aspect of the mind, and its causal efficacy or inefficacy is not a side issue. As long as we remain clueless about subjective experience, we have not even begun to understand mental causation properly, no matter what we can do about the rest of the problems.

The issue is this. Minded creatures have a first-person perspective on the world; their perceptions, thoughts and actions are in a nontrivial sense their own, and somehow known to be their own; and they typically have a unified set of subjective experiences. This, roughly, is what we mean by saying that minded creatures are conscious. Searle’s ‘connection principle’ (e.g. 1992), implying that mindedness implies consciousness, is generally considered too strong; yet, an aspect of subjective experience seems to be an essential part of the cognition of at least the higher animals.

Now, some philosophers of mind distinguish, in creatures capable of consciousness, mere states of awareness from ‘phenomenal’ states, better known as qualia. Phenomenal consciousness is said to differ from the former kind of mental states (‘access’ consciousness according to Block 1995; ‘psychological’ consciousness according to Chalmers 1996) in that it has a qualitative, ineffable, essentially-first-person, what-it-is-like character. Arguments for this view standardly appeal to the seeming conceivability of so-called ‘zombies’: imaginary beings that are functionally, or informationally, or causally, like us, but that do not have qualia. Also invoked are cases of inverted spectra: of subjects that behave as normal, but experience an abnormal, because inverted, colour spectrum.
In trying to account for such phenomenal consciousness (Why is it as it is? Why is it there at all?) we are often said to encounter an ‘explanatory gap,’ (e.g. Chalmers 1996). Indeed, claims like Searle’s, that qualia have a ‘subjective mode of existence’ (1997, 98), suggest that they are pretty mysterious entities. Their causal role is especially problematic: Jackson (1982) and Chalmers (1996) toy with epiphenomenalism, even in the face of the fact that epiphenomena are unknowable, cannot be referred to, and do not make biological sense. For, it is held, mere awareness can be ‘functionalized’ (i.e., analyzed in terms of its causal role), but it is hard to see how this can be done with qualia (see, for instance, Kim 1998a). So if we do have qualia, and if they cannot be functionalized, they have to be epiphenomena.

It will be clear that this issue directly concerns the third problem of mental causation: that of phenomenal states of affairs. On the one hand our common sense tells us that pains, tickles, tastes, etc. cause behaviour, verbal or otherwise, doing so precisely because of their subjective feel. On the other hand, if it is maintained that such behaviour can fully causally be accounted for by the merely ‘functional’ or ‘psychological’ aspect of conscious states, and not by the ‘phenomenal,’ the latter will be regarded as causally redundant.

I would now like to make four remarks that are intended to discourage a few assumptions about qualia. They are remarks, not arguments; and they will certainly not give us a theory of mental causation concerning subjective experience. But as the assumptions are a source of scepticism about ‘phenomenal’ mental causation, my remarks may help a little to dispel the idea that epiphenomenalism about subjective experience is inevitable.

1. Qualia are not entities in their own right, but have the ontological character of states of affairs. Locutions like ‘I feel a pain’ do not describe a relation between a subject and another entity (which is a pain), but merely the fact that someone is in pain, feels pain, has pain, or perhaps even ‘is painfully being appeared to.’ Qualia, it seems, are had rather than perceived. The consequence is that qualia might be causally efficacious without (impossibly) having to interact with physical entities.

2. Qualia need not be regarded as (non-functional) aspects of (functional) items that are themselves mental states. They may also be regarded as mental states themselves. It is not as if one is in a mental state that has both (functional) intentional and (non-functional) phenomenal aspects.

---

95 Putnam (1999) attacks the notion of such perceived qualia on the ground that it relies on an outdated epistemology of sense data.
15. The concept of mind

There is just an organism with phenomenal properties, such as being in pain, the latter constituting mental (c.q., phenomenal) states of affairs. The consequence here is, that qualia are not, at least not already by definition, screened off from causal efficacy by merely 'psychological' properties.

3. According to a number of theories, qualia are a species of intentional state. Among these theories of qualia that we could call 'representational' are, for instance, Damasio 1999, Dennett 1991, Dretske 1995, Edelman and Tononi 2000 and Tye 1995. Put a bit crudely, according to such theories qualia are the many-dimensional and compelling ways in which information, or misinformation, about an organism's environment and/or bodily state (or the relationship between these) is conveyed to that same organism, so as to steer the organism's behaviour in the dramatic way that we are so familiar with. Now, insofar as representational theories of subjective experience are on the right track, subjective experience does not raise a problem of mental causation over and above the other problems.

4. No purely intrinsic qualia can be demarcated. It is often held that phenomenal consciousness supervenes on local brain activity, while other mental phenomena, such as beliefs, do not. Yet, an extensive argument about allegedly possible inverted colour qualia in Hurley 1998 (Ch. 8), to mention one example, suggests that there is no way of demarcating the purely phenomenal and intrinsic from everything else. In this way, we could contest the dichotomy of 'psychological' and 'phenomenal', with the result that the former need not a priori be regarded as pre-emptive of the latter.

It is true that qualia are puzzling. It is tempting to think that somewhere in the brain all the information must come together to be picked up by an inner spectator (what Dennett 1991 has called the 'Cartesian theater'); and the relation between local brain activity and such a spectator will of course appear utterly mysterious. But the currents of information constituting qualia do not seem to have a centre; it rather seems that some form of self-organization is at work. One might persist in asking how qualia are supposed to arise from a tangle of physical processes; but in my view, the fact that we cannot readily imagine qualia arising from currents of information may as well be due to the mere complexity of the processes involved, and to the early state of the scientific study of consciousness, as to some unbridgeable ontological barrier.

96 See, for instance, Edelman and Tononi 2000 and their theory of consciousness as a 'dynamic core' of 'reentrant' neural connections.
I admit, then, that I am unable to explain how qualia, as standardly conceived, can be causally efficacious; but I suggest that we should not take standard qualia discourse for granted. The above remarks are intended to question the alleged gap between subjective experience and other intentional phenomena, and the inevitable epiphenomenalism that goes with it.

This concludes my rough and exploratory general sketch of mental reality. Undoubtedly, it will have left the reader with a number of unanswered objections, and it will also have begged a number of questions with regard to the more detailed and critical chapters below. I am afraid, however, that such a thing is inevitable. In discussing mind and mental causation we have to start somewhere in a field where every issue is linked up with pretty much every other issue. But I hope that at least the above suffices to explicate the notion of ‘mental’ as I use it in the critical arguments to follow.
16
Against the computational theory of mind

16.1. INTRODUCTION

In the present chapter, I want to take issue with a view of the mind as, most basically, a reasoning device. It is the view that the mind consists of formal operations over abstract items that are called symbols. The symbols, as well as the operations over them, are then assumed to be physically implemented in the brain. This view (or better: cluster of views) is widespread in current analytic philosophy of mind, and it has shaped a good deal of the debate on mental causation.

In one form or another, it is a conception of the mind that is already old, but in the second half of the past century it has taken on new significance (and its present physicalist form) with the advent of the fields of artificial intelligence and cognitive science. The view outlined here goes under names such as 'computationalism' or 'syntacticalism.' Contemporary philosophers who endorse, or have endorsed, this view include Dretske, Fodor, Kim, and Stich.

Syntacticalism is closely related to the view, or family of views, known as functionalism. The core doctrine of functionalism is that mental states are individuated by their causal roles, i.e., that they are specific dispositions of, or in, the brain. Now, as a causal role (or realizer of such a role) may implement syntactical operations over symbols, functionalism and computationalism are related doctrines. In the present chapter, however, my target will be syntacticalism, not functionalism in general. Functionalism as such will also come under fire, but later on, in the context of, first, the Theory Theory of propositional attitudes, and secondly, psychophysical identity doctrines.

We will see below that syntacticalism is a source of epiphenomenalism. This is, of course, one reason for me to criticize it: as long as we endorse it, we will not be able to understand what mental causation really amounts to. We will also see, however, that there are many independent grounds for suspicion of this view. Of course, there will be ways to accommodate or

---

97 One well-known precursor is Leibniz.
Part III. Mind

evade the various criticisms, to which I cannot here do justice. But I hope that those who will not follow me in my harsh verdict on syntacticalism will in any case appreciate that the view need not be taken for granted, and that there are plausible alternatives.

I will proceed as follows. I will first describe what the computational theory of mind amounts to. Next, I will criticize this theory on a number of points. I will then suggest an alternative view of the category of mental states, i.e., propositional attitudes, concentration on which may have lent syntacticalism its seeming plausibility. This alternative will be McGinn's 'theory of mental models.' I will also discuss, but not advocate, Dennett's and Davidson's interpretationism as another alternative.

16.2. THE COMPUTATIONAL THEORY OF MIND

The basic idea of syntacticalism is that the mind consists of a set of representations, or symbols, that are manipulated by way of formal operations. The representations are then thought to be implemented by physical structures, the causal relations among which mirror the abstract 'rational,' 'syntactic' or 'inferential' relations between the symbols. It is, in fact, a conception of the mind as software that is implemented by hardware. In the case of human beings, this hardware happens to be the brain, but in principle, any physical structure with the right causal organization may qualify.

In the syntacticalist camp, it is generally believed that the symbols, or strings of symbols, correspond with what in daily discourse are called 'beliefs' and 'desires', or what since Russell (1940) are known as the 'propositional attitudes.' The idea of a propositional attitude is that one can have a range of attitudes towards any proposition (such as 'it is raining,' or 'there is a fly in the room'): belief, desire, expectation, fear, hope, etc. Hence, propositional attitudes are items such as the belief that p, or the desire that q.

On the general view of the mind that I outlined in the introductory chapter, the computational theory of mind may not at all appear a strange view. We have seen that causal influence can constitute information, and that such information can be used by a minded organism. The information, when processed, then becomes intentional content. On this picture, thinking turns out to be content-processing: patterned causal processes are put
to use by an organism as a source of information about its (inner and outer) environment.

Now, syntacticalist or computational theories of mind assume that such information has a formal structure. This does not seem unreasonable. After all, an organism does not just seem to perceive As and Bs but, more typically that the A is F and the B is G. And an organism that does so also seems capable of perceiving that the B is F or that the A is G. In other words, mental content seems to allow of combinatorics that generates new content.

If so, the question arises how such a combinatorics might be implemented. The syntacticalist answer is, then, that the brain is essentially a symbol manipulator. Tokens of abstract symbols are physically implemented by brain states, just as units of information in a computer are implemented by machine states. And the brain states have a causal structure that mirrors the logical structure among the symbols manipulated. Pinker (1997, 25) writes:

The computational theory of mind (...) says that beliefs and desires are information, incarnated as configurations of symbols. The symbols are the physical states of bits of matter, like chips in a computer or neurons in the brain. They symbolize things in the world because they are triggered by those things via our sense organs, and because of what they do once triggered. (...) The computational theory of mind (...) allows us to keep beliefs and desires in our explanations of behavior while planting them squarely in the physical universe. It allows meaning to cause and be caused.

### 16.3. FIRST CRITICISM: EPIPHENOMENALISM

The above might seem an elegant account of the causal efficacy of beliefs and desires, at a first look; but a complaint lies ready to hand. For how can the causal influence that is information be causal influence in virtue of the fact that it is information? The computational theory of mind assumes that there is a pre-established harmony between causal brain structure and abstract logical symbol structure: species evolution and individual development and learning have brought about the harmony. But that does not seem to make information (or meaning, or content, or semantics) causally efficacious.

To put it differently: syntacticalism makes us think of mental states as, somehow, token physical structures, and it thus shifts all efficacy towards the physical properties of these tokens. In doing so, it sidelines any inten-
tional content, and thus any mental states qua mental. In this way, syntacticalism seems to imply epiphenomenalism about intentional content.

Syntacticalism also seems to imply epiphenomenalism about phenomenal content. For it motivates the dualism, as described above, between merely intentional and phenomenal states. The reason is that even if propositional attitudes can be understood as syntactic entities, it seems that qualia cannot. Thus, subjective experience cannot be causally efficacious in the same way as beliefs and desires; but syntacticalism does not seem to have the resources to explain how it can be causally efficacious in any other way.

16.4. SECOND CRITICISM: LACK OF PSYCHOLOGICAL REALISM

As a second point of criticism of the computational theory of mind, I want to suggest that the category of intentional states is much broader than that of the propositional attitudes. If, namely, it is the formal relations among the propositions towards which we take up our attitudes that suggest that our minds must somehow be of a syntactic character, it will diminish the credibility, or at least the scope, of syntacticalism if it can be pointed out that the mind consists of much more than just beliefs and desires.

Propositional attitudes, of course, form an important category of mental states, because it is mainly in terms of these that we appreciate each other’s reasons for action. But assimilating intentional states with propositional attitudes does seem to yield an unrealistic picture of the mind. Mental reality involves selfhood, emotion, feeling, mood, learning, intelligence, creativity, imagination, and a number of other things; and arguably, all these are fully fledged intentional phenomena. None of them, however, need involve content with anything like a propositional (subject-predicate) structure.

The propositional attitude-oriented picture of the mind generally ignores the fact that, no matter how special the human mind may be because of language and culture, intentional properties do not seem unique to human beings at all. A guinea pig seems to have beliefs, desires and perceptions as well as we do, only less sophisticated ones. Consider our late guinea pig’s passionate squeaking on the sounding of the green-grocer’s hooter: she definitely heard the hooting, definitely wanted endive, and definitely understood the correlation between the hooter and endive. It even
16. Against the computational theory of mind

seems to me that she was rational, even though she certainly did not reason herself into squeaking.

16.5. FURTHER CRITICISMS

Let me now briefly list a number of further worries about the computational theory of mind.

First, unlike this view suggests, thinking often does not follow anything like formal, syntactical rules. The relations among our mental states may be subject to normative constraints so that, for instance, we will tend to avoid contradictions in what we think. But much of our practical reasoning seems to be carried out without anything like serial processing of symbol strings. Decades of research in cognitive science seem even to suggest that real-time cognition along those lines is impossible, and must be implemented by different mechanisms.98

Let us also note that the formal reasoning that syntacticalism regards as constitutive of thinking does not at all come naturally even to us human beings. It is a learned skill, and it is a well-known fact that decision theory makes inadequate psychology for most of us. Thinking, figuring-out, scheming, deliberating, and so on, do come naturally; but then, the point is exactly that these need not have much to do with formal reasoning or logical inference, not, at least, below a certain level of sophistication. Normally, practical reason is not at all a matter of pure logic. Damasio (1994 and 1999, 41) points out that thinking requires an underlying continuous flux of emotions, and that when specific emotional capacities are damaged, proper practical reasoning suffers accordingly.

Secondly, while we do not typically carry out formal operations when we think, there also seems to be little empirical evidence that the brain does. We might, of course, just speak of 'computation' whenever some physical processing of information takes place. Such a liberal view of 'computation' is found, for instance, in the connectionist schools of cognitive science (for instance Churchland 1992), where it is held that our brains do something like vector transformation. One might even speak of computation without requiring there to be anything like formal operations or calculations at all (see Clark 1997, 159). If that is what we call 'computation,' then our brains

---

98 See, for instance, Dennett 1984.
Part III. Mind

may very well be said to be computing devices (but see Thelen and Smith 1994). But that is not the view we normally call computationalism.

Thirdly, it seems that computationalism falls prey to the category mistake of a 'vehicle/content conflation' (Hurley 1998, 27). Hurley writes (29-30):

(N)o one thinks that the representation of something green must itself be green, or that the representation of something round must itself be round. But it is easy to slip into this confusion for more complex, abstract or relational properties, such as simultaneity. And many have found compelling a certain structural application of this assumption: that the systematic conceptual structure of the contents of thought must be explained by isomorphic syntactic structure among vehicles of thought (…).

Fourthly, computationalism seems to be attractive mainly because it assimilates formal and logical, normatively constrained relations among mental content with causal-nomic relations among these. The initial attraction of this idea is due to the popular assumption that by figuring out what other beings' mental contents are and how these are logically related, we are making causal hypotheses at the same time. Normative-rational and causal-nomic constraints seem then to come together in intentional states, and computationalism seems to provide an explanation of how such a marriage can come about. This wrongly suggests, however, that the normativity of reasons is mainly a matter of rule-based inference. There is a Humean agenda behind this suggestion: one starts with beliefs and passions, and calculates one's way to satisfaction. We have seen earlier, however, that normativity is a matter of functions and goals (etiologically understood), rather than of mere rules connecting means and ends.

A fifth worry is that computationalism is much too preoccupied with cognition, and unwarrantedly narrows down mental reality in its full breadth. It is characteristic of computationalism that it allows for the possibility of simulating, or even implementing, mental states in digital computers. But McGinn (1989, 155n) notes:

It is significant that computers are commonly used to model only cognitive states, not affective or conative states. But this must strike us as biologically unrealistic, since beliefs exist in order to help organisms satisfy their desires (…). Intentionality begins with wanting, I suggest, not with thinking. To simulate intentionality properly you need to simulate the entire biological system in which it is found, not attempt to slice off the cognitive part and hope to find real intentionality in that part considered by itself.
A sixth worry can best be made explicit on the basis of the form of computationalism known as the ‘Language of Thought’ hypothesis (originally defended by Fodor 1975). Lycan (1990, 278) writes that according to the Language of Thought hypothesis,

(...)

physically realized thoughts and mental representations are “linguistic” in the following sense: (i) They are composed of parts and are syntactically structured; (ii) their atomic parts refer to or denote things and properties in the world; (iii) their meanings as wholes are determined by the semantical properties of their atomic parts together with the grammatical rules that have generated their overall syntactic structures; (iv) they have truth conditions, and accordingly truth-values determined by the way the world is; they bear logical relations of entailment or implication to each other.

The problem that lurks here is, that the capacity for language, or at least for applying grammar, is carried over from us, human beings with a specific life form, to our brains, a different kind of entity altogether; without any attention to the question what language, or grammar, is, or has evolved, for.

Now, in natural language sentences are the units of meaning, and words acquire their meanings in virtue of the way in which they are combined with other words, according to syntactic rules. According to Deacon (1997), such a system of representations is a mnemonic strategy that enables subjects to store vastly much more knowledge of the world than is possible with a merely - as he calls it - 'iconic' representational system. This sentential strategy, as we might call it, also exploits the fact that we can serialize our behaviour (vocal or gestural), and use channels like sound and light (or more permanent ones) to make the information it carries available to others - and to our own future selves as well.

Linguistic utterances are, then, information tools, the use of which has enhanced, and no doubt profoundly changed, the informational capacities and strategies of our species. But our ancestors must have been already minded before starting to produce and use these sophisticated tools. If language is seen in this light, it does not make sense to explain thought by assuming an inner language. The same mistake is made when, for instance, one explains vision by assuming inner pictures, or mathematical capacities by assuming an inner calculator.

Now, this is a point about the Language of Thought hypothesis. But insofar as computationalism implies that the brain carries out syntactical op-
operations over meaningful symbols, the point may well carry over to computa-
tionalism in general.  

Finally, I have a metaphysical worry: the intelligibility of computational-
ism seems to require token physicalism about mental states. This is the the-
thesis that any intentional state is identical with a computational, neural, and
physical state. Computationalism requires token physicalism, for it seems
that the causal role of state $a$ can only be mapped onto, or identified with,
the computational role of state $b$, when states $a$ and $b$ are identical (or at
least, when they are related as first- and second-order state). Token physi-
calism about states of affairs is deeply problematic, as will become clear
later on.

16.6. PROPOSITIONAL ATTITUDES AS MENTAL MODELS

If indeed syntacticalism, or computationalism, is as problematic as I have
depicted it, we need an alternative account of the propositional attitudes.
For although in my view the significance of propositional attitudes amidst
other types of mental states is vastly overrated, it seems counterintuitive to
deny that, first, ascribing them to each other is an important means for us
to appreciate each other's reasons for action, and that secondly, insofar as
this practice of 'folk psychology' works well, such ascriptions can be true.

I will now argue that McGinn's 'theory of mental models' about mental
content (1989) provides a satisfactory non-syntacticalist account of the
propositional attitudes. McGinn defends a theory of mental content that,
as he points out, was already put forward in 1943 by Kenneth Craik:

We construct mental representations of the world by constructing mental models of
it (...). Manipulating mental models thus constitutes the working machinery of cog-
nitive problem solving. A thinking system, we might say, is a simulation engine.
(1989, 176)

99 Apart from all this, what a Language of Thought is supposed to be has never been
very clear. Pinker (1997) mentions four formats of representation in the brain: the visual,
the phonological, the grammatical, and, indeed, 'Mentalese.' Now this may be a credible
form of the hypothesis, but then, it does not seem very 'computationalist' about proposi-
tional attitudes anymore. For when we apply commonsense (belief-desire) psychology, we
do not seem to distinguish the formats at all, which suggests that the level of explanation of
beliefs and desires is quite different from the level of explanation featuring these formats.

100 A range of powerful arguments against computationalism that I do not discuss here
can be found in Hendriks-Jansen 1996.
16. Against the computational theory of mind

This 'theory of mental models' (173) takes seriously the phrase, that McGinn attributes to Ramsey (177), that a belief is a 'map (...) by which we steer.' A related idea is Derksen's metaphor (1999, 44-6) of a belief as a 'picture-of-the-world.'

We saw earlier that on a naturalistic understanding, mindedness is a way of making a living, exploiting an ecological niche. Now, many species get along quite well without representing the world to themselves: a repertoire of fixed behavioural patterns often works fine. But the more complicated a species' way of life and its normal environment, the less reliable will such rough and ready receipts for action be. This is due to the circumstantial nature of causation: whether or not a specific way of acting will cause fortune or disaster depends heavily on circumstance.

This is where a model of the world becomes useful. We do not just act; we first simulate the situation we encounter with our brain. How we do so is a difficult enough question. McGinn says (175):

By a model we (...) mean any physical or chemical system which has a similar relation-structure to that of the process it imitates. (...) It is a physical working model which works in the same way as the process it parallels, in the aspects under consideration at any moment.

On a simplistic understanding of this idea, we have something like a toy world in our heads. Note, however, that any replay, or mirroring, or simulation of a situation can happen at different levels of abstraction, and in many different ways. I will not speculate on how we actually do this. But we can be quite confident that it is not very much like forming, retrieving, manipulating and deriving formal symbols or propositions.

McGinn also notes that 'sentences do not simulate what they describe and models do not describe what they simulate' (181), and:

Sentences have grammatical structure; models have the structure of the worldly things they simulate, which is not grammatical (unless these things happen to be themselves sentences). Sentences have semantic properties - truth conditions, reference, sense; models do not - any more than maps or tree rings do. (181)

The difference becomes clearest when we realize that models work by analog coding, while sentences in the head, propositions, or a language of thought, will be digital:
Part III. Mind

An analogue code consists of symbols whose properties vary as a function of the things represented, while a digital code is such that its features are independent of the properties of the thing represented (...). (A)nalogue codes need not literally resemble what they stand for (...). Analogue representation depends upon a certain kind of identity of abstract structure (...). (178-9)

Note that it is more typically a structure of relations that is modeled than it is individual things or events. The isomorphism of a mental model is second-order, not first-order (see 179n).

The question now arises how, on this theory of mental models, there can be propositional content. Here, McGinn invokes the notion of indexing:

In measurement, say of temperature, we index physical states of objects with abstract entities, viz. numbers; and there is no temptation here to suppose that the physical ground of the indexing relation somehow contains simulacra of numbers and numerical relations. Similarly, in indexing states of a subject's head with propositions we are not thereby committed to inserting logical structure into the subject's head (...). Corresponding to the analog features of a map will be a set of propositions which describe what the map models (...). Thus logical structure is not literally in the head (...), but that does not prevent it being correctly assigned to what is in the head. (182-3)

Thus, the propositions in ascriptions of propositional attitudes do not mirror the mental model of the believer; they just describe what in the world, among other things, the believer has a model of. Of course, the complexity and fine grain of the propositions and concepts used by the ascriber is also supposed to be present in the mental model. If we say, for instance, that Jones believes that the elm tree is beautiful, we will be wrong when Jones's model of the world does not distinguish elm trees from other kinds of trees. Even if the tree before Jones is in fact an elm tree, and even if Jones believes that it is beautiful, Jones cannot then be said to believe that the elm tree is beautiful.

An important consequence of the theory of mental models is that causal and logical structure of intentional states come apart: 'Beliefs have causal relations with each other, and they also have logical relations, but it is quite wrong to assimilate these relations to each other.' (188n) A mental model can resemble the world, or be coherent or incoherent, but aspects of it do not seem able to be true, to entail each other, or to be contradictory. For these notions come with language, while mental models may be nonlinguistic. Note that by the concept of 'indexing' the theory of mental
16. Against the computational theory of mind

models has an elegant explanation of the fact that we often ascribe propositional attitudes to nonlinguistic creatures ('The guinea pig thinks that it will get endive').

Let us finally note that mental models may not be internal brain states. Of course, models are physically realized, and it will be brain behaviour that does the modeling (but see the discussion of 'environmentalism' below). But 'being a model,' or 'modeling' are function categories, and these require an appropriate history and context. McGinn puts it this way:

(M)odels realize content only in as much as they are embedded in a certain background of goals, behavioural propensities and a network of causally related states. (...) (A) map only really represents because it is employed in a certain way. (199)

16.7. IN FAVOUR OF THE THEORY OF MENTAL MODELS

I consider McGinn’s theory of mental models much more plausible than the computationalist sentential model that conceives of mental content as abstract symbols, or propositionally structured items in the brain. Note that it accounts very well for the following facts:

1) We can believe that \( p \), desire that \( p \), assume that \( p \), and so on, because we can adopt these various attitudes towards anything our model of the world represents. True, when we desire that \( p \) the world is of course not, or not yet, such that \( p \). Desiring that \( p \) amounts to comparing our model of the actual world with one of a hypothetical world (which is such that \( p \)), and preferring the hypothetical world. If this is right, world-to-mind attitudes like desires are accommodated as well by the theory of mental models as mind-to-world attitudes like beliefs.

2) The concept of indexing makes it clear how it can be that, although thinking is not at all like making inferences from propositions, mental content is typically referred to in a propositional form. Linguistic creatures just use their linguistic capacities to describe how someone models the world. They do it in the same way as describing anything else in the world. How these linguistic capacities, with their combinatorial character, can themselves be explained, and whether such explanation can succeed without an appeal to syntacticalism is, of course, an empirical issue.\(^{101}\)

3) The theory of mental models is, in fact, more adequate to our practice of folk psychology than is the computationalist model. First, it easily accommodates the well-known fact that we can believe things that seem contradictory; a fact that is problematic for any inference- and rule-based account of propositional attitudes. Derksen (1999), for instance, describes the Lottery Paradox, a situation in which we do not believe of any of the participants of a lottery that he or she will win, but in which we also believe that there will be a winner.

On the theory of mental models, we could say that the Lottery Paradox is just a matter of beliefs about probabilities, and that insofar as probabilities belong to reality, we model them just like anything else. Note that if one's world model is indexed, and the incompatible beliefs of the Lottery Paradox result, it need not be the model that is at fault, but the inaccurate way in which it was indexed. But in general, insofar as our models are vague or partial, and the scene of ongoing reconstruction, filling in, sharpening and disambiguating, the theory of mental models will make us expect occasional incompatible beliefs.

Secondly, it is well-known that we can believe that $p$, and believe an indefinite amount of implications of $p$. If I believe that a certain stick is 2 inches long, I also believe that it is shorter than things of 2.5 inches, things of 60.5679 inches, things of a billion billion inches, things of two miles, and so on. The problem for a sentential conception of thought is that we could never harbour such a near infinite stock of (actually had) tacit beliefs. I do not want to claim that there is no way around this puzzle for computationalist views, but note that for the theory of mental models it does not even arise. Any proposition mentioning what is longer than the 2 metre stick is only manufactured when my mental model is indexed.

Thirdly, content ascription can be indeterminate. If I thirstily walk to the fridge, I can be said to believe that there is something drinkable in it, but also that there is a beer in it, that there is a Heineken in it, that there is a cylindrically shaped, fluid-filled metal object in it, or that es gibt etwas zu trinken im Kuhlschrank, and so on. In fact, I believe all of these things, or none of them in particular. For any proposition- or inference-based account of my beliefs it is awkward that there is not one proposition that uniquely states what I believe. For the theory of mental models it is not a problem: models may be indexed in an indefinite number of ways.
16. Against the computational theory of mind

16.8. PROPOSITIONAL ATTITUDES ACCORDING TO INTERPRETATIONISM

As an alternative to computationalism I have adopted McGinn’s theory of mental models. But there is also the alternative of interpretationism (the descendant of behaviourism), defended by Davidson (1980), Dennett (1987) and Child (1994). The core idea is not to look under the skull for intentional states, but just to follow the way in which they are ascribed, out in the open, on the basis of the observable behaviour of actors. On interpretationism, believing that p is just being *interpretable* as believing that p. This view is obviously immune to attacks like the above on computationalism.

My reason for still preferring McGinn’s theory of mental models, however, is that it allows for more realism about beliefs and desires than interpretationism. My worry with interpretationism is that it seems to conflate what beliefs are with what having them implies. It may be said that being interpretable as having the belief that p is having the belief that p, just as having TBC is being diagnosable as having TBC. But this violates our intuition that one’s having the belief that p makes it the case that one is interpretable as having the belief that p, without the reverse being the case. The second seems to depend on the first, the first not on the second. If there is just the fact of being interpretable-as, or being diagnosable-as, it is unclear what could justify the interpretation, or the diagnosis.¹⁰²

¹⁰² This, I believe is the central idea of Slors 1994: Dennett’s claim that believing that p is just a matter of being an appropriate object for ‘the intentional stance’ faces the problem that it is impossible to explain what adopting that stance amounts to without a further, more realist account of intentional states.
17

Against the Theory Theory of folk psychology

17.1. INTRODUCTION

Although interpretationism may be mistaken in assuming that propositional attitudes are essentially a matter of interpretative practice, it rightly stresses the tight link between these kinds of intentional states and such practice. After all, we saw that the very reason why propositional attitudes are so important is that attributing them to each other, and to ourselves as well, is the way in which we explicate each other's reasons for what we do.

We will now take a closer look at this interpretive practice, or 'folk psychology,' as it is often called. This will enable us to criticize a conception of intentional states that is the natural counterpart of syntacticalism: functionalism. On functionalism intentional states, such as beliefs and desires, are individuated by their causal profile: they are dispositions that mediate in their specific ways between possible inputs and outputs: inputs in the form of perceptual and intentional states; outputs in the form of intentional states and behaviours. Here, the inputs are supposed to consist of sensory stimuli and brain states, while the outputs are supposed to consist of brain states and bodily behaviour.

On this view, intentional explanation becomes primarily causal explanation: it consists of picking out the brain states that realize the above-mentioned causal profile (although this picking-out happens by way of an intentional, not by way of a neurological description), and specifying their

---

103 Although the term 'folk psychology' may be found to sound pejorative (it is used in such a way by the Churchlands), and to wrongly suggest that our making sense of each other is proto-science, the term as I use it is just synonymous with mutual commonsense explanatory practice.

104 To be precise, there is a choice in functionalism between regarding intentional states as first- or as second-order brain states: a token propositional attitude can be identified with a disposition or with the categorical base of that disposition, i.e., with a causal role or with whatever implements that causal role.

105 Functionalism is the heir of behaviourism. Behaviourism considered just sensory stimuli and bodily behaviour; functionalism considers the rich interplay of presumed brain states in between.
causal relations with other such states, perceptions, and actions (the in- and outputs). Coupled with a nomological view of causality, functionalism thus suggests that folk psychology attempts to explain human thought and behaviour on the basis of something like causal laws.

This suggestion is known as the ‘Theory Theory’ of folk psychology. This theory assumes that folk psychology, our mutual attribution of contentful mental states, is a causal-explanatory and predictive device; that we practice it on the basis of a body of implicit theory; and that the question can be raised whether this body of theory will turn out to consist of adequate categories and true hypotheses, in the face of mature neuro- or cognitive science. Whether or not the beliefs and desires of folk psychology (i.e., the intentional properties attributed by folk psychology) are causally efficacious, will then depend on whether the causal explanations of folk psychology can be vindicated by those of our best brain science, or perhaps cognitive-computational science.

I will below argue that this conception of intentional explanation and of mental states is fundamentally mistaken. Because a mistaken theory of intentional explanation and intentional states leads to a mistaken view of mental causation (in the case at hand, to the view that such a thing as mental causation is, strictly speaking, impossible) this is an important issue.

In the present chapter, I will criticize the functionalistic conception of folk psychology as a theory that deals in causal laws (the Theory Theory). The functionalistic conception of intentional states as causal role states will then be the subject of later discussion (i.e., on Kim’s ‘functional’ model of psychophysical reduction). I will here do the following: I will first describe the Theory theory in some detail. I will then raise my objections against it. After that I will contrast it with the alternative accounts of, first, interpretationism and, secondly, simulationism. My conclusion will be that folk psychology may have theoretical aspects, but that it should not be regarded as a precursor of scientific psychology.

Before I start, a word of caution. There are many versions of functionalism, and also many ways of dressing up the Theory Theory. My criticisms can only be general, and they will very probably not cover all varieties of functionalism. I do hope, however, to attack the general spirit of functionalism in such a way that the need for an alternative view becomes clear.
17.2. **THE THEORY THEORY**

The idea that folk psychology is a theory was originally put forward by Searlars (1956), and has more recently been defended by Churchland (1981). He writes:

> Seeing our commonsense framework for mental phenomena as a theory brings a simple and unifying organization to most of the major topics in the philosophy of mind, including the explanation and prediction of behavior, the semantics of mental predicates, action theory, the problem of other minds, the intentionality of mental states, the nature of introspection, and the mind-body problem. Any view that can pull this lot together deserves careful consideration. (2)

Churchland notes that people can explain and predict each other’s behaviour ‘with a facility and success that is remarkable’ (2); but he suggests that in order to do so they need to be, and indeed are, in command of laws. Examples are (5):

\[
(x)(p) \left( (x \text{ fears that } p) \lor (x \text{ desires that } \neg p) \right) \\
'(x)(p) \left( (x \text{ hopes that } p) \& (x \text{ discovers that } p) \right) \lor (x \text{ is pleased that } p)
\]

Churchland points to the parallel between quantification over numbers in physical theory and quantification over propositions in folk psychology; a parallel that we encountered earlier in McGinn’s claims about the ‘indexing’ of mental models. One should keep in mind, however, that doing folk psychology need not be about propositional attitudes: ‘is hungry,’ ‘needs help,’ ‘isn’t up to that task,’ ‘will be mad at you’ or ‘intends to G’ are folk-psychological characterizations that do not involve propositional attitudes.

Churchland claims that folk psychology is a weak and stagnant theoretical framework that will in time be replaced by neuroscience. He does not think that the propositional attitudes are ‘causal role states’, as func-

---

106 Note, however, that Churchland cannot be called a functionalist: his aim is not to account for propositional attitudes, but to attack the very conceptual framework in which they figure. But the Theory Theory is also defended by Fodor, Jackson and Pettit (among many others): philosophers who are realists about propositional attitudes.
17. Against the Theory Theory of folk psychology

tionalism assumes, because he is sceptical about the very existence of propositional attitudes. This makes him an eliminative materialist rather than a functionalist. Even so, he endorses the Theory Theory of folk psychology.

17.3. AGAINST THE THEORY THEORY

I will argue in a moment that the Theory Theory misconstrues folk psychological practice. I want first to point out, however, that although this theory forms a natural counterpart to syntacticalism, which view I criticized above, it need not presuppose the latter. That is, the plausibility of the Theory Theory does not seem to depend on the plausibility of syntacticalism. Indeed, it seems that the Theory Theory is compatible with McGinn's theory of mental models. We have seen that on this theory, ascribing propositional attitudes consists in indexing (i.e., describing aspects of) mental models. Now, when we ask how do we do this, or what enables us to do this properly, the Theory Theory may well provide an appropriate answer here. It would say that we hypothesize and theorize about each other's mental models on the basis of visible behaviour.

It also seems that the Theory Theory need not presuppose a nomological view of causation either, i.e., one that assumes there to be 'causal laws.' Earlier on, I have argued that the notion of a causal law is problematic. Yet, on the Theory Theory, one may argue that folk psychology is like marine ecology, psychopathology or economics: a scientific discipline that may not, like basic physics, identify laws of nature, but that is certainly concerned with reliable causal patterns, given specific structured circumstances such as the presence of a terrestrial atmosphere, minded creatures, or an economic system.

Let us now see what can be said against the Theory Theory. A first objection is that it inevitably gives a distorted picture of psychology as a science. It may either regard psychology as a search for causal principles relating propositional attitudes; but whatever psychologists are actually doing, it is nothing remotely like searching for causal generalisations about propositional attitudes. Or it may regard psychology as brain dynamics, fully ignoring ordinary mentalistic notions such as perception, planning, knowledge and, indeed, belief and desire; but present-day scientific psychological discourse is replete with such notions, without seeming ready to abandon them. Alternatively, it may regard psychology as some kind of computer
Part III. Mind

science, a study of mental calculus, or mental grammar. But although, for instance, Chomskyan linguistics may be considered a branch of psychology, it is just that: a branch.

A second objection is that the Theory Theory does not provide a realistic picture of either scientific explanation, or actual folk psychology. Goldman (1989, 79) writes:

At least three sorts of problems face the theory theory: vagueness, inaccuracy, and non-universality. When philosophers try to formulate the laws, their examples are typically larded with *ceteris paribus* clauses. This vagueness is a problem because it is hard to see how an interpreter could draw any reasonably definite interpretive conclusion using laws so vague. How could they tell when the *ceteris paribus* clauses are satisfied? Yet interpreters frequently do manage to make quite definite assignments of desires and beliefs. The second problem, the problem of accuracy, arises as follows. It is important for analytical functionalism that the laws be reasonably accurate. If the names of mental states work like theoretical terms, especially in the Ramsey-sentence account of theoretical terms, they do not name anything unless the theory (the cluster of laws) in which they appear is more or less true (...). It is doubtful, however, that ordinary interpreters do possess laws that are true. Third, the standard version of this approach assumes that a single set of laws or platitudes is shared by all competent users of mentalistic vocabulary. But this universality assumption is very dubious.

The third and most fundamental objection to the Theory Theory is that it seems to miss the fact that folk psychology is normative. Kim (1998b, 70-2), as it happens an adept of the Theory Theory himself, rightly claims that deliberating about what you should do is very different from theorizing about what you will do. And folk psychology is at least as much about the former as it is about the latter. In doing folk psychology, we do not merely predict or causally explain a person's actions: we also judge them. Folk psychology is about deliberating actors whose intentional states can be evaluated on the basis of considerations of rationality.

On the Theory Theory, it seems at least to be difficult to get reasons into their proper perspective. The Theory Theory does not miss reasons as causal factors, but it does seem to miss the normative, valuational dimension of intentional explanation. True, when it could be assumed as given that an actor is fully rational, finding out what that actor should do would be a reliable source of predictions about what she will do. But whatever it could mean for the Theory Theory, this correspondence between value and fact does not actually exist: rationality is an ideal or yardstick rather than an actual fact. We can at best be assumed to be partly rational. Possibly, no
actor could be fully rational because of natural constraints on what his brains can handle; many actors are far from rational; and some are even plain crazy.

The Theory Theory, then, seems to expect from us that we coolly calculate the behaviour of others, rather than - what is far closer to what everyday experience suggests we do - our taking up these others’ perspectives, and assess in their place what would be the best thing to do. Let me now discuss two alternatives to the Theory Theory that do accommodate this latter aspect.

17.4. INTERPRETATIONISM

First, interpretationism, as endorsed by Davidson. With its denial of intentional laws and its emphasis on rationality, coherence, and the interpreter’s judgement, Davidson’s interpretationism seems to present a credible alternative to the Theory Theory.

I do have, however, some worries about this approach. First, interpretationism rules out cases in which we ascribe to a person a set of beliefs most of which are false. Its ‘Principle of Charity’ dictates that we ascribe beliefs on the assumption that the interpretee is as rational as is compatible with the behavioural evidence. But for the case of a real madman, interpretationism would imply that the ascription of mental states is pointless. But that is counterintuitive: being mad is not the same as being mindless.

Secondly, and relatedly, Goldman (1989, 77) gives an example of a person holding obviously inconsistent beliefs: an author of a book who believes everything she claims in that book, but also believes that she is fallible and will have made some mistakes (the so-called Preface Paradox, a case rather like that of the Lottery Paradox, see above). Such cases are awkward for the interpretationist approach.

Thirdly, if we ask with Goldman (78) what norms of rationality we are supposed to apply on interpretationism, we do not seem to get a clear answer. For ideal rationality is just unrealistic while, given considerable differences among interpreters, it is doubtful whether an adjusted norm is available that applies across the board.

107 Not, for instance, by Dennett: his version seems to be compatible with the Theory Theory, hence not an alternative to it.
17.5. SIMULATION THEORY

Then there is simulation theory. Its idea is that in predicting and explaining what others do, say, or think, our brain simulates the situation of these others. Somehow we mentally put ourselves in the others' shoes and do some vicarious deliberating on their part: 'What would I do if I were him?' Simulation theory is defended by, among others, Heal (1986), Gordon (1986), and Goldman (1989).

Simulation theory seems to sit rather well with the theory of mental models that I defended earlier: this theory says that simulating is the way in which we think in general. But it is important to realize that simulating a person's mental state is different from just simulating a non-intentional situation. First, if our intentional states are aspects of a mental model, then what we are doing when ascribing intentional states to others is, to construct a model of a model, or, better: a model of an entire situation containing a model.

Secondly, if modeling were all there is to ascribing mental content, some version of the Theory Theory about it might still be true. For adepts of the Theory Theory need not assume something like a deductive-nomological account of explanation. They might as well hold that scientific explanation has much more to do with constructing models and carrying out simulations: a theory of explanation that is perhaps more appropriate to science in a computer age. And when Stich and Nichols (1992) defend the Theory Theory against the attacks of simulation theorists, much of their argument relies on a distinction between a broad from a narrow (sententially oriented, rule- or principle-based) conception of theory (135-7).

Endorsing the theory of mental models does not, then, seem to commit one to endorsing simulationism. But there may be other reasons to do so. Simulation theory may successfully explain 1) how we can make (rational, normative) sense of each other, without having to go interpretationist, and 2) how we can successfully predict and explain the actions and intentional states of others without having to claim that we use a theory. Now, is simulationism better than the Theory Theory?

Both Goldman (1989, 83-92) as well as his Theory Theory-committed critics Stich and Nichols (1992, 150-52), invoke various experimental results to back their relative positions. Stich and Nichols (1992, 142-4) argue that simulationism as well as the Theory Theory are very much underdetermined by empirical data. But it is significant that the latter authors have
to rely on a broad conception of theory (see above): on a narrow view, the Theory Theory is at a great empirical disadvantage relative to simulationism.

A second consideration is Goldman's emphasis on phenomena such as humour, beauty, personal charm or charisma, things that are often very well handled by ordinary people, but not, it seems, on the basis of any theory. Comedians, for instance, are very good at predicting the effects of their jokes on people, but not because they are good in making generalizations such as: 'an unexpected and naughty connection between seemingly unrelated situations will be found funny.' Such rules do not seem to be what enables comedians to play with their audience: what matters here is not cleverness but imagination, the capacity to simulate others' thoughts and feelings.

At this point we might make a choice for simulationism; but in my view it is more plausible to say that simulation and theorizing (understood broadly) are two interrelated aspects of one phenomenon. In fact, this seems already to have become a consensus view. As Goldman (1989, 83) admits, simulation is not our only way of attributing intentional states: behavioural regularities are exploited as well. The issue may now seem to be whether folk psychology is mainly theory or simulation; but then, this may not be a fruitful dichotomy to begin with. Folk psychology is mainly a matter of know-how, and applying such practical knowledge may not be well-described by either calling it theorizing or simulation. Stich and Nichols suggest that folk psychology may exploit a 'system [of internal representation] that is neither sentence-like nor rule-based' (134); but this can be taken, not as a reason for construing the notion of theory broadly, but for arguing that the role of theory in folk psychology is limited.

Folk psychology itself is, of course, scientifically studied (think of studies of autistic people, or of primates, and the question whether these have an intact - 'theory of mind'). Also, scientific disciplines such as history, cultural studies, study of literature, etc. do involve quite a lot of well-documented and systematic folk psychology. This is not surprising, for mental models are part of reality like anything else. They play a key role in phenomena such as wars, financial transactions, weddings, artistic styles or religious world-views, and are indispensable in scientific explanations of such phenomena.

108 In Davies and Stone 1995, for instance, we hardly find any 'pure' views about this issue.
But none of this suggests that in doing folk psychology we apply an implicit psychological theory with laws relating types of mental states and behaviours. We may think up such a theory for an individual person: indeed, many individuals can be said to be studied scientifically by historians. But this is not the way in which the Theory Theory assumes folk psychology to be (proto-) scientific; and such study is not likely to be vindicated or refuted by neuroscience.

The point of the present chapter was not to develop a fully fledged account of folk psychology, but to establish that folk psychology is not a causal theory that subsumes types of mental states under causal laws. If I have succeeded in doing this, the functionalist picture of mental states as causal role states becomes suspect, and with it the idea that mental states can, as causal tokens, be mapped onto neural, or physical states. Hence, an important part of the Standard View of mental causation has been under attack. I will continue this attack in Chapter 19, when discussing Kim's 'functional' model of reduction.
18
Against internalism

18.1. INTRODUCTION

Where is the mind? Descartes thought that it was in the head. He was thus an internalist. Strictly speaking, it is doubtful that his unextended mental substance could be anywhere at all, but it was said to make contact with physical substance in the brain. Physicalist philosophers have dropped the Cartesian mental substance, but many of them have not dropped internalism. They still locate the mind in the brain. They identify mental states with brain states.

In the present chapter I will argue that internalism is mistaken, and that there is no intelligible sense in which mental states can be said to confine themselves to the brains, or bodies, of minded creatures. Mental states of affairs, I will argue, are relational states of organisms. For mental states of affairs turn out to be always, and essentially, environment- and past-involving. Mindedness, we will see, is a way for a living creature to be embedded in its world, rather than a way for its brain to be active; although the first will imply the latter.

The expression 'mental property' may suggest that there is somehow something intrinsically yours that is mental. You may think that the idea of a working brain outside of its proper context makes sense, and that there would be mental states in such a brain. You may close your eyes and ears, let your mind roam, and think that you have shut out the surrounding world and the past. You may think that in ascribing mental states to someone, you do not presuppose very much about her past or environment. Considerations like these explain, perhaps, the persistent appeal of internalism. But we will see below that they lack any sound basis. It can hardly be denied, of course, that we perceive with our senses, think with our brains, and act with our bodies. But these intentional phenomena involve more than their enabling machinery, just as a television show involves more than cameras and televisions.

109 Such as the hackneyed 'brain in a vat' (for instance, Putnam 1981).
I will, then, advocate externalism. With regard to mental causation, externalism presents us with a problem as well as with an opportunity. The problem is our fourth problem of mental causation: if mental properties are relational in character, how can they have causal efficacy, given that nature does not seem to allow 'action at a distance'? I have earlier on argued, however, that on a proper understanding of causality, causation by relational properties is common enough. Especially in those cases where various particulars and states of affairs are causally connected to a past 'structuring cause,' the relational properties implied by such a connection can be argued to have tremendous causal impact. The fourth problem of mental causation should not, then, be an obstacle to seriously considering externalism.

The opportunity that externalism presents us with is, that as soon we realize that mental properties are higher-level relational properties, and that such properties are irreducible to more basic, merely physical properties (see Part I), we can see that mental states of affairs are distinct states of affairs from specific merely physical states of affairs. But if so, we can see that insofar as they are distinct, they can have their own causes and effects. This is why the issue of internalism versus externalism is of crucial importance to my argument.

Given what I have already been claiming about the mind, externalism seems at least a reasonable option. If having a mind essentially consists in representing the world to oneself, one can only have a mind insofar as there is a world that can be represented, acted on, and valued. And if representing is a way of biological functioning, and functioning in its turn presupposes etiology, only those beings can have minds that have an appropriate past.

But externalism comes in different kinds and strengths. Although without a doubt most present-day philosophers of mind endorse one or another form of externalism, many of them are reluctant to draw any radical consequences on its basis. For instance, if mental states have 'wide,' or relational content, then externalism seems to suggest that mental states are relational states, or relations. For they do not merely seem to require relations, but also to be constituted by relations. But this is strongly at variance with the popular view that mental states are token brain states; no matter whether these are often individuated relationally (if only by their causal role). So we often find that mental states are construed as local physical states with causal- or intentional-relational properties (for instance Kim); or we find
18. Against internalism

philosophers (such as Fodor) who distinguish 'broad' from 'narrow' mental content. Indeed, the idea of narrow content is of one piece with the conception of thinking as computation that I objected to earlier on.

I would like to show below that these attempts may well be futile: i.e., that mindedness is a relational phenomenon no matter from what angle we consider it. I will below distinguish six different kinds of externalism: semantic externalism, interpretationist externalism, social-linguistic externalism, etiological externalism, environmentalism, and reasons externalism. These views are all distinct, but not, mostly, incompatible. I will discuss all six forms in turn. I will argue that only the latter four are really important; but these will together be sufficient to undermine any form of internalism.

18.2. SEMANTIC EXTERNALISM

Our first form of externalism about mental content was originally propounded by Putnam (1975b) in the form of the famous Twin Earth, or H2O/XYZ thought experiment. This form of externalism, called 'semantic externalism' by Putnam, relies on the causal theory of reference defended by Kripke (see, for instance, Kripke 1980) and, for a time, by Putnam himself.

The idea of the causal theory of reference is, that in having the concept of a natural kind, such as water, we are supposed to have had causal contact (directly or indirectly, via intermediate language users) with at least one instance of that natural kind. In virtue of this causal link with the stuff originally baptized 'water,' our word 'water' rigidly designates water: if it was the natural kind H2O that was called 'water,' this fixed H2O as the referent of all subsequent word-tokens of 'water.'

This causal theory of reference is often supposed to have considerable consequences for our conception of mental content. For on the causal theory whether or not we are thinking of water appears to depend, not just on whatever seems the case for us ourselves, but also on past states of affairs (about which we are largely ignorant), and the world outside our skin.

Here we are asked to imagine Twin Earth, a world exactly like Earth, except that the substance that everybody calls 'water' is not H2O, but XYZ. If I am unknowingly transported to Twin-Earth, I will think of XYZ that it is water, and I will be wrong. But Twin-I, not physically different
Part III. Mind

from me, has a true belief. Ergo, mental content is not fixed by intrinsic, physical make-up.

What we will think about this kind of externalism depends, of course, on whether or not we accept Kripkean reference theory. Now, on the one hand it seems unlikely that thoughts, or linguistic utterances, can acquire the function of referring to a natural kind like water without appropriate causal links to instances of that kind. But on the other hand, it may well be doubted that such a causal link decides what we refer to each time we say or think ‘water’. For it is either not obvious that the term ‘water’ refers to H2O, and not to anything that is sufficiently H2O-like; or it is not obvious that the meaning of ‘water’ was the same before and after the discovery that typical watery stuff is H2O.110 If so, it is not obvious that my water-beliefs, or water-utterances on Twin Earth are false. In that case, however, it is also not clear that mental content is fixed by factors outside the head.

Two further objections concern the thought experiment. First, any alleged difference between Earth and Twin Earth will either have or not have causal consequences. If there are causal consequences the thought experiment is spoiled, depending as it does on physical indistinguishability between I and Twin-I. If there are no causal consequences, any relevance to referential practice is precluded, and the argument will become lame, assuming differences that do not make a difference.

Secondly, there is ample reason to mistrust thought experiments like the above. The idea of a duplicate world to which we can be transferred is so exotic that we cannot, as the thought experiment asks us to do, rely on any of our intuitions. Twin Earth arguments teach us as much about real-world meaning and reference as Disney comic books teach us about real-world ducks and mice.

Finally, the scope of the Kripke-Putnam argument is limited to cases in which intentional content is constituted by linguistic meaning. From the argument presented nothing follows about so-called ‘nonconceptual content.’ For even if the way in which the word ‘water’ refers is rigid designation, the way in which our brains represent water may not be something like rigid designation at all. If so, semantic externalism does not apply to intentional content in general.

110 The difference is between what Chalmers (1996, 57) calls ‘primary intention’ and ‘secondary intention.’
18. Against internalism

18.3. INTERPRETATIONIST EXTERNALISM

Interpretationism (see above) says that having intentional properties amounts to being interpretable as believing that \( p \), desiring that \( q \), and so on. If so, the notion of mental content only makes sense in the context of a community of interpreters and interpretees. True, a lone agent can on this view still have beliefs and desires, for one can be interpretable as believing that \( p \) without actually being interpreted as doing so. But insofar as interpretationism precludes beliefs or desires-that-\( p \) in abstraction from any actual interpretive practice, it implies a form of externalism.

Because, however, the theory of mental models to which I am committed does not support interpretationism, I do not consider this form of externalism relevant to my purposes. On the theory of mental models, beliefs and desires can be had (i.e., the world can be modeled) regardless of any practice of content ascription.

18.4. SOCIAL AND LINGUISTIC EXTERNALISM

Whatever we think of interpretationist externalism, there are propositional attitudes that arguably could not be had outside the context of a community. The belief, for instance, that the elections will be on Wednesday, could not be had in the absence of a society in which elections are held and in which time is tracked by means of a calendar. In the spirit of Searle (1995) we could say that there exist institutions that are constituted by the speech acts of community members, so that when we think about the institutions, these speech acts help determine the contents of these thoughts.

This kind of externalism applies only to the human mind, but it is not a side issue. A little dramatically, we could say that our minds are largely shaped by the social entities that our speech acts help constitute. True, we are different from brain cells or ants that cannot, as individuals, reflect and react on the wholes they collectively realize. But we humans are so much socially embedded that even our least articulated moods and feelings, up to the tickles in our toes, may be laden with socially constituted meanings.

One variant of social externalism is linguistic externalism. Burge (1979) has argued that for a number of terms meaning is fixed by experts, so that whenever we use such terms non-idiosyncratically, we do so in deference to these experts. If, for instance, we have thoughts about arthritis, what exactly it is our thoughts are about, and whether or not these are true, de-
Part III. Mind

pends on what is collectively established as the meaning of the term 'arthritis.' If someone thinks he has arthritis in his thigh, he may be right in one possible context and wrong in another.

In general, we could say that social-linguistic externalism follows from the fact that a language is a community’s collective possession. Whether or not we hold that strictly private meaning is impossible, in a number of cases meaning is certainly not private. If, then, such non-private meaning constitutes intentional content, such content cannot be fixed by a local basis (e.g., brain activity).

The present case of externalism applies only to the human mind; but we can, if we want, regard it as a corollary of the two kinds of externalism that I will discuss next, and that apply more generally. First, linguistic devices are items with proper functions, different from but continuous with the functionality of items ordinarily called biological (etiological externalism). Secondly, the linguistic decisions of experts can be regarded as an extension of our own cognition (environmentalism). If so, the following two kinds of externalism are the really crucial ones.

18.5. ETIOLOGICAL EXTERNALISM

We have seen earlier that mental states are, essentially, representings to cognizers. Now, if the informational/teleological account of representation is correct, any mental state of a subject presupposes: 1) a personal history of development and learning; 2) a remote history of natural selection shaping the species that the subject belongs to.

This, in effect, is what we might call etiological externalism: the claim that without the right kind of past an organism cannot have intentional states, or harbour specific intentional content. First, mental states presuppose representational capacities, and a capacity can only be representational if it has a function, rather than being a mere disposition. Secondly, precisely what content an intentional state has will be a matter of how, and in what environment, the organism’s representational capacities have been put to use.

Intentional properties, then, are fixed by two stages of an organism’s past. The first is the phylogeny of the species that the organism is a member of. This stage has provided a present organism with the right kind of biological equipment (the properly functioning hardware, as we might put
it) to be the subject of intentional states in the first place. The second stage is the organism’s individual developmental and learning history. If an organism is to have a properly functioning nervous system, the development of that particular nervous system arguably must have taken place in a way that is normal for members of that species. And the specific contents of that organism’s intentional states presuppose a very specific history of causal intercourse with parts and aspects of its environment.

This etiological externalism implies that mental properties are historically-relational properties. The free lunch principle will even say that, because such properties supervene on historical-causal relations, there is nothing more to mental properties than those relations. The consequence is that in ascribing an intentional state to a subject, what we are doing is, in fact, recapitulating or abbreviating a part of its developmental and learning history, or at least (because we do not know the actual historical details) specifying a range of possible such histories.

This is etiological externalism at its starkest: the claim that one’s belief is (an aspect of) one’s history. In the spirit of functionalism, one might object that this is counterintuitive, and that believing that p rather amounts to being predisposed to do F in circumstances c, to do G in circumstances d, etc. For is the identification of such dispositions not the very point of ascribing attitudes, enabling interpreters to explain and predict actions?

But first, if mental states are taken to be just dispositions, their aspects of function, goal-directedness and meaning become obscure. And secondly, identifying a subject’s believing, or desiring, that p with a specific kind of history of that subject already implies the relevant dispositions. For instance, given my specific history (one that shaped me and to which present normal circumstances are causally related), I could not fail to have the disposition to say F when c, to do G when d, to think H when e, and so on.

I admit that often, when ascribing beliefs or desires, we do not care one bit about a subject’s history. We say things like ‘That person wants to get in’ on the basis of mere current behaviour. For all we care, the person might have materialized a few moments ago. But terms like ‘wants to F’ cannot be expected to distinguish between normal cases and freak cosmic accidents, because we just need not, and have never needed, reckon with the latter. If so, the question whether a person-like item without a history could really want to F has no answer: our conceptual apparatus was never meant to deal with such cases. In ascribing beliefs and desires we just assume that an appropriate history is present.
Let us also note that once we switch to less crude ascriptions of mental states (e.g. ‘That person wants to enter his house’), the significance of history is more obvious. After all, many mental states are only meaningfully ascribed in the context of one’s autobiography. Slors (1997, 156) writes:

(E)very mental content that makes sense only within a framework of memories - which is almost every mental content of significance for us (...) - requires implicit reference to the past for its individuation and description.

If Slors is right here, cases in which we abstract from etiology are the exception, while cases in which we include it are the rule.

18.6. ENVIRONMENTALISM

The forms of externalism discussed so far are all about mental content. But we can also adopt an externalist view of the mechanisms underlying mental content. Following Rowlands (1999) I will call this position environmentalism. While content externalism says that mental content does not supervene on current and local bodily, or brain properties, environmentalism says that in thinking we employ more than just our bodies, or brains. According to Rowlands (1999, 22), who has made an extensive and convincing case for environmentalism, this position consists of two claims:

The Ontological Claim: Cognitive processes are not located exclusively inside the skin of cognizing organisms.

The Epistemological Claim: It is not possible to understand the nature of cognitive processes by focusing exclusively on what is occurring inside the skin of cognizing organisms.

To understand this position we should start with Dawkins’s notion of ‘extended phenotype’ (1982). Dawkins urges that we regard things like birds’ nests, beavers’ dams, or spiders’ webs not as incidental add-ons to what in essence are individual organisms, but as essential constituents of the life forms that these organisms represent. They are genuine biological equipment. A corollary of this view is that in order to study an animal’s behaviour, making a sharp distinction between the animal’s body and its environment is a mistake. And indeed, when we substitute ‘biological’ and ‘living’ for ‘cognitive’ and ‘cognizing,’ Rowlands’s Epistemological Claim seems to hold good for the study of birds, spiders, bees, etc.
18. Against internalism

As can be expected, the notion of an extended phenotype applies to humans even more than to any other species. Clothes, tools, buildings and symbols are all part of the human life form, and if we studied human behaviour in abstraction from these, we would entirely miss the point of that behaviour (try to study the Inuit while disregarding their coats, kayaks, igloes, and icy surroundings). Apart from that, it sometimes is not even clear where the body ends and the environment begins. For instance, are hairstyles and tattoos artifacts, or changes of the body? What about artificial limbs and pacemakers?

It seems, then, that extensions of the body in the form of nests, burrows, boomerangs, eye-glasses, factories, and so on, are an integral part of the causal processes that constitute the behaviour of animal or man. Millikan (1993, Ch. 8) argues that the only way to distinguish behaviour in its proper (i.e., biologically relevant) sense from anything else that happens is to consider its function; and that the notion of function is inextricably bound up with that of an environment. She even goes so far as to claim that there is no unambiguous boundary that separates organism from environment. The difference is made by what is relatively in or beyond the control of a system; and that does not give us a sharp division.

The notion of an extended phenotype applies a fortiori to cognition. Just as extensions of the body can enable more efficient ways of foraging, hunting, fighting or transportation (and even may constitute entirely new modes of behaviour, such as grinding an axe or playing golf), they can also enable more efficient cognitive behaviours. Chalmers and Clark (1998), for instance, discuss the computer game Tetris:

In Tetris, falling geometric shapes must be rapidly directed into an appropriate slot in an emerging structure. A rotation button can be used. David Kirsch and Paul Maglio (1994) calculate that the physical rotation of a shape through 90 degrees takes about 100 miliseconds, plus about 200 miliseconds to select the button. To achieve the same result by mental rotation takes about 1000 miliseconds. Kirsch and Maglio go on to present compelling evidence that physical rotation is used not just to position a shape ready to fit a slot, but often to help determine whether the shape and the slot are compatible. The latter use constitutes a case of what Kirsch and Maglio call an 'epistemic action.' Epistemic actions alter the world so as to aid and augment cognitive processes such as recognition and search. (...) Epistemic action, we suggest, demands spread of epistemic credit. (...) Cognitive processes ain't (all) in the head! (8)
And using a rotation button while playing Tetris is just one of the vast amount of ways in which we might undertake epistemic action. We may count on our fingers, write things down in notebooks, use calculators (or supercomputers), call the help of experts, or think aloud. Now, some of these may not require equipment other than our bodies (or the bodies of other people), but others do. With Donald (1991) we may consider the application of external ('extra-cranial') storage of information a major evolutionary shift compared with the cognition of our ancestors, one that did not merely change our behaviour, but the very structure of our minds. Donald (1991, 309) writes:

In the traditional view of psychologists, the mind has clear biological boundaries. "External storage" might be seen as just another term for the culture or civilisation within which the individual exists. The individual picks and chooses, acquires skills and knowledge from society, but nevertheless exists as an easily identifiable unit within that society. (...) But external memory is not simply coextensive with culture in general, or with civilization. Culture and civilization are broader concepts, including material products, such as technologies and cities, and many aspects of human life that are not cognitive. External memory is best defined in functional terms: it is the exact external analog of internal, or biological, memory, namely a storage and retrieval system that allows humans to acquire experience and knowledge.

And he draws the rather sweeping conclusion that '(v)isuographic inventions of all kinds should be classified as hardware changes' (311). Such inventions, Donald claims, have had tremendous consequences for what we, and our minds, are:

Individuals in possession of reading, writing, and other visuographic skills (...) become somewhat like computers with networking capabilities; they are equipped to interface, to plug into whatever network becomes available. And once plugged in, their skills are determined by both the network and their own biological inheritance. (311)

We can push environmentalism to its limits, and say that all channels of information, not just artifacts, may constitute cognitive infrastructure. Rowlands (1999, 106-13) argues that this is the way in which we should understand the Gibsonian concept of 'optic array': the information is out there in the distribution of light, ready to be used by sighted organisms. In the same vein, Dennett (1991, 195-7) argues that thinking aloud amounts to constructing an external information channel connecting subsystems of
the brain that are not, perhaps, connected in the right way by an internal channel. If this is indeed what we do when thinking aloud, the air surrounding us is just as much part of our cognitive machinery as our neurons.

In spite of all this, one might keep insisting that cognition proper takes place in the head. One might admit that external processes are required (perhaps even always required) for us to do our thinking, but that they are not themselves part of the cognitive processes. Clark (1997, 106), however, argues that at some point this may no longer be credible. This will be the case if one cognitive task requires multiple causal loops through the environment, and if these loops are not synchronous. As an example, we may think of a conversation, where multiple sensory modalities are involved. Demarcating the process of cognition by distinguishing what takes place inside and outside the head will then be just as arbitrary as talking about the biological processes that take place in the right half of one's body.

On environmentalism, then, the brain may be the headquarters of the cognitive process to which all lines of communication converge and from which they all diverge; but the causal pathways and loops will come from and extend to the far wide world, rather than stop at the senses or muscles. But if this is right, mental properties will typically not be intrinsic properties of organisms, even if the issue of intentional content is ignored. Regarding only what is spatially internal to an organism, or brain, as 'properly psychological'\(^\text{111}\) (encouraged by such principles as 'methodological solipsism,' 'individualism,' or 'psychological autonomy') will then be on the wrong track. The physical basis of mental properties will not just consist of brain processes, but also of tangles of ongoing interactions with the environment.

### 18.7. Reasons Externalism

I will finally discuss a form of externalism that is not concerned with mental content or cognitive processes, so much as with what we may call normative significance, or the domain of reasons. Probably, this form of externalism is not really distinct from either the social or the etiological form of externalism discussed above. Yet, reasons are a subject in their own right,

---

\(^{111}\) Jackson 1988, 382.
and our conception of reasons will be crucial to the way in which we will approach our second problem of mental causation, the problem of explaining how actions can come about by way of motivation, or how we can be moved by something like 'normative force'.

In the introductory chapter, we have seen that 'mind' is a thoroughly normative notion. Correspondingly, the concept of a propositional attitude is normative, and it is so in the following three ways. First, connections among propositional attitudes are subject to criteria of rationality; secondly, although beliefs are concerned with what is the case, desires are concerned with what ought to be the case; and thirdly, beliefs and desires are subject to criteria of rightness: beliefs are supposed to be true, desires are supposed to be good. These normative aspects all contribute in making them not just causes of, but also reasons for actions (or further propositional attitudes).

I would now like to give a brief (and inevitably simplistic) sketch of what internalism about reasons amounts to: the Humean-functionalist conception of reasons that can be found in the writings of, among others, Dretske, Fodor, and Kim. This conception of reasons does not merely acknowledge a close connection among propositional attitudes and reasons, but it virtually assimilates them. It locates reasons in the heads, or at least in the minds, of individuals, and it can therefore be called an internalist view of reasons.

Let us see how functionalism accounts for the normativity of propositional attitudes that seems to turn them into reasons. For although functionalism pretends that it can construe them in merely causal-dispositional terms, it does have a way to capture the three-way normativity described above. First, the normative connection between, say, a belief that $p$ and a belief that $q$, where the proposition $p$ entails $q$, is accounted for by saying that the causal-syntactic structure of the brain mirrors (as a result of evolutionary, developmental and learning processes) rational connections such as $p$'s implying $q$.

Secondly, the 'ought'-character of desires is accounted for in terms of past reinforcements and acquired dispositions; which, in the face of our rich motivational profiles, are of course supposed to be extremely complex, interconnected, many-sided, indirect, etc.

---

112 There is, perhaps, no reason why Humeanism and functionalism should come together, but that is what we find: they form a stable alliance in contemporary philosophy of mind.
Thirdly, on the view at issue it does make sense to say that beliefs are supposed to be true, but much less to say that desires are supposed to be good. That is, this view does not distinguish desires from value judgements. We encounter here the Humean strand in the view: desires are goals that a creature happens to have, while beliefs are receipts for attaining these goals. Beliefs are on this view concerned with the state of the world as well as with the rules and principles one has to follow in order to have one’s desires satisfied; but not with good and bad, for these are the province of the desires. Beliefs and desires are in this way both the causes of and the reasons for a subject’s actions: in ideal circumstances (i.e., when the beliefs are true and the desires can be satisfied) they justify what they cause, and cause what they justify.

I will below suggest an alternative to this internalist picture of normativity and rationality. To do this, let me contrast my own view with the Humean-functionalist one on each of the three points outlined above. My view of the first aspect, the rational connections among propositional attitudes and the alleged mirroring of rational by causal-syntactic structure will be clear. Earlier on, one of my criticisms of computationalism was that neither our thoughts (at least, many of them), nor the processes in our brains consist in syntactic string-processing. And on the theory of mental models (see above) we will say that any rational coherence of propositional attitudes is constructed by an interpreter (the one, remember, who indexes a mental model by his linguistic means), rather than being mirrored in the subject’s own head by the causal flow of his thoughts. The rational coherence of our propositional attitudes will then consist in the coherence of our mental model, while logical connections among propositions are an aspect of an interpreter’s linguistic capacity, if it functions properly.

On the second aspect of propositional attitude-normativity, the ‘ought’-character of desires, I am largely in agreement with the Humean-functionalist model. Our desires seem to be shaped by our biological design as well as individual development and learning; these factors are what make us goal-directed beings.

The third aspect, however, looks different on my alternative view. I will now suggest that desires can be justified by value judgements, and that the latter are made true or false by states of affairs of which the normative significance is to some extent independent of the beliefs and desires of specific individuals. This externalist perspective on reasons is inspired by Stoutland
(1998), and by Millikan's teleological notion of 'proper function' (see above).

What reasons do we have for believing in normatively significant states of affairs, or external reasons? First, we may argue, with Stoutland, that in a social environment there are situations that we may not think about very much, but to which we nevertheless reliably respond in accordance with specific rules. For instance, we shake hands when introduced to strangers, we halt before stop signs, we pay the right amount of money for our groceries, etc. In such cases, we may construe our reasons for acting as we do as involving desires to follow specific rules in those situations; but we may perhaps more plausibly follow Stoutland in claiming that we just have acquired the know-how to deal with those situations in an appropriate manner. If, however, we act on know-how in such situations, rather than on beliefs and desires, it seems that the reasons for our ways of acting must be sought elsewhere than in our heads.

Secondly, we have seen that a teleological view of biological functioning provides us with an objective, although primitive, notion of good and bad. Good is what enhances proper functioning, bad is what hinders it. Now, I admit that there is quite a gap between our everyday value judgments and this elementary notion of good and bad. It seems, however, that many valuational terms are quite straightforwardly linked to at least some notion of proper functioning. Misunderstandings, diseases, bankruptcies, accidents, crimes, etc. all seem to imply some form of malfunction. And in applying such terms we seem to unreflectively ascribe normative significance to situations, rather than to express specific desires about them. Our desires may even conflict with the normative significance expressed by our terms, for instance when we hope that our enemy will make a mistake.

I would now like to suggest that it is also possible for us to perceive, or fail to perceive, the normative significance of situations. This will be possible if perceiving such normative significance can be understood by analogy with perceiving causal relations. Earlier on I agreed with Searle that once we realize that causality is not object but content of perception, we may hold that causal relations can be perceived as such, just as anything can be perceived as an X or as an Y.

Now, just as it seems plausible to argue that we living creatures are primed by natural selection to appreciate the causal character of a wide

---

113 Note that this is not circular, at least not when we explicate proper function in terms of 'history of use'.

240
range of types of situations, it also seems plausible that we living creatures are primed by natural selection, or by education, to appreciate the normative significance of a wide range of types of situations.

On this externalist account of reasons, reasons need not be mentally represented as beliefs or desires; so that one can have a reason without knowing it. We may here distinguish between reasons for and reasons of a subject, where the former may or may not be mentally represented, while the latter must be so represented. The advantage of this is that we can easily make sense of irrationality: one is irrational if the reasons for one are not one's own reasons. For it seems that one can be irrational even with a coherent set of beliefs and desires.

The above makes the causal efficacy of reasons a more complicated affair than the Humean-functionalist account. Suppose that we are outraged by a war crime. On the latter view we will say that an action or situation caused a desire (say, for a trial), and also, perhaps, that the situation was such that we were predisposed to desire that \( p \) in such a situation. What we may call the 'normative force' of the situation will then just be the causal efficacy of our desire (our reason), which means that the second problem of mental causation is not a separate problem at all.

In my view, however, this does not adequately account for a number of cases. Often, we encounter the normative significance of a situation (e.g., the injustice of slavery, discrimination, etc., or the beauty or awkwardness of a piece of art) as a fact that we can understand or fail to understand, that we can discover or fail to discover, of which we can try to convince others, and so on. This is why I do regard the causal efficacy of 'external' normative significance as a real issue.
19
Against reductionism

19.1. INTRODUCTION

In the above chapters, I have sketched a naturalistic conception of the mind, and criticized a cluster of views that has been influential in shaping the debate on mental causation. It should now be possible to explicate how mental reality, as discussed in the previous chapters, relates to the ontological picture of Part I. How is the mind physically realized? What is the ontological character of perception, thought, and action? This, of course, is the issue that will decide which form our account of mental causation should take on.

The outline of my position will already be clear: I hold that mental properties are higher-level relational properties of particulars. The particulars are physical entities, and may be described as either lumps of matter, systems, organisms, animals, actors, intentional subjects, and so on. Some of them are also persons. The properties consist mainly in myriad relations between these particulars and both their environment and their past.

These relations are of types absent in the domain of the merely physical, supervening as they do, among other things, causal relations at lower ontological levels. Because causal relations are external in character, and thus involve 'addition of being,' and because causal relations at the mental and biological level are different in character from those at the merely physical level (for instance, remembering, feeling, biting, or mating are different causal relations from pushing, falling, or breaking), identity between mental or biological states of affairs on the one hand, and merely physical ones on the other hand, seems precluded (all this was argued in Part I).

Physicalism implies, of course, that mental reality supervenes on merely physical reality, and is therefore nothing over and above physical reality. Even so, mental states of affairs in particular, in supervening widely and holistically, cannot be identical with merely physical states of affairs in particular, or sets of these. The point is that a relation of identity requires a one-to-one relation among the alleged relata, and that for reasons outlined
19. Against reductionism

in Part I, there does not seem to be such one-to-one relation among merely physical and higher-level states of affairs.

An account such as this, however, diverges sharply from other physicalistic views, ones that are far more standard. These are the psychophysical identity theories, among which we may distinguish between theories of type identity and token identity. If we want to understand mental causation, it is crucial that we should have a closer look at such identity theories. For if type identity is plausible, we will say that mental causation is, in fact, physical causation. If token identity is plausible, it is not easy to escape the conclusion that there may be causation by mental tokens, but that mental properties are causally inert. Only if neither is plausible, it seems, can we believe in autonomous mental causation.

In the present chapter I will consider a few current type identity theories (Nagel-reductionism and Kim's functional reductionism), as well as eliminativism (taken as a 'disappointed' version of type identity theory). I will argue that these views all face serious problems. This will probably not exhaust all reductionist options, but I think that my discussion below does suggest that the prospects for psychophysical type identity theories are rather dim. Theories of token identity will be the subject of the next chapter.

19.2. NAGEL-REDUCTIONISM

Let us start with Nagel-reductionism, called after Ernest Nagel. Nagel's program for attaining a unity of science, outlined in *The Structure of Science* (1961), tries to relate theories of the special sciences to each other and, ultimately, to physics, by means of so-called bridge laws. Bridge laws were supposed to translate the vocabulary of the special sciences, such as psychology, into the language of physics.

As famously pointed out by Fodor (1974), the main trouble for Nagel-reductionism is that many special-science properties are multiply realizable. And after having himself struggled with the issue of multiple realizability for many years, Kim (1998a) criticizes Nagel-reductionism on three other points - rightly, as far as I can see. First, Kim points out (95-6) that finding bridge principles cannot be the end point of the reductive endeavour, but only a starting point. For the bridge laws are themselves in need of explanation. Brute mind-body correlations, for instance, would not explain
anything about mental states. The brute fact that there are such correlations in itself is just puzzling.

Secondly, (96-7) Kim complains that Nagel-reductions would not provide any conceptual or ontological simplification. This complaint, if justified, is rather devastating, in that achieving conceptual or ontological simplification is the very point of philosophical reductionism. If Nagel-reduction could generally be carried through, it would in a sense be true that all laws in the world are physical laws. But there would be a staggering amount of these: the bridge laws would themselves all be irreducible and basic, so that they would have to be added to the physical laws already known.

Kim finally points out (97) that Nagel-reductionism would be fully compatible with mind-body theories such as epiphenomenalism, pre-established-harmony theory, double-aspect theory, emergentism, or even Cartesian substance dualism. If so, Nagel-reductionism does not entail any metaphysically interesting claims. In the light of all this, it is not surprising that Nagel-reductionism is now widely considered history.

19.3. FUNCTIONAL REDUCTIONISM

Kim himself, however, has proposed a model of psychophysical reduction that 1) deals with multiple realization; 2) explains any allegedly lawful connections among the physical and the mental (instead of just stating them); 3) provides ontological simplification; and 4) is unambiguously physicalistic.

Let us see how Kim accomplishes all this. Of multiple realizability, Kim points out (1998a, 106-12) that ontologically, it need not be an obstacle to reduction at all. On the contrary, the very idea of physical realization presupposes, Kim argues, that each and every mental property, even though its realizer varies from species to species, or from organism to organism, or even from time to time within an organism, does indeed have a physical realizer. And that may be all the reductionist needs. Kim relies here on a thesis of token identity: he assumes that the multiple-realization thesis shows us that every token mental event or state is identical with a token physical event or state.

On Kim’s ‘property exemplification’ account of events, token identity between mental and physical events is possible only if the properties
constituting the events are also identical. That is, on the Kimian event conception there can only be psychophysical token identity if there is psychophysical type, or property identity. This, presumably, is why Kim holds that reductionism is the only way to salvage physicalism.

Yet, in outlining his program for reduction, Kim already relies on mental-physical token identity. The first reductive step is the 'functionalization' of a mental property $M$ (say, a belief that $p$ or a pain). Kim writes (1998a, 98):

\[ (T) \text{To reduce a property } M \text{ to a domain of base properties, we must first "prime" } M \text{ for reduction by construing, or reconstruing, it relationally or extrinsically. This turns } M \text{ into a relational/extrinsic property.} \]

In his 1996, (105-6) Kim suggests that such 'priming' is achieved by way of the Ramsey-Lewis method. The idea here is to provide 'functional definitions of mental kinds': roughly, to specify mental state types (or whole clusters of these) in terms of their causal roles, that is, in terms of what causes and effects they typically have in various circumstances (for the details, see Kim 1996, 104-13).

Such a functional specification is obviously relational, because the state at issue is related to its actual and possible causes and effects. But note that this kind of relationality is not of the metaphysically ambitious kind advocated by my own relationism. Kim's view is firmly intrinsicalistic.

Now, a causal-role property is a disposition, and dispositions have categorical bases. Kim, then, regards the relational and functional causal role properties to which $M$ is already reduced (or better: Ramseified) as second-order properties:

So $M$ is now the property of having a property with such-and-such causal potentials (…). (1998a, 98)

The final step in the reduction is to claim that, while second-order $M$ is relational, the first-order property, call it $P$, is intrinsic and, indeed, physical. $M$ will then be identical with $P$, for:

$M$ is the property of having $P$. But in general the property of having property $Q = \text{property } Q$. It follows then that $M$ is $P$. (99).
It can now be seen how Kim has met the requirements on a model of reduction, not just of dealing adequately with multiple realization, but also the other three that made the Nagelian model fail. First, in Kim's model there is no need to explain allegedly lawful correlations between mental and physical properties, for the two are identical. Secondly, it would be an understatement to say that ontological simplification has been achieved: the Kimian model in fact allows just physical properties. This also makes it clear that Kim-reduction, in contrast to Nagel-reduction, is only compatible with physicalism, not with epiphenomenalism, dualism, double-aspect theory, etc.

In order to see how radical Kim's proposal is, let us see how he answers the following, obvious question:

Isn't it incoherent to think that a property could be both first-order and second-order, both extrinsic/relational and intrinsic, both a role and its occupier? (103)

He deals with this problem by introducing a sparse conception of properties, according to which 'distinct properties must represent distinct causal powers' (103). Obviously, M does not have distinct causal powers from P, so the conclusion reached by Kim is, that it is not really a property at all! Having made his point, Kim writes (104-5):

(I) it is less misleading to speak of second-order descriptions or designators of properties, or second-order concepts, than second-order properties. (...) From the ordinary epistemic and practical point of view, the use of second-order property designators probably is unavoidable, and we should recognize that these designators introduce a set of useful and practically indispensable concepts that group first-order properties in ways that are essential for descriptive and communicative purposes.

Kim has been warning his readers that physicalism comes at a steep price (59-60), and here it is: arguing from functionalist premises on which mental properties are dispositional, or causal role properties, mental properties do not have any autonomous reality at all. For convenience, we do group physical properties in ways that we need to talk about each other: we do apply psychological concepts. But such concepts do not correspond to real properties. If this is not eliminativism, it does come quite close.

But how good is Kim's functional model of reduction? We might have our worries about Kim's claim that a property like M can be declared identical with P once functionalization has been carried out; after all, property M is multiply realizable while P is not. Slors (1998), for instance,
pointing out that many allegedly functionalizable properties are functional properties (in the teleological sense), asks: ‘How can one function be defined in different specific causal terms and yet be the same in each case?’

More worrisome about Kim’s model of reduction, however, I find its assumption that mental properties (and higher-level properties in general) can be functionalized, i.e., analyzed as dispositional or causal-role properties. We have seen earlier that this functionalist account of mental properties leaves out the entire normative dimension of mental phenomena, and also (in the context of externalism, for instance) that although mental properties will certainly involve dispositions to kinds of behaviour, such dispositions are not what mental properties consist in.

A related worry is that Kim’s model of functional reduction must assume mental-physical token identity in order to get off the ground. It presupposes that we can select a cluster of mental states of affairs and their causes and effects, select a cluster of merely physical states of affairs and their causes and effects, and map them onto each other. If, however, higher-level causes and effects are indeed different from merely physical ones (as I have argued earlier on), such a mapping is impossible. Of course, a functional reduction of mental properties would imply psychophysical token identity, but to get at that point such identity must already be presupposed. Without it, no ‘Ramseifying’ or functionalization of properties can be accomplished.

19.4. ELIMINATIVISM

Let us now consider a view that is even more radical than Kim’s reductionism: eliminativism, or ‘eliminative materialism.’ Eliminativism is a view that is reductionist in spirit, but despairs of the possibility of psychophysical reduction. Assuming that irreducible concepts are not scientific, and therefore not about the real world, eliminativism denies the reality of psychological reality, at least as we think we know it. Eliminativism is an extreme position that is not very popular among philosophers. The best-known eliminativist philosophers are P. M. and P. S. Churchland (e.g. 1981 and 1986 respectively) and Stich (e.g. 1983).114

114 Feyerabend (1963) and Rorty (1965) have also put forward eliminativist views, but they have not been involved in any recent debate about this issue.
Eliminativism makes two claims. First, it endorses the Theory Theory (discussed earlier), suggesting that our mutual interpretive practice, or folk psychology, is proto-science. Secondly, it says that the scientific study of the human mind ought to concern itself with the neural machinery that realizes all our mental capacities and steers our bodily behaviour; and that this study has no use for mentalistic, or folk-psychological concepts such as 'belief' and 'desire.' In this vein, Stich (1983, 5-6) writes:

As I see it, the notion of "content" or the folk psychological strategy of identifying a mental state by appeal to a "content sentence," despite all its utility in the workaday business of dealing with our fellow creatures, is simply out of place when our goal is the construction of a scientific theory about the mechanisms underlying behavior.

According to this view, then, belief-desire discourse is in need of wholesale replacement by neuroscientific discourse.

Because of its counterintuitive implications eliminativism has been criticized extensively and from many sides. Horgan and Woodward (1985), for instance, argue that folk psychology does a much better explanatory job than Stich or the Churchlands suggest it does; that their demands on folk psychology are unreasonably severe; and that, contrary to what it says, there may very well be physical, or causal role states of the brain that can be called beliefs and desires. An often-heard argument is also that eliminativism is self-defeating in that being an eliminativist implies having eliminativist beliefs. Eliminativism is also said to overestimate the possible explanatory and predictive powers of neuroscience.

Not all of these criticisms are bad, but because in my view they do not raise the central issues, I will myself concentrate on two other points of criticism. First, following Baker (1995, 85-9), I would like to stress that for beliefs and desires to count as real, belief-desire discourse does not need any 'vindication' from natural science. The Theory Theory misconstrues folk psychological practice: for its own purposes the general 'method' of folk psychology (taking up one another's perspective and indexing mental models by means of natural language) is perfectly legitimate, albeit mostly not 'scientific.'

Secondly, eliminativism seems to presuppose internalism: it falsely assumes that by studying the 'mechanisms underlying behavior' we also study the mind. This, at least, is the source of Stich's 'principle of psychological autonomy' (see his 1978). The idea is that action is identical with observable bodily behaviour, hence that only events internal to the organism can
19. Against reductionism

causally explain action. On externalism and relationism, however, this view must be mistaken.

19.5. REDUCTIONISM AS METHOD

I have argued earlier that multi-leveled relationality precludes the ontological reduction of states of affairs from higher-level domains to merely physical states of affairs. In the present chapter I have made my criticism of reductionism more specific by discussing two forms of psychophysical reductionism (Nagel-reductionism and functional reductionism) and eliminativism. One might ask, however, how this antireductionism squares with the stark physicalism of the free lunch principle. If physical reality is all of reality, should we not expect there to be some way of reducing everything to the physical?

Let me finish by pointing out that reductionism as a scientific principle is indeed plausible. Scientific reductionism can be considered the direct consequence of a physicalist worldview, and the fact that reductionist science works so well is the best evidence for the truth of physicalism. It is mainly reductionism in this sense that we find in Wilson (1998):

The cutting edge of science is reductionism, the breaking apart of nature into its natural constituents. (...) It is the search strategy employed to find points of entry into otherwise impenetrably complex systems. (...) (R)eductionism is the primary and essential activity of science. But dissection and analysis are not all that scientists do. Also crucial are synthesis and integration (...). (54)

The idea is that in studying a complex system, be it an atom, a cell, an organism, or an ecosystem, a scientist ignores the relations of the system with the rest of the world and concentrates on the capacities that enable it to enter such relations. One then looks primarily at the system's internal structure. It is not as if the system's embeddedness, context, or relational properties are unimportant: the system will be part of a wider system, and as long as this is not appreciated the system's microstructure makes no sense. But the context can be temporarily abstracted from.115 One takes the system apart (physically or in texts, drawings, diagrams, or computer models), studies the constituents (they can also be taken apart), determines

115 See, however, Clark 1997, 113-19.
their causal roles, and finally puts them together again. In the process a
great deal is learned about the system’s behaviour in its wider context.
Unfortunately, Wilson does not stop here. On the next page a sound
scientific principle is turned into a dubious philosophical agenda:

Behind the mere smashing of aggregates into smaller pieces lies a deeper agenda
that also takes the name of reductionism: to fold the laws and principles of each
level of organization into those at more general, hence more fundamental levels. Its
strong form is total consilience, which holds that nature is organized by simple uni-
versal laws of physics to which all other laws and principles can eventually be re-
duced. (55)

If we believe that all of reality is physical, and that higher-level domains supervene on lower-level ones, we may well expect that all of science, from
basic physics to the humanities, will one day be put into one coherent, unified or, as Wilson says, ‘consilient’ worldview. On the other hand, there
is no reason whatsoever to expect laws, theories or sciences to ‘reduce’ to
anything lower-level. Relationality at level a cannot, after all, be captured by concepts appropriate to level b: mere chemistry will never capture the function of a certain protein, and mere biology will never capture the meaning of a grimace.

There is also no reason at all to expect, say, biological patterns and regularities to be mirrored in, or mappable onto, or made intelligible by, chemical patterns and regularities. They just have their own character. Even if the dependency of level a on level b becomes fully transparent, the b-level vocabulary will typically remain unsuitable to even describe (let alone explain) the a-level states of affairs (let alone the patterns and regularities). I can see no possibility of ‘reduction’ in all this beyond the analysis-synthesis reduction described above (and the physicalism that goes with it). That is, I can find nothing that would motivate Wilson’s ‘deeper agenda.’
20

Against token physicalism

20.1. INTRODUCTION

Let us now consider psychophysical token identity theory. First, let me declare myself a hearty psychophysical token identity theorist, or token physicalist, when the tokens at issue are concrete particulars. Persons, for instance, are in my view physical tokens with biological, mental and social properties. The person that is I is token identical with the organism and lump of matter that is I. For standard psychophysical token physicalism, however, the tokens at issue are events and states of affairs. And I will argue below that this form of psychophysical token physicalism is problematic.

Token physicalism in this latter sense (let us call it token event physicalism) is often regarded as the one and only alternative to either reductionism or substance dualism (see, for instance, Kim 1989a or Macdonald 1990), and therefore as the only available form of nonreductive physicalism. On the ontological picture that I have defended, however, it is not. I consider this an important result, for all varieties of token event physicalism (be they Davidsonian or Kimian) seem to run afoul of mental causation.

Below, I will do the following. I will first discuss two possible charges against psychophysical token identity theory that I do not myself endorse: one based on meaning holism, another one based on the normativity of mental states. Next, I will raise what I do consider a problem for token event physicalism: the fact that there is no criterion available for preferring one possible state or event token-identification to another. I will then argue that the way in which token event physicalism tries to handle externalism (the position called 'weak externalism') is defective. Finally, I will point out how, without token event physicalism, we can improve our thinking about actions.
20.2. MENTAL HOLISM AND UNCODIFIABILITY

Although Davidson is the best-known champion of both token event physicalism and interpretationism, interpretationism might provide two arguments against claims of psychophysical token identities.

The first argument derives from intentional holism. On interpretationism, intentional states can only be attributed in clusters. Belief content is said to be fixed holistically: whether or not someone can be said to believe that $p$ is both a matter of what the proposition $p$ is supposed to mean and of whether or not the entire set of that person's beliefs and desires has a minimum degree of rationality.

Now, if it is held that content is constitutive of, rather than (as Davidson himself holds) merely had by some intentional event or state, such states can only be individuated along with a great many other intentional states. We can also say that belief clusters are glued together by rational connections (entailment, justification, etc.), and that no ascription of intentional states outside of such a cluster makes sense.

The trouble for token event physicalism will then be that merely physical events or states are not in this way holistically individuated. If, then, we take a physical state and say that it is a belief that $p$ (of course given its context of other intentional states), it seems difficult to explain why this physical state should be the belief that $p$, rather than some other intentional state. After all, the intentional states come only in clusters. There just does not seem to be a basis for singling out a brain state as the realizer of this, rather than that belief, because the brain states lack the rational connections of the intentional states that they realize.

I am hesitant, however, to myself sustain this charge to token event physicalism. For on McGinn's theory of mental models, any Davidsonian mental holism will be with the (linguistically endowed) interpreter rather than with the intentional subject. I do expect mental models to be so holistic as to make it impossible to physically isolate specific beliefs within it; but that is another kind of holism than the interpretationist variety.

A different, although related charge against token event physicalism derives from an emphasis on the normative aspect of intentional states (remember that intentionality and rationality are conceptually linked). This is the point raised by Child (1994, Ch. 2).

Child takes up the Davidsonian theme of mental anomalism in arguing that while we could perhaps exhaustively and determinately specify the
causal/nomic relations among physical states, we cannot so specify the rational connections among propositional attitudes. The problem is that rationality is uncodifiable: we cannot construct a set of rules that guarantees that if we follow these, what we will do will be good or what we will come to believe will be true. Child argues (60):

If rationality is uncodifiable, there is no system of principles from which we could derive, given a specification of an agent’s physical properties, a statement of what it would be rational for her to believe, desire, and do. If there are no such principles for deriving a statement of what would be rational for S, there are no such principles for deriving a statement of which attributions of mental properties make the best sense of S; and if there are no principles for deriving that, there are no principles for deductively deriving a specification of S’s mental properties.

On Child’s view, then, what propositional attitudes an actor has is indeterminate: it depends on how we make sense of the actor, for which there are no strict rules and which can therefore be done in more than one proper way. Normative-intentional structure, then, is not as determinate as physical-causal structure, which makes it impossible to map the one on the other. So:

The sort of general isomorphism required by the common-sense functionalist’s project, between psychological causal roles and physically specifiable causal roles, is incompatible with the uncodifiability of rationality. (80)

On the theory of mental models, however, intentional content does not depend on how an actor is made sense of by an interpreter, and therefore is not indeterminate. The fact that the rationality on the part of the interpreter is uncodifiable will then be irrelevant to the individuation of intentional states.

20.3. THE PROBLEM OF ARBITRARINESS

Although adoption of the theory of mental models seems to neutralize the interpretationist objections to psychophysical token identity theory, it will not save this view from another objection. This objection derives from the fact that when we try to establish token identities between higher-level and lower-level events or states of affairs, we find that we can do so only in an arbitrary way. My claim is that if one declares a specific mental event token
identical with physical event $a$, one will find no good reasons for holding that it should be $a$ rather than $b$, $c$, etc. If so, this will make claims of token identity either indeterminate or arbitrary.

And this would be a problem for token event physicalism. For first, claims of token identity ought to be determinate. The reason is that any credible psychophysical mapping will have to be a causal mapping. If a mental event $m$ and a physical event $a$ are to be token identical, they ought to have the same causes and effects. But to achieve such a causal mapping requires a very precise and unambiguous delimiting of the physical event, for the smallest physical differences can have the biggest consequences.

Secondly, claims of token identity ought to be nonarbitrary. Otherwise no justification of token event physicalism can be given and that, of course, will render claims of token event identity hollow.

Let me now argue that psychophysical event token mappings must indeed be either indeterminate or arbitrary. First, remember my discussion of states of affairs and, later, of events. If, as I have argued, events and states are constituted by properties, and if higher-level properties, because of their relational character, cannot be reduced to merely physical properties, any psychophysical dependencies will be many-many. On this view it is impossible to identify specific intentional with specific physical states of affairs. Of any intentional state of affairs we can only specify a wide physical base; and this base will be the realizer of many more intentional states of affairs. Taking, however, any narrower physical base and assigning it to one intentional state of affairs in particular would lack justification and would therefore be arbitrary. Note that much depends here on my rejection of an ontology on which events are concrete particulars.

Secondly, however, the indeterminacy/arbitrariness charge does not seem to hinge on the peculiarities of my own position: it is arrived at from different angles by different authors. Van Gulick (1995, 251), for instance, notes that

> the criteria for applying special science predicates may be anchored in our discriminatory cognitive abilities in ways that make them sufficiently indeterminate to prevent any exact match-up with precisely specified sets of physical properties (...).

This argument does not seem to depend on any relationist metaphysics. One could of course still object that the discussion is about events and not about properties. But consider Horgan and Tye (1985, 429-30):
20. Against *token physicalism*

(T)he only way to find the neural event which is allegedly identical with a particular mental event $e$, in a creature who has the neural hardware to subserve such a mental event, is to locate the neural event which has all the causes and effects which the correct psychology of that creature attributes to $e.$ (...) If there are several neural events that fit the bill, then there will be no basis for saying that one of them rather than another is identical with $e.$

To rescue token event physicalism, one would have to argue here that for any mental event there is precisely one neural event that has the same causes and effects. I will not try to prove here that this cannot be done, but from Horgan and Tye (1985) we learn that the obstacles are considerable. Note that their argument grants event discourse, as well as the assumption that mental events could be brain events.

Finally, let us note that the problem of indeterminacy/arbitrariness is not limited to psychophysical mappings, but shows up in much simpler cases as well. Consider, for instance, Sebastian's stroll. On precisely which physical event can it be mapped? What comes to mind is the movements of Sebastian's body. But bodily movements alone do not make a stroll: a cosmonaut might make like movements in his spaceship (say, as an exercise), and that would not be a stroll. Perhaps we might say that Sebastian's movements are a stroll given their context, but it will then be hard to determine where the movement ends and the context begins. For instance, does Sebastian's fastening his shoelaces belong to the stroll? There is nothing in the concept of a stroll that would favour one answer over another.

The same considerations apply to non-actions such as earthquakes, fires, breakings, etc. The general point is that, in spite of what most event ontologies make us expect, no event can be assigned a spatiotemporal boundary that is both determinate and nonarbitrary. As was argued earlier, the correct diagnosis of this shortcoming is that events are not individuated spatiotemporally at all, but narratively.

### 20.4. **Weak externalism**

My next objection to psychophysical token identity theory is that it cannot adequately handle externalism. In my view, externalism implies that mental events involve higher-level relations (e.g., biological and intentional ones). And for reasons stated earlier (see Part I), such relations cannot be reduced to merely physical properties and relations. Many token physical-
ists, however, hold that although externalism is fine for mental content, mental states (or events) can still be regarded as token identical with nonrelational physical states (mostly brain states).

The idea is that beliefs and desires are such physical states, but that these states do have relational properties. The latter are assumed to determine the states' 'wide content' (while opinions diverge on the question whether they also have 'narrow content'). This view, called 'weak externalism' is defended by, among others, Dretske, Fodor, Kim, and Jackson and Pettit. Kim, comparing it with two relationist alternatives, describes it as follows (1996, 200-1):

The third possibility is to consider beliefs to be wholly internal to the subjects who have them but consider their contents as giving relational specifications of the beliefs. On this view, beliefs may be neural states or other types of physical states of organisms and systems to which they are attributed. Contents, then, are viewed as ways of specifying these inner states; wide contents, then, are specifications in terms of, or under the constraints of, factors or conditions external to the subject (...). (W)hen we specify Jones' belief as the belief that water and oil don't mix, we are specifying this belief relationally, in terms of water and oil, but this doesn't mean that water and oil are constituents of the belief or that the belief is itself a relation to water and oil.

To better understand this view, we might compare beliefs to the states of an abacus. Any configuration of beads on the abacus can be described without reference to the abacus' context or history. The configurations are the abacus's 'physical states.' Of course, when the abacus is used, the configurations serve a purpose. Suppose that we use the abacus for encoding messages. We then have a system that can take on various physical states, where any of these states has a meaning. The representational properties of the states will then be relational, but that does not make the states themselves relational in character.


The property of being a planet is an externally individuated property in the sense that, (i) specification of what it is to be a planet involves reference to such things as stars and orbits, (ii) one cannot know what a planet is unless one knows what stars and orbits are, (iii) it is of the essence of a planet that it be related to stars and orbits, and (iv) one could not master the concept of a planet unless one had also
mastered the concept of a star and orbit. Nevertheless, even though the property of being a planet is externally individuated in the sense described above, this does not mean that a token of this type, an individual planet, is located, even partly, where its orbit and sun are located. (...) In a similar vein, mental states can be individuation-dependent upon environmental objects and properties, hence be externally individuated in the above sense, without being located, even in part, where those environmental objects and properties are located.

Macdonald (1990) herself argues that it may be that instances (or better, as she herself says, instancings) of mental properties have physical and nonrelational redescriptions. This could be the case, she argues, if the relational properties are not constitutive of the mental states, but are merely ‘characterizing’ properties. On such a scenario,

instancings of mental properties are instancings of physical types whose individuation conditions do not entail existential claims about objects other than themselves and their subjects (401).

So Macdonald argues (402-3) that events that we characterize relationally may not be relational essentially. She says, for instance, that events that are in fact thinking-of-Vienna may not be thinking-of-Vienna essentially.

20.5. AGAINST WEAK EXTERNALISM

Let me now argue that the above attempts to square psychophysical token identity theory with externalism fail: that in spite of the suggestions above to the contrary, intentional states are not plausibly regarded as internal physical states of mental subjects.

A first objection concerns the analogy of mental states with items like coins (see Dretske 1998), planets, or fathers (see Macdonald 1990). It is pointed out that even though we can specify these entities relationally, they are nevertheless perfectly local, and not constituted by any external relata at all. There is here, however, a glaring disanalogy with mental states (e.g., beliefs), in that the latter are not concrete particulars but states of affairs. Now, while concrete particulars can be described either relationally or nonrelationally, states of affairs cannot. At least they cannot if they have a relation as a constituent. One cannot, for instance, nonrelationally specify Jack’s pushing John. And we have already seen that mental states of affairs do indeed seem to have relations as their constituents.
A second objection to weak externalism concerns its picture of intentional states as representational ‘internal’ states of a system. This picture may at first seem quite plausible. One might argue that representational systems must do their representing by means of a range of possible internal states. After all, information comes to us by means of local causal processes, and it will only become mental content when previous cases of like information have effected internal changes in us that have endowed us with the right behavioural dispositions.

My problem with this picture is, however, that it conflates the producer with the consumer of the information. It may well be true that internal, local and physical brain states represent the world. But that does not make the brain states beliefs, or the brain a believer. For the brain is itself not the consumer of its representations. The representations are not there (at least not primarily) for the good of the brain, but for the good of the organism whose brain it is. Remember from my introductory chapter that mere information, or indication, or ‘natural meaning,’ is not enough for representation, or genuine, underived intentionality. What is vital is that the information be used, or consumed. And it is arguably the survival and reproduction of the organism (or the propagation of its genes) that all the representing that is going on in the brain is for.

Brain representations, then, are to whole organisms, while beliefs are of these organisms. It is, then, a mistake to identify representing brain states with beliefs. An intentional state, we might say, is a state of being-represented-to, or of consuming information. But such a state cannot be merely internal: it is an instance of a teleo-functional and hence (if etiological externalism is right) historical-relational property.

My conclusion so far is that, because weak externalism fails, token event physicalism is not up to the challenges of externalism.

20.6. ACTIONS

As a third and final argument against psychophysical token identity theory, let me show that if it is replaced by relationism, much better sense can be

---

116 Although there are schools in cognitive science that reject internal representations. See, for instance, Hendriks-Jansen 1996.
117 Baker (1995) also holds that beliefs are states of whole persons instead of brain states, but not on etiological-externalistic grounds.
made of actions, especially of the so-called 'accordion effect,' the fact that we do things by doing other things.

My argument will run counter to Davidson's (1969, 164-5). Davidson suggests that it is precisely token event physicalism that enables us to make sense of the accordion effect. His idea is that when events are considered particulars, we can talk about one event (an action) under different descriptions. We say, for instance, that if Brutus killed Caesar by stabbing him, Brutus's stabbing of Caesar was his killing of Caesar. After all, it made no sense for Brutus to ask: 'I have stabbed him, but shall I also kill him?' It was just one action.

But what event, then, was that action token identical with? Davidson here (1971b, 49) introduces the concept of a 'basic' or 'primitive' action. Basic actions are things that are done simpliciter, and not by doing something else. According to Davidson, basic actions are best regarded as the actor's bodily movements (taken liberally so as to include 'standing fast' or 'deciding'). If we do so we can say, for instance, that when I write a letter to greet a friend, my hand movement is a greeting, even though the friend is not greeted before days after the hand movement. There is just a simple local event, and depending on context it may deserve various (relational) specifications.

This account sounds quite plausible\(^{118}\) (note that we do not need a Davidsonian account of events-as-particulars to accept it), until we start to ask what event the basic action is exactly. If, indeed, we could help ourselves to a solid account of events, the account would be fine. But because of the indeterminacy-arbitrariness problem, pointed out above, it falls apart. This can be appreciated when we consider an argument against token physicalism by Hornsby (1997, 64-6).

Hornsby argues that when we want to get at a basic action as Davidson defines it, we must look at the causal chain formed by the 'accordion' that an action typically is, and subtract the effects one by one. So, for instance: a friend is greeted, a letter comes into existence, my pen moves, my hand moves, my muscles contract, and so on. But where are we supposed to stop? Why, for instance, should the first link in the chain be a muscle-contraction? Any criterion for basicness, it seems, could only be arbitrary.

Hornsby claims that when we follow the subtractive procedure as described, we end up somewhere in the brain. But we then realize, she notes,
that ordinary action discourse is not precise enough by far to pick out brain events. She also notes that it has a normative dimension that fails to map onto neuroscientific discourse. The conclusion is that no physical events can be found that are plausibly declared token identical with basic actions.

To me, the conclusion that an action must take place somewhere in the brain is a reductio of the above way of identifying a primitive action: subtracting effects until one finds the basic action. The procedure just leads nowhere, and the notion of a basic action, supposedly token identical with a physical event, is hollow. When asked what the basic action is, we either cannot say, or we just say something arbitrary. If anywhere, the letter-writing takes place at my desk, certainly not somewhere in my brain. But it is just me, my pen and my sheet of paper at the desk, and not also some letter-writing event.

Let us now see how we can account for the accordion effect from a relationist point of view. Here, we should just remind ourselves that actions are causings by intentional states of affairs. If I greet a friend, for instance, this means that some intentional pro-attitude on my part non-waywardly\(^{119}\) causes a perception on the part of the friend.\(^{120}\) In this way, action sentences like: 'I move my pen' can be taken at face value: the action is a causal relation involving me and my pen.

It is true that we can now no longer say that the moving of the pen is the greeting of the friend. The former is a causal relation between me and my pen; the latter between me and my friend. This means that in order to account for the accordion effect, we need an alternative for talk about 'basic' actions and event token identities. For although my writing and my greeting are not identical, they are not distinct either. I did the latter by doing the former, and we have to tell how this can be.

At this point, various options are available. We might say that the writing 'constitutes' (Baker 1997) or 'generates' (Goldman 1970) the greeting. We may also say that the greeting is the writing plus a causal relation. While part of the greeting is then identical with the writing, the former involves

---

\(^{119}\) That is, by way of a speech act. To avoid circularity, we may suppose 'non-waywardly' to be unpacked with an appeal to normal circumstances, and normal circumstances with an appeal to proper function.

\(^{120}\) Or perhaps the action by which I greet. Note that on my account of supervenient causation above, which precludes downward causation, it would be false, strictly speaking, to say that the pro-attitude causes a mere bodily movement. Although it does constrain, or enable, the bodily causes of such a movement.
something extra with regard to the latter. This explains why in a sense I did just one thing, rather than two.

In regarding actions as causal relations we reverse the Davidsonian picture. What on this account was thought to be a mere necessary condition for the action to be a pen-movement or a greeting (namely a relation to pen or friend) becomes constitutive of the action; and what formerly was thought to be constitutive of the action (namely muscle contractions or hand movements) becomes a mere necessary condition for it.

Now, if my reversal succeeds in accounting for the structure of actions, relationism enables us to do what token event physicalism can do, but to do it without the problems besetting talk of basic events and token identities.
Conclusion: mental causation

The above concludes my series of criticisms of received conceptions of mental properties and states of affairs. I had to mount these criticisms in order to forego the dead ends faced by received accounts of mental causation, and to create room for an alternative account. I have, of course, been unable to offer a fully fledged (and nonreductionist) alternative to the mammoth alliance of syntacticalism, functionalism, internalism and token event physicalism. But the fact that I have in Part III been able to mainly rely on existing views clearly suggests that such an alternative is not a castle in the air. And the brief sketch that I did give there, combined with the points made in Parts I and II about ontology and causality, should suffice to now give us an idea what mental causation really amounts to.

Let us, then, revisit our five problems of mental causation in turn: the problem of mental anomalism, the problem of normativity, the problem of phenomenal states of affairs, the problem of externalism, and the problem of causal exclusion. The reader may then decide for herself to which extent, or in what respects, mental causation has ceased to be or remains a philosophical problem.
The five problems of mental causation once again

21.1. MENTAL ANOMALISM

Our first problem, mental anomalism, was that our practice of mutually ascribing mental states to each other in order to explain each other's actions does not seem to be a matter of applying any laws. Yet, explaining action definitely seems to be causal explanation, and causation and natural law do seem to be intimately linked.

Indeed, our subsequent discussion of causality and folk psychology confirmed our worst fears on these points: yes, causality and natural lawfulness are virtually the same thing; no, folk psychology is in no important sense scientific, theoretical, or concerned with laws. Given this, how can our beliefs and desires ever be causes of our actions?

The answer is that something a's having property F at t1 can very well be the cause of something b's having property G at t2 without there being a law connecting F and G. For we saw that, because of causation's circumstantial character, there are no causal laws to begin with. If my analysis of causal relations is correct it is enough that, first (and barring pre-emption), a's having property F at t1 is a circumstantially necessary condition for b's having property G at t2, and that secondly, the particular a is physically connected to (or disconnected from) the particular b in such a way that this (dis-)connection is a circumstantially necessary condition for the dependency among a's having property F at t1 and b's having property G at t2.

My analysis of causation in terms of counterfactual dependencies and underlying physical infrastructure explains why, although there are no causal laws, natural lawfulness is all the same essential to all causal relations; for the required physical connections making up the infrastructure are by definition constituted by physical laws. Anomalism, then, turns out to be a phenomenon characteristic of any higher-level property, but it in no way precludes something's having property F's causing something's having property G.
21. The five problems of mental causation once again

21.2. NORMATIVITY

Our second problem was that, because we typically act for reasons, and reasons are by definition normative, some causes of our actions seem to have an essential normative aspect. At least in normal cases, and allowing for occasional wickedness, self-destructiveness or weakness of will, we do what we do because doing it seems good to us. Apparently, certain circumstances can carry 'normative force': when, for instance, a species threatens to go extinct, or human rights are violated, we find that people take action because they judge these things bad. But how can there be such a thing as motivation (with, occasionally, very tangible results) while natural law is blind and indifferent to the normative significance of any situation?

Let me first point out a respect in which, on the basis of my arguments, we will have to conclude that normative factors are indeed causally sidelined. It may seem that the veridicality of a perception, the truth of a belief, the rationality of a thought process, or the praiseworthiness of a desire must have causal consequences. The veridicality of a perception, it seems, might make the difference between life and death; if I falsely (instead of truly) believe that my paintings will be appreciated by the public, my exhibition will be a fiasco instead of a success; and so on.

In spite of this, however, the truth, rationality, praiseworthiness (etc.) of a subject's intentional states are not causal factors over and above the other states of affairs that they directly supervene on, and are therefore causally redundant. The truth of a belief that \( p \), for instance, just consists in someone's believing that \( p \), along with the fact that \( p \). And if you believe that \( p \) and it is a fact that \( p \), these two factors do the causal work. Truth is in this way not an extra factor, and therefore does not add causal efficacy.

The same, I believe, can be argued of veridicality, rationality, and praiseworthiness: these, too, have their causal efficacy screened off by their direct supervenience base. This conclusion seems to be backed by common sense: no matter how distorted a person's perceptions, wicked his desires, misguided his beliefs, and idiotic his deliberations, such a person acts on them anyway. If this were not the case and the normative properties of intentional states were causally relevant, our world would be a better place.

'Normative force', then, must be sought elsewhere. Note first that although the normative aspects of perceptions, beliefs, desires and thought processes are causally redundant, those of utterances and actions are not. It seems that the reasonableness of a stated argument, the truth of a claim, or
the praiseworthiness of an action, does sometimes make a causal difference, namely if it is appreciated by an audience. Because your utterances and actions are public and can be observed, people might declare you a sage or a dunce (or a hero or a coward, or an artist or a bungler) and treat you accordingly, which may of course have very tangible effects. Even thoughts, and their reasonableness or unreasonableness, can be observed in a sense, namely by introspection. One might, for instance, become disgusted by one's own lewd or hateful thoughts, and decide to resist them. In this respect, utterances, actions and thoughts are in the same category as situations, which may also have a normative significance (i.e., be dangerous, sad, dramatic, ideal, etc.) that can be observed and motivate action.

The crucial question then becomes how such normative aspects of situations, utterances, actions and thoughts can indeed be causally efficacious in motivating those who perceive them. To answer this question, let us first remember that at least some minimal notion of normativity can be rooted in a naturalized teleological conception of proper function. Let us supplement this with a notion of normativity that is derived from the pro-attitudes of individuals and groups, social conventions and institutions, common goals, a shared culture, etc. (We need not go into details as to how both sources of normativity may be related). Let us finally note that these notions of normativity should suffice for covering a good part, perhaps even all, of the ways in which we can be motivated, from running for our lives to observing etiquette.

If the latter claim is right, we can be confident that normative significance is generally a historical-relational phenomenon. Whether we find things good or bad is then not a coincidence: we are primed to find things good or bad by structuring causes in the past, whether these consist of natural selection among our ancestors, personal development and experience, or of some peculiarities of our culture and a bit of social conditioning.

Now, suppose that this historical-relational account of normativity is correct. Can we perceive situations as normatively significant, so that we may be motivated? I have argued earlier on that such perception is indeed possible. We do not perceive the normative significance of a situation in the same way as we perceive concrete particulars. But we do perceive situations as dangerous, unjust, etc. The normativity does not hit our retinas or ear-drums, but it is in the content of our perception. This is not at all exceptional: perception is always perception-as. Our perceptions can have such normative content, I have suggested, because we are primed to per-
21. The five problems of mental causation once again

ceive types of situations as good or bad by a past (of natural selection, development, and learning) in which specific reactions to specific perceptions were normally advantageous.

In order to be causally effective, the normative significance of an action or situation must become our own reason. Mere normative significance seems to be causally inert. Perhaps there are reasons for us that are too subtle to appreciate, and on which we therefore will never act; such alleged reasons do not make a causal difference. But in causing specific beliefs and desires, i.e., in becoming reasons of us, reasons for us can have causal efficacy, i.e., normative force.

21.3. PHENOMENAL STATES OF AFFAIRS

Our third problem of mental causation was about phenomenal, first-person states of affairs. On the one hand it seems undeniable that the ways things look, feel, smell, taste and sound to us are the primary determinants of our behaviour. It often takes quite an effort of the will to resist them: to enter an evil-smelling room, to keep a secret under torture, to leave cigarettes or unhealthy foods untouched, etc. On the other hand, the causal order does not seem to have any truck with this subjectivity at all. The laws of physics are blind, and the chemical-electric brain processes that steer our limbs do not seem to contain much of a 'qualitative' or 'what-it-is-like' character. How, then, can subjective experience have any causal efficacy?

To this problem I have not given a straightforward answer, because the issues about phenomenal consciousness are still too controversial. If qualia are indeed states of affairs that have to be regarded as distinct from all other mental states of affairs, in the sense that the subjective aspect of mental states can be separated from these states' representational aspects, they do raise a daunting problem of mental causation. On such a view, all the causal work is done by the mental states qua 'functional,' intentional, or merely psychological, but not qua phenomenal. A solution to the problem of phenomenal causation is then precluded a priori.

For reasons that I outlined earlier on, I am myself very mistrustful about a conceptualization of qualia along these lines. I would myself rather go along with those who regard qualia as representational mental content. But the issue is controversial, and we just know too little, yet, about consciousness, to speak with confidence.
21.4. EXTERNALISM

Our fourth problem had to do with the relational character of intentional states. We have found that our minds are relational in every dimension that we can think of: even the underlying physical machinery of our mental processes is not confined to our bodies, let alone our brains. This relationality seems to conflict, however, with the plausible assumption that there cannot be immediate ‘action at a distance.’ This assumption seems to many philosophers to imply that causally relevant factors must be local and intrinsic, so that any causal efficacy of the relational mental states of affairs seems precluded. For can our limbs, when we act, be steered by anything else than local bodily processes?

I have had to attack this problem from various sides. First, I have tried to undermine the intrinsicalist assumption that only local and intrinsic properties of particulars can be causally efficacious. Crucial to this endeavour has been my distinction between causal dependencies and causal connections, and my corresponding criterion of causal efficacy. I have spent quite some effort on arguing that without such a distinction no adequate analysis of causation is possible. If, however, we make the distinction, we see that there is every reason to take the causal role of relational properties seriously.

How such properties can indeed be causally efficacious was, secondly, explained with the help of the concept of a structuring cause, together with the idea that causation is circumstantial. If a particular has a relation to an event, state or episode in the past that has not only shaped its present intrinsic properties but also its circumstances, ones that determine which present states of affairs involving the particular cause which other states of affairs, then in being so related, the particular has a present historical-relational property that constitutes causes, hence is causally efficacious.

Thirdly, by endorsing an etiological conception of function and an informational conception of mental content, I have construed mental properties as, indeed, such historical-relational properties. In my view, such a construal is about the only way in which we can make naturalistic sense of mental properties, at least if we do not want to ignore their normative dimension.

Such a view does not seem to sit well with mental internalism, or its weaker offshoot, the functionalist position called weak externalism. Because of this, however, we avoid the many ontological hurdles of token
event physicalism that go with these views. And, even more importantly, the epiphenomenalism. For our relationist conception of intentional states of affairs, including perceptions, thoughts and actions (as a tight, conceptually linked cluster) precludes the identification of actions with mere bodily movements. It thus suggests that mere bodily movements and actions have their own causes each.

21.5. CAUSAL EXCLUSION

Our fifth problem was that, given that 1) mental states of affairs supervene on physical ones (i.e., are nothing over and above physical reality), 2) the physical domain is causally closed (i.e., all merely physical states of affairs have a complete merely physical cause), and 3) no state of affairs can have more than one complete cause (i.e., have more than one true complete and independent causal explanation), there does not seem to be room for causal efficacy of higher-level states of affairs such as mental ones.

The first step towards solving this problem has been to establish an anti-reductionism about higher-level states of affairs. This was achieved by insisting on the importance of relations, the holistic character of higher-level relational properties, and the indirectness of physical realization.

The second step was to establish that mental states of affairs, being higher-level relational states of affairs, cannot be picked out by describing merely physical states of affairs. Their physical realizers are wide, and also the physical base of many other states of affairs. This suggests that it is impossible to individuate a mental-level causal relation at the merely physical level.

The third step was to argue that there are indeed such higher-level causal relations. For I analyzed causation as counterfactual dependence plus underlying physical infrastructure, and I argued that there are indeed higher-level counterfactual dependencies (such as that between monetary value and change of ownership) that have an underlying physical infrastructure. If this has been right, there is room for irreducible higher-level causation, and intentional states of affairs, among which actions, can then be seen to have their own causes and effects. Intentional causal efficacy need not, then, be pre-empted by merely physical causal efficacy. The totality of such efficacy at one moment can perhaps be said to pre-empt eve-
Conclusion: mental causation

...rything higher-level; but then, we have found reasons to deny that it can be called the cause of any higher-level state in particular.

We have seen that this leaves us with a problem about merely physical effects of actions. How, for instance, can my intentionally kicking a stone cause ripples in water? While causal closure seems to preclude this, it seems that without downward causation, any causal efficacy of intentional states of affairs is trumped by the efficacy of a set of lower-level states of affairs. It seems, however, that a metaphysically harmless version of 'downward causation' can to a large extent accommodate our intuitions: if we say that higher-level states of affairs constrain lower-level causal relations, we may say that our intention to kick constrained the options for the merely physical level in such a way that ripples were caused. The causal efficacy is then at the merely physical level, but in being exploited it still belongs, in a way, to us minded creatures.

21.6. CONCLUSION

This is, then, what my reconstrual of mental causation has to offer. In my view its various parts, whether they concern ontology, the philosophy of causality, or the philosophy of mind, can be defended on grounds that are largely independent of their relevance to the issue of mental causation. And together they seem better able to make mental causation intelligible than the received views, which seem to imply some form of epiphenomenalism. One may think that epiphenomenalism is a momentous philosophical discovery, so that demonstrating mental causation is useless. But I rather find epiphenomenalism extravagant.

I have also not developed a theory of mental causation in order to put to rest anybody's worries that it may all be an illusion. Hardly anybody really worries about that, at least outside the philosophers' studies. In my view, the use of constructing a theory of mental causation is that in the process there arise various interesting issues, and much can be learned.
References


273
References

References

References

References

Kim, J. (1990b) 'Supervenience as a philosophical concept' reprinted in Kim (1993), 131-60.
References

Lewis, D. (1979) 'Counterfactual Dependence and Time’s Arrow,' Nous 13, 455-76.
Regenboog.
References


References

References

References


Index

Anscombe, G. E. M. 47, 49, 107, 111
Armstrong, D. M. 5, 6, 7, 8, 10, 11, 13, 16, 21, 22, 47, 49, 87, 102, 103, 109, 117, 118, 123, 167
Austin, J. L. 48
Baillargeon, R. 113
Baker, L. R. xviii, xxiii, 8, 9, 49, 118, 164, 249, 259, 262
Bennett, J. 144, 201
Blakemore, R. P. 194
Block, N. 202
Boden, M. A. 29
Brand, M. 126
Burge, T. 232
Buskens, C. J. J. viii, 113, 198
Cartwright, N. 106, 107
Chalmers, D. J. xvii, 202, 203, 231, 236
Child, W. 198, 218, 253, 254
Chisholm, R. M. 126
Chomsky, N. 223
Churchland, P. M. xxii, 210, 221, 248, 249
Churchland, P. S. 248, 249
Clark, A. 29, 210, 236, 238, 250
Corbí, J. E. 39, 180
Craik, K. 213
Crane, T. 17
Crucius, Ch. A. 10
Cuyper, S. E. ix, 63
Damasio, A. R. 204, 210
Davidson, D. xi, xii, xiii, xvi, xvii, xxvi, 8, 19, 46, 53, 63, 103, 107, 117, 118, 122, 126, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 198, 201, 207, 218, 224, 252, 253, 260, 262
Dawkins, R. 235
Dawson, Jr., J. 116
Deacon, T. W. 212, 216
Dennett, D. C. 196, 198, 204, 207, 210, 218, 224, 237
Derksen, A. A. viii, 214, 217
Descartes, R. xi, xii, xiii, 17, 20, 24, 28, 30, 185, 186, 204, 228, 245
Donald, M. 216, 237
Dowe, P. 48, 62, 71, 74, 158, 82, 87, 88, 91
Dretske, F. xvii, xviii, xxvi, 102, 118, 155, 156, 168, 169, 170, 172, 173, 191, 193, 194, 195, 204, 206, 239, 257, 258, 260
Ducasse, C. J. 48, 54, 107, 126
Eberhardt, M. E. 91, 92
Edelman, G. M. 204, 205
Ehring, D. 48, 50, 60, 65, 91
Fair, D. 48, 55, 71, 107
Fales, E. 102, 110, 117, 118, 154
Feigl, H. xii
Feyerabend, P. 248
Fodor, J. A. 33, 118, 164, 165, 206, 212, 221, 230, 239, 244, 257
Francescotti, R. 14
Frankel, R. B. 194
Frege, G. 128
Gasking, D. 109
Gibson, J. J. 237
Gödel, K. 116
Goldman, A. I. xxvii, 166, 223, 224, 225, 226, 262
Gordon, R. M. 225
Gould, S. J. 119
Grice, H. P. 194

283
<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harré, R.</td>
<td>48, 55, 71, 111</td>
</tr>
<tr>
<td>Hart, H. L.</td>
<td>48</td>
</tr>
<tr>
<td>Heal, J.</td>
<td>225</td>
</tr>
<tr>
<td>Hegel, G. W. F.</td>
<td>14</td>
</tr>
<tr>
<td>Hempel, C. G.</td>
<td>72, 117</td>
</tr>
<tr>
<td>Hendriks-Jansen, H.</td>
<td>213, 259</td>
</tr>
<tr>
<td>Hitchcock, C. R.</td>
<td>72, 80</td>
</tr>
<tr>
<td>Holland, J. H.</td>
<td>28, 183</td>
</tr>
<tr>
<td>Honderich, T.</td>
<td>59</td>
</tr>
<tr>
<td>Honoré, A. M.</td>
<td>48</td>
</tr>
<tr>
<td>Horgan, T.</td>
<td>249, 256</td>
</tr>
<tr>
<td>Hornsby, J.</td>
<td>260, 261</td>
</tr>
<tr>
<td>Humberstone, I. L.</td>
<td>14</td>
</tr>
<tr>
<td>Hume, D.</td>
<td>10, 31, 47, 50, 51, 52, 53, 55, 58, 70, 71, 73, 75, 82, 109, 110, 111, 112, 113, 114, 117, 211, 239, 240, 242</td>
</tr>
<tr>
<td>Hurley, S. L.</td>
<td>200, 204, 211</td>
</tr>
<tr>
<td>Jackendoff, R.</td>
<td>216</td>
</tr>
<tr>
<td>Jackson, F.</td>
<td>165, 166, 172, 184, 203, 221, 238, 257</td>
</tr>
<tr>
<td>Jacob, P.</td>
<td>194</td>
</tr>
<tr>
<td>Johnson, M.</td>
<td>149</td>
</tr>
<tr>
<td>Juarrero, A. xi</td>
<td>148, 183, 191, 192, 200</td>
</tr>
<tr>
<td>Kenny, A.</td>
<td>148</td>
</tr>
<tr>
<td>Kim, J. xii, xvii, xviii, xx, xxviii, 3, 6, 7, 19, 20, 25, 26, 27, 30, 31, 36, 37, 38, 51, 53, 56, 101, 117, 118, 133, 134, 139, 140, 144, 145, 154, 155, 158, 167, 189, 203, 206, 220, 223, 227, 229, 239, 244, 245, 246, 247, 248, 252, 257</td>
<td></td>
</tr>
<tr>
<td>Kincaid, H.</td>
<td>39, 40</td>
</tr>
<tr>
<td>Kirsch, D.</td>
<td>236</td>
</tr>
<tr>
<td>Kitcher, P.</td>
<td>75, 76, 81, 85</td>
</tr>
<tr>
<td>Kripke, S.</td>
<td>230, 231</td>
</tr>
<tr>
<td>L. Paul</td>
<td>158</td>
</tr>
<tr>
<td>Lakoff, G.</td>
<td>149</td>
</tr>
<tr>
<td>Langton, R.</td>
<td>14</td>
</tr>
<tr>
<td>Leibniz, G. W.</td>
<td>14, 206</td>
</tr>
<tr>
<td>Leslie, A.</td>
<td>113</td>
</tr>
<tr>
<td>Lewis, D. xv</td>
<td>14, 47, 50, 54, 56, 65, 95, 117, 118, 126, 127, 131, 145, 201, 246</td>
</tr>
<tr>
<td>Looijen, R.</td>
<td>ix, 28</td>
</tr>
<tr>
<td>Loux, M. J.</td>
<td>5</td>
</tr>
<tr>
<td>Ludwig, K. A.</td>
<td>7, 166</td>
</tr>
<tr>
<td>Lycan, W. G.</td>
<td>212</td>
</tr>
<tr>
<td>Macdonald, C. A. 4</td>
<td>9, 135, 252, 257, 258</td>
</tr>
<tr>
<td>Macdonald, G. F.</td>
<td>135</td>
</tr>
<tr>
<td>Madden, E. H.</td>
<td>48, 55, 71, 111</td>
</tr>
<tr>
<td>Maglio, P.</td>
<td>236</td>
</tr>
<tr>
<td>Malebranche, N.</td>
<td>20</td>
</tr>
<tr>
<td>McGinn, C. xviii</td>
<td>xix, xxvii, 161, 173, 207, 211, 213, 214, 215, 216, 218, 221, 222, 253</td>
</tr>
<tr>
<td>Meijers, A. W. M.</td>
<td>viii, 5</td>
</tr>
<tr>
<td>Mellor, D. H. 17</td>
<td>62, 117, 118, 120, 122, 128, 130, 153, 154</td>
</tr>
<tr>
<td>Menzies, P. 7</td>
<td>109, 118, 123, 124, 128, 130</td>
</tr>
<tr>
<td>Mill, J. S.</td>
<td>56, 130</td>
</tr>
<tr>
<td>Millikan, R. G. xxvi</td>
<td>2, 155, 192, 193, 236, 241</td>
</tr>
<tr>
<td>Mourelatos, A. P. D.</td>
<td>148, 149</td>
</tr>
<tr>
<td>Nagel, E.</td>
<td>244, 245, 247, 250</td>
</tr>
<tr>
<td>Nagel, T.</td>
<td>xvii</td>
</tr>
<tr>
<td>Newton, I.</td>
<td>xi</td>
</tr>
<tr>
<td>Nichols, S.</td>
<td>225, 226</td>
</tr>
<tr>
<td>O'Connor, T.</td>
<td>109</td>
</tr>
<tr>
<td>Oppenheimer, P.</td>
<td>72</td>
</tr>
<tr>
<td>Papineau, D.</td>
<td>50, 60, 65, 71</td>
</tr>
<tr>
<td>Pettit, P. 165</td>
<td>166, 172, 184, 198, 221, 257</td>
</tr>
<tr>
<td>Pinker, S.</td>
<td>186, 208, 213, 216</td>
</tr>
<tr>
<td>Place, U. T.</td>
<td>xii</td>
</tr>
<tr>
<td>Plotkin, H.</td>
<td>113, 198</td>
</tr>
</tbody>
</table>
Index

Popper, K............................................. 50
Prades, J. L........................................ 39, 180
Putnam, H. xii, 33, 204, 228, 230, 231
Quine, W. V. O............................... 128, 198
Ramsey, F........................................ 128, 143, 214, 223, 246
Reichenbach, H.................................. 50, 72, 73
Rorty, R............................................. 248
Rowlands, M..................................... 235, 236, 237, 258
Russell, B........................................... 7, 14, 48, 72, 104, 207
Ryle, G.............................................. xi
Salmon, W. C. xxv, 45, 48, 50, 55, 62, 71, 72, 73, 74, 75, 76, 77, 79, 80, 81, 82, 83, 85, 86, 87, 88, 90, 92, 109, 112, 126, 127, 136, 146
Sanford, D. H...................................... 118, 123
Schaffer, J........................................ 72, 77, 78, 83, 88
Schilder, J........................................... 200
Searle, J. R. 28, 29, 30, 32, 114, 128, 175, 196, 202, 203, 232, 241
Sellars, W........................................... 221
Shoemaker, S................................... 4
Slors, M. V. P. viii, 38, 159, 218, 235, 248
Smart, J. J. C................................... xii
Smith, L. B...................................... 211
Sosa, E........................................... 133, 135
Spelke, E......................................... 113
Stalnaker, R..................................... 21
Stephan, A....................................... 28
Steward, H. 3, 118, 129, 140, 142, 143, 146, 147, 148, 149
Stich, S. P....................................... 206, 225, 226, 248, 249
Stoutland, F..................................... 241
Suppes, P......................................... 62
Thelen, E........................................ 211
Tononi, G......................................... 204, 205
Tooley, M......................................... 48, 50, 102, 111, 136
Tye, M............................................ 204, 256
Vallentyne, P.................................... 14
Van Gulick, R................................. 183, 255
Vendler, Z........................................ 148
Von Wright, G. H............................. 109
Waldrop, M. M................................. 3
Wiener, N........................................ 192
Wilson, E. O.................................... 250, 251
Wittgenstein, L............................... 7
Woodward, J................................... 249
Wouters, A. G.................................. viii, 168
Yablo, S......................................... 158, 179
Nederlandse samenvatting (Dutch summary)

Inleiding

Mentale veroorzaking


De volgende redenering lijkt aan te tonen dat mentale veroorzaking onmogelijk is. 1) Wij zijn fysische wezens in een fysische wereld. 2) Ons waarnemen en handelen kan niet anders dan dankzij fysische processen verlopen. 3) Fysische processen beantwoorden aan strikte wetmatigheden. 4) Mentale toestanden, of processen, zijn in belangrijke opzichten anders van karakter dan louter fysische: ze zijn niet strikt wetmatig, ze hebben een normatief karakter (ze kunnen bijvoorbeeld rationeel of irrationeel zijn), ze zijn subjectief (noodzakelijkerwijs van iemand), en ze hebben intentionaliteit (ze ‘gaan ergens over’). 5) Deze kenmerken kunnen de blinde fysische wetmatigheid van onze wereld niet beïnvloeden. 6) Strikt genomen wordt het pakken van de paraplu, een handeling die immers ook een fysische gebeurtenis is, niet veroorzaakt door wat ik denk en wil, maar door louter fysische processen in mijn lichaam.

Ik ga ervan uit dat 1) t/m 5) juist zijn, maar 6) niet. Het doel van mijn proefschrift is, dit laatste aan te tonen. 6) wordt door verschillende filosofen in de analytische traditie beschouwd als onontkoombaar. Maar wanneer we de realiteit van mentale veroorzaking ontkennen, wordt het een mysterie hoe we verantwoordelijk kunnen zijn voor wat we doen, waarom we in de eerste plaats dingen denken en willen, en waarom we de moeite nemen om met elkaar te communiceren. De wereld houdt simpelweg op begrijpelijk te zijn, inclusief het denken zelf dat tot de ontkennning leidt. We kunnen beter de betreffende redenering

I
wantrouwen dan de absurde, wellicht zelfs incoherentere conclusie aanvaarden.

De lezer vraagt zich misschien af waarom ik 1) en 2) hierboven accepteer. Waarom immateriële entiteiten en energieën uitsluiten? René Descartes heeft inderdaad beweerd dat stof en geest onderscheiden ‘substanties’ zijn die met elkaar interacteren. De aanname is dat de geest een soort immaterieel iets is dat krachten uitoefent op de materie, een extra invloed naast de fysische, een ‘geest in de machine’ (Gilbert Ryle).

Dit idee lijkt echter onmogelijk juist te kunnen zijn. Stel, we spreken van een immateriële oorzaak c van een gebeurtenis e. Als c de oorzaak is van e, treedt e niet toevallig op: veroorzaking van e (anders dan: zomaar optreden van e) lijkt dan een intacte natuurlijke wetmatigheid te vereisen. Interventie van een ‘psychische’ energie zou echter de doorbreking van zo’n wetmatigheid inhouden. Dit ontnemt ons de reden om te spreken van een causale relatie tussen c en e: de relatie zou er hoogstens een van opeenvolging in de tijd kunnen zijn. Men kan protesteren, en ontkennen dat causaliteit natuurlijke wetmatigheid vereist. Elke analyse van het begrip ‘oorzaak’ lijkt dan echter uitgesloten.

Dualisme is verder verdacht door het soort vragen dat het oproept: Waarom zouden alleen levende wezens met hersenen een geest hebben? Waarom tasten hersenbeschadigingen of chemische stoffen de geest aan? De fundamentele vergissing van het lichaam-geest-dualisme lijkt te zijn, dat het ons ‘hebben’ van een geest beschouwt als een relatie tussen twee dingen, geest en lichaam. Een alternatief is het hebben van een geest te zien in analogie met het hebben van een gezicht of een stem: het hebben van een verzameling eigenschappen, niet van een ding.

Tegenwoordig worden 1) en 2) (zie boven) vrij algemeen aanvaard, althans in de analytische filosofie. Dit heet fysicalisme. Iets nauwelijkerig: fysicalisme is de aanname dat het geheel aan fysische feiten in de wereld een voldoende voorwaarde is voor alle verdere feiten, zoals biologische, mentale of sociale. We noemen dit wel het ‘gelaagde’ wereldbeeld: ‘bovenop’ het basis-domein van louter fysische feiten bestaan er, door meerdere niveaus van complexiteit, domeinen van biologische, mentale en sociale feiten. Deze feiten van ‘hogere’ niveaus zijn echter in een belangrijke zin geen toevoeging ten opzichte van het geheel aan fysische feiten: zijn de laatste gegeven, dan zijn de eerste noodzakelijkerwijs ook gegeven.
Inleiding

VIJF PROBLEMEN VAN MENTALE VEROORZAKING

Binnen dit fysicalistische raamwerk doet zich het probleem van mentale veroorzaking echter hernieuwd voor. Mijn eigen diagnose is dat er vandaag de dag binnen de analytische philosophy of mind niet minder dan vijf problemen van mentale veroorzaking bestaan:

1) Het probleem van mentaal ‘anomalisme.’ Terwijl causaliteit natuurlijke wetmatigheid vereist, lijkt de relatie tussen denken en handelen van een actor vaak wel aan rationele maatstaven te beantwoorden, maar nauwelijks aan strikte wetmatigheden.

2) Het probleem van normativiteit. Onze beoordeling van een situatie als gevaarlijk, onrechtvaardig, etc., kan tastbare gevolgen hebben voor hoe wij handelen. Maar hoe kunnen de normatieve aspecten van een door ons waargenomen situatie verschil maken voor wat er gebeurt? Hoe kunnen redenen oorzaken zijn?

3) Het probleem van fenomenaal bewustzijn. Ons handelen wordt in belangrijke mate bepaald door onze subjectieve ervaringen, zoals kleurgewaarwordingen, geuren, etc. De biochemische processen in ons lichaam lijken echter een kwestie van blinde causaliteit: een extra subjectief, of ‘fenomenaal’ aspect lijkt cauusal niets toe te kunnen voegen.

4) Het probleem van ‘extrinsieke’ mentale eigenschappen. De inhoud van onze gedachten lijkt vaak niet alleen te berusten op processen in onze hersenen. Iemand kan bijvoorbeeld niet aan het dalen van de beurskoersen denken zonder ingebed te zijn in een specifieke, c.q., sociale context. Dit suggereert dat de mentale eigenschappen van een actor relationeel kunnen zijn, inplaats van intern, of intrinsiek. Relaties tot zaken ver weg, of in het verleden, lijken echter geen verschil te kunnen maken voor iemands handelen hier en nu: causale invloed is locaal en er is geen directe ‘werking op afstand.’

5) Het probleem van causale uitsluiting. Dit probleem ontstaat wanneer we, naast het boven beschreven fysicalisme, de volgende aannamen doen:

a. Causale geslotenheid van het fysische domein: Fysische feiten hebben een complete fysische oorzaak. Als deze aanname niet juist was, zou van fysische wetenschap geen sprake kunnen zijn; het klaarblijkelijke succes van de hedendaagse fysica weerspreekt dit echter.

b. Uitsluiting van meerdere complete oorzaken: Geen enkel feit kan meer dan één complete oorzaak hebben. Causale ‘overdeterminatie,’ in
Nederlandse samenvatting

strikte zin, bestaat dus niet. Als we dit niet aan zouden nemen, zouden we oorzaken niet meer kunnen onderscheiden van bijkomstigheden.


Geconfronteerd met deze problemen zouden we mentale feiten kunnen zien als ‘epifenomenen’ van de fysische feiten. Mentale feiten worden dan wel fysisch veroorzaakt, maar hebben zelf geen causale consequenties. Maar zoals reeds aangegeven, een onkenning van mentale veroorzaaking leidt tot absurditeiten. Zo’n ontkening ondergraaf ons hele kader van intentionele begrippen: termen als ‘kennis,’ ‘waarneming,’ ‘denken,’ ‘intelligentie,’ ‘communicatie,’ ‘verantwoordelijkheid,’ ‘wil,’ etc. zijn inhoudsloos wanneer de zaken waar ze naar verwijzen geen causale rol hebben. We zouden kunnen ophouden met denken en communiceren zonder dat dit enig tastbaar verschil zou maken. Wat mijzelf betreft: liever de vijf problemen, hoe lastig ook, dan die conclusie.

**DE OPZET VAN DE TEKST**

In de hedendaagse discussies over mentale veroorzaaking vallen de volgende drie dingen op:


2) Dit probleem doet zich verreweg het hinderlijkst voor bij het begrip ‘veroorzaaking,’ of ‘causaliteit.’ In de discussies bestaat geen helder begrip van causaliteit waar bij verschillen van opvatting op teruggegrepen kan worden.
Inleiding

3) Voor de meeste partijen staat wel vast dat mentale toestanden hersentoestanden zijn.

Deze drie punten hebben mij doen besluiten om mijn onderzoek op de corresponderende terreinen uit te voeren: ontologie, causaliteit, en filosofie van het mentale. Deze onderwerpen komen achtereenvolgens aan de orde in deel I, II, en III.

Ik had me graag tot één van de gebieden (dan wel tot één van de vijf problemen) beperkt, en ik heb me inderdaad hoofdzakelijk gericht op de kwestie van causaliteit. Maar er heeft zich in de afgelopen decennia zoiets ontwikkeld als een ‘standaard opvatting’ van mentale veroorzaeking, een brede consensus over de richting waarin oplossingen gezocht moeten worden. Die standaard opvatting berust op een aantal verschillende aannamen die afzonderlijk niet eens zo overtuigend zijn, maar als combinatie een verbazend krachtig geheel vormen. Kritiek op afzonderlijke elementen van de standaard opvatting heeft hierdoor de neiging te stranden op het intact blijven van weer andere elementen. In mijn proefschrift heb ik nu een poging gedaan om de standaard opvatting op verschillende plaatsen tegelijk te ondergraven, en een alternatieve totaalvisie op het onderwerp voor te stellen. De meeste punten van kritiek en alternatieve theorieën lagen in de literatuur al klaar voor gebruik, bij wijze van spreken: waar het op aankwam was, ze op de juiste manier te combineren, en hier en daar wat gaten op te vullen.

Deel I: Ontologie

Laten we teruggaan naar de drie genoemde punten. Het eerste punt maakt duidelijk dat we een onderzoek naar mentale veroorzaeking moeten baseren op een duidelijk ontologisch raamwerk. In deel I van mijn proefschrift bespreek ik bovengenoemde ontologische categorieën, en verdedig ik een niet-reductionistische versie van fysicalisme. Ik probeer hier begrijpelijk te maken hoe mentale feiten, d.w.z. instantiaties van mentale eigenschappen (of beter, hoger-niveau-eigenschappen in het algemeen), een ontologische autonomie kunnen hebben (d.w.z. niet te reduceren zijn tot louter fysische feiten), terwijl ze evengoed fysisch gerealiseerd zijn. Hiertoe ontwikkel ik een visie die ik ‘relationisme’ noem.

V
Nederlandse samenvatting

DEEL II: CAUSALITEIT

Het tweede punt maakt dat het voor de hand ligt dat we bij een onderzoek naar mentale veroorzaking helderheid scheppen over causaliteit. Wat wil het zeggen dat x oorzaak is van y? In de discussie over mentale veroorzaking wordt soms verondersteld dat er een 'causale wet' w moet bestaan zodanig dat we, gegeven x, een aantal randvoorwaarden, en w, langs logische weg y moeten kunnen afleiden.

Behalve deze 'neo-humeaanse' opvatting van oorzakelijkheid vinden we in de literatuur echter een aanzienlijke hoeveelheid alternatieve theorieën. En het is verbijsterend om te zien hoezeer deze uiteenlopen. In deel II probeer ik een visie op causaliteit te ontwikkelen die a) dit uiteenlopen verklaart; b) recht doet aan een aantal gangbare intuities over causaliteit (bijvoorbeeld dat er geen 'werking op afstand' is, of dat causaliteit en natuurlijke wetmatigheid nauw verbonden zijn); en c) opgewassen is tegen een aantal notoire problemen (zoals het maken van onderscheid tussen causale en niet-causale afhankelijkheid, of tussen feitelijke oorzaken en mogelijke 'ondervangers').

DEEL III: DE STANDAARD OPVATTING VAN MENTALE VEROORZAKING, EN EEN ALTERNATIEF

Het derde punt, tenslotte, geeft aanleiding tot wat we de 'standaard opvatting' van mentale veroorzaking kunnen noemen: het idee dat de causale effectiviteit van overtuigingen, verlangens, intenties, etc. feitelijk de causale effectiviteit is van hersentoestanden. Het probleem met dit idee is dat mentale eigenschappen, en daarmee mentale feiten, causaal overbodig lijken. Wanneer bijvoorbeeld een bepaalde hersentoestand veroorzaakt dat ik een paraplu pak, lijkt het er causaal niet toe te doen dat die hersentoestand een overtuiging is (bijvoorbeeld dat het regent), of een verlangen (bijvoorbeeld om droog te blijven). De intentionele inhoud van die toestand, d.w.z. waar ze 'over gaat,' voegt causaal niets toe aan de biochemische processen in mijn hersenen. Alle vijf genoemde problemen van mentale veroorzaking lijken van toepassing.

Vanuit een bepaalde opvatting over mentale eigenschappen kan men wel enigszins afdingen op de ernst van dit feit. Ik doel hier op wat bekend staat als 'functionalisme.' Een functionalist zal zeggen dat mentale eigenschappen disposities van de hersenen zijn. Zo'n dispositie is een
toestand zodanig, dat wanneer de hersenen input van het type $x$ krijgen, output van het type $y$ het gevolg zal zijn. Eén zo'n mogelijke toestand zou te beschrijven kunnen zijn als de overtuiging dat het regent. De overtuiging dat het regent wordt op die manier een hersentoestand die (ongetwijfeld in combinatie met andere hersentoestanden) causaal ‘bemiddelt’ tussen specifieke typen input en output. Een mentale toestand is zo de belichaming van een ‘causale rol’ in de hersenen. Die causale rol is uiteraard fysisch, maar om haar te typeren is mentale terminologie wellicht toch onmisbaar.

We kunnen hierbij een parallel trekken met een computer. Een computer heeft ook ‘interne toestanden’ zodanig dat input van type $x$ (bijvoorbeeld de toetsencombinatie Ctrl+o) output van type $y$ tot gevolg heeft (bijvoorbeeld een dialoogvenster op het scherm). Bij een computer zijn deze interne toestanden voorgeprogrammeerd door personen. Een functionalist zal echter zeggen dat de interne toestanden van hersenen ook ‘voorgeprogrammeerd’ zijn, namelijk door ontwikkelings- en leerprocessen in het verleden. Inderdaad wordt de computermetafoor door veel filosofen serieus genomen. Veel functionalisten hebben beweerd dat de manier waarop onze hersenen informatie verwerken evenals bij computers bestaat in het uitvoeren van formele bewerkingen op ‘symbolen’ die beantwoorden aan een natuurlijke ‘codering’ van informatie.

Een functionalist kan afdingen op de ernst van de problemen van mentale veroorzaking, omdat het bovenstaande suggereert dat in ieder geval onze beschrijving van hersentoestanden in intentionele termen (dus als overtuigingen, verlangens, intenties, etc.) voor verklarende doeleinden onmisbaar is. Door elkaar overtuigingen, etc. toe te schrijven (iets wat wel ‘folk psychology’ wordt genoemd) kunnen we elkaars gedrag tot op zekere hoogte voorspellen en verklaren, zelfs op de lange termijn. Dit rechtvaardigt een gebruik van mentale terminologie, en tot op zeker hoogte een geloof in - zo niet causale ‘effectiviteit,’ dan toch een soort causale ‘relevantie’ van mentale toestanden.

We kunnen met het bovenstaande genoegen nemen. Ikzelf zou echter willen blijven staan op het belang van mentale veroorzaking in de letterlijke zin van het woord. In deel III stel ik bovendien vast dat de boven beschreven opvatting van mentale toestanden als hersentoestanden om een aantal redenen onhoudbaar is. Ik bepleit hier een alternatieve visie op mentale toestanden, een visie die we ‘radicaal externalisme’ zouden kunnen noemen. Dit externalisme sluit aan op de relationistische ontologie.
Nederlandse samenvatting

uit deel I: het is de opvatting dat mentale toestanden geconstitueerd worden door relaties van een organisme met zijn ontwikkelings- en leergeschiedenis, en via deze met zijn natuurlijke en sociale omgeving.

Het idee is als volgt. Waarnemen, denken en handelen zijn onlosmakelijk met elkaar verbonden, en het is allemaal te beschouwen als 'gedrag,' of 'activiteit.' Ruwweg: waarnemen is het oppikken van informatie, denken het verwerken ervan, en handelen het benutten ervan. (Onder 'informatie' wordt hier weinig anders verstaan dan omgekeerde causaliteit; het feit dat een gebeurtenis e plaatsvindt de mogelijkheden inperkt betreffende wat er elders aan de hand is - bijvoorbeeld het ruiken van brandlucht en een brand.) In contrast met het functionalistische input-output-model: waarnemen is geen input, maar een actief proces (dat echter wel input vereist); de informatieverwerking die we denken noemen kan deels plaatsvinden buiten de hersenen (zoals wanneer we pen en papier gebruiken bij het maken van een som); en elke handeling omvat zowel in- en output als interne processen (we sturen en begeleiden onze bewegingen).

De activiteit van zintuigen zal altijd, direct of indirect, gekoppeld zijn aan verdere activiteit (het benutten van de informatie), en wel zodanig dat het de overleving en voortplanting van het betreffende organisme bevordert (dit laatste lijkt de enig begrijpelijke betekenis van 'benutten' te zijn).

Hiermee komen we op een belangrijk punt: mentale, of intentionele, categorieën zijn functiecategorieën. Functiecategorieën introduceren teleologie: ze betreffen niet zozeer de fysische structuur van een organisme (zoals 'heeft drie uitsteeksels,' 'bevat 70% water,' etc.), als wel het doel waar iets toe dient, 'waar het voor is.' Denk hier aan biologische begrippen als 'paren,' 'broeden,' etc. Deze verwijzen niet naar louter causale processen, maar naar typen gedrag dat onder normale leefomstandigheden het overleven of voortplanten bevordert.

Ruth Millikan heeft nu betoogd dat ook intentionele categorieën zoals waarnemen, denken en handelen functiecategorieën zijn. En inderdaad: zien, bijvoorbeeld, is niet slechts het worden geraakt door licht (anders zou een blinde muur ook kunnen zien), maar het benutten van licht als informatiebron. Fylo- en ontogenese, en individuele leerprocessen, hebben hierbij iemands fysische kenmerken (bijvoorbeeld de 'bedrading' van de hersenen) zodanig gestructureerd dat men zich gegeven enigszins normale leefomstandigheden, zo goed en zo kwaad als het gaat, en op zijn eigen
Inleiding

soortspecifieke en/of individuele manier, kan redden. Bij functiekenmerken kunnen we zo van een 'gebruiksgeschiedenis' spreken: de kenmerken in het heden zijn ergens 'voor' dankzij het feit dat ze zich in het verleden gevormd hebben, door aanpassing van het organisme (of diens voorgangers) aan de leefomgeving: een leefomgeving die nog steeds min of meer intact is, en de context vormt voor het zich manifesteren van de betreffende functie.

Het toekennen van mentale eigenschappen veronderstelt zo dat er sprake is van een voorgeschiedenis. Wanneer we dus van iemand beweren dat hij gelooft dat het regent, schrijven we, volgens bovenstaande redenering, die persoon impliciet een verleden toe van ontwikkelings- en leerprocessen in een normale leefomgeving. Dit is de belangrijkste door mij aangehangen vorm van externalisme. Mentale eigenschappen zijn dan geen hier-en-nu-eigenschappen van de hersenen, maar (met name causale) relaties van een organisme met zijn ontwikkelings- en leergeschiedenis, en via deze met zijn normale (natuurlijke en/of sociale) omgeving. Het hebben van een verlangen of overtuiging, het doen van een waarneming, of het uitvoeren van een handeling, wordt zo een relationele eigenschap van een levend wezen als geheel. De juiste 'interne toestanden' zijn vereist, maar niet constitutief; de juiste relaties zijn constitutief, niet slechts vereist. Dit is radicaal externalisme. (Volgens de meeste functionalisten liggen de zaken precies omgekeerd.)

Het is waar dat niet alleen mensen, maar ook andere, vaak primitieve, levende soorten informatie verwerken. Een vlieg verwerkt bijvoorbeeld visuele informatie, maar we zullen aan een vlieg nauwelijks mentale eigenschappen toeschrijven. Dit suggereert dat we pas mentale terminologie gaan toepassen wanneer de informatieverwerking in zekere mate indirect en complex is. Er lijkt geen reden te zijn om te geloven dat hier een scherpe grens te trekken is tussen wezens die geen, en wezens die wel een geest hebben.

Het is ook waar dat van veel informatie die wij verwerken moeilijk volgehouden kan worden dat ze het overleven en voortplanten bevordert. De weg van primitieve reflexen naar complex menselijk gedrag is lang en grillig. Toch staat waarschijnlijk niets van wat zich in onze geest afspeelt los van onze belangen als levend wezen. Misschien heeft het schrijven van een gedicht geen overlevings- of voortplantingswaarde. Maar taalvermogen en inventiviteit hebben dat wel degelijk. Vanwege de doorgaans indirecte relatie van ons alledaagse denken enerzijds en primitieve biologische noodzaak anderzijds, heeft Millikan een belangrijke uitbreiding van ons
begrippenapparaat voorgesteld met termen als 'afgeleide functie' en 'aangepaste functie.'

De vraag is nu hoe we met behulp van de causaliteitstheorie uit deel II moeten oordelen over de causale rol van mentale toestanden. Voor ik hierop inga zal ik eerst mijn stellingnamen in deel I, II en III in wat meer detail schetsen.

**Deel I. Ontologie**

**STANDEN VAN ZAKEN**

In deel I betoog ik, in het voetspoor van David Armstrong en anderen, dat de werkelijkheid wordt geconstitueerd, niet zozeer door objecten als wel door 'standen van zaken.' Een stand van zaken is het hebben van een eigenschap op (of gedurende) een bepaalde tijd, door een concrete particuliere entiteit (een tijd-ruimtelijk gesitueerd object), of het staan in een relatie (een meerplaatsige eigenschap) van zo'n entiteit tot een andere. Wanneer ik in deze samenvatting 'feit' of 'toestand' schrijf, doel ik op 'standen van zaken' in deze zin. Een 'mentaal feit' is dus het hebben van een mentale eigenschap (bijvoorbeeld: gelooft dat het regent) door een object, c.q., organisme, gedurende een bepaalde tijd. Deze opvatting geeft aan dat er niet zomaar objecten bestaan, maar dat er noodzakelijkerwijs aangaande elk object dingen 'het geval zijn.' Immers, objecten kunnen in verschillende opzichten op elkaar lijken of van elkaar verschillen.

**FYSICALISME**

Vervolgens beschrijf ik de positie die we 'fysicalisme' noemen als de opvatting dat er een eenzijdige afhankelijkheidsrelatie bestaat tussen fysische feiten enerzijds, en biologische, mentale en sociale feiten anderzijds. Ik karakteriseer hier fysische feiten louter negatief: die feiten die we niet biologisch, mentaal, of sociaal zouden noemen (het zal duidelijk zijn dat het hier gaat om feiten waarbij particuliere entiteiten betrokken zijn, en niet om feiten zoals 1 + 1 = 2). Dit doet mijns inziens het meeste recht aan de niet-dualistische basisintuïtie: mentale feiten (en hoger-
niveau-feiten in het algemeen) zijn slechts een kwestie van complexe fysische processen, en niet van een extra substantie, of energie, die aan de fysische werkelijkheid wordt toegevoegd. Met het oog op de hedendaagse stand van kennis over met name de hersenen, en de impasses van een dualistische visie, lijkt een serieus alternatief voor althans dit ‘minimale’ fysicalisme niet voorhanden.

**FYSISCHE REALISATIE**

De vraag is nu hoe we de relatie moeten denken tussen de fysische en de verdere feiten (Ik zal mij vanaf hier voorlopig beperken tot mentale feiten, maar het betoog gaat ook op voor biologische en sociale feiten. Merk op dat er in de allereerste plaats geen scherpe scheiding tussen deze categorieën bestaat). Wat moeten we ons voorstellen bij een ‘gelaagde’ werkelijkheid? Een mogelijk antwoord, reductionisme, is de opvatting dat mentale feiten identiek zijn met fysische feiten. In dat geval moeten mentale eigenschappen (zoals de overtuiging dat het regent) echter te begrijpen zijn als fysische eigenschappen (zoals biochemische eigenschappen van de hersenen). Een dergelijke reductie lijkt echter, om een veelheid van redenen waarop ik in deze samenvatting niet nader zal ingaan, principieel onmogelijk.

Behalve als identiteit wordt de relatie tussen fysische en mentale feiten ook wel gekarakteriseerd als superveniëntie, emergentie, of veroorzaking. De eerste twee termen verklaren echter niet veel, en van de derde valt aannemelijk te maken dat hij misplaatst is. Zelf betoog ik dat we hier moeten spreken van ‘realisatie.’ Deze term drukt uit dat mentale feiten misschien niet identiek zijn met fysische feiten in het bijzonder, maar ook geen toevoeging inhouden ten opzichte van het geheel aan die fysische feiten. Om dit te begrijpen kunnen we denken aan eigenschappen en objecten: de vorm van een tafel, bijvoorbeeld, is geen ‘toevoeging’ aan die tafel, maar is er ook niet identiek mee (en kan evenmin gereduceerd worden tot andere eigenschappen ervan, zoals de moleculaire structuur of de kleur). Op dezelfde manier zijn mentale eigenschappen eigenschappen van fysische objecten naast andere (c.q., louter fysische) eigenschappen.
Nederlandse samenvatting

RELATIONISME

Nu is echter de vraag hoe er in de allereerste plaats andere dan louter fysische eigenschappen kunnen bestaan. Wat maakt eigenschappen tot 'hoger-niveau-eigenschappen'? Een antwoord dat voor de hand ligt is dat zulke eigenschappen hun bestaan danken aan een gelaagde complexiteit in fysische processen, van subatomaire deeltjes en organische moleculen tot complete organismen en gemeenschappen. Deze complexiteit ontstaat op elk niveau voornamelijk doordat systemen met elkaar interacteren: dus door causale relaties. Wij kunnen ook vaststellen dat er typischervijze hele netwerken aan causale relaties aanwezig moeten zijn: biologische eigenschappen doen zich slechts voor in de context van een ecosysteem, en sociale eigenschappen doen zich slechts voor in de context van een gemeenschap. Ook het hebben van mentale eigenschappen veronderstelt een causale inbedding, namelijk in de vorm van ontwikkelings- en leerprocessen in een leefomgeving.

Het bovenstaande suggereert dat hoger-niveau-eigenschappen typischervijze relationele eigenschappen zijn. Dit levert het beeld op van vele lagen van verschillende typen causale inbedding: corrosie, catalyse, etc. (chemisch); stofwisseling, camouflage, voortplanting, etc. (biologisch); waarneming, herinnering, beslissing, etc. (mentaal); en wetgeving, economie, onderwijs, etc. (sociaal). Een consequentie is dat hoger-niveau-eigenschappen niet 'intrinsiek' aan objecten zijn: het hebben van zulke eigenschappen bestaat in het (met name causaal) gerelateerd zijn aan andere objecten en gebeurtenissen. Dit maakt dat deze eigenschappen een belangrijk historisch aspect hebben; immers, causale relaties nemen (anders dan bijvoorbeeld ruimtelijke relaties) tijd in beslag.

Wanneer we ons nu afvragen hoe zulke eigenschappen fysisch gerealiseerd zijn, merken we dat veel ervan een brede tijd-ruimte-regio moeten beslaan. Men kan bijvoorbeeld alleen maar politieagent zijn in relatie tot een rechtssysteem, en daarmee tot een hele maatschappij. Deze theorie over hoger-niveau-eigenschappen, en de niet-locale manier waarop ze fysisch gerealiseerd zijn, noem ik 'relationisme.' Zij speelt een cruciale rol in mijn alternatief voor de standaard opvatting van mentale veroorzaking: als mentale feiten handelingen veroorzaken, dan zal dit volgens de relationistische positie de causale effectiviteit moeten zijn van relationele feiten, en niet van locale hersentoestanden.
Deel II. Causaliteit

INLEIDING

Het lijkt vanzelfsprekend dat een theorie over mentale veroorzaking alleen een kans van slagen heeft wanneer ze gebaseerd is op een heldere analyse van oorzakelijkheid. Maar oorzakelijkheid is een zeer lastig te analyseren begrip, en er bestaat weinig overeenstemming over de mogelijkheid en eventuele vorm van zo'n analyse.

Mijn betoog over causaliteit valt uiteen in enerzijds een analyse van de causale relatie (Wat is veroorzaking?); en anderzijds een ontologie van causale relata (Wat voor soort zaken zijn oorzaken en gevolgen?). In dit centrale en veruit meest omvangrijke deel van mijn dissertatie laat ik de problematiek van mentale veroorzaking grotendeels buiten beschouwing. Om namelijk de legitimiteit te verkrijgen om later een causaliteitstheorie toe te passen die afwijkt van de meer gangbare, wil ik deze theorie eerst op onafhankelijke gronden ontwikkelen en verdedigen.

Wat de vraag 'Wat is veroorzaking?' betreft stel ik een fusie voor tussen twee bestaande typen causaliteitstheorieën: de ‘contrafactische’ theorie (van bijvoorbeeld John Mackie en David Lewis) die oorzaken analyseert als noodzakelijke voorwaarden; en de ‘mechanistische’ theorie (van bijvoorbeeld Phil Dowe en Wesley Salmon) die oorzakelijkheid analyseert als behoud en voortplanting van een fysische grootheid, zoals energie. (De term ‘contrafactisch,’ of ‘tegenfeitelijk,’ geeft aan dat we feitelijke met louter mogelijke situaties kunnen vergelijken, bijvoorbeeld: als ik mijn paraplu gepakt had, was ik niet nat geworden. Pas daardoor kunnen we nadenken over noodzakelijke voorwaarden.)

De eerste theorie stelt veroorzaking voor als een asymmetrische afhankelijkheidsrelatie tussen feiten of gebeurtenissen; de andere als het voortbestaan van en/of de interactie tussen objecten. Ik beargumenteer dat we deze twee zaken moeten beschouwen als complementaire aspecten van causale relaties; ik ontwikkel een theorie over hoe die aspecten gerelateerd zijn; en ik suggereer dat deze geïntegreerde theorie opgewassen is tegen een aantal problemen waartegen de contrafactische en mechanistische theorie geen stand houden wanneer ze op zichzelf blijven staan. Bovendien suggereer ik dat deze analyse krachtiger en intuïtiever is dan andere
soorten causaliteitstheorieën, zoals neo-humeaanse (of nomologische), probabilistische, particularistische, en antropomorfistische.

OORZAKEN ALS NOODZAKELIJKE VOORWAARDEN

Een contrafactische causaliteitstheorie stelt dat oorzaken noodzakelijke voorwaarden zijn voor hun effecten, gegeven de omstandigheden. De omstandigheden kunnen hier elk afzonderlijk ook oorzaken genoemd worden, zodat gebeurtenissen typischwijze een veelvoud aan (samenwerkende) oorzaken hebben. Een ongeluk bijvoorbeeld kan als oorzaak de dronken toestand van een bestuurder hebben, maar ook de aanwezigheid van een boom langs de weg, de gladheid van het wegdek, etc. Elk van deze factoren kan een oorzaak genoemd worden mits de factor (gegeven de andere factoren) 'vereist' was voor het ongeluk, d.w.z. dat het ongeluk in de feitelijke omstandigheden niet had plaatsgevonden wanneer één van de factoren had ontbroken.

Vaak wordt binnen deze theorie eveneens gesteld dat alle factoren tezamen een voldoende voorwaarde vormen voor het effect, dat wil zeggen dat oorzaken hun gevolgen noodzakelijk maken. Deze aanname heeft vaak aanleiding gegeven tot het idee dat er causale wetten bestaan. Er zijn echter voorbeelden te bedenken van gevallen waarin oorzaken hun gevolgen alleen maar mogelijk, niet noodzakelijk maken: wanneer een terrorist een bom plaatst waarvan het ontsstekingsmechanisme in werking treedt bij het uiteenvallen van één specifieke atoomkern, en wanneer de bom explodeert, dan zullen we zeggen dat de terrorist de explosie heeft veroorzaakt, ook al was niet zeker dat deze op zou treden. In het algemeen kunnen we zeggen dat het zoeken naar causale wetten hoe dan ook tevergeefs is, omdat elke zogenaamde causale wetmatigheid wel verstoord kan worden door een onvoorziene omstandigheid. En het lijkt een hachelijke zaak, causale wetten te formuleren die immuun zijn voor zulke verstoorders.

Om deze redenen beschouw ik oorzaken als noodzakelijke, niet als voldoende voorwaarden. Wanneer het echter bij deze stelling blijft, schiet de theorie schromelijk tekort. Immers, ze verklaart dan niet wat oorzakelijkheid en natuurlijke wetmatigheid met elkaar te maken hebben. Evenmin kan ze overweg met ogenschijnlijk niet-causale noodzakelijke voorwaarden, bijvoorbeeld het schrijven van tweemaal '1' achtereen om
Deel II. Causaliteit

'lolly' te schrijven. Verder is niet duidelijk gemaakt waarom van oorzaken en gevolgen verondersteld wordt dat ze aangrenzend zijn in ruimte en tijd, of althans met elkaar verbonden via een keten van aangrenzende oorzaken en gevolgen.

Het meest serieuze probleem is misschien wel, dat veel oorzaken helemaal geen noodzakelijke voorwaarden zijn. Als er twee krachtige explosies vlakbij huis plaatsvinden en de ruiten springen, dan lijkt geen van beide explosies afzonderlijk vereist (gegeven de omstandigheden) voor het springen van de ruiten. Hoogstens zijn de explosies als disjunctie vereist (d.w.z. óf de ene, óf de andere); maar het is vreemd om van disjunctieve oorzaken te spreken. We zullen in zo'n geval zeggen dat één van de explosies (de eerste), of het geheel aan explosies (als ze tegelijk plaatsvinden) de oorzaak is. Een op zichzelf staande contrafactische theorie van veroorzaking kan echter niet aangeven waar dit op gebaseerd is.

**Fysische infrastructuur**

Om deze onvolkomenheden van de contrafactische benadering te ondervangen, moeten we haar aanvullen met een mechanistische causaliteitstheorie, één die causaliteit benadert in termen van wat we zouden kunnen noemen een fysische infrastructuur. Inderdaad wordt causaliteit vaak in verband gebracht met kracht, overdracht van energie, en het over tijd blijven voortbestaan en interacteren van objecten. De bekendste hedendaagse exponenten van zo'n benadering zijn Wesley Salmon en Phil Dowe. Zij analyseren causaliteit in termen van zogenaamde 'causale processen,' die gedefinieerd worden als tijd-ruimte-trajecten van fysische grootheden waarvoor een behoudswet geldt, zoals energie en impuls.

Zo'n mechanistische benadering verschilt radicaal van de boven beschreven contrafactische benadering. Want terwijl de laatste zich concentreert op afhankelijkheidsrelaties tussen gebeurtenissen of standen van zaken, concentreert de eerste zich op voortbestaan en interacties tussen objecten. We mogen ons om te beginnen wel verwonderen over het feit dat twee dermate verschillende benaderingen van causaliteit naast elkaar kunnen voortbestaan.

Evengoed kunnen we vaststellen dat beide benaderingen hun tekortkomingen hebben. De bezwaren tegen de contrafactische benadering
Nederlandse samenvatting

zijn al genoemd. Aan de mechanistische benadering kleeft vooral het bezwaar dat de causale relevantie van elke andere eigenschap van een object dan een behouden fysische grootheid ontkend moet worden. Dit is tegen-intuïtief: het twee meter lang zijn van een basketballer lijkt een causaal relevante eigenschap van die persoon te zijn, maar het is geen behouden fysische grootheid. Ook is recentelijk door Jonathan Schaffer het probleem van 'veroorzaking door disconnectie' opgeworpen: een arts kan het sterven van een patiënt veroorzaken door deze van het infuus af te halen, maar dit bestaat er eerder in dat een fysische connectie beëindigd wordt dan dat er een wordt gecreëerd. Wie causaliteit louter begrijpt in termen van tijd-ruimte-trajecten van fysische grootheden zal moeten ontkennen dat hier sprake is van een causale relatie, wat echter vrij absurd lijkt.

EEN GEÎNTEGREERDE THEORIE

Wat ik in deel II bepleit is het samenvoegen van de contrafactische en de mechanistische benadering, en wel op zo'n manier dat de problemen van elke benadering afzonderlijk worden opgeheven. Dit lijkt een vorm van filosofisch opportunisme, maar merk op dat afhankelijkheden tussen standen van zaken en gebeurtenissen enerzijds, en fysische connecties tussen objecten anderzijds ten nauwste gerelateerd zijn. Denk aan een 'stand van zaken' zoals eerder beschreven: in het kort, het hebben van een eigenschap door een een object op een bepaald moment. De wereld op een bepaald moment kan misschien 'mechanistisch' opgevat worden als een systeem van fysische objecten waarvan de ontwikkeling over tijd wordt bepaald door fysische wetmatigheden. Onvermijdelijk mee-gegeven zijn echter allerlei feiten omtrent de objecten: de standen van zaken. En dankzij het feit dat de ontwikkeling van het hele systeem wetmatig verloopt, zullen latere feiten, of standen van zaken, 'contrafactisch' afhangen van eerdere. Contrafactische afhankelijkheid tussen standen van zaken veronderstelt, of vereist, daarom een fysisch-wetmatige infrastructuur tussen de objecten die bij die standen van zaken betrokken zijn. Bedenk ook dat wanneer we 'fysische infrastructuur' en 'contrafactische afhankelijkheid' gezegd hebben, we impliciet 'causaliteit' gezegd hebben. Causaliteit is niet iets anders, iets extra's, ten opzichte van die twee zaken.
Deel II. Causaliteit

Maar hoe moeten we contrafactische afhankelijkheid en fysische infrastructuur in een theorie samenvoegen? Hoe komen we tot een geïntegreerde definitie van veroorzaking? Mijn suggestie is

1) dat elke causale relatie een contrafactische afhankelijkheidsrelatie is, waarbij in gevallen als dat van de twee explosies de afhankelijkheid disjunctief kan zijn (in welk geval met behulp van 2) hieronder uitgemaakt kan worden welke van de disjuncten de oorzaak is);
2) dat, wil een contrafactische afhankelijkheidsrelatie causaal genoemd kunnen worden, zij contrafactisch-afhankelijk moet zijn van een welbepaald stuk fysische infrastructuur. (Is de afhankelijkheid disjunctief, dan kunnen we zien welke van de disjuncten fysisch met het gevolg verbonden is).

Een voorbeeld kan dit wellicht verduidelijken. Stel, iemand kijkt naar een rijpe en radioactieve tomaat, en heeft een visuele impressie van rood. Hij loopt ook oogschade op. In dit voorbeeld is sprake van twee contrafactische afhankelijkheidsrelaties (gegeven omstandigheden): 1) Het rijp zijn van de tomaat is een noodzakelijke voorwaarde voor de visuele impressie van rood; 2) Het radioactief zijn van de tomaat is een noodzakelijke voorwaarde voor de oogschade. Er is ook sprake van twee 'paden' fysische infrastructuur: licht en radioactieve straling. Welk stuk fysische infrastructuur hoort nu bij welke contrafactische afhankelijkheidsrelatie? Het lijkt duidelijk dat het licht bij 1) hoort, en de straling bij 2). Waarom? Het enige zinnige antwoord lijkt hier te zijn, dat het licht (gegeven de omstandigheden) een noodzakelijke voorwaarde is voor 1), en de straling voor 2). Immers, zonder licht maar met straling geen rood-impressie, maar wel oogschade; en andersom met licht maar zonder straling. De notie van contrafactische afhankelijkheid lijkt ons zo een eenvoudige, intuitieve en algemene manier te bieden om de mechanistische en de contrafactische aspecten van veroorzaking te verenigen.

In deel II beargumenteer ik dat met boven beschreven definitie van veroorzaking de problemen van de contrafactische theorie enerzijds, en die van de mechanistische theorie anderzijds, uit de wereld zijn. Ik zal in deze samenvatting niet ingaan op de details van dat argument.

Bovenstaande analyse van veroorzaking is bedoeld om een geloofwaardig criterium te vinden voor 'causale effectiviteit,' dat we
Nederlandse samenvatting

vervolgens kunnen toepassen op mentale eigenschappen. Het criterium dat ik voorstel is het zogenaamde 'criterium van locaal verschil.' Wanneer we inderdaad aannemen dat een causale relatie een infrastructuur vereist die bestaat uit fysische signalen, dan vereist een oorzaak een meetbaar verschil aan tenminste een van de objecten die betrokken zijn bij het vermeende effect.

Hier is een voorbeeld van een afhankelijkheidsrelatie die niet causaal is volgens dit criterium: door met iemand te trouwen kan ik een bepaalde persoon die zich onbereikbaar ver weg bevindt tot zwager maken. Mijn trouwen heeft het zwager-zijn van die persoon misschien 'teweeggebracht,' maar niet veroorzaakt, omdat locaal geen verschil aan die persoon te meten is ten opzichte van een hypothetische situatie waarin ik niet getrouwd was. Wanneer ik een persoon door te spreken op andere gedachten breng is er wel sprake van causale effectiviteit: een locaal verschil is ongetwijfeld meetbaar bij, bijvoorbeeld, de trommelvliezen van de betreffende persoon. Het is uiteraard aan de criticus om uit te maken of dit bevredigende resultaten van het criterium zijn.

Een subtiel maar belangrijk punt hier is dat, wanneer we de doctrine van 'causale geslotenheid' (zie boven) aanhangen, moeten zeggen dat het 'locale verschil' een complete fysische oorzaak heeft, en dus zelf niet veroorzaakt kan zijn door een hoger-niveau-feit zoals spreken. Het spreken kan evengoed wel causaal effectief zijn: als zijn 'eigen' gevolg heeft het een mentale verandering bij de gesprekspartner (een hoger-niveau-feit). Maar in zoverre als spreken iets anders is dan lucht in beweging zetten kan het, strikt genomen, geen trillingen van een trommelvlies veroorzaken (een louter fysisch feit). Mijn criterium berust echter op het feit dat er evengoed (gegeven de omstandigheden) zonder spreken geen (of andere) trillingen waren geweest. De trillingen zijn immers deel van de fysische infrastructuur van de causale relatie spreken - mentale verandering.

WAARNEMING VAN CAUSALITEIT

Kunnen we causaliteit waarnemen? Om kennis te hebben van een relatie van contrafactische afhankelijkheid moeten we een actuele met een denkbeeldige, louter mogelijke situatie vergelijken (hoe de zaken ervoor hadden gestaan wanneer de veronderstelde oorzaak niet had plaatsgevonden), maar een louter denkbeeldige situatie kunnen we niet
Deel II. Causaliteit

waarnemen. Toch lijkt het alsof we wel degelijk direct kunnen waarnemen, bijvoorbeeld, dat dingen botsen, dat een dier iets eet, of dat iemand gekwetst raakt. En dit zijn causale relaties.

Ik ga in op deze kwestie, omdat een antwoord op de vraag of, en zo ja hoe, we causale relaties kunnen waarnemen, het ons verderop mogelijk maakt te verklaren hoe we de normatieve significantie van een situatie kunnen waarnemen (cruciaal voor het derde probleem van mentale veroorzaking). Het antwoord vinden we bij John Searle. Searle heeft betoogd dat we nooit domweg dingen waarnemen, maar de dingencds zus-of-zo. We kunnen ook zeggen dat neutrale waarneming niet bestaat, omdat waarneming (het oppikken van informatie door een organisme) altijd gekoppeld is aan de biologische belangen van een soort.

Het is nu plausibel om te veronderstellen dat evenals onze soort het vermogen heeft ontwikkeld om, bijvoorbeeld, bepaalde dieren te herkennen als gevaarlijk, zij ook heeft geleerd om de opeenvolging van bepaalde gebeurtenissen te herkennen als causale afhankelijkheid (namelijk, omdat dit ingrijpen mogelijk maakt, met alle voordelen van dien). De causaliteit is hier geen object, maar inhoud van waarneming. Ons vermogen om direct causaliteit waar te nemen is dus een vorm van ons biologisch voorgeprogrammeerd zijn om (in een normale omgeving) situaties als zus-of-zo waar te nemen.

DE ONTOLOGIE VAN OORZAKEN EN GEVOLGEN

Een belangrijk punt in het debat over mentale veroorzaking is de vraag wat voor entiteiten (als dat het goede woord is) oorzaken en gevolgen nu eigenlijk zijn. Waar het in deze discussie om gaat is de precieze aard van wat we ‘gebeurtenissen’ of ‘toestanden’ noemen. Immers, we moeten zien uit te maken in welk opzicht mentale gebeurtenissen of toestanden oorzaken en gevolgen zijn, en of (en zo ja op welke manier) mentale eigenschappen hierbij betrokken zijn. Wanneer we bijvoorbeeld kunnen zeggen dat mentale oorzaken identiek zijn met fysische oorzaken, dan is er voor mentale eigenschappen vrijwel zeker geen causale rol weggelegd.

Er lijken nu ruwweg twee opvattingen te bestaan over gebeurtenissen: de ene opvatting ziet gebeurtenissen als geconstitueerd door eigenschappen. Volgens deze opvatting zijn mentale gebeurtenissen niet-identiek met fysische wanneer de constituerende eigenschappen ervan andere zijn dan
die van fysische gebeurtenissen; dat wil zeggen, wanneer mentale eigenschappen niet te reduceren zijn tot louter fysische. Dit is een opvatting van gebeurtenissen als ondersoort van de 'standen van zaken' uit deel I; een opvatting die ikzelf ondersteun, in het voetspoor van auteurs als David Armstrong en Jaegwon Kim.

De andere opvatting van gebeurtenissen is verdedigd door Donald Davidson. Volgens deze opvatting zijn gebeurtenissen concrete fysische entiteiten, zoals objecten, en zijn ze slechts *dragers* van eigenschappen. Davidson's critici hebben aannemelijk gemaakt dat zijn opvatting van gebeurtenissen alleen fysische, geen mentale veroorzaking toelaat. De opvatting lijkt echter ondersteund te worden door de manier waarop wij in de taal naar gebeurtenissen verwijzen: wanneer we het over 'de oorlog,' 'het feest,' etc. hebben, lijken we eerder naar concrete entiteiten te verwijzen dan naar eigenschappen van objecten.

Ik beargumenteer nu 1) dat de Davidsoniaanse opvatting van gebeurtenissen op verschillende punten niet strookt met breed ondersteunde aannamen betreffende oorzakelijkheid (met name de centrale rol van eigenschappen hierin); en 2) dat we het zojuist genoemde punt, over onze verwijzing naar gebeurtenissen, ook, en beter, kunnen verklaren door aan te nemen dat gebeurtenissen niet tijd-ruimtelijk, maar narratief worden afgebakend.

**RELATIONELE FEITEN ALS OORZAKEN**

Dit maakt het ons mogelijk om vol te houden dat oorzaken en gevolgen het ontologische karakter hebben van 'standen van zaken,' en dat eigenschappen (zoals mentale eigenschappen) constitutief kunnen zijn voor oorzaken. Een volgende vraag is dan of relaties, of relationele eigenschappen ook een dergelijke constitutieve rol kunnen vervullen. Dit wordt vaak ontkend, daar veel relaties niet locaal zijn, terwijl causale invloed (denk aan boven beschreven mechanismistische kijk op veroorzaking) wel locaal is. Ik beargumenteer echter dat tijd-ruimtelijke en causale relaties, en standen van zaken die geconstitueerd worden door zulke relaties (dat wil zeggen, zo ongeveer alle hoger-niveau-standen van zaken) wel degelijk constituerend voor oorzaken, ofwel causal effectief, kunnen zijn.
Deel II. Causaliteit

Ik doe dit door het voorbeeld van de eigenschap *geldelijke waarde* te bespreken, een relationele eigenschap van bijvoorbeeld muntstukken. Deze eigenschap is door Fred Dretske bestempeld als causaal ineffectief. Dit lijkt aannemelijk genoeg: zijn het niet de fysische eigenschappen (gewicht, vorm) van zulke munten die bijvoorbeeld cola-automaten in werking zetten, in plaats van de weinig tastbare relaties van die munten met de nationale bank, of het wettige fabricageproces ervan in het verleden?

Toch deugt deze redenering niet. Ten eerste: van het hebben van geldelijke waarde kunnen we volhouden dat het zijn ‘eigen’ hoger-niveau-effect heeft: een verandering in eigendomsrelatie betreffende een stuk handelswaar. En op grond van onze ontologie van gebeurtenissen kunnen we aannemelijk maken dat die verandering van eigendom (een sociaal-relationeel feit) niet identiek kan zijn met het vallen van een blikje cola (een locaal fysisch feit).

Ten tweede: de geldelijke waarde van de munten die we inwerpen maakt wel degelijk locaal verschil: het blikje was niet gevallen wanneer we niet het juiste bedrag hadden ingeworpen. Hiermee voldoet de eigenschap geldelijke waarde aan het criterium van locaal verschil.

Het is waar dat we met valse munten bij cola-automaten hetzelfde kunnen bereiken als met echte. Dit punt is hier echter irrelevant, want het vals-zijn van een munt is evengoed een relationele eigenschap als het echt-zijn. Het is ook waar dat toevallig gevonden platte voorwerpen, of buitenlandse muntstukken, soms hetzelfde resultaat geven als geldige munten. Dit doet echter niets af aan de causale effectiviteit van geldelijke waarde in gevallen waarin toevallig werkende vervangers voor echte munten ontbreken; en dat zijn verreweg de meeste gevallen.

Het belang van dit voorbeeld is, dat geldelijke waarde een historisch-relationele eigenschap is: in deel III beargumenteer ik dat ook mentale eigenschappen als zodanig begrepen moeten worden. Een argument voor de causale effectiviteit van geldelijke waarde is zo impliciet een argument voor de causale effectiviteit van mentale eigenschappen.

Mijn verklaring voor de causale effectiviteit van historisch-relationele eigenschappen is de volgende. 1) Veroorzaking vindt altijd plaats bij de gratie van omstandigheden. (Merk op dat dit boven een reden was om het bestaan van causale wetten te ontkennen). 2) Die omstandigheden zijn het gevolg van gebeurtenissen uit het verleden. 3) Die gebeurtenissen hebben zo de huidige causale structuur van de wereld mee vorm gegeven (d.w.z., de omstandigheden waaronder specifieke feiten specifieke andere feiten
Nederlandse samenvatting

veroorzaken), en kunnen daarom 'structurerende oorzaken' genoemd worden (de term is van Dretske).

4) Een relatie van een object met zo'n structurerende oorzaak in het verleden bepaalt mede welke feiten aangaande dat object in het heden wel en niet veroorzaakt. Mist een muntstuk bijvoorbeeld de juiste relatie met de structurerende oorzaak die het geldig verklaren ervan is, dan zal het (anders dan bij toeval) cola-automen niet in werking kunnen zetten; aangezien cola-automen zo gemaakt zijn dat ze doorgaans alleen op geldige muntstukken reageren. 5) Historisch-relationele eigenschappen als geldelijke waarde, en mentale eigenschappen, bestaan precies in zo'n relatie met een structurerende oorzaak. Ze zijn daarmee causaal effectief.

Deel III. Het mentale

Wij hebben hierboven al kennis gemaakt met de 'standaard' opvatting van mentale toestanden; ook heb ik al geschetst welke alternatieve opvatting ik hiervoor in de plaats wil stellen. In deel III voer ik een aantal argumenten aan tegen verschillende aspecten van de standaard opvatting, en probeer ik te laten zien, vanuit bestaand werk in de philosophy of mind, dat er goede alternatieven beschikbaar zijn. De verschillende aspecten die ik bekritiseer zijn de volgende:

A. Computationalisme

Dit is de computermetafoor van de menselijke geest, waar we al mee kennisgemaakt hebben. Het basisidee is dat een denkend organisme via de zintuigen informatie ontvangt, deze causale input verwerkt, en zo de juiste causale output produceert. Hierbij wordt de input in de hersenen omgezet in betekenisvolle 'symbolen.' Deze beantwoorden aan een syntaxis: afhankelijk van welke theorie men aanhangt wordt deze syntaxis voorgesteld als analoog aan die van natuurlijke taal, logica, of wiskunde. Van deze innerlijke syntaxis wordt dan doorgaans verondersteld dat die het resultaat is van evolutie.

Een probleem met deze opvatting is dat ze zich concentreert op cognitie die te vertalen valt in redeneerprocessen, terwijl ze weinig te zeggen heeft over niet-conceptuele cognitie, bijvoorbeeld over de rol van emotie in onze
Deel III. Het mentale

mentale processen. En zelfs wat conceptuele cognitie aangaat is de opvatting niet zonder problemen: enkele decennia van onderzoek in kunstmatige intelligentie hebben bij velen het vermoeden gewekt dat menselijk denken en gedrag, dat flexibel en contextgevoelig is, en onder tijdsdruk werkt, zich van nature niet leent voor implementatie in een ‘klassieke’ (d.w.z. symboolverwerkende) computer of robot.

Een belangrijke consequentie is, dat wanneer we ons afvragen hoe mentale processen fysisch gerealiseerd zijn, we niet moeten zoeken naar causale processen in de hersenen die syntactische operaties implementeren, of naar signalen in de hersenen die ‘symbolen’ zouden kunnen zijn. Het is van belang om dit op te merken, omdat het de these van ‘token’ fysicalisme (zie onder) van een belangrijk argument berooft.

B. DE THEORY THEORY

Wanneer we het hebben over ‘mentale toestanden,’ dan bedoelen we doorgaans zaken als overtuigingen en verlangens. Spreken hierover, en elkaars gedrag in termen ervan verklaren (folk psychology), lijkt de veronderstelling te vereisen dat ze echt bestaan. De vraag is echter wat dit inhoudt. De theorie die we al eerder tegenkwamen als ‘functionalisme’ vat onze folk psychology op als een proto-wetenschappelijke theorie (ter voorspelling en verklaring van elkaars gedrag) met causale ‘wetten’ als: ‘Wanneer persoon x verlangt dat q en gelooft dat p niet-q impliceert, dan zal hij ceteris paribus trachten niet-p tweedeg te brengen.’ We noemen dit de ‘Theory Theory.’ Aanhangers verschillen van mening over de juistheid van de ‘wetten’ van onze alledaagse folk psychology. Doorgaans meent men dat een toekomstige neurowetenschap ze in elk geval voor een deel wetenschappelijk respectabel zal maken; soms wordt beweerd (bijvoorbeeld door Paul Churchland) dat onze folk psychology als theorie niet te redden zal zijn.

Een alternatieve theorie, simulationisme, suggereert dat de Theory Theory fundamenteel onjuist is. Onze wederzijdse pogingen om elkaar te begrijpen hebben volgens deze theorie meer het karakter van inleven dan van theoretiëren. Hiertoe simuleren we dan andermans mentale toestand bij onszelf. Het debat tussen aanhangers van de Theory Theory en van het simulationisme is onbeslist; wel lijkt echter duidelijk dat de Theory Theory slechts te verdedigen is met een nogal ruime opvatting van ‘theorie.’ In elk
Nederlandse samenvatting

geval, ook wanneer onze praktijk van *folk psychology* een element van
theorie in zich heeft, lijkt dit niet het soort theoretische kennis en inzicht
dat verworpen of bevestigd kan worden door neurowetenschap.

Wat betreft de aard van mentale toestanden zoals overtuigingen neem
ik de suggestie van Colin McGinn over dat mentale toestanden zoals
overtuigingen geen symbolische structuren zijn (zoals het
computationalisme veronderstelt), maar aspecten van een model. Laten we
hierbij bedenken dat zo'n model geen innerlijke structuur in onze hersenen
is, maar een activiteit, een vorm van informatieverwerkend gedrag.
Volgens McGinn hoeft iemand, wanneer hij gelooft dat *p*, niet de zin, of
propositie *p* in het hoofd te hebben. De persoon modellerst, of
representeert de wereld slechts op een zodanige manier, dat een aspect van
dit model beschreven kan worden met de zin *p*.

C. INTERNALISME

De standaard opvatting van mentale toestanden is, dat deze interne
toestanden van een organisme, ofwel hersentoestanden zijn. In de
hedendaagse *philosophy of mind* wordt deze veronderstelling echter op
verschillende manieren ondergraven. Het resultaat is dan externalisme, de
opvatting dat mentale toestanden mede geconstitueerd worden door
factoren buiten het denkende organisme. Het bekendste externalistische
argument is afkomstig van Hillary Putnam, die heeft betoogd dat wanneer
we aan natuurlijke soorten denken, zoals water, het al of niet juist zijn van
zulke gedachten wordt bepaald door de gebruiksgeschiedenis van de term
‘water.’ Voorts heeft Tyler Burge betoogd dat bepaalde soorten gedachten
een sociale context vereisen, namelijk gedachten die begrippen bevatten
waarvoor de toepassingscriteria vastgesteld zijn door experts (zoals
‘artritis’).

Zelf meen ik dat het opvatten van mentale categorieën, in navolging
van Millikan, als functiecategorieën (zie boven) een veel omvattender en
doorslaggevender argument is voor externalisme. Het hebben van een
mentale eigenschap bestaat dan in het gerelateerd zijn aan een
ontwikkelings- en leergeschiedenis, en via die geschiedenis aan normale
leefomstandigheden in het heden (wat hier en nu aan omstandigheden
‘normaal’ is wordt dan bepaald door wat de omstandigheden waren waarin
we onze functiekenmerken ontwikkelden of aanpasten).
Deel III. Het mentale

Nu wordt binnen het functionalisme wel de opvatting van ‘zwak externalisme’ aangehangen, de opvatting dat mentale toestanden wel interne hersentoestanden zijn, maar dat zij hun betekenis-inhoud ontlenen aan relaties met een externe context. Ik betoog dat deze opvatting ongeloofwaardig wordt in het licht van argumenten (bijvoorbeeld van Andy Clark) die aantonen dat veel mentale processen causale ‘lussen’ buiten het lichaam om vereisen, zoals het uitvoeren van een moeilijke berekening met behulp van pen en papier. Externalisme betreft dan niet slechts mentale inhoud, maar evengoed mentale processen.

Nog een vorm van externalisme die ik verdedig, in navolging van Frederick Stoutland, betreft redenen. We kunnen een reden die iemand heeft om F te doen opvatten als het verlangen om F te doen, of de overtuiging dat F goed is. Maar misschien is meer realisme gerechtvaardigd ten opzichte van goed en slecht, gepast en ongepast, mooi en lelijk, etc. We kunnen dan beweren dat situaties normatieve aspecten hebben, in zeker opzicht buiten ons om, die we al of niet kunnen onderkennen (denk aan verkeerstekens). Redenen zitten dan niet slechts in ons hoofd, maar zijn aanwezig in de wereld buiten ons. Millikans opvatting van functies (en daarmee van normativiteit) ondersteunt deze opvatting. Normativiteit wordt dan een historisch-relationele aangelegenheid.

D. REDUCTIONISME

De twee resterende aspecten van de standaard opvatting die ik bekritiseer betreffen de manier waarop mentale toestanden fysisch gerealiseerd zijn. Volgens reductionistische theorieën kunnen mentale feiten begrepen worden als fysische feiten, namelijk doordat mentale eigenschappen beschouwd worden als in principe reduceerbaar tot fysische.

Ik bespreek twee reductiemodellen in detail: dat van Ernest Nagel en van Jaegwon Kim. Ook bespreek ik het eliminativisme van Paul Churchland, dat we kunnen opvatten als een 'teleurgesteld' reductionisme. Ik betoog dat psycho-fysische reductie volgens de modellen van Nagel en Kim om verschillende redenen weinig kans van slagen heeft, en dat dit vooral een gevolg is van het relationele karakter van mentale eigenschappen. We hebben echter eerder gezien dat deze relationistische vorm van antireductionisme zeer wel verenigbaar is met fysicalisme.
Nederlandse samenvatting

Tenslotte maak ik een onderscheid tussen reductionisme als succesvol gebleken wetenschappelijke onderzoeksstrategie (simpelweg: dingen uit elkaar halen om te zien hoe ze werken) en als metafysische doctrine. Tegen het eerste hoeft dan weinig bezwaar te worden gemaakt; wel echter tegen het tweede.

E. TOKEN FYSICALISME

De afgelopen tientallen jaren is het meest in het oog springende obstakel voor reductionisme geweest de ‘meervoudige realiseerbaarheid’ van mentale toestanden. Pijn, bijvoorbeeld, lijkt bij zeer ongelijksoortige organismen voor te komen. Hieruit hebben veel filosofen geconcludeerd dat typen mentale toestanden niet identiek kunnen zijn met typen fysische toestanden. Evengoed meent men vaak dat individuele mentale toestanden of gebeurtenissen wel identiek zouden kunnen zijn met individuele fysische toestanden (tokens); vandaar token fysicalisme. Vanuit deze positie is moeilijk te begrijpen hoe mentale eigenschappen ooit caucaal effectief kunnen zijn, daar van elke fysische toestand de fysische eigenschappen alle vereiste ‘causaal werk’ doen.

Deze opvatting lijkt echter geen stand te houden wanneer we serieus de vraag opwerpen wat toestanden en gebeurtenissen nu precies zijn. Met deze vraag komen we weer aan bij de boven beschreven discussie over de ontologie van oorzaken en gevolgen. Volgens de eerder verdedigde opvatting van ‘standen van zaken’ uit deel I worden toestanden en gebeurtenissen geïndividueerd door eigenschappen, en is zodoende token identiteit alleen te verdedigen via type identiteit, dus psycho-fysische reductie. Wanneer dit laatste niet haalbaar is, is token fysicalisme het ook niet.

Afgezien van dit ontologische punt lijkt het hoe dan ook een hachelijke zaak om waarnemingen, gedachten en handelingen te beschrijven in fysische termen. De voornaamste moeilijkheid is, dat er geen criteria voorhanden zijn om de ene fysische afbakening te prefereren boven de andere. Waar in de hersenen begint en eindigt een overtuiging-dat-p? Voorzover dit al een zinnige vraag is, lijkt het antwoord erop alleen maar arbitrair te kunnen zijn, ongeacht de stand van onze neurologische kennis.
Mentale veroorzaking

FENOMENAAAL BEWUSTZIJN

Hierboven is de kwestie van fenomenaal bewustzijn, dat het derde probleem van mentale veroorzaking vormt, niet aan bod geweest. Binnen de standaardopvatting wordt vaak beweerd dat fenomenaal bewustzijn het laatst overgebleven mysterie in de *philosophy of mind* is. Dit is gebaseerd op een dichotomie van enerzijds louter 'functionele' mentale eigenschappen (waaronder geloven-dat en verlangen-dat) en anderzijds 'fenomenale' eigenschappen (hoe dingen smaken, ruiken, voelen, etc.). Van de eerste wordt verondersteld dat ze kunnen worden beschreven in termen van hun causale rol, en zo worden opgevat als hersentoestanden. De laatste worden geheel gekenmerkt door hun subjectieve kwaliteit, en worden 'qualia' genoemd. Men redeneert dat het denkbaar is dat een wezen onze 'functionele' eigenschappen met ons gemeen heeft, maar niet onze qualia. Hiermee wordt de causale effectiviteit van fenomenaal bewustzijn *a priori* uitgesloten, en ook elke mogelijkheid om er wetenschappelijk grip op te krijgen.

Nu moeten we inderdaad vaststellen dat we het verschijnsel subjectief bewustzijn nog niet goed begrijpen. Dit maakt het echter riskant om er nu al een filosofisch oordeel over te vellen. De dichotomie van 'functioneel' versus 'fenomenaal' is bijvoorbeeld nogal twijfelachtig. Er bestaan ook theorieën volgens welke subjectieve ervaringen een soort representaties zijn. Wanneer dit juist is, vormt fenomenaal bewustzijn geen extra probleem ten opzichte van de andere problemen van mentale veroorzaking. Het zal duidelijk zijn dat, zo lang we het bewustzijn zo slecht begrijpen, een oplossing voor het derde probleem van mentale veroorzaking buiten ons bereik ligt.

**Mentale veroorzaking**

Laten we terugkeren naar de vijf problemen van mentale veroorzaking waarmee we begonnen zijn, en zien wat precies bereikt is.
1) HET PROBLEEM VAN MENTAAL ANOMALISME

We hebben in het voorgaande geen enkele reden gevonden om aan te nemen dat er een causaal-wetmatige relatie bestaat tussen denken en handelen van een actor. Evengoed suggereert mijn analyse van veroorzaaking dat causaliteit inderdaad natuurlijke wetmatigheid vereist. Is mentaal anomalisme daarmee een probleem?

Het antwoord blijkt nee te zijn. Dat er geen causale wetten bestaan ten aanzien van mentale feiten is niets uitzonderlijks: we hebben gezien dat veroorzaaking in de allereerste plaats geen kwestie is van causale wetten, d.w.z., wetten van causale opeenvolging. Oorzakelijkheid vereist wel degelijk een strikte fysische wetmatigheid, maar deze geldt voor de fysische infrastructuur (in en tussen objecten) die noodzakelijk is voor causale afhankelijkheden (tussen feiten). De laatste beantwoorden niet aan wetten, omdat ze afhankelijk zijn van incidentele omstandigheden. Mentaal anomalisme en het feit dat veroorzaaking (fysische) wetmatigheid veronderstelt zijn dus zonder meer met elkaar verenigbaar.

2) HET PROBLEEM VAN NORMATIVITEIT

Hoe kunnen redenen oorzaken zijn, d.w.z., hoe kunnen de normatieve aspecten van een situatie ons aanzetten tot handelen? Ik heb deze vraag langs een omweg proberen te beantwoorden door, in navolging van Searle, te beweren dat we causale relaties als zodanig kunnen waarnemen. Dit op grond van het feit dat de causaliteit geen object, maar inhoud van onze waarneming is (iets wat met eigenschappen en waarnemen-al altijd zo is).

Hiermee is verklaard hoe we een situatie als zus-of-zo kunnen waarnemen, waarbij dit 'zus-of-zo' ook op normatieve eigenschappen kan slaan. Althans, ik heb in navolging van Stoutland gepleit voor 'externe redenen,' ofwel het objectief aanwezig zijn van normatieve significantie (zoals gevaar, onrecht, gezondheid, etc.) in onze omgeving. Het gaat hier om historisch-relationele eigenschappen, d.w.z. kenmerken van situaties die te maken hebben met onze ontwikkelings- en leergeschiedenis. Wanneer er inderdaad sprake is van 'externe redenen,' en wanneer de inhoud van onze waarnemingen kunnen zijn (en het is aannemelijk dat levende wezens situaties allereerst leren onderscheiden op hun normatief significante kenmerken) kunnen ze vervolgens, evenals willekeurig welke
Mentale veroorzaking

andere waarnemingsinhoud, gevolgen hebben voor hoe wij, de waarnemers, handelen.

3) HET PROBLEEM VAN FENOMENALAAL BEWUSTZIJN

Hoe kunnen onze subjectieve ervaringen met hun 'kwalitatieve' aspect causal nog iets toevoegen aan de 'blinde' biochemische processen in ons lichaam? Een serieuze poging om dit probleem op te lossen is pas zinnig wanneer we een volgroeide wetenschappelijke theorie over het bewustzijn hebben. Maar dit is duidelijk nog lang niet het geval.

Het derde probleem van mentale veroorzaking berust echter op een aanvechtbare dichotomie tussen 'functioneel' en 'fenomenaal.' Voorzover deze dichotomie ongegrond is, kunnen we volhouden dat fenomenaal bewustzijn een intentioneel verschijnsel is naast andere, en daarmee dat er met betrekking tot mentale veroorzaking waarschijnlijk geen apart probleem van fenomenaal bewustzijn bestaat.

4) HET PROBLEEM VAN EXTRINSIEKE MENTALE EIGENSCHAPPEN

Hoe kunnen mentale feiten geconstitueerd worden door relaties (bijvoorbeeld met eerdere gebeurtenissen, of met zaken op grote afstand), en evengoed causal effectief zijn, gezien het lokale karakter van causale invloed, ofwel de onmogelijkheid van directe 'werking op afstand'?

Volgens de radicaal externalistische theorie over het mentale die ik aanhang als alternatief voor de standaard theorie (die mentale toestanden begrijpt als hersentoestanden), en volgens mijn 'relationistische' theorie over fysische realisatie, zijn mentale feiten inderdaad relationele feiten. Echter, zoals we zagen kan op basis van mijn analyse van veroorzaking beargumenteerd worden dat relaties causaal effectief kunnen zijn.

Ten eerste hebben we gezien dat behalve intrinsieke eigenschappen ook relaties causal gerelateerd kunnen zijn. Ten tweede hebben we kunnen vaststellen dat in ieder geval tijd-ruimtelijke en causale relaties causaal nonredundant (dus echte noodzakelijke voorwaarden) kunnen zijn, en voldoen aan het 'criterium van locaal verschil' (en dus een fysische infrastuctuur vereisen). Ten derde hebben we gezien hoe, met behulp van het begrip 'structurerende oorzaak,' begrijpelijk kan worden gemaakt hoe causale relaties in het verleden effecten hebben in het heden.
HET PROBLEEM VAN CAUSALE UITSLUITING

Hoe kunnen, met fysicalisme als aannname, mentale feiten causaal effectief zijn gegeven:

a. Causale geslotenheid van het fysische domein, en
b. Uitsluiting van meerdere complete oorzaken?

Immers, volgens deze aannamen hebben de realiseerders van alle feiten een complete fysische oorzaak, kunnen ze niet meer dan één zo'n oorzaak hebben, en zijn alle hoger-niveau-feiten automatisch met deze realiseerders mee-gegeven, en dus ook mee-veroorzaakt. Wat voor 'causaal werk' is er voor mentale feiten over?

Het antwoord op deze vraag begint bij de vaststelling dat fysische realisatie geen één-op-één, maar een veel-op-veel-relatie is. Vanwege het relationele karakter van hoger-niveau-feiten, waaronder mentale feiten, kunnen de laatste niet beschouwd worden als identiek met welk fysisch feit, of verzameling van zulke feiten, dan ook. We hebben gezien dat noch doctrines van type-identiteit (reductionisme), noch van token-identiteit, houdbaar zijn. Dit suggereert dat gedachten en handelingen (als hoger-niveau-feiten) hun 'eigen' oorzaken en gevolgen kunnen hebben.

Het zij toegegeven dat de totaliteit van alle voorafgaande fysische feiten, en van hun huidige gevolgen, een voldoende voorwaarde vormt voor welk feit dan ook (dit volgt uit de aanname van fysicalisme). Wanneer we de totaliteit aan fysische oorzaken van de realiseerder van bijvoorbeeld een handeling h zouden beschouwen als de oorzaak van h, dan zou bijvoorbeeld de beslissing die aan h voorafging causaal overbodig zijn.

Echter, de veronderstelde oorzaak zou hier niet slechts de oorzaak van h zijn, maar van talloze extra feiten. En het ontbreken van (type- of token-) identiteitsrelaties tussen h en specifieke fysische (verzamelingen) feiten (of
Mentale veroorzaking

toestanden, of gebeurtenissen) verhinderd dat een fysische oorzaak af te bakenen valt van h (dat wil zeggen: h in het bijzonder, of: h en niets anders dan h, of: specifiek h). Handeling h-in-het-bijzonder heeft wel degelijk zijn eigen oorzaak, maar dat is een mentale, geen louter fysische oorzaak. De handeling in het bijzonder kan slechts causaal verklaard worden door de voorafgaande beslissing als oorzaak te noemen. Noemen we louter fysische feiten, dan noemen we of teveel, of te weinig factoren. Dit suggereert dat h zijn eigen hoger-niveau-oorzaak heeft, een oorzaak die niet wordt ondervangen door louter fysische feiten.

Met deze theorie moeten we echter ontkennen dat mentale feiten 'neerwaarts' fysische feiten veroorzaken. Wanneer we bijvoorbeeld beslissen een steen in het water te gooien en dit vervolgens doen, lijkt onze beslissing de oorzaak van de kringen in het water. De doctrine van causale geslotenheid suggereert echter dat dit strikt genomen niet juist kan zijn: die kringen hebben een complete fysische oorzaak (waaronder biochemische processen in ons lichaam, maar geen beslissing), en niet meer dan één. Dit lijkt echter tekort te doen aan de causale rol van de beslissing.

We kunnen dit rechtzetten door te stellen dat de beslissing een beperking inhoudt van de verzameling mogelijke fysische oorzaken van de kringen in het water. Het 'causale werk' is inderdaad geheel dat van louter fysische processen, maar het wordt 'benut' door een systeem dat in de allereerste plaats een organisatie is van zulke fysische processen. Zolang we mentale veroorzaking beschouwen als komend vanuit een centraal punt (een Cartesiaans zelf, of de hersenen) zullen we deze processen zien als extern aan onszelf. Wanneer we ons 'zelf' echter beschouwen, niet als een controlecentrum, maar als een belichaamd, zelf-organiserend en (ecologisch en sociaal) ingebed organisme, kunnen we deze processen zien als evengoed een deel van onszelf als onze gedachten.
Hierbij wil ik u van harte uitnodigen tot het bijwonen van mijn promotie.

Titel van het proefschrift:

Dependencies, connections, and other relations. A theory of mental causation.

Plaats:
Aula Katholieke Universiteit Brabant
Warandelaan 2, Tilburg, gebouw A

Tijd:
Vrijdag 18 oktober 2002, 10.15 uur
Aansluitend receptie

Paranimfen:
Robert-Jan Breeman
Henk van der Velden

Wim de Muijnck

Koningin Wilhelminaalaan 4-114
4205 ET Gorinchem
cbrm

Natis
Dependencies, connections, and other relations. A theory of mental causation explores the philosophical problems raised by the commonsensical assumption that when we do things for reasons, our beliefs, desires, expectations, fears, etc. are causes of our behaviour. How can that assumption be true? Is our body not already moved by nerve impulses and muscle contractions, and do these not have their own 'physical' causes? What difference can be made by what is on our minds?

Yet, denying mental causation invalidates core notions about our existence as perceiving, thinking, and acting subjects. Notions such as will, freedom, autonomy, and responsibility all presuppose mental causation. So it quite unsettling that in contemporary analytic philosophy of mind we find widespread doubts about the very possibility of mental causation. In spite of the rise of disciplines such as artificial intelligence and neuroscience, the philosophical debate on this issue has remained very much inconclusive.

Dependencies, connections, and other relations. A theory of mental causation questions a number of assumptions that have so far shaped this debate, and that have delimited its range of outcomes. In doing so it covers, in its subsequent parts, ontology, the metaphysics of causation, and the philosophy of mind. It provides a firm theoretical basis for believing that in our all-physical world mental causation is perfectly real, and that it can be understood.