Mental Causation and the Metaphysics of Action

Andrea Suzanne White

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion and History of Science

February 2018

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I owe greatest thanks to my supervisor Helen Steward. I have greatly enjoyed my PhD and this is, in no small way, down to Helen. My research has benefited enormously from her astute criticism and sage advice.

I would also like to thank the University of Leeds for granting me the 110 Anniversary Scholarship and making this research possible.

I am also grateful to the organisers of the 2016 Warwick MindGrad Conference for an exciting and interesting conference. I am especially grateful to Johannas Roessler for providing extremely helpful feedback the paper I presented there. I also wish to thank the organisers of the 21st Oxford Graduate Philosophy Conference for their excellent conference. I particularly want to thank Ursula Coope for the constructive comments she offered in response to my paper, and for her encouragement.

Warm thanks go to David Hillel Ruben for running a weekly metaphysics seminar in London. Many of the philosophical works which have influenced this dissertation were brought to my attention by this stimulating seminar.

I also wish to thank the many friends and colleagues at Leeds and elsewhere who provided comments and suggestions on draft chapters of this dissertation.

I am greatly indebted to Emily Paul and Pete Blockley for so frequently granting me use of their spare room! I am sure I would not have managed my many trips between London and Leeds without their kindness and generosity.

Heartfelt thanks go to Dan Hinge not only for proofreading my dissertation, but for his inexhaustible patience and unfailing support.

Abstract

The subject of this dissertation is the problem of mental causation: the problem of how the mental is able to causally interact with the physical. I show that the problem of mental causation, as it is presented in contemporary philosophy of mind, is a pseudo-problem. My claim is that contemporary philosophy of mind has misidentified what it is about mental causation that we need, but struggle, to understand. This is because contemporary philosophy of mind labours under a misapprehension of what mental causation is supposed to be.

In most discussions of the problem of mental causation, mental causation is presented as a cause-effect relation between mental and physical entities. I call this understanding of mental causation the relational understanding of mental causation:

> **Relational understanding of mental causation**: mental causation is a matter of mental items (events, processes or states) standing in causal relations to physical events, e.g. bodily movements.

The relational understanding of mental causation is widely endorsed largely because it is thought essential to our conception of ourselves as agents who act intentionally and who bear moral responsibility. I argue that while intentional action does entail the existence of causation which involves mentality – something which is worthy of the name 'mental causation' – the mental causation intentional action presupposes ought not to be understood in relational terms. When we say that someone acted intentionally because of what she believed, desired or intended, the concepts *belief*, *desire* and *intention* do not refer to items which stand in causal relations to bodily movements. I will defend this thesis by examining metaphysics of action and the nature of agency.

Contents

Table of Figures

1

The Problem of Mental Causation

1.1 The Relational Understanding of Mental Causation

The problem of mental causation is the problem of how that which is mental is able to causally interact with that which is physical. In this dissertation, I will show that the problem of mental causation, as it is presented in contemporary philosophy of mind, is a pseudo-problem. This is not to say that there aren't aspects of mental causation which are very hard to understand. Rather, my claim is that contemporary philosophy of mind has misidentified what it is about mental causation that we need, but struggle, to understand. This is because contemporary philosophy of mind labours under a misapprehension of what mental causation is supposed to be, or so I contend.

The problem of mental causation is usually presented as a 'how possibly' question, that is, a question about how mental causation could exist. It is usually accepted as *prima facie* true that mental causation does exist; the difficulties arise only when we try to understand how this could be, given certain assumptions about the fundamental nature of reality. As Peter Menzies states, 'philosophical questions about mental causation revolve around […] how it is possible in the first place in the light of certain metaphysical assumptions and principles' (2013, p.58). The metaphysical assumptions and principles which seem to make the existence of mental causation puzzling concern the apparent physicality of the causal world. For example, Jaegwon Kim tells us that the problem of mental causation is 'to explain how mentality can have a causal role in a world that is fundamentally physical' (2005, p.1). The existence of mental causation is thought to be especially difficult to reconcile with the principle of

the causal closure of the physical world, which says that 'at every time at which a physical effect has a cause it has a sufficient physical cause' (Gibb, 2013, p.2).[1]

Few are willing to deny the existence of mental causation. This is because mental causation is thought to be essential to our conception of ourselves as agents who act intentionally and bear moral responsibility. Consequently, much contemporary philosophy of mind is devoted to showing how the existence of mental causation can be reconciled with the principles that seem at first to forbid its existence. Modern metaphysics of mind are often judged according to how well they deal with the problem of mental causation.

For example, much attention has been paid to establishing whether views which fall under the broad epithet 'non-reductive physicalism' can avoid the mental causation problem. The most popular kind of non-reductive physicalism grants that there are events, processes and states which are mental, but insists that mental events, processes and states are 'nothing over and above' certain physical events, processes and states. Exactly what it is for one entity to be 'nothing over and above' another is a matter of debate. For some, it is enough for mental entities to be nothing over and above physical entities if the former

[1] There are many alternative formulations of the principle of causal closure. For example: 'any physical state or change, if it has a cause or explanation, has a physical cause or explanation' (Hopkins, 1978, p.223); 'For all physical events and states there are necessary and sufficient physical conditions, their "explanations" or "causes"' (Skillen, 1984, p.514); 'Every physical effect has its chance fully determined by physical events alone' (Noordhof, 1999, p.430); 'All physical effects are fully determined by law by prior physical occurrences' (Papineau, 2001, p.9). These various formulations are obviously not equivalent. Furthermore, the principle of causal closure is supposed to be derived from, and supported by, the findings of scientific investigations into causal processes, but it is a contentious question which, if any, of the various formulations of causal closure enjoy such support. I will set such issues aside in this dissertation. See Lowe (2000) for a discussion of issues relating to the principle of causal closure, and see Papineau (2001) for a defence of the principle.

supervene on the latter (Hellman and Thompson, 1975; Haugeland, 1982). For others, the nothing-over-and-above relation must be something stronger than supervenience (Horgan, 1993; Wilson, 2005). And some have suggested this relation is one of composition or constitution, or an analogue of composition or constitution (Pettit, 1993). Regardless of how the nothing-over-and-above relation should be spelled out, several philosophers have argued that non-reductive physicalist views are unable to explain how there can be causation by mental entities.[2] Non-reductive physicalists have offered many counterarguments.[3]

The existence of mental causation is also used as a core premise in a number of arguments which seek to establish some form of physicalist identity theory about mentality. Three such arguments are the 'causal argument for physicalism' championed by David Papineau (1993, 2001, 2002),[4] Donald Davidson's (1970) argument for Anomalous Monism, and Kim's (1993, 1998, 2001) 'causal exclusion argument'.

As Papineau (2001, p.9) presents it, the causal argument for physicalism has three premises:

1. 'All mental occurrences have physical effects'
2. 'All physical effects are fully determined by law by prior physical occurrences' (this is Papineau's formulation of the principle of causal closure)
3. 'The physical effects of mental causes are not all overdetermined'

From these three premises, it is concluded that 'Mental occurrences must be identical with physical occurrences' (p.9). (The causal argument is used to establish a physicalist identity theory about mental *occurrences* or *events*. However, proponents of the causal argument typically assume that events,

[2] For example, Kim (1989, 2005), Crane (1995) and Heil (2013).
[3] For example, Bennett (2003), Árnadóttir and Crane (2013), Shoemaker (2013) and List and Menzies (2009).
[4] This argument is also known 'the causal overdetermination argument' (Crane, 1995; Gibb, 2013) and 'the overdetermination argument' (Sturgeon, 1998; Noordhof, 1999).

processes and states are not significantly different in nature, and so what goes for mental *events* goes for mental *processes* and mental *states* (or at least 'token states') as well.) Importantly, the first premise is understood by proponents of the causal argument as simply equivalent to the claim that mental causation exists.

Davidson's argument for Anomalous Monism is another example of an argument for identifying mental events with physical events which takes the existence of mental causation as a premise. Anomalous Monism asserts that every individual mental event is identical with some physical event, but mental kinds are distinct from physical kinds. Davidson's (1970/2001, p.208) argument for this view involves three premises:

1. 'At least some mental events interact causally with physical events' (The 'Principle of Causal Interaction')
2. 'Where there is causality, there must be a law: events related as cause and effect fall under strict deterministic laws' (The 'Principle of the Nomological Character of Causality')
3. 'There are no strict deterministic laws on the basis of which mental events can be predicted and explained' (The 'Anomalism of the Mental')

Again, the first premise of this argument is typically understood as an assertion that mental causation exists.

Kim objects to Davidson's Anomalous Monism on the grounds that 'on anomalous monism, events are causes or effects only as they instantiate physical laws, and this means that an event's mental properties make no causal difference' (1989, pp.34-35). In other words, Kim thinks that Anomalous Monism suffers a mental causation problem. On Kim's view, events are causes in virtue of the properties they involve, and not every property of an event involves causally matters, unless there is causal overdetermination. Kim argues that the Nomological Character of Causality and Anomalism of the Mental imply that it is always an event's *physical* nature which is causally relevant; whatever *mental* properties an event may involve are excluded from being causally relevant by physical properties which enjoy superior candidacy for this status. This is Kim's

'causal exclusion argument', which is yet another example of an argument for a particular metaphysics of mind – this time a physicalist identity theory which identifies mental *properties* with physical *properties* – which focuses on the problem of mental causation.

The problem of mental causation is thus central to many debates in philosophy of mind. However, philosophers of mind rarely question whether the conception of mental causation that gets these debates started in the first place is the right way to think about the place of mentality in causation. I agree that some form of mental causation is essential to our conception of ourselves as agents who act intentionally and bear moral responsibility. I agree that where there is intentional action there is causation that somehow involves mentality. However, I think it is less clear that this mental causation should be thought of in the way prescribed by the literature on the problem of mental causation in contemporary philosophy of mind.

In most discussions of the problem of mental causation, mental causation is presented as a cause-effect relation between mental and physical entities; mentality and physicality are presented as two sides of a causal exchange. As Jenifer Hornsby puts it, mental causation, as it is discussed in contemporary philosophy of mind, is something 'we are supposed to think of as causation *by* the mental' (2015, p.129). Or as Tim Crane puts it, 'the arguments for physicalism must assume that the labels "mental" and "physical" as applied to causation are really transferred epithets – what is mental and physical are the relata of causation, not the causation itself' (1995, p.219).

In many cases, mental causation is presented as a causal relation between mental and physical *events*. Recall Davidson's Principle of Causal Interaction: 'at least some mental events interact causally with physical events'. Consider also the following formulations of the first premise of the causal argument: 'all mental occurrences have physical effects' (Papineau, 2001, p.9); 'we think of mental and physical events as causally related' (Hopkins, 1978, p.223); 'Mental events have physical effects' (Noordhof, 1999, p.430). Sometimes mental causation is presented as a causal relation that can hold between *states*, as well

as events. For example, consider Anthony Skillen's formulation of the first premise of the causal argument:

> Of some physical events and states, mental events and states are causes. For example, because of my desire to give an example I move a pen – a physical object moves in space, a change which would not have occurred but for that desire. (1984, p.514)

Often, events, states and processes are thought of as being very similar in nature, so that there is no need to treat mental events, mental states and mental processes differently when considering their candidacy as causal relata. For example, when David Armstrong proposes that mental states are states which are 'apt for bringing about a certain sort of behaviour', he notes that his use of the word 'state' is 'not meant to rule out "process" or "event"' (1968, p.82). In most discussions of mental causation, events, states and processes are thought of as three sub-classes of the same general ontological category. Members of this general ontological category – I will call them *items* – are typically thought of as particulars, where particulars are unrepeatable, concrete individuals. So, even where mental causation is not presented as a causal relation between *events* – or not only between events – it is still presented as a causal relation between items which are mental and items which are physical. I call this understanding of mental causation the relational understanding of mental causation.

> **Relational understanding of mental causation**: mental causation is a matter of mental items (events, processes or states which are conceived of as particulars) standing in causal relations to physical events, e.g. bodily movements.

Central to the relational understanding of mental causation is the idea that the mental causation is a cause-effect relation between mental and physical items – mental phenomena are thought of as links in causal chains.

This understanding of mental causation has become the standard way of thinking about mental causation because it is widely thought that intentional or voluntary human action is possible only if mental items stand in causal relations

to physical events such as bodily movements. For example, consider Kim's remarks on why it is important that mental causation is real:

> First and foremost, the possibility of human agency, and hence our moral practice, evidently requires that our mental states have causal effects in the physical world. In voluntary actions our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged. (2005, p.9)

Here Kim endorses the idea that the possibility of voluntary action depends on beliefs and desires, or intentions and decisions, standing in causal relations to bodily movements. Kim also states that:

> […] it seems plain that the possibility of psychology as a science capable of generating law-based explanations of human behaviour depends on the reality of mental causation: mental phenomena must be capable of functioning as indispensable links in causal chains leading to physical behaviour, like movements of the limbs and vibrations of the vocal cord. A science that invokes mental phenomena in its explanations is presumptively committed to their causal efficacy; if a phenomenon is to have an explanatory role, its presence or absence must make a difference – a causal difference. (2005, p.10)

I agree that the worth of psychology as a science and as a means by which we can predict, explain and control each other's behaviour requires that people's behaviour can be causally explained by what they think, feel, believe and want. However, this claim is much weaker than the claim Kim makes. Kim claims that the possibility of psychological explanations of human behaviour requires that 'mental phenomena must be capable of functioning as indispensable links in causal chains leading to physical behaviour, like movements of the limbs and vibrations of the vocal cord'. Thus, Kim thinks that the possibility of voluntary action and psychological explanation presupposes that mental phenomena, like believing that one ought to brush one's teeth, or wanting to make a cup of tea, are *links in causal chains*. I disagree.

1.2 Mental Causation in the Philosophy of Action

I accept that when we act intentionally, there is mental causation. However, I deny that the mental causation on display here should be relationally understood. This is the thesis that I shall be arguing for in this dissertation.

This thesis does not entail that mental items *never* stand in causal relations to physical events. The following examples seem to be cases where a mental event stands to a physical event as cause to effect:

i. A football fan witnesses his team score a goal, and this seeing event causes the football fan to cheer.
ii. I suddenly remember that I need to put the bins out for collection, and this recollection causes me to get up and go outside.
iii. I imagine a frightening scene, and this successful exercise of my imagination causes the hairs on the back of my neck to rise.

I am not seeking to show that *these* instances of mental causation should not be understood relationally (although I think that even these sorts of cases are often misunderstood in contemporary philosophy of mind.) The focus of this dissertation is the mental causation that is on display when human agents act intentionally. It is precisely this sort of mental causation which I want to argue should not be understood relationally. The aim of this dissertation is to show that Kim is wrong to claim that 'In voluntary actions our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways' (2005, p.9). When we say that someone acted intentionally because of what she believed, desired, intended or decided, the concepts *belief*, *desire*, *intention*, perhaps even *decision*, do not refer to items which stand in causal relations to limb movements.

The view I will be arguing for has much in common with a view put forward by Gilbert Ryle. Ryle rejects the 'para-mechanical theory of the mind' (1949, pp.19-23, p.64), where to say that someone did something intentionally is to say that some 'mental thrust' caused some muscular movements. Ryle argues against the idea that 'mental conduct verbs' – like 'knowing', 'believing', 'intending' and 'desiring' – signify or denote episodes in a person's 'secret history'

or ‘stream of consciousness’ (1949, pp.16-17). In Ryle’s view, mental conduct verbs do not denote inner causal events. So, when such verbs are employed to explain why an agent acted as she did they do not designate inner causes of the actions they explain.

Some philosophers (e.g. Wittgenstein 1958, Anscombe 2000, Tanney, 2009) who are sympathetic to Ryle’s view, have sought to reject the idea that beliefs and desires stand to actions as causes to effects by arguing that explanations of intentional actions are not usually causal explanations. These philosophers, whom I shall call ‘non-causalists’, argue that rationalising explanations of intentional actions, i.e. explanations which work by showing why what the agent did seemed, to the agent, like a rational or sensible or good thing for the agent to do, are not causal explanations. Non-causalists assume that to show that the mental concepts employed in explanations of intentional actions do not signify items which are causes of the actions they explain *just is* to show that these explanations are not causal. Non-causalists reject the view that as-a-cause is how we should understand the place of mentality in intentional action, but they also reject the idea that when we say that an agent acted as she did because of what she believed or desired we are giving causal information. In other words, non-causalists deny that, as a matter of conceptual necessity, every instance of intentional action is an instance of mental causation relationally understood, but they also deny that, as a matter of conceptual necessity, every instance of intentional action is an instance of mental causation simpliciter.

In this dissertation, I hope to arrive at the same destination – namely concluding that concepts like *belief*, *desire* and *intention*, do not refer to items which can stand in causal relations to actions – via a different route, one that does not require denying that there is anything worthy of the name ‘mental causation’ necessarily on display when human agents act intentionally. The route I will take involves examining the metaphysics of action and the nature of agency.

To act is to bring about change, to affect the world. In short, there’s causation – which the agent is somehow involved in – wherever there is action.

This much seems to be implied by the causative nature of the verbs we use to report actions. The verb 'raise', for example, is causative: to raise my arm is to cause it to rise. An important metaphysical theory of action is the causal theory of action. The causal theory of action holds that actions are events, namely certain bodily movements, and bodily movements count as actions when and only when they are caused, in the right way, by certain mental states of the agent or mental events involving the agent.[5] Proponents of the causal theory of action typically do not regard the distinction between states and events to be relevant to the plausibility of their theory. The mental causes of actions are typically said to be (the onset of) beliefs, or (the onset of) desires, or (the onset of) intentions – i.e. (the onset of) the kind of states which the concepts employed in rationalising explanations are taken to refer to.

The fundamental ontological claim of this theory is that the agent's involvement in the causality of her action can be reduced to the involvement of her mental states in the causation of her action. In other words, what it is for an agent to affect the world is for certain event-event causal relations to obtain: agent-causation can be reduced to event-causation. Part of what motivates causal theorists to propose this reduction is certain presuppositions about the metaphysics of causation. Causal theorists, implicitly or explicitly, assume that causation is always, everywhere a relation between events. I call the understanding of causation assumed by causal theorists the Humean approach to causation.[6] I will argue that agency cannot be located within a worldview that assumes a Humean metaphysics of causation. This is because agency is misconstrued when it is taken to be something that can be identified with a certain kind of event-causation. I will argue that to properly understand agency

[5] See Bishop (1989, pp.40-44), Davidson (1963/2001, pp.3-21 and 1971/2001pp.43-63), and Smith (2012).
[6] It should be noted that it is unclear whether Hume himself endorsed the view I'm calling the Humean approach to causation. See chapter 3.

one must accept what the causal theory of action denies, namely that the exercise of agential power cannot be reduced to a special kind of event-causation.

Hornsby (2004a, 2012) has argued that what's needed to properly understand agency is a metaphysical framework which provides intellectual space for thinking of causation as something other than a relation between events. Without such a framework, it is impossible to see how the causality of action might be something other than a causal relation between mental event and action, as the causal theory of action supposes. In line with this directive, I will present a non-Humean way of thinking about causation, which denies that causation is always a relation, and holds instead that causation can be a determinable process which substances engage in. This non-Humean theory of causation depends upon a novel ontology which denies that processes belong in the same ontological category as events, and instead takes processes to be a special kind of universal. I suggest that engaging in a process is analogous to instantiating a property, and that events are related to processes by an instancing relation – or to put it more naturally, events are instances of processes. I believe that this radically non-Humean theory of causation, and its associated process ontology, has several advantages, but my aim in this dissertation is to show how this new approach enables us to put together a more successful account of agency, one which does not take the causality of action to consist in any sort of cause-effect relation.

A key part of this account is the thesis that to exercise power is to engage in a process. I argue that to understand the concept of agency it is vital to distinguish between the different sorts of power a substance can exercise. I think there are two distinctions crucial to our concept of agency: the distinction between active and passive powers; and the distinction between one-way and two-way powers. As I will show, agency does not reduce to the exercise of active power, nor does it reduce to the exercise of two-way power. My view is that agency is a complex concept which incorporates both distinctions. Some substances' agential powers are one-way; these substances manifest their agency

when they are active, but not when they are passive. Other substances' agential powers are two-way; these substances manifest their agency when they are active, but also sometimes when they are passive.

The views on causation and agency I advance allow me to argue that even though it should not be understood relationally, there is something worthy of the name 'mental causation' which is necessarily on display when human agents act intentionally. I will argue that to act intentionally is to engage in a process, and as such is to exercise a power – but a power of a special sort. I will suggest that to act intentionally is to exercise one's two-way powers in accordance with some form, structure or pattern. In other words, the power to act intentionally is a power to structure one's own activities so that they demonstrate a pattern – a pattern which is revealed by attributing mental states to the agent. So, when an agent acts intentionally, they engage in the process of causation. I will suggest that the process they engage in counts as *mental* causation in virtue of the fact that in engaging in this process the agent is manifesting a special power, namely a power to organise their activities into a pattern that can be made sense of by appeal to mental concepts.

In this way, I hope to show that even if our concept of intentional action entails that there is mental causation, there is no reason to think that this mental causation should be *relationally* understood. The mental causation that is entailed by the existence of intentional action is not causation by mental items. It is no part of our concept of intentional action that causal relations between mental items and physical events obtain. Consequently, the usual way of presenting the problem of mental causation, as a question of how it could possibly be that mental items could stand in causal relations to physical events, is misconceived. We have no *a priori* reason to think that mental items *do* stand in causal relations to physical events. The real mystery of mental causation is not how mental-physical causal relations are possible, but how it is that we – that is, human beings – have the unique capacities we do, in particular our capacity to act for reasons. How have such capacities evolved? From what simpler powers do such capacities emerge? I will not have much to say in answer to these

questions – indeed, I do not think they are purely philosophical questions – my aim is rather to demonstrate that the real problem of mental causation should be articulated in terms of *these* questions, and not in terms of how causal relations between mental items and physical events are possible.

1.3 Summary of Remaining Chapters

In chapter 2, I examine the causal theory of action. I put forward my reasons for thinking that the main thesis of the causal theory of action, that agency is nothing over and above some special kind of event-causation, cannot be right. I argue that any attempt to reduce agency to special sort of event-causation misconceives agency by failing to accommodate two basic facts about agency: 1) that the agency concept covers the doings of inanimate objects and animals as well as the intentional doings of human beings, and 2) that, for some agents, their agential power can be demonstrated in passivity as well as in activity (for example, I can demonstrate my agential power when I deliberately refrain from doing something I am capable of doing). I argue that to explain how these two facts about agency can be true, one must accept that the exercise of agential power cannot be reduced to a special kind of event-causation.

In chapter 3, I articulate the Humean metaphysics of causation presupposed by the causal theory of action. I will explain what the key commitments of this Humean approach to causation are, and outline an alternative, non-Humean, approach which rejects these commitments. According to my preferred non-Humean approach, causation can be a process rather than a relation, of which processes like breaking, crushing, bending etc are more determinate species. My proposal will be that causation is on display not only when events make the difference to the occurrence of other events, but also when substances exercise causal powers. What it is for a substance to exercise causal power is for there to be an entity, i.e. process, in which the substance engages.

In chapter 4, I explain what a process is. I propose a novel ontology which denies that processes belong in the same ontological category as events, and

instead holds that processes are a special kind of universal. I suggest that engaging in a process is analogous to instantiating a property, and that events are instances of processes. On this proposal, a substance's engagement in a process is a special sort of state of affairs, namely a *dynamic* state of affairs.

In chapter 5, I argue that my alternative, non-Humean, way of thinking about causation, and the ontology that permits it, allows us to put together a more successful theory of agency, one that avoids the problems facing the causal theory of action. I also compare my account of agency with some other alternatives to the causal theory of action, namely agent-causation-based theories of action, and argue that my account has advantages over these theories.

In chapter 6, I turn to intentional action, and specifically the question of whether our concept of intentional action presupposes that mental items stand in causal relations to actions, or physical events. I outline Davidson's (1963) argument that rationalising explanations are causal explanations and consider objections to the idea that explanations of intentional actions which cite reasons are made true by mental events or states standing in causal relations to actions. I will focus on Julia Tanney's (2009) argument which aims to show that the mental concepts employed in rationalising explanations do not discharge their explanatory role by designating causes of the actions they explain.

In chapter 7, I argue against an assumption made by both Davidson and his opponents, namely the assumption that causal explanations are precisely those explanations whose explanandum designates an effect and whose explanans designates an item which is the cause of that effect. I will discuss counterexamples to this view. And, in line with the non-Humean theory of causation advanced in chapters 3 and 4, I will suggest that some causal explanations may be made true by facts about dynamic states of affairs. I will also investigate the proposal that rationalising explanations are a special kind of disposition-citing explanations and outline what this means for our understanding of intentional action.

Chapter 8 concludes.

2

The Causal Theory of Action

The causal theory of action holds that actions are events, namely certain bodily movements, and bodily movements count as actions when and only when they are caused, in the right way, by certain mental states of the agent or mental events involving the agent. This theory has its roots in Davidson's (1963) argument that rationalising explanations of actions – explanations which explain why an agent acted as she did by giving the agent's reason for acting as she did – are causal explanations. Davidson claimed that the explanantia of rationalising explanations are facts about what the agent wants to do (or what the agent has an urge to do, or what the agent has an ambition to do) and facts about what the agent believes about how to do it.[7] Davidson calls the composite of a desire to perform some type of action and a belief about how performance of that action may be achieved 'the *primary reason* why the agent performed the action' (1963/2001, p.4). Davidson argued that when we say the agent acted as she did *because* she wanted to do something, or *because* she believed that something was the case, this 'because' implies causality. From this, Davidson concluded that states of desiring and states of believing – or, at least, events suitably related to states of desiring and states of believing, such as the onset of the desire or the onset of the belief – are causes of the actions they explain.[8] Causal theorists

[7] I think that Davidson is broadly correct on this point, however, I will consider issues facing this understanding of rationalising explanation in chapter 6.
[8] I will examine this argument, and objections to it, in more detail in chapter 6. The argument in this chapter does not depend on refuting Davidson's (1963) argument.

assume that an adequate account of what it is to act can be extracted from Davidson's anatomisation of rationalising explanations. On the basis of Davidson's argument, causal theorists, like Michael Smith, have proposed that:

> […] actions are those bodily movements that are caused and rationalised by a pair of mental states: a desire for some end, where ends can be thought of as ways the world could be, and a belief that something the agent can just do, namely move her body in the way to be explained, has some suitable chance of making the world the relevant way. Bodily movements that occur otherwise aren't actions, they are mere happenings (Smith, 2004, p.165)

Famously, attempts to analyse intentional action in terms of causation by certain mental states have been impeded by the existence of counterexamples which involve 'deviant causal chains'. Davidson himself doubted that a reductive analysis of intentional action could be developed from the idea that states of desiring and states of believing are causes of the actions they explain because of deviant causal chain cases. The example Davidson uses to illustrate this problem is that of a mountain climber holding another man by a rope, whose desire to rid himself of the weight, and belief that he can do so by letting go, causes him to become so nervous that he lets go of the rope:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. (Davidson, 1973/2001, p.79)

In this example, the climber has an end he wants to achieve, namely to rid himself of the weight and danger of holding the other man, and the climber reasons that loosening his hold is the best means to achieve this end. This belief-desire pair causes a bodily movement of a type that is rationalised by the belief-desire pair, just as causal theorists allege it would in an ordinary case of intentional action. But in this case the causal route from belief-desire pair to bodily movement involves an intermediary state of nervousness which 'robs the

climber of control', as John Bishop (1989, p.134) puts it. In this example, the climber didn't let go intentionally.

Many have argued that a reductive causal analysis of intentional action is still possible, Davidson's nervous climber example notwithstanding. Davidson's example shows that the original causal theory failed to specify jointly sufficient necessary conditions for intentional action, but this doesn't mean that a more sophisticated version of the causal theory will also fail.[9] However, even if we cannot specify, in neutral terms, jointly sufficient necessary conditions for intentional action, this doesn't mean that the ontological component of the causal theory of action – that action is nothing over and above some special kind of event-causation – is false (although it might weaken the case for thinking that it is true). The fact that intentional action cannot be analysed in terms of causation by mental events does not refute the claim that intentional actions are really nothing over and above events caused to happen, in the right way (whatever that may be), by certain mental events. Indeed, Torbjörn Tännsjö (2009) has argued that because the set of intentional actions is very diverse, we should not expect to be able to say, in completely general terms, what causal requirements are needed for intentional action to take place. But this doesn't matter, because 'in relation to specified actions types we can distinguish between right and deviant causal chains' (p.473), and that is enough reason to be confident in the causal theory of action's ontological thesis is correct.

It is the ontological element of the causal theory of action that I am interested in. So, for my purposes here, it does not matter whether the problem of causal deviance can be solved or not. I will attempt, in this chapter, to explain why I think the idea that action is nothing over and above some special kind of event-causation cannot be right. My target is the claim that actions are bodily

[9] More sophisticated versions of the causal analysis of intentional action have been offered by Peacocke (1979), Bishop (1989), Mele (2003) and McDonnell (2015).

movements which are caused, in the right way (whatever that may be), by certain mental states of the agent or mental events involving the agent. I will argue that any attempt to reduce agency to a special sort of event-causation misconceives agency by failing to accommodate two basic facts about agency.

My argument will proceed as follows. In section 2.1, I will outline two distinctions which I believe are crucial to our concept of agency. In section 2.2, I will show why the causal theory of action wrongly implies that the agency concept does not cover the 'doings' of inanimate objects and animals. I will suggest that this error isn't considered as devastating to the causal theory of action as it should be because the causal theory of action fails to keep the two distinctions outlined in section 2.1 apart. In section 2.3, I will show why the causal theory of action cannot accommodate the fact that, for some agents, their agential power can be demonstrated in passivity as well as in activity (for example, I can demonstrate my agential power when I deliberately refrain from doing something I am capable of doing). In section 2.4, I will explain why the source of these two errors is the causal theory of action's reductive ambitions. I will argue that to explain how it can be that a) the agency concept covers the 'doings' of inanimate objects and animals as well as the 'doings' of human beings and b) some agents' agential powers can be demonstrated in passivity as well as in activity, one must accept that the exercise of agential power cannot be reduced to a special kind of event-causation.

2.1 Two Distinctions Crucial to our Concept of Agency

The causal theory of action is sometimes presented as a theory of what marks the difference between things that one does, and things that befall one. This way of presenting the question a theory of agency is supposed to answer is problematic, as the verb 'do' is very imprecise. There are many things that I can be said to do which are not actions of mine, but are instead things that I undergo or suffer. It is perfectly legitimate, in certain contexts, to speak of forgetting or falling over as things that one has done, even though forgetting and falling over are not, in any sense, actions of mine. Reflex behaviours, like blinking or

sneezing or ducking to avoid a projectile, are also commonly said to be things that we do, but are not usually considered to be appropriate targets of the causal theory of action.

The 'doings' of inanimate objects, like the stove's heating the soup or the acid's melting the beaker, are also not typically seen as targets of the causal theory of action. Some proponents of the causal theory of action also deny that the 'doings' of animals, like the casting of a web by a spider, or the pouncing of a cat, are targets of the causal theory of action.[10] The 'doings' of inanimate objects and the 'doings' of animals are not *actions* in the required sense, the causal theorist might say. Even though we might describe these examples as things that are done (by us, or inanimate objects or animals), and even though we might not ordinarily describe these examples as things that are suffered, the causal theorist would insist that these 'doings' are not sufficiently different from suffering change to qualify as targets of the causal theory of action. The target of the causal theory of action is something that involves a greater, more sophisticated kind of agential control. Davidson also indicates that not every case where it would be correct to use a verb of action to describe what is going on qualify as targets of the causal theory of action. Davidson uses as his example a case where I spill my coffee because you jiggle my hand (1971/2001, p.45) – even though I *spill* my coffee in this example ('spill' being a verb of action), this sort of case is not the proper target of the causal theory of action. If this case describes an action of spilling at all, then it is an action only in a weak sense. Whatever agency is on display in this case, it is not sufficiently distinct from passivity to qualify as the target of the causal theory of action.

[10] For example, Velleman (1992) distinguishes between 'full-blooded human action' and animal behaviour and thinks that only the former '[provides] the philosophy of action with its distinctive subject matter' (p.465). Similarly, Bratman (2001) thinks there is a distinction between 'merely motivated behaviour', which animals may be able to demonstrate, and 'full-blown agency', and it is the latter that we are seeking an account of.

So, what 'doings' *are* legitimate targets of the causal theory of action? Davidson claims that cases where I spill my coffee *intentionally* – perhaps I hate coffee and want to indicate my disgust – are legitimate targets of the causal theory of action. Cases where I spill my coffee unintentionally, but only because I have made a certain kind of mistake – for instance, when I mistakenly think my coffee is tea and spill it to express my disgust for *tea* – also qualify as targets of the causal theory of action. This is because, in the latter example, even though I do not spill the coffee intentionally, I intentionally do *something*, namely, spill the contents of my cup (Davidson 1971/2001, p.46). In this thought lies a way of delineating the target of the causal theory of action. When an agent does something which can be 'described under an aspect that makes it intentional' (Davidson 1971/2001, p.46), then we have the sort of thing the causal theory of action aims to give an account of. In my view, this way of navigating through the various contrasts that have something to do with agency to get to the proper target of the causal theory of action represents a confusion of two different distinctions, both of which are crucial aspects of our concept of agency.

2.1.1 Activity and Passivity

The first distinction is the distinction between activity and passivity. Activity is the exercise of an active power, i.e. a power to wreak change. Passivity is the manifestation of a passive power, or a liability, i.e. a power to undergo or suffer change. Active powers are powers to change, and passive powers are powers to be changed. Substances which exercise active powers are agents, and substances which manifest passive powers are patients. An instance of an exercise of active power is an action. An instance of a manifestation of passive power could be called a passion. As John Hyman points out, the difference between agent and patient is not a difference between two different kinds of substance, it is rather a difference between two different roles substances can adopt (2015, p.35). This is demonstrated by the fact that one and the same substance can be an agent at one time, and a patient at another time – for example, when I push you, I am the agent, when you push me back, I am the patient. It is also possible for one and

the same substance to be both agent and patient at the same time – for example, as Hyman notes, a victim of suicide is both agent and patient.

The active–passive distinction is thrown into doubt when we consider the fact that in many cases when an intuitively active power is manifested, the manifestation of this power involves the possessor of the power suffering change as well as producing it. For example, when salt is dissolved in water, we may intuitively class the power of the water to dissolve the salt as active: the water is producing change in the salt. However, the water is also changed by the dissolution process, and necessarily so – if the water were not liable to become uniformly salty when salt was added to it, then it wouldn't be possible to dissolve salt in water. So, it seems that the intuitively active power of water to dissolve salt is *also* passive. It seems like the distinction between the exercise of active power and the manifestation of passive power, and hence the distinction between activity and passivity is spurious. At best, the distinction is a matter of there being two alternative ways to describe the very same sort of eventuality.

The solution to this problem is, I think, to reject the idea that for a substance to exercise an active power the substance must, in exercising this active power, be 'purely active', that is, suffer no change at all. Similarly, it is not the case that a substance exercising a passive power needs to be 'purely passive'. Erasmus Mayr suggests that 'the distinction between active and passive powers is one of degree, with all powers situated on a more or less continuous spectrum of more or less active and passive powers' (2011, p.204). What this means is that some powers are such that when they are exercised the substance in possession of the power produces much more change than it undergoes. For example, when I squash a grape, the grape is drastically changed, whereas I remain much the same. Other powers are such that when they are exercised the substance in possession of the power undergoes as much change as it produces – as in the case of the water dissolving the salt. The power of the water to dissolve salt is, as it were, less active than my power to squash a grape.

The danger with this solution is that it means that the distinction between activity and passivity is not absolute; it also makes the activity-passivity

distinction less than fully objective. Whether what a substance is doing is activity or passivity is relative to the degree of change it is wreaking and/or undergoing, and assessing how much change a substance is wreaking and/or undergoing may not be a fully objective matter. How much change one thinks the water undergoes when salt dissolves into it may depend on one's views about the nature of water. How big an issue this is depends on what work the activity–passivity distinction is put to. I will be putting the distinction to work in this dissertation when I come to articulate my own view on what agency is in chapter 5, so I will address this issue in due course.

2.1.2 Settling and Non-Settling

The second distinction is more controversial than the first. The second distinction is a distinction between two kinds of agency. On the one hand, we have the agency of substances which are in control over what is going on. The first kind of agency I will call 'self-movement', and the substances involved in self-movement 'self-movers'. On the other hand, we have the agency of substances which are not in control over what is going on. I will call these substances 'moved-movers'. The former sort of agency I will call 'settling' and the latter I will call 'non-settling'.

I am borrowing the semi-technical term 'settling' from Helen Steward (2012). For Steward, when an agent acts, certain open questions, like whether the agent will φ, how the agent will φ, where the agent will φ, when the agent will φ etc, come to have answers. This is settling. In Steward's view action is settling; what it is to act is to settle some matter. Furthermore, Steward argues, the existence of settling is inconsistent with universal determinism. I do not claim that action is settling, or that settling is incompatible with universal causal determinism. However, the concept 'settling' is still suitable for my purposes. I intend the term 'settling' to capture the idea of something's being left up to the agent. Agents capable of settling, i.e. self-movers, are agents for whom some of what goes on with them is up to them.

To help illustrate the distinction, consider the following examples. When a stone is thrown at a window with sufficient force, there is no sense in which it is *up to the stone* whether or not it breaks the window. If the conditions are right, i.e. the stone is heavy enough and the glass is thin enough, the stone will break the window (provided nothing comes along and interferes, e.g. no-one snatches the stone out of the air before it hits the window). The stone may well be the thing that is breaking the window, in this way the stone is a 'mover' (or more precisely, a 'breaker'), but the stone was 'moved' to do so, that is, the stone was directed to break the window by some other thing (whatever threw it). A robot like Honda's ASIMO is also, in my opinion, a moved-mover and not a self-mover. This might seem counterintuitive because, unlike a stone, ASIMO can move around and perform various tasks without human intervention. However, ASIMO's movements are strictly governed by his construction and programming. To illustrate: ASIMO has two cameras, a laser sensor, an infrared sensor and an ultrasound sensor. When information recorded by these sensors conflicts with information in ASIMO's pre-loaded map of navigable paths (e.g. by signalling that there is an obstacle in one of these paths) ASIMO cannot but move around the obstacle (American Honda Motor Co. Inc. 2017). ASIMO is moved to move around the obstacle by his component parts. It is not up to ASIMO what goes on with his legs.[11]

Often when human beings and many (perhaps all) animals act, there is a sense in which what they do is up to them. To help make the idea vivid, imagine my friend Amy really wants me to get up and make tea, so she makes sure I'm thirsty by giving me something salty to eat, puts a cup and some teabags nicely in view, then says "why don't you have some tea?". The conditions are right for me to get up and make tea. But, Amy's plan might not come to fruition because,

[11] Although ASIMO is, in my opinion, a moved-mover and not a self-mover, there may be (or might be in the future) other robots which are self-movers.

even though the conditions are right, I don't have to get up, I could still stay seated. Amy might install some clever machinery to manipulate my brain and nervous system and use that to make me get up (in the manner of the character Black from Harry Frankfurt's (1969) thought experiment), but in that case, I would cease to be a self-mover and I would become a moved-mover. Amy would be taking control over what goes on with me; it would no longer be up to me what happens with my body.

I believe that the distinction between settling and non-settling should be spelled out in terms of two-way and one-way powers. I endorse Kim Frost's definition of a two-way power as one which has 'two fundamental, mutually exclusive kinds of exercise', whereas a one-way power has only one fundamental kind of exercise (2013, p.612). The easiest way to spell out this idea is by means of an example. In the right circumstances my power to get up is two-way.[12] What this means is that if I do end up getting up, I am manifesting my two-way power; but if I end up *not* getting up (which might involve actively doing something else, but might not – it might involve continuing an activity already in progress, or letting something happen to me), I am *also* manifesting my two-way power. Thus, my power to get up, because it is two-way, is sometimes manifested by getting up, and sometimes manifested by *not* getting up. The

[12] While I do not think it is possible to give necessary and sufficient conditions for possession of a two-way power, it is possible to say something about what must be the case for an agent to possess a two-way power. A necessary condition for having a two-way power to φ at a time t, is to be able both to φ and not φ at t, and to have the opportunity both to φ and not φ at t (Alvarez, 2013, p.108). If agent A has the ability to φ, then she has the right attributes for φing and knows how to φ (for example, A only has the ability to wave her arms if she has arms and knows how to wave them). If A has the opportunity to φ, then there is nothing preventing her from φing (for example, she is not tied up or injured) – c.f. Kenny (1975, p.133).

power has two mutually exclusive kinds of exercise, which I will call positive and negative, and only one of these (the positive) is the activity the power is specified as a power to do. In the case of one-way powers, when the conditions are right for the manifestation of a one-way power, the activity the power is a power to do will be engaged in; whereas in the case of two-way powers, when the conditions are right for the *positive* manifestation of a two-way power, the two-way power may *not* be exercised positively – it may be exercised negatively – and thus the activity the power is a power to do may not be engaged in. [13] It is important to note that while one-way powers can be distinguished into those which are active and those which are passive, the active-passive distinction does not have application in the case of two-way powers. This is because two-way powers are powers to act *or refrain*, so they are all powers to be active in a certain way, or not (which might be to be active in a different way, or might be to be passive).

Steward (2013b) finds the conception of two-way powers as powers with two distinct fundamental kinds of manifestation problematic. For Steward, a power to φ is two-way just in case the agent who possesses the power to φ *also* possesses the power not to exercise their power to φ (2013b, p.691). Steward argues that a conception of two-way powers like mine (and Frost's) has

[13] I believe possession of a two-way power is compatible with universal causal determinism, where universal causal determinism is the thesis that 'every event is necessitated by antecedent events and conditions together with the laws of nature' (Hoefer, 2016). Compatibilism of two-way powers and determinism can be established by distinguishing between two sorts of possibility: *physical* possibility and *agential* possibility. Once physical and agential possibility are distinguished, one can argue that while causal determinism entails that past events (perhaps together with the laws of nature) close off alternative physical possibilities, they leave open alternative agential possibilities, and the latter is all that is entailed by the claim that some substance has a two-way power. If agent A has the two-way power to φ then both φing and not φing must be open agential possibilities for her, but this doesn't mean that it must be both physically possible that the future contains a φing and physically possible that the future does not contain a φing. Common to most compatibilist arguments of this type is the idea that facts pertinent to what is an agential possibility for some agent are not the same as facts pertinent to what is a physical possibility in that situation. C.f. Kenny (1975), Campbell (2005), Kapitan (2011), Berofsky (2011), Frost (2013) and List (2014).

counterintuitive consequences (2013b, p.691). As Steward notes, it seems to entail that in not singing right now while I'm working on this chapter, I am exercising my power to sing, albeit negatively. I accept that it is counterintuitive to think that in not singing right now, I am exercising my power to sing. It is more intuitive to think that my power to sing is dormant while I am working on this chapter: it is not being exercised at all. I thus acknowledge that not *every* case where an agent does not φ counts as a negative exercise of a two-way power to φ; not every case of not doing something is a case of *refraining* from doing it. However, I think a conception of two-way powers as powers with two mutually exclusive kinds of exercise is compatible with the fact that not every case of not doing something is a case of *refraining* from doing it.

As long as one can say something about how to distinguish cases where a two-way power to act is exercised negatively from cases where the power to act is just not exercised at all, then one is permitted to claim that there's more to exercising a two-way power to φ negatively than simply not φing. I doubt that there is a completely general way to distinguish cases where an agent exercises her two-way power to φ negatively from cases where an agent's not φing does not count as a negative exercise of her two-way power to φ. This is because what it takes for some instance of not acting in a certain way to count as refraining from acting in that way might depend on the type of action in question. For example, the fact that I am consciously aware of my cup of coffee might be sufficient for my not reaching for the cup to count as a negative exercise of two-way power to reach for it. But for my not singing right now to count as a negative exercise of two-way power to sing, I may need indexical knowledge that the circumstances I am in are circumstances in which I could (or should) be singing. In all cases of refrainment, I think some sort of awareness of what one could be doing is required, but precisely what sort of awareness is required differs depending on the type of action in question.

These two distinctions – between activity and passivity, and between settling and non-settling – are, I believe, fundamental to our conception of agency. Competence with the agency concept demands that one have some

grasp of these two distinctions. The agency concept has something to do with the idea of agents as things which intervene, as determiners of what goes on, as bringers-about of change. It seems to me that the concept *agent* is kindred with *causation, production, activity,* and *action.* The activity–passivity distinction is also a way of clarifying the distinction between what one does and what happens to one, which is the most mundane and common way to expressing the distinction between actions and 'mere happenings'. At the same time, it seems to me that one hasn't really mastered the concept of agency until one has recognised the difference between things that just lie there until something else comes along and prods them into action, and things which, sometimes with effort, move themselves about. This is because, as well as being kindred with concepts like *causation*, *agent* is associated with ethical concepts like *responsibility* and *blameworthiness*. As Hyman (2015) puts it, some instantiations of agency have an 'ethical dimension' as well as a 'physical dimension'. It is of great ethical significance that some things are, as it were, victims of their circumstances: their actions are not up to them, whereas other things are somewhat independent from their circumstances. I do not think the settling–non-setting distinction is one and the same as the free–unfree distinction, because I think it takes a lot more for an action to be free than for it to be up to one. If a mugger threatens to kill me unless I hand over my wallet, then, when I acquiesce, I am not acting freely, even though I had the ability and opportunity not to hand over my wallet and face the dire consequences. But having a two-way power to φ is a necessary condition for φing freely: if the movements of your body are being controlled by some other substance, then you are not freely making them. It is therefore very important to us that the settling–non-settling distinction is real.

The causal theory of action fails to keep these two distinctions apart. It confuses settling with activity, and non-settling with passivity. In conflating these two distinctions, the causal theory of action ends up failing as an account of agency, a failure which is demonstrated by two key errors made by the theory. Firstly, the core claim of the causal theory of action entails that some

examples of activity fail to count as agency at all, thus the link between agency and activity is broken. Secondly, the theory cannot accommodate the fact that self-movers can exercise their agential powers by remaining passive, thus the link between agency and settling is also broken.

2.2 Not All Agents Are Human

The fundamental claim of the causal theory of action is that *the agent's* involvement in the causality of action can be reduced to the agent's *mental state's* involvement in the causation of an action. Proponents of the causal theory of action can grant that to be an agent is to be something which brings about change, they only insist that for an agent to be something which brings about change is for the mental states of that agent – or mental events involving that agent – to cause an event. The causal theory of action thus ties agency to mentality. However, there seem to be cases of agency which do not involve causation by a mental state or mental event, or at least not by the kind of mental state which could rationalise the action it supposedly causes. These cases fall into three inexact groups: actions of inanimate objects, actions of animals and human actions which are either 'sub-intentional' or spontaneous expressions of emotion.

2.2.1 Actions of Inanimate Objects

If what it is to act is for one's mental states to cause a bodily movement, then substances which do not possess mental states cannot be agents. This means that inanimate objects cannot be agents. So, for example, when the stove heats the soup, its heating of the soup is not really an action. As Hyman points out, denying that inanimate objects can act is at odds with the language we use to report actions. We typically report actions by means of causative verbs like 'melt', 'burn', and 'pump'. But we say things like 'the acid melted the beaker' or 'the poker burnt the cloth', 'his heart pumped blood', just as readily as we say 'the cook melted the butter', or 'the criminal burnt the evidence', or 'the man pumped the water'. And, as Hyman (2015, pp.30-31) has argued, it is implausible to think that these verbs have different meanings when they are

used to report what inanimate things have done and when they are used to report what human beings have done.

2.2.2 Actions of Animals

The causal theory of action's focus on mentality may also rule out animal agency. Since the causal theory of action is motivated by Davidson's (1963) argument that rationalising explanations are causal explanations, if the causal theory is correct, then the possibility of animal agency would seem to depend on whether animals' bodily movements can be causally attributed to mental states such as believing or desiring – mental states whose contents could rationalise the kind of bodily movement the animal performs. However, it is not obvious that animals possess mental states sophisticated enough to serve as rationalisations of their actions. If it is not true that an animal acts because it wants to achieve something and believes that so acting is a way to achieve its aims, then we have no basis on which to conclude that the animal's action was caused by its belief or desire. Perhaps such scepticism about the rational capacities of animals is unwarranted. However, the important point here is that while we might be unsure about whether animals can act on beliefs and desires, it is a much greater test of credibility to deny that animals act at all.

2.2.3 Sub-intentional Action and Expressions of Emotion

Brian O'Shaughnessy delineates a class of actions he calls 'sub-intentional'. Sub-intentional actions include actions like 'tapping my feet to the music' and 'idly moving my tongue in my mouth' (1980, p.61). Other examples may include shifting one's position, automatically scratching an itch, or fiddling with one's hair. Whether such examples can really be regarded as lacking intentionality is open to question. However, what does seem right is that actions like unthinkingly tapping one's foot to music, or shifting one's position, or fiddling with one's hair do not seem to be preceded by or accompanied by (and hence not caused by) an intentional state such as believing that performing the action is a good idea, or wanting to achieve something by means of the action. O'Shaughnessy acknowledges, rightly I think, that sub-intentional actions are

subject to psychological explanations. For example, O'Shaughnessy suggests that sub-intentional actions might be explained in terms of feelings of restlessness (p.61). When I shift my position, it is usually because I feel uncomfortable – my action can thus be explained in terms of a feeling. Nevertheless, sub-intentional actions are not performed to achieve any thought-of goal. They cannot be rationalised by facts about what the agent wants to do and what the agent believes about how to do it, because they are not actions which seem, to the agent at the time of performing them, like sensible, or rational or good things to do. At the time of the performing a sub-intentional action, the agent is often not aware that she is performing the action at all. Because sub-intentional actions are not accompanied by an intentional state, and cannot be rationalised, they will not count as actions at all according to the causal theory of action.

Another class of human actions which are potential counterexamples to the causal theory of action are spontaneous expressions of emotion (Hursthouse, 1991). Like sub-intentional actions, it is doubtful that spontaneous expressions of emotion can be rationalised: when we embrace a loved-one, or cry upon hearing bad news, we do not do these things because it is sensible, or rational, or good to do so. Such actions do not seem to be accompanied by a mental state which could be considered a Davidsonian 'primary reason'. However, spontaneous expressions of emotion are certainly actions.

2.2.4 Weakening the Conditions

There are three sorts of reply a causal theorist could offer in response to the three sorts of counterexample I have just described. First, the causal theorist might weaken the conditions which must be met for a bodily movement to count as an action. The causal theorist might suggest that rather than require that a bodily movement be caused by a mental state of the sort that could rationalise the bodily movement, a bodily movement only needs to be caused by some mental state or other to count as an action. This strategy works best in response to the third class of counterexamples: sub-intentional actions and spontaneous expressions of emotion. This is because both sub-intentional action and spontaneous expressions of emotion are (plausibly) subject to psychological

explanation, if not rationalising explanation. It is plausible to suggest that even though sub-intentional actions and spontaneous expressions of emotion are not accompanied by (and hence not caused by) intentional states, they are accompanied by (and hence possibly caused by) other sorts of psychological state.

However, for this response to be viable, we need a new argument for the conclusion that sub-intentional actions and spontaneous expressions of emotion are not only *explained* by psychological states but are also stand to such states as effect to cause. Davidson's (1963) argument provided reason to think that actions subject to *rationalising* explanations are caused by the mental states whose contents rationalise the action.[14] But Davidson's (1963) argument concerned rationalising explanations specifically – not psychological explanations more generally. Someone who wants to hold that sub-intentional actions and expressions of emotion are actions because they are caused by psychological states must provide a *new* argument showing that the existence of sub-intentional actions or expressions of emotion conceptually entails the existence of causation of a bodily movement by a psychological state. This cannot be simply assumed. Steward suggests that one might be tempted to assume that sub-intentional actions and spontaneous expressions of emotion, if they really are actions at all, stand to mental states as effect to cause, because one thinks that 'unless there is some reason to suppose that a movement is in some sense the product of something mental, there can be no reason to think it should be associated in any special way with the self, with the agent' (2009, p.303). But reducing the agent's involvement in the causality of their action to the agent's mentality's involvement is the core thesis of the causal theory of action, hence it must be argued for.

[14] Again, this is not to say that Davidson's argument is sound – see chapter 6.

Another problem with this type of reply is that even after weakening the conditions for action so that bodily movements count as actions just in case they are caused (in the right way) by some psychological state or other, actions by inanimate objects will still constitute counterexamples to the causal theory of action as these actions are not subject to psychological explanation. Animal actions may also still constitute counterexamples, as long as it remains plausible to argue that the actions of animals are not subject to psychological explanation.

2.2.5 Rejecting the Counterexamples

A second sort of reply available to the causal theorist is to deny that the behaviours described in my counterexamples qualify as actions – or at least not 'full-blooded actions'. If such behaviours count as actions at all, then they are 'actions' of a lesser kind, and not the sort of actions which the causal theory of action seeks to give an account of (c.f. Velleman, 1992 and Bratman, 2001). Adherents of the causal theory of action may deny that the examples given in sections 2.2.1 to 2.2.3 are cases where we have an action that nevertheless fails to satisfy the conditions for action proposed by the causal theory, because they will deny that these examples are actions at all. This reply is most intuitively plausible in response to the first sort of counterexample, and so I will assume that if it can be shown that causal theorists are wrong to deny that inanimate objects can act, one can take it for granted that it is wrong to deny that animals can act, and wrong to deny that sub-intentional actions and spontaneous expressions of emotion are real actions.

To show that it is wrong to deny that inanimate objects can act, let me first point out that if the behaviours of inanimate objects are not *really* actions, then our everyday practice of extending the agency concept to them must be anthropomorphic or metaphorical or careless. I do not think it is anthropomorphic or metaphorical or careless to speak of inanimate objects as agents. I shall argue in this section that the actions of inanimate objects only *seem* like they are not *real* actions when the activity-passivity distinction is confused with the settling-non-settling distinction, or more precisely, when the

exercise of one-way power is confused with passivity and exercise of two-way power is confused with activity.

Many philosophers have confused activity and settling. For example, Locke argued that the origin of our idea of passive power comes from our observations of bodies undergoing change. Passion, i.e. the exercise of passive power, is something we can directly perceive, but action, which Locke rightly took to be the exercise of active power, couldn't be perceived. Locke thought that we couldn't directly observe the active production of change. According to Locke, the idea of *active* power comes from our experience of bringing things about because we choose to. However, for Locke, the power we discover only by observing it at work in ourselves is 'the power to begin or forbear':

> This at least I think evident, That we find in our selves a *Power* to begin or forbear, continue or end several actions of our minds, and motion of our Bodies, barely by a thought or preference of the mind ordering, or as it were commanding the doing or not doing such or such a particular action. (1975, p.236)

It seems from this passage that it is in fact a *two-way power*, or a power to settle what one does, which Locke is claiming can only be observed in ourselves when we act voluntarily, and cannot be observed in material bodies.

Locke may be right to think that our concept of *two-way power* is not a concept that can be gained from perception alone – acquisition of this concept may well require 'reflection on what passes in ourselves' (1975, p.235). Consider again the case where I get up to make tea of my own accord, as compared with the case where Amy uses her brain-manipulation device to remotely control the movements of my body: it is possible to imagine that Amy's device is so sophisticated that the case where she controls my body is perceptually indistinguishable from the case where I get up of my own accord. This shows that exercises of two-way power do not come with a distinctive perceptual marker (which is not to say that, in most cases, there aren't perceptual clues, or rules of thumb, which enable us to work out whether a two-way power is exercised or not). On the other hand, there is certainly a phenomenological

difference between cases where I move my body voluntarily and cases where my body moves due to some involuntary spasm – this phenomenological difference may be the source of our idea of two-way power.

However, even if Locke was right to think that acquisition of the concept of *two-way power* requires 'reflection on what passes in ourselves', Locke was wrong to conclude from this that we cannot directly perceive *active* powers being exercised by material bodies. Active power and two-way power are not the same thing. Furthermore, activity and passivity are two sides of the same coin: a substance cannot manifest a passive power or liability unless some substance is manifesting an active power or ability. It is plausible to think that if we can directly perceive one, we can directly perceive the other. It seems that Locke confused active power with two-way power.

In my opinion, it is not only seventeenth century philosophers who make this confusion. Many twentieth century philosophers assume that substances which are 'moved to move' are not *really* active. In other words, they assume that when a substance is directed to cause some change by some other substance acting upon it, as the stone is when someone throws it, this substance is not active, but passive.

Agent-causationists, who maintain that an irreducible notion of agent causation is essential for understanding agency, explicitly make this assumption. For example, Richard Taylor commits himself to the view that inanimate objects are never truly *active* and are never agents: 'a man is sometimes an agent who originates a change, and is not, like a match, merely a passive object which undergoes change in response to other changes' (1966, p.122). Taylor denies that a match can be an agent because a match cannot 'wreak changes in itself', what a match does is always a response to the circumstances it is in and what's acting upon it. A man, in contrast, '*can* bring about such a change as a motion of his arm quite by himself' (p.122). I think the notion of 'wreaking changes in oneself' is parallel to the notion of self-movement. So, Taylor is claiming, rightly, that inanimate objects like matches are not self-movers. However, Taylor takes this to imply that inanimate objects are not agents and are not active.

Furthermore, Hyman (2015) suggests that twentieth century philosophers have confused activity with *voluntariness*. For example, Hyman quotes Ryle:

> Very often we oppose things done voluntarily to things suffered under compulsion. Some soldiers are volunteers, others are conscripts; some yachtsmen go out to sea voluntarily, others are carried out to sea by the wind and tide. […] So sometimes the question 'Voluntary or involuntary?' means 'Did the person do it or was it done to him?' (Ryle, 1949, pp.73-74)

Granted, an action's being voluntary is not the same as an action's being an exercise of two-way power. Hyman defines voluntariness in the following way: 'an act is voluntary if it is due to choice as opposed to ignorance or compulsion' (2015, p.7). So, ignorance and compulsion cancel voluntariness. However, ignorance and compulsion may not strip an agent of two-way power. Whether ignorance or compulsion strips an agent of two-way power depends on the nature of the ignorance or compulsion. For example, if I kick off the covers in my sleep, I do not do so knowingly, but in this case, my total lack of awareness strips me of a two-way power to kick off the covers – if my leg moves the right way I will kick off the covers, if it doesn't, I won't; in either case, it won't be up to me. However, if I do not realise I am tapping my foot because I am not paying attention, it is less clear that this ignorance strips me of two-way power to tap my foot even though it does render my tapping non-voluntary. Similarly, if I am compelled to spill my coffee because you jiggle my hand, you strip me of the power not to spill my coffee – this kind of compulsion robs me of my two-way power. However, when a mugger compels me to give me his wallet by threatening to kill me, I do not lose my two-way power, even though when I hand over my wallet, I do not do so voluntarily. So, an action can be an example of settling without being an example of voluntariness. However, voluntariness is not always distinguished from settling: sometimes 'voluntary' is taken to be synonymous with 'up to the agent'. When Ryle states that 'sometimes the question 'Voluntary or involuntary?' means 'Did the person do it or was it done to him?" the sense of 'voluntary' he has in mind is supposed to mean 'not forced to' in the way that a yachtsman might be forced out to sea, and not in the way a conscript is forced to join the army (p.74). So, it is plausible that, at least

sometimes, when philosophers contrast voluntariness with passivity, they are also contrasting *settling* with passivity, and hence presupposing that anything that is not up to us is something we undergo.

Evidence that causal theorists also assume that substances which are 'moved to move' are not *really* active can be found in their claim that actions are bodily movements. Causal theorists, following Davidson, take actions to be movements of one's body. Agential power is, first and foremost, a power to move one's body. If 'one' and 'one's body' are taken to be one and the same, then defining actions as movements of one's body already assumes that all action is self-movement. And even if 'one' and 'one's body' are distinguished – for example, if 'one' is taken to be a person considered as something which *has* a body but which is not identical with it – assuming that action is movement of one's body rules out the possibility that substances which cannot be said to have bodies can act. So, implicit in the causalist's claim that actions are bodily movements, is the assumption that activity is self-movement, and moved-movement is passivity.

If one thinks that only two-way powers are active powers, and non-settling action is really passivity, then one will naturally suppose that the target of theory of action will not include actions by inanimate objects, and hence, one will not think these examples are counterexamples to the causal theory of action. However, it is a fallacy to conclude from the fact that some substance in certain circumstances cannot exercise *two-way* powers that this substance cannot exercise *active* powers. The causal theory of action assumes from the outset that the target of a theory of agency, or action as such, needs to be delineated via the concept of intention, or 'intentional under a description', because the actions of inanimate objects, animals, and (perhaps) some sub-intentional action and spontaneous expressions of emotion are not sufficiently distinct from passivity to qualify as the target of a theory of action. But it is wrong to presuppose that when a substance does not personally control what goes on with them this is not activity, and hence it is wrong to exclude the examples given in sections 2.2.1 to 2.2.3 as targets of a theory of action.

2.2.6 Limiting the Scope of the Causal Theory of Action

A third way a causal theorist might respond to the three sorts of counterexample described above is to accept the counterexamples but deny that the causal theory of action was ever meant to be a theory of action in its most general sense. Causal theorists can claim that the causal theory of action is meant only as a metaphysical account of a *special kind* of action. Indeed, the causal theory of action is usually presented as a theory of specifically *intentional* action.[15] According to this response, the causal theory of action does not seek to answer the question 'what makes an event an instance of agency?', rather it seeks to answer the more specific question 'what makes an event an instance of *intentional* agency?'. The causal theory of action claims only that *intentional* agency is nothing over and above some special kind of event-causation, and the agent's involvement in the causality of *intentional* action can be reduced to the agent's mental state's involvement in the causation of an (intentional) action.

Insisting that it is only *intentional* agency, and not agency more generally, which reduces to a *special kind* of event-causation (namely the kind that involves mental events and states), is not necessarily to give up on the idea that agency in general is nothing over and above event-causation. I have already noted that the fact that the agency concept cannot be analysed in terms of event-causation doesn't rule out the possibility that the worldly phenomenon covered by the agency concept is nothing over and above event-causation. It is possible for adherents of the causal theory of action to maintain that, ontologically speaking, all agency is nothing over and above event-causation, but conceptually speaking, agency-in-general is a primitive notion. That is, causal theorists could maintain that agency as such is not associated with any *special kind* of event-causation,

[15] As indicated by its title, *Springs of Action: Understanding Intentional Behavior*, Mele's 1992 book on the causal theory of action specifically concerns *intentional* activity. In his 2003 book, *Motivation and Agency*, Mele seeks only to defend 'a popular causal perspective on intentional action', not action more generally (p.5).

even though every instance of agency is nothing but some sort of event-causation or other. The concept of agency cannot be understood in terms of any special kind of event-causation – but whenever agency is demonstrated, that eventuality is nothing more than some form of event-causation.

The trouble with this response is that it weakens the case for thinking that agency reduces to event-causation. If there is just no saying what sort of event-causation is agency and what sort is not, why should we believe that agency reduces to event-causation at all? One reason might be to avoid ontologically profligate metaphysical theories. However, in this case ontological economy comes with a significant explanatory cost, which is that we are forced to concede that agency is a primitive concept, that is, we can say nothing about what it is for something to be an agent.

2.3 Intentional and Voluntary Passivity

The second error of the causal theory of action is its failure to accommodate the fact that, for some agents, their agential power can be demonstrated in passivity as well as in activity. The agents who can demonstrate their agential power in passivity as well as in activity are self-movers, i.e. agents who possess two-way powers to act. As stated above, two-way powers are powers with both a positive and negative manifestation. A two-way power to φ is manifested positively when the agent φ's and is manifested negatively when the agent does not φ. This means that two-way powers can be manifested by agents who remain passive. As mentioned above, not *every* case where an agent does not φ counts as a negative exercise of her two-way power to φ. Some important examples of failures to act which do count however are cases of intentionally refraining from performing some action, and voluntarily allowing something to be done to one.

Hornsby and Hyman have noted the importance of intentional refrainment and voluntary passivity. Hornsby has also suggested that intentional refrainment fits poorly within the causal theory of action's account of agency. Hornsby uses examples of intentional refrainment to show that agency can be demonstrated even when there is no action:

> But someone can do something intentionally without there being any action that is their doing the thing. Consider A who decides she shouldn't take a chocolate, and refrains from moving her arm towards the box; or B who doesn't want to be disturbed by answering calls, and lets the telephone carry on ringing; or C who, being irritated by someone, pays that person no attention. Imagining that each of these things is intentionally done ensures that we have examples of agency in a sense that Davidson's claim brought out. But since in these cases, A, B and C don't move their bodies, we have examples which the standard story doesn't speak to. (2004a, p.5)

In these three cases, the agent's demonstration of agential power is not an action in the sense of a 'positive performance'. In these examples, there is no action, so what makes these cases demonstrations of agency cannot be expressed in terms of the mental causation of an action. This means that the causal theory of action is unable to explain what makes these cases examples of agency.

Hyman (2015, pp.10-11) uses an example of a child being picked up by a parent to show that sometimes passivity is voluntary – with respect to being picked up, the child is passive, but being picked up is voluntary for the child. I think that this example is also an example of a two-way power being manifested negatively: specifically, the child is manifesting her two-way power to resist being picked up (e.g. by pushing away the parent) negatively. So, even though the child is, so to speak, not doing anything, but rather, letting something happen to her, she is demonstrating an agential power. In this case, there is an action, but it is the action of the parent not the child. So, it seems impossible to explain how this action is an exercise of *the child's* agential powers by pointing to the action's standing in a causal relation to the child's mental states, as the causal theory of action would have one do, while at the same acknowledging that the action is not an action of the child. The correct way to describe the agency of this example is, I think, as follows: a single event, the child's being picked up by its parent, is an instance of an exercise of active power on the part of the parent, and of passive power on the part of the child, and it is an instance of the parent's two-way power to pick a child up, and of the child's two-way power to resist being picked up. In other words, to correctly describe the agency

demonstrated in this example, one needs to talk about the various causal powers which are exercised by the two substances involved; one cannot only talk about what events are causally related to what other events.

Cases like Hornsby's and Hymans's are not counterexamples to the causal theory of action, rather, they are examples which indicate that the causal theory of action cannot possibly tell the whole story about agency in terms of causation of an action by a mental event. As Hornsby puts it, 'perhaps that view – of causality operating through items linked in causal chains – is the correct view of causal truths in some areas', but examples of agency where there is no 'positive performance' suggest that 'the truths that make up the phenomenon of agency seem not to belong in a world in which causality operates only in such a manner' (2004a, p.10). The basic point here is that by attempting to reduce the agent's involvement in the causality of action to the agent's mental state's involvement in the causation of an action, the causal theory of action cannot accommodate the fact that self-movers can exercise their agential powers by remaining passive, and hence by not bringing about an action.

A causal theorist might respond to this objection by once again limiting the scope of their theory. They may claim that the causal theory of action is only an account of *action*, and not voluntary passivity or intentional refrainment. However, as Hornsby points out, this response will not work as long as the adherent of the causal theory of action takes agency to be delineated by its appropriateness for receiving rationalising explanations. This is because: 'A's, B's and C's cases count as agency on this reckoning, because one can construct tales of what each of them believed and desired which will appropriately explain their doing their things – not moving, letting the phone ring, not paying attention to X' (2004a, pp.6-7). And, it would do no good to insist that rationalising explanations of actions come in two sorts, because 'when we ask why someone did something, expecting to learn about what they thought or wanted, we don't always need to consider whether or not there was a positive performance on their part; explanation can carry on in the same vein, whether there was or not' (p.7).

2.4 The Disappearing Agent

I have argued that the causal theory of action cannot accommodate two facts about agency. First, the causal theory of action wrongly entails that inanimate objects, and animals, and human beings when their actions cannot be causally traced back to any intentional state, do not act. I argued that causal theorists have failed to see this as an error of their theory, because they have confused activity with self-movement – that is, with the agency of substances which personally control what is going on with them – and taken the agency of substances which do *not* personally control what they do to be no different from passivity. Second, the causal theory of action cannot accommodate the fact that, some agents, namely agents who possess two-way powers to act, can exercise their agential power by remaining passive. In this section, I will explain why the source of these two errors is the causal theory of action's reductive ambitions. To explain how it can be that the agency concept covers the 'doings' of inanimate objects and animals as well as the 'doings' of human beings and that for some agents, their agential power can be demonstrated in passivity as well as in activity, one must accept what the causal theory of action denies, namely that the exercise of agential power cannot be reduced to a special kind of event-causation.

Several philosophers have raised an objection against the causal theory of action which has come to be known as 'the disappearing agent problem'. According to this objection, it is an essential part of our concept of agency is that, in acting, the agent herself brings about changes, but on the causal theory of action the agent is merely the area within which mental states or events cause bodily movements. In other words, in the causal theory of action's metaphysical articulation of the causality of action, the agent herself 'disappears', and this cannot be right, because a world where agents themselves play no causal role in bringing about the results of their actions is a world where there are no actions. For example, Abraham Melden complained that:

> It is futile to attempt to explain conduct through the causal efficacy of desire – all that can explain is further happenings, not actions performed by agents.

> The agent confronting the causal nexus in which such happenings occur is a helpless victim of all that occurs in and to him. There is no place in this picture of the proceedings either for rational appetite or desires, or even for the conduct that was to have been explained by reference to them. (1961, pp.128-129)

Thomas Nagel suggested that when causal reality is viewed as nothing but chains of causally related events, 'everything I do or that anyone else does is part of a larger course of events that no one "does" but that happens' (1986, p.113), and 'the agent and everything about him seems to be swallowed up by the circumstances of action; nothing of him is left to intervene in those circumstances' (1986, p.114). And Hornsby has argued that 'agency cannot be portrayed in a picture containing only psychological states and occurrences and no agent making any difference to anything' (2004a, p.12). As she puts it elsewhere:

> The role of agents in a world of events is evident only when it is appreciated that agents cause things – things that ensue from their actions. It seems unthinkable that agency should be manifest from any point of view from which it is impossible to locate agents. (2004b, p.176)

It might seem to an adherent of the causal theory of action that the disappearing agent problem is begging the question. The core proposal of the causal theory of action is that agency *can* be located in a world where causal reality consists of nothing but chains of causally related events. Causalists theorise that for an agent to be something which brings about change is for the mental states or events of that agent to cause an event. The disappearing agent problem can seem like a straightforward denial of the causal theory of action's core proposal.

However, I do not think that the disappearing agent problem begs the question against the causal theory of action. Rather, the disappearing agent problem is the correct diagnosis of the difficulties the causal theory of action faces, such as wrongly entailing that agency cannot be found where

intentionality is lacking, and being unable to accommodate the fact that some agential powers can be manifested in passivity.

If one assumes that causal reality is nothing more than a chain of causally related events, and therefore that the causal truths about agency are truths concerning causation of and by certain events, then any distinction crucial to our conception of agency must be a distinction between different types of event-causation. This is why the distinction between intentional actions and other events becomes very important, because there is some plausibility to the idea that causation by mental states or events is key to understanding *this* distinction.[16] But even if the distinction between intentional action and non-intentional action is best understood via the notion of causation by mental states or events, it is a mistake to think that understanding this distinction is to understand the distinction between what is agency and what is not – because not all agency is intentional. However, if agency is not understood in terms of event-causation of a special kind, then the causal theorist, because of his metaphysical commitments, is forced to say that the agency concept cannot be understood in other terms at all. The metaphysics of causation presupposed by the causal theory of action compels an action theorist to seek to either understand agency in terms of a distinction between different types of event-causation or to admit that agency is a primitive concept which cannot be understood in other terms at all.

Furthermore, if one assumes that for a substance to be something which exercises agential power is for the mental states or events of that substance to cause an event, then there appears to be no way to explain how it can be that sometimes substances exert personal control over what goes on even while they remain passive. If being a substance capable of agency is just to possess mental state which cause actions to happen, then how can we account for the fact that

[16] Although I will question this proposal in chapter 7.

some substances demonstrate their agency even when they do not perform an action, but instead allow things to be done to them?

To understand how it can be that the agency concept covers the 'doings' of inanimate objects and animals as well as the 'doings' of human beings, and that for some agents their agential power can be demonstrated in passivity as well as in activity, it is necessary to make the two distinctions I described in section 2.1. To understand how it can be that both the stove, when it heats the soup, and the child, when his parent picks him up, are demonstrating agential power, one must distinguish between active and passive powers *and* between one-way and two-way powers. The idea that understanding agency requires making multiple distinctions between different types of power fits poorly with the causal theory of action's ambition to understand agency via a single divide: between event-causal sequences that involve intentional states and those which do not.

What this suggests is that seeking to understand agency in terms of a distinction between different types of event-causation cannot be done without misconstruing the agency concept. To avoid misconstruing the agency concept one must accept that the exercise of agential power cannot be reduced to a special kind of event-causation. Instead, one must accept that, as Hornsby puts it, 'human beings are actually ineliminable from an account of their agency' (2004b, p.182). One must acknowledge that, in acting, the agent herself plays a causal role and attempting to reduce the agent's involvement in the causality of her action to her mental states' involvement in the causation of her action is to render the agent merely the setting for events to cause other events.

Hornsby (2004a, 2012) has argued that what's needed to properly understand the causality of agency – and in particular to recognise the essential role of the agent in the causality of action – is a metaphysical framework which provides intellectual space for thinking of causation as something other than a relation between events. Without such a framework, it is impossible to see how the causality of action might be something other than a causal relation between mental event and action, and instead something that casts the agent as a causal

player, rather than merely the setting for events to cause other events. In other words, Hornsby suggests that to properly understand agency we must make a radical departure from the Humean approach to causation which the causal theory of action presupposes. I agree, and it will be the aim of chapters 2-5 to outline an alternative to the Humean approach to causation and show how this enables a more successful understanding of agency.

3

The Humean Approach to Causation

The causal theory of action presupposes a metaphysics where causation is always, everywhere a relation between events. This approach to causation compels the causal theorist to seek to understand agency in terms of a distinction between different types of event-causation. However, I argued in the previous chapter that the causal theory of action misconceives agency. I suggested that an adequate theory of agency would have to accept what the causal theory of action denies, namely that the exercise of agential power cannot be reduced to a special kind of event-causation. In other words, presupposing that causal reality is nothing more than a chain of causally related events, and then trying to locate agency within this worldview, will not succeed; a worldview where causal reality is nothing more than a chain of causally related events is one which eliminates agency.

The purpose of this chapter is to articulate the approach to causation presupposed by the causal theory of action. I call this approach to causation the 'Humean' approach because it can be seen as inspired by Hume's discussion of causation. In section 3.1, I will describe three essential commitments of the Humean approach, and show how two theories often described as Humean, the regularity theory of causation and David Lewis's counterfactual theory of causation, abide by these commitments. In section 3.2, I will discuss some prominent theories of causation which reject some, but not all, of the commitments which characterise the Humean approach. In section 3.3, I will outline my preferred non-Humean view, which rejects all three commitments of the Humean approach.

3.1 The Humean Approach

A theory of causation follows what I'm calling 'the Humean approach to causation' if and only if the theory incorporates and embraces commitment to the following three theses:

i. **Reductivism** – causation, as it exists in the world independently of our thinking about it or knowledge of it, is exhaustively constituted by non-causal states of affairs.

ii. **Denial of Powers** – a primitive concept of power (that is, one that cannot be analysed or understood in other terms) is not needed to understand the nature of causation.

iii. **Relationalism** – causation is always and everywhere a relation; the worldly phenomenon which is referred to by our concept 'causation' is not ontologically diverse.

Although I have called the approach to causation characterised by commitment to these three theses 'Humean', it is actually unclear whether Hume himself really subscribed to them. For example, Galen Strawson (1989) argues that Hume was not a reductivist about causation. This interpretation of Hume is controversial,[17] but it is worth remembering that, although I'm calling the approach to causation characterised by commitment to Reductionism, Denial of Powers and Relationalism 'Humean', Hume himself might not have endorsed a 'Humean' theory.

Hume thought that the proper aim of empirical science was to systematise and codify observable events by devising laws which summarise general patterns. However, Hume doubted that this system of laws provided a basis for even probable conjectures about how events will play out in the future. Laws of nature do not provide us with any kind of insight which might tell us why certain events regularly follow on from others, which we could use to predict

[17] See Millican (2007) and Beebee (2007).

how unobserved events will unfold. Hume acknowledged that we instinctively think unobserved cases will be like observed cases, but he argued that this thought has no rational basis – it is just a psychological habit of ours. In particular, Hume denied that we have knowledge of any underlying powers possessed by things, which might explain why certain events *must* follow from certain others.

The idea that there are necessitating connections in nature, whereby an object with certain powers 'must' behave in certain ways in certain conditions, is not 'in any instance, attained by reasonings *a priori*' (1975, p.27) because, as Hume rightly recognised, it is never *logically* or *metaphysically* necessary that an object with certain powers will behave in certain ways, even when the conditions are right for the object's powers to be manifested, because it is always possible that something could intervene and prevent the object from producing its usual effect. For example, it is not *logically* or *metaphysically* necessary that when I drop my pen it will hit the ground, even though that is what I expect will happen, because, as unlikely as it may be, a sudden gust of wind could pick the pen up and deposit it on the table, or my cat might run to it and catch it before it hit the ground. Interventions are always possible, so we cannot deduce *a priori* what an object will do from what the object is like.

Hume also argued that efficacy, or power, or necessitation between cause and effect, is not an aspect of the external world which we can experience. Powers which actuate the operation of things are never directly experienced, according to Hume. We can perceive a thing's properties, what it is like, but not what it is capable of doing. The necessary connections the operation of such powers would give rise to are also never perceived. According to Hume: 'one event follows another; but we never can observe any tie between them. They seem conjoined, but never connected' (1975, p.74). Knowledge of 'ties' in nature, if it is to be had at all, is not derived from experience. Hume concluded that the idea of power or 'necessary connexion' comes neither from 'outward sense' nor the 'mere operation of thought'. Instead, Hume suggested that when we repeatedly experience events of one type being followed by events of another

type, we come to *expect* an event of the second type when we experience an event of the first, and this internal feeling of expectation is the impression from which this idea of necessitation between cause and effect arises.

Hume's argument that we can have no knowledge of things' underlying powers, which might explain why events unfold in regular ways, is questionable. Nevertheless, powers and natural necessity have long been regarded as epistemically suspicious and ineffable. In response to Hume's arguments, philosophers with empiricist sympathies have either denied that causation exists in the external world, independently of our thinking of it (e.g. Russell, 1912), or they have claimed that even if causation does exist in the external world, we can have no knowledge of it, or they have sought to reduce causation, as it exists in the external world, to something more empirically respectable. As Helen Beebee puts it, empiricists have sought to 'show that in using the term 'causation' we are not claiming to be referring to some mysterious, ineffable, empirically unrespectable feature of the world at all, but to something that passes the empiricist credibility test (whatever that might be)' (2006, p.510). So, we find in Hume inspiration for Reductivism, the view that all causal facts are translatable (perhaps with the aid of *a posteriori* identities, or metaphysical reductions) into facts about non-causal states of affairs, and Denial of Powers, the view that a primitive concept of power is not needed to understand the nature of causation.

Hume also spoke of 'cause and effect' as a 'relation' or a 'connexion' (1975, pp.26-27), so here we find inspiration for Relationalism. Most philosophers working on causation, not only those with empiricist or Humean sympathies, accept Relationalism, at least implicitly. Most have assumed that providing a theory of causation is a matter of explaining what a relation must be like to be a causal relation. The possibility that causation may not fit into a

single ontological category is rarely taken seriously.[18] So, few have recognised that there is anything particularly 'Humean' about Relationalism.

In the remainder of this section, I will present two examples of a Humean theory of causation: the regularity theory of causation and Lewis's counterfactual theory of causation.

3.1.1 The Regularity Theory of Causation

The paradigm example of a theory of causation which follows the Humean approach is the regularity theory of causation. The regularity theory holds that causation, as it exists in the world independently of our thinking about it or knowledge of it, is exhaustively constituted by certain relations of spatiotemporal contiguity which obtain with regularity. More specifically, the regularity theory holds that causation is a relation of spatiotemporal contiguity between two events, *c* and *e*, where *c* occurs before *e*, and where all events of the same type as *c* are regularly followed by events of the same type as *e*. The regularity theory as stated above faces problems and in response more sophisticated versions of the regularity theory have been proposed.[19] However, for my purposes here, I do not need to examine these more sophisticated versions of the regularity theory – the most basic version will suffice as a demonstration of a theory of causation which observes the three Humean commitments.

The regularity theory is a reductive theory of causation, and so embraces Reductivism. As Stathis Psillos puts it:

> [The regularity theory] is typically seen as offering a *reductive* account of causation. As with all reductive accounts, causal talk becomes legitimate, but it does not imply the existence of a special realm of causal facts that

[18] Notable exception: Steward (2012, see especially pp.212-216).
[19] For example, Mill (1843), Mackie (1974), and Baumgartner (2008) have all offered more sophisticated versions of the regularity theory.

> make causal talk true, since its truth conditions are specified in non-causal terms, that is, in terms of spatiotemporal relations and actual regularities. (2002, p.4)

Indeed, the regularity theory's promise to offer a reductive account of causation is part of its appeal. The main argument for adopting a regularity theory is that the theory accounts for causation while at the same time avoiding ontological commitment to anything ineffable, or otherwise empirically suspect, like natural necessitation or power. The idea is that the regularity theory of causation – or at least a suitable worked-up version of it – provides everything we would want from a theory of causation, without positing the existence or powers or a *sui generis* kind of necessity, and as the best metaphysical theories are the theories which serve our explanatory aims without bestowing unnecessary ontological burdens, this is reason enough to prefer the regularity theory.

According to the regularity theory, what ascriptions of power, or statements about what a thing can do, actually *mean* (if they are not false or nonsense) is that the behaviour of the object to which the 'power' is attributed is regular in a certain way. That is, it might be true to say some object has a power, but what makes such a statement true will be some fact about the arrangement of the spatiotemporal mosaic of instantiations of intrinsic, qualitative, categorical properties. Thus, the regularity theory denies that *power* is a primitive concept, i.e. it accepts Denial of Powers.

The mosaic metaphor is David Lewis's means of describing the Humean metaphysics presupposed by the regularity theory. In more detail this metaphysics says:

> [...] in a world like ours, the fundamental relations are exactly the spatiotemporal relations: distance relations, both spacelike and timelike, and perhaps also occupancy relations between point-sized things and spacetime points. And it says that in a world like ours, the fundamental properties are local qualities: perfectly natural intrinsic properties of points, or of point-sized occupants of points. Therefore it says that all else supervenes on the spatiotemporal arrangement of local qualities throughout all of history, past and present and future. (Lewis, 1994, p.474)

As Lewis puts it in the introduction to his *Philosophical Papers (vol. II)* 'all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another (1986, p.ix). Jonathan Schaffer describes the Humean worldview slightly differently: Schaffer writes that, for the Humean, the world is 'history' i.e. 'the fusion of all events throughout space-time' (2007, p.83).

Lewis's metaphysics, which he calls 'Humean Supervenience', is, as Beebee puts it, 'extraordinarily austere' (2006, p.513). It admits very little as fundamental. Just points, the properties instantiated by (at?) those points and the relations between those points (and for Lewis properties and relations are themselves nothing more than classes of points). On Lewis's picture, even facts about what substances exist are supervenient on the mosaic, because, on Lewis's view, substances are four-dimensional objects composed of temporal parts, which are each themselves collections of space-time points. Schaffer's version of the Humean worldview seems, at first, to be a little less austere than Lewis's, as Schaffer seems to admit that there are such things as *events*, i.e. things that *occur*, and not just instantiations of properties by points. As Schaffer puts it:

> Each individual event is a concrete particular with an intrinsic nature – what occurs in some region of space-time. History is the whole of this – it is what occurs in all of space-time. History is the total pattern of events. Each event is like a bit of a frame in the movie, and history is the whole picture. (2007, p.83)

However, Schaffer uses, as an example of an event, 'an instance of red' (2007, p.88) – so it is not clear that Schaffer thinks there is a significant metaphysical distinction between events and property instances or property instantiations.

The regularity theory can also be assumed to be committed to Relationalism, if the regularity theory is intended to be a comprehensive and complete account of causation, since the regularity theory proposes that causation is a relation, namely a relation of spatiotemporal contiguity which obtains with regularity. Most proponents of the regularity theory do not explicitly claim that their theory is supposed to be comprehensive and complete, but similarly, they do not explicitly claim that it is not. In a way, the regularity

theory is committed to Relationalism by default because proponents of the regularity theory simply do not consider whether causation might not be a relation.

3.1.2 Counterfactual Theory of Causation

Another example of a Humean approach to causation is Lewis's (1973a, 1973b) counterfactual theory of causation. Lewis's theory of causation analyses causation in terms of counterfactual dependence. This theory exploits the intuition that causes are that which made the difference to the occurrence of the effect; that is, had the cause not occurred the effect wouldn't have occurred either. Lewis developed this idea by analysing the causal relation as the ancestral of a counterfactual dependence relation. So, an event *c* stands in a causal relation to another event *e* if and only if *e* counterfactually depends on *c,* or *e* counterfactually depends on an event which counterfactually depends on *c*, or *e* counterfactually depends on an event which counterfactually depends on an event which counterfactually depends on *c*, etc. As with the regularity theory, Lewis's counterfactual theory has been modified in light of objections raised against the original version of the theory.[20] However, again, I do not need to consider the more sophisticated versions of the theory for my purposes here.

Lewis's counterfactual theory's status as Humean depends, in part, on Lewis's theory of modality. Lewis opts for a possible world semantics for counterfactuals. So, a counterfactual like 'if *c* had not occurred, then *e* would not have occurred' is true if and only if *e* does not occur at the closest possible world where *c* does not occur. How 'close' a possible world is to the actual world depends on how *similar* that world is to the actual world. For Lewis, similarity

[20] See, for example, Lewis (2000), McDermott (2002), Ganeri, Noordhof and Ramachandran (1996), and Sartorio (2005).

between two possible worlds is determined by what particular states of affairs obtain at the two worlds and what the laws of two worlds are. So, world w^1 is more similar to world w^2 the more states of affairs w^1 has in common with w^2 and the more laws w^1 has in common with w^2.

If one went along thus far with Lewis's semantics for counterfactuals, but thought that laws of nature were brute facts about what powers things have, or facts about primitive 'necessitation' relations holding between universals, then even if one opted for an account of causation where causation is reduced to counterfactual dependence, the resultant theory of causation would not be Humean. This is because, on such a view, the truth of counterfactual conditionals depends on similarity rankings of possible worlds which in turn depends on brute facts about powers, or a *sui generis* form of necessity. Such a view would thus seem to contravene Denial of Powers.

However, Lewis gives an account of laws of nature which does not presuppose the existence of powers or anything over and above the spatiotemporal mosaic of instantiations of intrinsic, qualitative, categorical properties. For Lewis, laws of nature are simply regularities which are deducible from axioms in an explanatory system that best balances simplicity and strength. An explanatory system picks as few general truths as possible to serve as axioms – the fewer, the simpler – then deductively derives further general truths from these. The more general truths the system deductively entails, the stronger the system.

As Beebee (2006) points out, because Lewis seeks to analyse causation without assuming the existence of any kind of worldly necessitation, and ends up turning to regularities in order to fulfil that mandate, Lewis's counterfactual theory of causation has a lot in common, metaphysically speaking, with the regularity theory. On both theories, the worldly structures that make true causal claims are, in the end, regularities. And, just like the regularity theory, Lewis's counterfactual theory does not posit any kind of entity, or deeper fact (like facts about what powers things have or what is a natural necessity), which grounds or

explains why regularities hold, or why certain counterfactual conditionals are true.

Lewis's counterfactual theory of causation, insofar as it attempts to reduce causation to counterfactual dependence, and provides a semantics of counterfactuals which does not at any point appeal to irreducibly modal or causal facts, embraces Reductionism. It also embraces Denial of Powers, as only the concept of counterfactual dependence is needed to understand causation, and counterfactual dependence in its turn can be understood without a primitive concept of power. What about Relationalism? Again, Lewis's counterfactual theory endorses Relationalism by default as Lewis analyses causation in terms of a relation, and seems neither to confirm nor deny that his theory is intended as comprehensive.

There are, I suspect, other theories of causation which adhere to the three Humean statutes which I have not discussed,[21] but I hope my brief discussion of the regularity theory and counterfactual theory of causation has been sufficient to make clear what a Humean account of causation is.

3.2 Partially Non-Humean Approaches

I will call a theory of causation 'partially non-Humean' if the theory, explicitly or implicitly, rejects some but not all of the three Humean theses. I have already mentioned one partially non-Humean approach: a kind of counterfactual theory of causation, which attempts to reduce causation to counterfactual dependence, but then maintains that counterfactual conditionals depend for their truth on brute facts about what powers things have, or a *sui generis* form of necessity. This theory would observe Reductionism and Relationalism, but not Denial of Powers. A similar kind of partially non-Humean approach which observes

[21] For example, Davidson's (1967) nomological account and versions of the probability account of causation (Pearl, 2000; Hitchcock 1993).

Reductionism and Relationalism, but not Denial of Powers, is a theory which reduces causation to a relation of spatiotemporal contiguity that is an instance of a law-backed regularity, but which gives an account of laws that makes essential reference to power or necessitation (e.g. Armstrong, 1999). However, most partially non-Humean theories of causation defended in the literature on causation are theories which reject Reductionism.

3.3.1 Manipulability Theories of Causation

Manipulability accounts of causation are partially non-Humean. However, precisely which of the three Humean theses manipulability accounts reject depends on the type of manipulability account.

Take, for instance, Georg Henrik von Wright's agency-based account of causation. Von Wright believed that an event *c* is the cause of event *e* if and only if bringing about *c* is a way for an agent to bring about *e*, that is only if *e* can be considered the result of the action of bringing about *c*:

> […] to think of a relation between events as causal is to think of it under the aspect of (possible) action. It is therefore true, but at the same time a little misleading to say that if *p* is a (sufficient) cause of *q*, then if I could produce *p* I could bring about *q*. For *that p* is the cause of *q*, I have endeavoured to say here, *means* that I could bring about *q*, if I could do (so that) *p*. (1971, p.74)

An important objection to von Wright's theory is that it is problematically circular because agency is a causal notion: *producing* and *bringing about* are causal concepts, hence agency-based theories purport to analyse causation in terms of causation. Von Wright denied that this account of the causal relation is circular, because he held that the relation between an action (e.g. cutting of the cake) and its result (the cake's coming to be cut) is not a causal relation, it is rather a logical one (if the cake doesn't come to be cut, then no-one cut it – the cutting-of-the-cake action did not take place):

> I am anxious to *separate* agency from causation. Causal relations exist between natural events, not between agents and events. When by *doing p* we *bring about q*, it is the *happening* of p which *causes* q to come. And p has this

> effect quite independently of whether it happens as a result of action or not. (1974, p.49)

Insofar as von Wright takes himself to be providing an analysis of causation, understood as a relation, in terms of a non-causal relation (between an action and its result), he is committed to Reductivism.

However, even though von Wright denies that action is itself causation, his account of causation relies on the modal concept: what an agent *could* bring about. Understanding causation also requires that we possess a concept of how things would have been had no person intervened. On von Wright's account, causation exists where and when human beings 'interfere with the course of the world thereby making true something which would not otherwise (i.e., had it not been for this interference) come to be true of the world at that stage of history' (1974, p.39). In acting, we switch the world from the course it would have followed to another course. So, von Wright's account helps itself to modal concepts such as how things would have been (had no-one interfered) and how things can be (if someone interferes), and doesn't appear to offer any analysis of these modal concepts. As J.L. Mackie puts it, 'the natural necessity, the power, the counterfactuality which are among the most puzzling aspects of causation – especially if we start with atomic states or events – *are being accepted, without analysis,* as located partly in human action, partly in the non-causal persistence of "normal" states of affairs' (1976, p.215, emphasis added). So, von Wright's view seems to reject Denial of Powers, in that his account of causation appeals to a primitive concept of how things can be.

I think von Wright is correct to insist that the relation between an action and its result is logical and not causal. I also think von Wright is right to sharply distinguish between agency on the one hand and causal *relations* on the other – von Wright is correct to claim that to demonstrate agency is not for an agent to stand in a causal relation to an event. However, I do not think, as von Wright does, that this entails that agency is not a causal phenomenon. Von Wright does not recognise this because he subscribes to Relationalism: the view that causation is always, everywhere a relation. Von Wright's view can be thought of

as abiding by the following reasoning: causation is the relation between cause and effect, agency is not a relation between cause and effect, therefore, agency is not causation. This argument is sound only if Relationalism is true. So, von Wright accepts Reductionism and Relationalism, but rejects Denial of Powers.

Another kind of agency-based manipulability account of causation was put forward by Peter Menzies and Huw Price: 'an event [*c*] is cause of distinct event [*e*] just in case bringing about the occurrence of [*c*] would be an effective means by which a free agent could bring about the occurrence of [*e*]' (1993, p.187). And an event *c* is an effective means by which a free agent could bring about occurrence of *e*, just in case the probability of *e* occurring given that *c* was brought about by a free agent is greater than the unconditional probability of *e* occurring.

The circularity objection that von Wright faced can be directed against Menzies and Price's view as well. Menzies and Price respond to the circularity objection in the following way:

> The basic premise is that from an early age, we all have direct experience of acting as agents. That is, we have direct experience not merely of the Humean succession of events in the external world, but of a very special class of such successions: those in which the earlier event is an action of our own, performed in circumstances in which we both desire the later event, and believe that it is more probable given the act in question than it would be otherwise. To put it more simply, we all have direct personal experience of doing one thing and thence achieving another. […] It is this common and commonplace experience that licenses what amounts to an ostensive definition of the notion of 'bringing about'. In other words, these cases provide direct non-linguistic acquaintance with the concept of bringing about an event; acquaintance which does not depend on prior acquisition of any causal notion. An agency theory thus escapes the threat of circularity. (1993, pp.194-195)

Unlike von Wright, Menzies and Price do not deny that agency is a causal phenomenon. What they deny is that acquiring the agency concept requires that one has already acquired the concept of causation. For Menzies and Price, even

though agency itself is an essentially causal phenomenon, the *concept* of agency is one that can be understood and grasped independently of the *concept* of causation, and because it can be independently understood, it can be used to analyse causation.

So, unlike von Wright, Menzies and Price do not intend to show how causation, as it exists in reality, is exhaustively constituted by non-causal states of affairs. Price (2017) claims that Menzies and Price (1993) aren't seeking to tell us what real structure in the world causation ought to be identified with, instead their aim is to tell us how our concept of causation is pieced together. Menzies and Price's theory thus appears to reject the Humean thesis of Reductionism. But in some respects, their account is still quite Humean: like Hume, Menzies and Price offer a psychological story about where our concept of causation comes from, and are at best agnostic about whether there is any structure existing in reality, independently of our thinking of it, to which our causation concept refers.

James Woodward (2003) argues that Menzies and Price's view is unacceptably anthropomorphic and subjectivist.[22] Because Menzies and Price invoke a concept of agency which we grasp via direct experience of our own agency at work, their theory faces a difficult problem concerning causes which cannot be manipulated by human agents. To take an example from Menzies and Price (1993, p.195), it seems to be true that movement of tectonic plates caused the 1989 San Fransisco earthquake, but it is not true that movement of tectonic plates was an event which could have been an effective means by which a human agent could have brought about the earthquake. Manipulating tectonic plates is just not within our power.

Woodward (2003) offers his own manipulability theory of causation which avoids this problem by using the concept of an intervention to analyse the

[22] See especially pp.124-125.

causal relation, rather than manipulation by a human agent. Woodward contends that a variable *c* is causally related to a variable *e* if and only if intervention on *c* leaves the relationship between *c* and *e* invariant but changes the value of *e*. An intervention is any event which 'surgically' causes the value of *c* to change, that is, by blocking all causal influence over the value of *c* the usual causal antecedents of *c* have and without causally influencing the value of *e* except through *c*. An intervention is any event which has certain causal characteristics; an intervention need not involve human agency at all (although no doubt many interventions do involve human agency).

Because Woodward's account of what it is for *c* to cause *e* involves appeal to a causal relation between an intervention and *c*, Woodward's account does not attempt to reduce causation to *non-causal* state of affairs. So, Woodward rejects Reductionism. But Woodward argues that this is not problematic, as although his account of causation is not reductive, it is still illuminating. Furthermore, it is not circular because the causal relationships to which one appeals to explain what it is for *c* to cause *e*, are not the causal relationships that explain what it is for an intervention to cause *c* to take on a certain value.

Woodward's theory is a kind of counterfactual theory of causation: whether two variables are causally related to each other depends on how the relationship between those variables would change if certain interventions were made. However, there are key differences between Woodward and Lewis when it comes to the semantics of counterfactual conditionals. The most important difference is that in Lewis's account of how we should evaluate counterfactual conditionals in causal contexts it is never necessary to appeal to causal facts. By contrast, in Woodward's account of how we should evaluate counterfactual conditionals in causal contexts we are supposed to imagine that the antecedent of the counterfactual is made true by the occurrence of an intervention, which presupposes that certain causal facts obtain.

To illustrate this point with an example, suppose event *c* caused e_1 and e_2, and e_1 and e_2 are not causally related to each other. Because counterfactual

dependence is sufficient for causation, we would want the following counterfactual to come out false:

(a) If e_1 had not occurred, e_2 would not have occurred

But in a world where e_1 does not occur, we might suppose that this was because it was not caused by *c*, i.e. because c did not occur – but in that case, e_2 would not have occurred either. This world – where e_1 does not occur because *c* does not occur – is therefore the wrong world to turn to when evaluating the truth of the counterfactual in a causal context. Lewis recommends that when we evaluate counterfactuals in a causal context we forbid 'backtracking' – i.e. we are forbidden from imagining that prior events and circumstances were also changed so as to cause the antecedent of our target counterfactual to be true. When we evaluate (a) we must imagine that a small miracle makes it the case that e_1 does not occur. So, the world we should use to evaluate the truth of (a) is a world where *c* still happens, but then, miraculously, e_1 does not occur – in such a world e_2 would still occur (because *c* would still cause it), and therefore (a) comes out false.

Woodward achieves this same result using the notion of an intervention, rather than the notion of a 'small miracle'. For Woodward, when we evaluate (a) we are supposed to imagine that an intervention occurred to make it the case that e_1 did not occur – and such an intervention, by definition, leaves all causal relationships, except those which have e_1 as effect, unchanged. Evaluating the truth of (a) thus requires assuming certain other causal relations in the situation under discussion obtain.

Even though Woodward's and Lewis's theories differ in this important way, it is not part of Woodward's theory that the truth of counterfactual conditionals depends on brute facts about powers, or a *sui generis* form of necessity. Thus, Woodward's theory is consistent with the view that counterfactual dependence can be understood without a primitive concept of power, and is therefore consistent with Denial of Powers. As for Relationalism, just as with the regularity theory and Lewis's counterfactual theory, Woodward's

theory appears to embrace Relationalism by default, as the question of whether causation could be something other than a relation is not considered.

3.3.2 Realist Theories of Causation

Strawson articulates a conception of causation which he calls Causation with a capital 'C'. To believe in the existence of Causation is to believe: 'a) that there is something about the fundamental nature of the world in virtue of which the world is regular in its behaviour; and b) that that something is what causation is, or rather it is at least an essential part of what causation is' (1989, pp.84–85). Strawson thus advocates a view which takes causation to be an entity which *grounds* the world's regularities, but cannot be reduced to regularities, or indeed any aspect of the Humean 'mosaic'. Strawson therefore rejects Reductionism. Michael Tooley (1990a) also rejects Reductionism about causation. More precisely, Tooley argues against views which hold that 'causal relations are […] logically supervenient upon non-causal properties and relations' (1990a, p.217).

The point I wish to emphasise is that Tooley is arguing specifically against attempts to reduce the causal *relation* to some non-causal relation. Furthermore, when Tooley discusses alternatives to Reductionism, it is specifically 'realism with regard to causal relations' which he considers (1990a, p.233). Similarly, Strawson (1989) is concerned to show that we should believe there is something more to the relation between cause and effect that regular succession. Elsewhere, Strawson (1987) argues that realism with respect to the external world rationally requires belief in the existence of 'Producing Causation', and Strawson takes Producing Causation to be a view about what it means to say that 'some object-involving event A caused other object-involving event B' (1987, pp.254-255). So, both Strawson and Tooley seem to be committed to Relationalism.

Tooley outlines two different sorts of realism with regard to causal relations. The first sort of realism takes causal relations to be directly observable 'not only in the everyday sense of that term, but in a much stronger sense which entails that concepts of causal relations are analytically basic' (1990a, pp.233-

234). This sort of realism would maintain, in contradiction to Hume, that causation cannot be reduced to non-causal states of affairs, but it is nevertheless something in the external world which we can observe. Tooley cites Elizabeth Anscombe as a philosopher who upholds a view like this. Anscombe (1971) suggested that we come by our primary knowledge of causality when we learn to speak and come to associate the linguistic representation of a causal concept with its correct application. An example of such a causal concept which Anscombe provides is 'infect'. Others include 'scrape, push, wet, carry, eat, burn…' (1971, p.9). She suggests that causal activities like scraping and pushing (though perhaps not infecting) are activities which we can directly perceive. Tooley argues that the fact that we know by perceptual observation that, for example, something is pushing something else, 'would not seem to provide adequate grounds for concluding that the relevant concepts are analytically basic' (1990a, p.234). Tooley's thought is that even if Anscombe is right that we know by observation that one thing is pushing another (for example), this doesn't show that what it is about the events we're seeing, which makes it the case that they are causally related, is something irreducible which we can nevertheless observe – it might be that we *infer*, from what we perceive, that causation is there.

I think that Tooley has misconstrued what Anscombe is claiming in her 1971 lecture *Causation and Determination* from which Tooley cites. What Anscombe suggests we directly perceive is not a special relation between cause and effect, but substances exerting causal power over other substances. We do not observe a cause causing an effect, we observe an agent acting on a patient. Anscombe is suggesting that an agent acting on a patient is causation, and this is in spite of the obvious truth that agent and patient are not related to each other as cause and effect. Anscombe's point is that we come by knowledge of causality when we directly perceive agents pushing patients and correctly associate what we see with the inherently causal concept *pushing*. Tooley might be right that the fact that we directly perceive agents pushing patients (for example) may not be enough to show that we directly perceive a connection between the events

which makes it the case that they are causally related. But why can't the fact that we directly perceive an interaction, like an agent pushing a patient, be enough to show that we directly perceive causation? Tooley construes Anscombe's claim incorrectly, I think, because of his commitment to a version of Relationalism which says that causation is a relation between events.

The sort of causal realism which Tooley endorses, treats 'causal concepts as theoretical concepts, so that causal relations can only be characterised, indirectly, as those relations that satisfy some appropriate theory' (1990a, p.234). The appropriate theory, Tooley (1990b) proposes, is one which includes claims about the formal properties of causal relations, and which tells us what a law must be like to be a causal law. Causal relations are thus relations which have the right formal properties and 'whose presence in a law makes that law a causal one' (1990b, p.303). Tooley shares Armstrong's view about laws of nature (of which causal laws are a subset), that is, he thinks that laws are necessitation relations between universals. So, it would seem that Tooley's account of causation, in virtue of its appeal to causal laws, makes use of a *sui generis* form of necessity. In this way, Tooley's view can be seen as rejecting Denial of Powers.

Although Strawson (1989) argues that causation, as it is in reality, is regular succession plus something extra, which explains why events unfold in a regular way, he is non-committal on what this extra element is. According to Strawson's Hume, we can attain no contentful conception of what this extra element is, although we can be sure that it is there. Strawson (1987) suggests that this additional element could be the presence of 'objective forces – e.g. the "fundamental forces" postulated by physics' which 'govern the way objects behave and interact' (p.254), and adds:

> I will avoid speaking of "natural necessity", or of "laws of nature" (understood in a strong, non-Regularity-theory sense), or of the "causal powers" of objects. It is very difficult to keep control of these rival terminologies. But here the notion of objective forces is being understood in such a way that accounts of causation given in terms of these other notions may be supposed to reduce naturally to the account in terms of forces. For example: (1) if objects have causal powers, they have the powers they do

> wholly in virtue of the nature of the forces informing (and so governing) the matter of which they are constituted. (1987, p.255)

It is possible, therefore, that one could be a Strawsonian realist about causation without thinking that a primitive concept of power was necessary for understanding causation (and so without rejecting Denial of Powers).

3.3.3 Powers-based Theories of Causation

The starting point of powers-based approaches to causation is rejection of Denial of Powers. On a powers-based theory of causation, a primitive concept of power is essential for understanding the nature of causation. Facts about what powers things have, or what things can do, cannot be analysed as claims about what events regularly follow on from what others. On powers-based theories of causation, just like on realist theories of causation, causation is something in nature which constrains the ways in which events can unfold, and which therefore *grounds* regularity. In other words, worldly events unfold in a regular way *because* causation exists. On a powers-based theory, causation is the exercise of power, and worldly events unfold in a regular way because what can occur is limited by what powers entities possess: an entity with certain powers must behave in certain ways when the conditions for the manifestation of the power arise, provided there is nothing interfering with the entity and thereby blocking the manifestation.

Steven Mumford (2009) argues that no powers-based account of causation can be reductive, because *power* is a causal notion. For example, it is impossible to understand what it is to have the power to intoxicate, without having some grasp of the phenomenon of intoxication, which is a causal process. Thus, powers-based theories deny Reductivism. However, Mumford (like Woodward) insists that an account of causation can be informative without being reductive. That is, an account of causation can give some insight into the nature of causation without telling us what non-causal structures exhaustively constitute causation. However, given that the powers-based theory takes causation to be the exercise of power, without saying more about what an

exercise of power is, this account is in danger of seeming uninformative, perhaps even circular. What is missing from powers-based theories of causation is a suitable ontology which tells us what an exercise of a power is, what sorts of entities possess and exercise powers, and what sorts of relations those things stand in when they exercise their powers. There is no consensus regarding how these questions should be answered.

Steven Mumford and Rani Lill Anjum have proposed a powers-based theory of causation which is premised on the idea that 'the world is a world containing real powers' (2011, p.4). In other words, Mumford and Anjum hypothesise that powers are real entities, and causation is powers tending towards their manifestations. In slightly more detail, Mumford and Anjum hold that 'causation happens when powers do their work' (2011, p.30). What's more, powers do not work alone (except in exceptional cases). Most effects are the upshot of multiple powers manifesting themselves. For example, for a light bulb to burn me, the filament needs to be manifesting its power to get hot, the glass needs to be manifesting its power to propagate this heat, and my hand needs to be manifesting its liability to be burnt. Each power has a contribution to make to the coming-about-of the effect. Each power, in its own way, pushes towards an effect. When many powers make their contributions, these contributions add together, and after they reach a certain threshold the effect has been produced. (It is this *contribution* towards the coming-about-of some effect, not the effect that eventually comes about, which Mumford and Anjum take to be the power's manifestation. This is because Mumford and Anjum want to maintain that powers are individuated by their manifestations, so *distinct* powers cannot have the *same* manifestation, and one and the same power cannot have a different manifestation in different contexts, so Mumford and Anjum distinguish a power's manifestation from the effect of the power's manifesting itself.)

Does this make causation a relation between a power and the effect it makes a contribution towards producing? Or between the set of powers which have accumulated and the effect their accumulation has produced? Or between the power and its manifestation, i.e. the contribution it makes towards an effect?

Mumford (2009) indicates that causation is 'the whole process going from power to exercise and from contribution to event' (p. 108):

> The dispositionalist, instead of seeing causation as a matter of clearly distinguishable cause and effect, with the appropriate relation between them, sees causation as almost always complex, involving multiple powers combining to produce something together through a process. Only in the idealised laboratory conditions would we theoretically have an event produced just by one power acting alone. Instead of discrete, externally related causal relata, we have a process of interconnected powers. Given that a manifestation is a part of the essence and identity of a power, then if the power and its manifestation exists, any such causation would be an internal relation. (2009, pp.108-109)

Elsewhere, Mumford and Anjum state that 'We argue that causation is a single, unified, and continuous event or process rather than a relation between distinct and discrete events, that causes and effects are simultaneous and that causes tend towards their effects without necessitating them' (2013, p.554). Mumford and Anjum also describe causation as the passing or shifting of powers from one substance to another. So, when fire heats a person, the power to heat possessed by the fire is passed to the person, and when a stone breaks a window the power to cut that the window comes to possess after this causal transaction was drawn from powers possessed by the stone. Mumford and Anjum also suggest that this process of passing around powers is more fundamental than the substances which possess the powers (p.555). In holding that causation is a process, Mumford and Anjum seem to reject Relationalism – which would actually make Mumford and Anjum's theory a fully non-Humean theory of causation, although one which is a rival to the fully non-Humean theory of causation I will espouse in section 3.3.

Misgivings about Mumford and Anjum's metaphysics have been raised by Jennifer McKitrick (2013). McKitrick objects that Mumford and Anjum's theory of causation has nothing to say about how dormant powers become active, or come to be exercised. Mumford and Anjum's view identifies causation with a continuous process of powers pushing towards an effect, but this presupposes

that the powers are already being exercised – they are already making their contribution to an effect. Mumford and Anjum respond to this objection by claiming that 'when a power is not doing its work, it is not part of the causal story, so it is not something we should be trying to include' (2013, p.556), and by insisting that they do have something to say about how a dormant power could become active: a dormant power's becoming active could be the effect of a causal process, it could be something which resulted from the addition or removal of some other active power. However, I think that Mumford and Anjum underestimate the seriousness of McKitrick's complaint. According to Mumford and Anjum, causal effects are achieved by the accumulation of many powers manifesting themselves reaching a certain threshold. One may wonder how, on this picture, anything is really *produced*. On this picture, powers tend towards an effect, and once this 'tending' reaches a certain magnitude, the effect has come into being. The effect seems not to be causally produced so much as constructed, in the same way that bringing together the various parts of a statue is a way of bringing a statue into being. Mumford and Anjum deny that their view entails that causation should be thought of as a kind of ontological construction, but this denial seems inconsistent with their proposal that causation is the culmination of power-exercises adding together as opposed to the transition from a power being dormant to a power being exercised. Anna Marmadoro (2013) has also noted that Mumford and Anjum's view seems to involve two distinct metaphysics which are in tension with each other.

An alternative take on a powers-based theory of causation makes substances, or powerful particulars, central (e.g. Harré and Madden, 1975). On such a view, powerful particulars stand in production relations to events, and coming to stand in such production relations to events *is* exercising a power. For example, when a rock breaks a window, it comes to stand in a production relation to a window-breaking event. Thomas Reid thought that causation was the production of change by the exertion of power and 'that which produces a change by the exertion of its power we call the *cause* of that change; and the change produced, the *effect* of that cause' (1788, pp.12-13). Causes, on Reid's

view are therefore the bearers of powers (rather than the powers themselves as on Mumford and Anjum's view). E.J. Lowe endorses a similar view, 'a causal power, as I shall construe this term, is one whose manifestation or "exercise" consists in its bearer's acting on one or more other individual substances (or sometimes on itself) so as to bring about a certain kind of change in them (or it)' (2013, p.158). There are two relations highlighted here: the relation that holds between the powerful substance and the event it produces in exercising its power (i.e. the change it induces) and the relation between the powerful substance and the substance it is producing a change in. It is unclear whether Lowe's view entails that the exercise of power, and thus, causation, can be identified with the first sort of relation, or the second, or neither. Reid seems to identify causation with the relation that holds between a powerful particular and the event it produces in exercising its power. Insofar as powers-based approaches attempt to spell out the notion of an exercise of power in terms of one or other of the relations mentioned, it seems to me that even powers-based approaches succumb to the Humean intuition that giving an account of causation is a matter of explaining what a *relation* must be like to be a causal relation. Such powers-based approaches would accept Relationalism.

3.3 A Fully Non-Humean Approach

In this last section, I will sketch my preferred non-Humean approach to causation. This approach is characterised by two core ideas which set it apart from the previous theories discussed. First, the approach rejects Relationalism. Causation is not always and everywhere a relation, and giving a full account of causation is not merely a matter of explaining what a relation must be like to be a causal relation. Put positively, I maintain that causation can be a process rather than a relation, of which processes like breaking, crushing, bending etc are more determinate species. This view is in line with Anscombe's (1971) suggestion that causation is a 'highly general', determinable concept, which is an abstraction from the plethora of more specific causal concepts represented by verbs of action. I also agree with Anscombe that we come by this concept of causation when we directly perceive substances exerting causal power over other

substances and associate what we see with the appropriate specific causal concept.

Hornsby has described views like mine as 'Neo-Aristotelian':

> Neo-Aristotelians do not treat cause as everywhere a relation — neither as a relation between two events, nor between two objects, nor between an object and an event […] They take an object's powers to tell us what kinds of processes the object can engage in, so that they connect our understanding of causality with our recognition of the display of the potentialities of things by the things having those potentialities. Thus they defend a metaphysics in which a substance ontology belongs, and to which such notions as powers, capacities, liabilities are central. […] Causality, then, is present in the world inasmuch as something is actually exercising its powers, perhaps affecting something else in doing so. (2015, pp.131-132)

Obviously, it is no good saying that causation can be a process rather than a relation without saying what a process is. In the next chapter, I will provide a metaphysical framework which includes processes in its ontology – I will, as Hornsby puts it, 'defend a metaphysics in which a substance ontology belongs, and to which such notions as powers, capacities, liabilities are central' – and thereby explain exactly what it means to say that causation is a process.

The second core idea of my preferred non-Humean approach is *pluralism*. On my preferred view, even though causation is sometimes a process rather than a relation, this is not to deny that there is a distinctive sort of relation which answers to claims like '*c* is the cause of *e*'. To reject Relationalism, it is only necessary that one deny that causation is *exhaustively* constituted by a special sort of relation. One need not claim that we never think of causation as a relation between cause and effect. In my view, we think of causation in two ways: as an exercise of causal power and, separately, as a relation obtaining between cause and effect.

Insofar as my view grants that causation can be an exercise of causal power, my view has a lot in common with powers-based theories. I also maintain that *power* is a primitive concept, i.e. one which cannot be analysed in other terms. So, one cannot say, in other terms, what is meant by 'can' in statements of

what a thing can do. Thus, I reject Denial of Powers. However, unlike Mumford and Anjum, I do not think that powers are *entities*. Powers do not exist in concrete reality; they are not, to borrow a phrase from Lowe (2005b, p.35), 'elements of being'. As Anthony Kenny puts it 'a power must not be thought of as a thing in its own right' (1975, p.10). And as Ryle accepts:

> Potentialities, it can be truistically said, are nothing actual. The world does not contain, over and above what exists and happens, some other things which are mere would-be things and could-be happenings. (1949, p.119)

In agreement with Ryle, I deny that ascriptions of powers to things report 'limbo facts'. Ascriptions of power do not ascribe to things strange nearly-properties. But as Ryle puts it, 'the truth that sentences containing words like 'might', 'could' and 'would…if' do not report limbo facts does not entail that such sentences have not go proper jobs of their own to perform' (p.120). The concept *power*, it seems to me, is best thought of as a way of thinking about how substances are connected to the processes they engage in, not just currently, but possibly in the future and in circumstances which may never come to pass. As Ryle contends, the job of ascriptions of power is to allow us to make inferences about what substances can, will and would do.

My view also rejects Reductionism. On my view causation is a determinable process, but not all processes are examples of causation. Only those processes which are (to some degree) ways for substances to be effecting change, are species of causation. However, the notion of 'effecting change' is clearly a causal notion, hence my account cannot be reductive. However, I deny that my account is circular: this is because we are acquainted with the determinate forms of causation (like breaking and crushing) via direct observation. Rom Harré and Edward H. Madden (1975) argue that we directly perceive processes in which causal powers are manifested. They argue that Hume's denial that we directly perceive powers being exercised is based on the false assumption that our perceptual experience is primarily atomistic. Hume assumes that what we directly experience are 'punctiform', 'atomistic' sensations. Once this assumption is made, it follows that it is impossible that a single

impression could be the experiential origin of our idea of causal power, and hence some story must be told about how the idea of causal power arises from multiple impressions. However, why assume that our singular impressions are all and only 'puntiform', 'atomistic' sensations? Why assume that we directly perceive the leaf as green and, later, the leaf as brown, but we do not perceive the leaf *changing* from green to brown? Anscombe objects to Hume's idea that we cannot observe causality in the individual case, by pointing out that 'someone who says this is just not going to count anything as "observation of causality"' (1971, p.8). Anscombe is, I think, making a very similar point to Harré and Madden. If one assumes from the outset that perceptual experience is primarily atomistic, then of course it will turn out that 'all we find' are impressions of events which 'seem entirely loose and separate' (Hume, 1975, p.74), but that's because 'the arguer has excluded from his idea of "finding" the sort of thing he says we don't "find"' (Anscombe, 1971, p.8).

My commitment to pluralism, however, sets my view apart from many powers-based theories of causation. On my view, even though causation is sometimes a process rather than a relation, we *also* think of causation as a distinctive sort of relation. The distinctive sort of relation can be characterised as 'difference-making'; it is the relation that obtains between an effect and that which made the difference to the effect's occurring or obtaining. Exactly what the nature of the cause–effect relation is, beyond difference-making, is not a question I shall seek to answer here. Pluralism leaves this question open – it says nothing specific about the nature of the cause–effect relation. It only insists that difference-making is distinct from the relation that obtains between a substance exercising a power and the event the substance produces in exercising that power. To give this latter relation a name, I will call it *the agency relation* – the relation that obtains between an agent and the event that agent is an agent of. To be clear, I am not denying that the agency relation cannot truthfully be called causal, in the sense that it has something to do with causation. What I am denying is that this relation is a relation between 'cause' and 'effect'. The agent relatum of an agency relation (which in my view will always be a substance), is

not the cause of the event they bring about in exercising their causal power. Substances can be, as it were, *causers* – but this is just to say they can be, for example, movers, or breakers, or crushers, or scrapers; that is, substances can be things which engage in causal processes. But, substances cannot be *causes*, given the way I want to use that word. This is because substances cannot be that which made the difference to an effect's occurring, as I shall presently show. My view also denies that the relation substances exercising causal power bear to substances they are acting upon, a relation I will call *the agent-patient relation*, is causation. Again, we may sometimes call the agent-patient relation 'causal' because we want to indicate that there is causation going on when the relation obtains, but the agent-patient relation is not what causation is.

Substances cannot be that which made the difference to an effect's occurring because difference-makers must be *dated entities.* Substances, as I understand them, are entities which exist at more than one time by 'enduring' – entities which, as it were, 'sweep through' time. They exist at multiple times (most of them anyway), but not by having *temporal parts* located at each time. Proper temporal parts are parts of an object which are cut out of the object along temporal dimensions but not spatial dimensions. So, temporal parts are parts which can be described as 'earlier than' or 'later than' other parts but not 'to the left of' or 'to the right of' other parts. On the view I endorse, substances don't have temporal parts at all; they only have spatial parts. Because substances exist at more than one time by enduring, this means that substances cannot instantiate properties 'atemporally'. To take an 'atemporal' perspective on the world is to think about how the world is while ignoring the distinction between past, present and future. It is not to think about the world as it is now, or as it was in the past, or will be in the future; it is to think about the world as it is independently of what time is 'now'. On the endurantist view of substances, substances do not instantiate (at least temporary) properties independently of what time is now. If you think about how the world is while ignoring the distinction between past, present and future, it will be impossible to say what properties substances have. It will be impossible to say, for example, whether I

have blonde hair or brown hair – this is because I *had* blonde hair in the past and *now* I have brown hair. In contrast to substances, events are paradigmatic dated entities. Events which exist at more than one time do not 'sweep through' those times, they are rather 'spread out' across those times. That is, events which exist at more than one time exist at those times by having temporal parts at those times. Events can instantiate properties atemporally. That is, events instantiate properties independently of what time is now. For example, the passage of time has made no difference to Roger Bannister's record-breaking mile-run taking 3 minutes 59.4 seconds. In 1954, this event took 3 minutes 59.4 seconds, and today, 3 minutes 59.4 seconds is still how long the event took. Difference-makers must be dated entities because, in looking for that which made the difference to the occurrence or obtaining of an effect, we are looking for a part of the history of the world which stands in a relation to another part of the history of the world atemporally.[23]

I believe pluralism is the best way to do justice to the diversity of our causal thinking. When it comes to explaining why the relation between the collision with the iceberg and the sinking of the ship, or the relation between the fluttering of the flag and the bull's charging, are instances of causation, appeals to powers and their exercise may not provide the answer. (Appeals to powers and their exercise may explain why such relations exist, without explaining what the relations actually are.) On my view, there is no demand to provide a semantics for *all* causal discourse in terms of powers. I can allow that the conceptual scheme that relates the concepts *power*, *substance*, and *process*, may not (and, I suspect, cannot) be sufficient to clarify the content of *all* our causal claims.

[23] Fales offers a similar explanation for why events, and not substances, are the relata of causal relations (1990, p.54).

The idea that we have more than one way of thinking about causation is not such a novel idea. Brian Skyrms has suggested that rather than being a single concept, causation is an 'amiable confused jumble' of concepts (1984, p.254). My view honours this suggestion: on my view the concept *causation* covers an ontologically diverse 'jumble', including a distinctive cause-effect relation and a special kind of determinable process, which is in turn associated with two distinctive sorts of relation, the agency relation and the agent-patient relation.

Ned Hall (2004) has also suggested that we have two concepts of causation, one which he calls 'counterfactual dependence' and another which he calls 'production' – although, Hall thinks of production as a relation holding between events, rather than a process which substances can engage in. Elliot Sober (1984) also argues that we have two concepts of causation, however, my view has little in common with Sober's. Sober argues that our two causation concepts are 'property causality', which relates properties and can be analysed in terms of causal factors raising the probability of their effects, and 'token causality' which is a 'physical thing' connecting token events which has as a paradigm case the relation between a parent's genetic makeup and its offspring's genetic makeup.

My view is most similar to a position put forward by Taylor (1966). In his introduction to *Action and Purpose*, Taylor distinguishes between two meanings which have been attached to the words 'cause' and 'causation'. On the one hand, there is a notion of causation which is tied up with notions of power, which was once regarded as a 'basic' concept 'more obvious and more clear than any concepts by means of which one might try to describe or define it' (p.16). On the other hand, there is the notion of causation as a 'complex relationship between changes or events, analysable in terms of other familiar relations such as constant conjunction and not, in any case, one that can be understood only in terms of some further primitive notion of active power, or the power to make things happen' (p.16). Taylor then points out that the plethora of 'theories of causation' which don't make any essential appeals to the notion of power, do not

show that the notion of causation which is tied up with the notion of power is not 'a basic, clear, and unanalysable concept' (p.17).

A potential objection to the view I am proposing is that the idea that we think of causation in two distinct ways, as a process and separately as a cause-effect relation, is inconsistent with the fact that we use just one word, 'causation', to cover the worldly phenomenon. As Randolph Clarke presents the objection:

> To say that entities of both these categories [substance and event] can be causes is to say that causation can work in two dramatically different ways. Causation would then be a radically disunified phenomenon. It may be claimed, with some plausibility, that this cannot be so. (2003, p.208)

I think this objection can be dealt with by acknowledging that even though we think of causation in two different ways, our two causation concepts are not entirely disconnected from each other. One way to spell out this claim is to offer a plausible story of how one of the two causal concepts may have grown out of the other. The story I find most plausible runs as follows. As I noted in section 3.3.3, if substances possess and exercise causal powers, then substances with certain powers must behave in certain ways when the conditions for the manifestation of the power arise, provided there is nothing interfering. In other words, when a power is properly triggered, it will manifest itself in 'canonical ways', as Nancy Cartwright puts it (2009, p.144). The exercise of powers will therefore be the source of regular and stable relations between trigger-events and manifestation-events. We can use knowledge of these relations to change how powerful substances behave. For example, if one knows that being near flowers triggers an allergic reaction, then one can prevent the allergic reaction by avoiding flowers; similarly, if one knows that a release of luteinising hormone by the pituitary gland triggers ovulation, then one can prevent ovulation by preventing the release of luteinising hormone. From this we get the idea that events, particularly (but not exclusively) trigger-events, can be *devices* for manipulating later events and can *produce* later events. However, this is a metaphor: events are not literally devices, and cannot literally produce events,

because they are not the right sort of thing to be devices or produce events – only substances can literally play these roles. This is because producing an event is a process. A trigger-event cannot produce a manifestation-event because the manifestation-event occurs after the trigger-event is over and done with – the trigger-event is in the past when the manifestation-event begins to occur, hence the trigger-event is not around at the right time to produce it. Only something which endures for the occurrence of an event can produce it. However, that talk of events as devices or producers is a metaphor doesn't mean there aren't conditions under which use of this metaphor is correct and conditions under which use of this metaphor is incorrect, just as the fact that feelings can only metaphorically, and not literally, be hurt doesn't mean that it is never incorrect to say my feelings have been hurt. This metaphor is thus the source of the idea that there is a special sort of relation between events which is causation.

Because my view has a lot in common with powers-based theories of causation it risks falling foul of the same objections. For example, Schaffer (2007) objects to the idea that worldly events unfold in a regular way because what can occur is limited by what powers entities possess. According to Schaffer such a view places implausible limits on what can be. Schaffer regards the view that 'anything can coexist with anything else, at least provided they occupy distinct spatiotemporal positions' (Lewis, 1986, p.87), as a 'plausible principle about what is possible' (2007, p.85). The idea that what can happen is limited by what powers things possess entails 'implausible limitations on recombination' for example: 'if *c* is accorded the basic property of *causing e*, then the intuitive possibility of *c* without *e* is lost' (p.85). However, I do not think the limits on what can be entailed by the admission that things with certain powers must behave in certain ways when the conditions for the manifestation of their powers arise are implausible. To borrow an example form Harré and Madden (1975), if fire has the power to burn a person, and the conditions for the manifestation of this power are met (e.g. a person has stepped into the fire), what this means is that, *unless something interferes*, the person will get burnt. Is that an implausible limitation on what can be? And as for Schaffer's own

example, if some substance is engaged in the process of causing *e*, this does not imply that the possibility of the substance existing without *e* occurring is lost. While the substance is engaged in the process whose completion eventually constitutes occurrence of *e*, *e* has not yet been caused, and may never be caused: something could interrupt the process, and *e* may never come to be.[24] As I have already said: interventions are nearly always possible, so the manifestation of a thing's power is never metaphysically necessary, not even when the conditions are right for the manifestation, because the manifestation can be blocked by an intervention.[25] So, this objection of Schaffer's fails.

Another common objection to powers-based theories of causation, is that such theories are ontologically profligate. That is, they posit the existence of fundamental sorts of entity, or make use of unanalysable concepts, to no explanatory advantage. Schaffer suggests that theories like mine involve a 'terrible metaphysical price for a relatively flimsy intuition' (2007, p.89). It is important to be clear on what the metaphysical price of my theory is. The metaphysical price of my theory involves an ideological and an ontological component. The ideological element is the primitive power concept which I think we need to understand causation: I am maintaining that there are facts about what substances can do, which we can discover, where the notion of 'can' here cannot be analysed in other terms. The ontological element is the process ontology I am proposing: I am positing the existence of processes; as well as the history of events, there is also the bringing about of those events. The task of the next chapter will be spell out this process ontology in detail. I will leave it to the reader to judge whether the metaphysical price of this theory is terrible.

And what do we get for this price? One of the desiderata a theory of causation should satisfy is, I claim, that the account of causation it provides

[24] This point will, I suspect, seem clearer after I have explained what processes are in chapter 4.
[25] C.f. Mumford and Anjum (2010).

facilitates a satisfactory theory of agency. It is my view that Humean approaches to causation fail to meet this desideratum. One of the conclusions of the previous chapter was that to properly understand the causality of agency we need to see it as something that casts the agent as a causal player, rather than merely the setting for events to cause other events. As the argument in chapter 2 demonstrated, a theory of causation which takes causation to be, always and everywhere, a relation between events seems to prevent this – to recognise the essential role of the agent in the causality of action, we need to think of causation as something other than a relation between events. So, the explanatory advantage of my non-Humean metaphysics is a theory of agency which recognises the essential role of the agent in the causality of action. I will leave it to the reader to judge whether the intuition that agency disappears in a Humean metaphysics of causation is flimsy.

4

Events and Processes

In this chapter, I will propose a novel process-ontology which permits us to think of causation in the non-Humean, or 'neo-Aristotelian' way I described in the previous chapter. In chapter 3, I contended that causation can be a process rather than a relation, of which processes like breaking, crushing, bending etc are more determinate species. My proposal was that causation is on display, not only when events make the difference to the occurrence of other events, but also when substances exercise causal powers, and what it is for a substance to exercise causal power is for there to be an entity, i.e. a process, in which the substance engages. The process-ontology outlined in this chapter gives content to this proposal by explaining what a process is.

The orthodox view of processes is that if there are any differences between processes and events, they are not significant enough to warrant treating them differently in theories of causation. Many philosophers who have written extensively on causation have not paid the distinction between events and processes much, or any, attention.[26] Others have considered the distinction, but have explicitly rejected its metaphysical significance.[27] I propose a theory of

[26] For example, Kim, Davidson and Bennett. See especially Kim (1976), Davidson (2001), and Bennett (1988).

[27] An exception may be Salmon (1984), who does take the distinctive features of processes to be important in understanding causation. However, for Salmon 'the main difference between events and processes is that events are relatively localized in space and time, while processes have much greater temporal duration, and in many cases, much greater spatial extent' (p.139).

processes which denies that processes belong in the same ontological category as events, and instead holds that processes are a special kind of universal. I suggest that engaging in a process is analogous to instantiating a property, and that events are instances of processes. On this proposal, a substance's engagement in a process is a special sort of state of affairs (namely a *dynamic* state of affairs).

In section 4.1, I will summarise an argument put forward by Alexander Mourelatos (1978) which shows that an important subclass of verbal predications, which Mourelatos calls process predications, do not implicitly quantify over particulars which have (or will have) happened. In section 4.2, I will suggest that what these predications implicitly quantify over are processes, which are a special kind of universal. In section 4.3, I will consider and respond to some important objections to my process ontology.

4.1 Verbal Predications and Progressive Aspect

Drawing on earlier work by Zeno Vendler (1957) and Kenny (1963), Mourelatos (1978) argues that predications can be distinguished into three semantic classes, *event*, *process* and *state*, the predications in each class reporting a different sort of situation or eventuality. Examples of sentences reporting events include "The sun went down," and "Roger has run a mile." Examples of process predications include "The plant is growing," and "Roger was running." And sentences that report states include "He knows Paris is in France," and "Leo loved Lauren."

Merely considering these examples is enough to afford an intuitive grip on the differences between Mourelatos's three classes. However, Mourelatos offers a more rigorous account of the features of predicative sentences which determine which of his three classes a predication falls into. Mourelatos suggests that, when it comes to working out what sort of eventuality a sentence reports, the most illuminating feature is the grammatical aspect of the main verb. (However, semantic and lexical features also play a part.) In Mourelatos's view, process predications typically involve verbs with progressive aspect. In English, the progressive is formed by combining the 'present' or 'ing' participle of the verb

with the auxiliary verb 'be' as in "She is swimming" or "He was walking".[28] An important feature of sentences involving progressive verbs is that these sentences do not necessarily imply that the eventuality reported has or will come to an end. For example, neither "Roger was running," nor "Roger was running a mile," necessarily imply that Roger has finished, or will finish, his task. For all the first sentence reports Roger may still be running and for all the second sentence reports Roger may still be running a mile. In the present tense, this is even clearer: "Wendy is walking," obviously does not imply that Wendy has finished walking, it implies the reverse: what Wendy is doing, walking, is still going on. Contrast this with sentences such as "Roger ran a mile," which does not have progressive aspect. This sentence necessarily implies that Roger has finished his activity. It is because the progressive is often used to indicate that something is or was in progress that it is such a reliable indicator of process predications.

The fact that progressive sentences do not necessarily imply that the eventuality reported has or will come to an end allows us draw a conclusion with metaphysical import: process predications do not implicitly quantify over particulars which have (or will have) happened. If process predications implicitly quantify over anything, what they implicitly quantify over are not particulars, or countable items. We can see this if we transform process predications into sentences that involve explicit quantification over the eventuality reported. Mourelatos calls this kind of transformation a 'nominalisation transcription' (1978, p.425). For example, if we nominalise the process predication "Roger was running," we get "There was running by Roger". This nominalisation does not include an indefinite article. Similarly, the gerund "running" couldn't be preceded by a word like "few" or "many" and yield a sensible sentence. In these

[28] There is no consensus among linguists as to whether grammatical aspect is a universal feature of languages; it also appears to be encoded differently in different languages. For further discussion of grammatical aspect see Filip (2012), de Swart (2012) and Gvozdanović (2012).

respects, the sentence “There was running by Roger,” is akin to sentences like “There is snow on the roof,” or “There is sand in the bucket,” which involve mass nouns. Sentences like “There is snow on the roof,” don’t involve quantification over countable items; instead they involve quantification over stuff, or ‘mass quantification’. The similarities between the nominalisations of process predications and quantifications over stuffs suggest that the quantification involved in “There was running by Roger” is also not quantification over countable items. As Hornsby points out, the sentence “There was running by Roger” ‘tells us that something […] was going on. But it does not say of any event, nor of any particular of any other sort, that it was going on’ (2012, p.236). What the nominalisation of a process predication says there is (or was), is not a particular, and hence not an event.

In this way, process predications stand in contrast to sentences like “Roger ran a mile.” Recall that “Roger ran a mile,” necessarily implies that Roger has completed the mile. When we nominalise this sentence, we get “There was a running of a mile by Roger”. This nominalisation does involve quantification over particulars, and the gerund “running” refers to a particular event. “Roger ran a mile,” does say that an event (at least one) has occurred, namely Roger’s running of a mile. It is for this reason that sentences like “Roger ran a mile,” are classed as event predications by Mourelatos. This is also why it is plausible to argue (as Davidson 1967 does) that the sentences Mourelatos classes as event predications involve implicit quantification over events.

4.2 Processes as Universals

Mourelatos’s (1978) observation that sentences reporting processes do not report the occurrence of any specific event, and involve mass quantification when they are nominalised, shows that we have a *concept* of a type of entity which is not particular, and hence not an event, but which exists by unfolding

over time.[29] I submit that what this concept refers to is a special kind of universal. Processes, on my theory, are universals. So, running, singing, respiring and melting are single repeatable entities; when Usain Bolt is running the very same entity is present, or going on, as when Roger Bannister was running. More specifically, I propose the following ontological scheme: *process*, *event* and *substance* are three distinct ontological categories; processes are engaged in by substances and events are instances of processes, where *engagement* and *instancing* are primitive concepts.

The distinctness of the categories *process*, *event* and *substance*, is indicated by the differing existence conditions for members of each of these three categories. A particular event *e* exists, or rather *happens* (happening is an event's mode of existing), only when a substance engages in the process *e* is an instance of, and then completes the process. A process P exists, or rather *goes on* (going on is a process's mode of existing), only when, and for as long as, a substance engages in the process. Substances do not depend for their existence on their engaging in any processes or their involvement in any event (although substances might need to engage in certain processes to count as a substance of

[29] Although one must attend to verbal predications with progressive aspect to establish that English-speakers have a concept of an entity which is not particular and which exists by unfolding over time, the presence of a process concept may be less hidden in other cultures. For example, Wang (2013) notes that 'it is well known that Chinese thought lays great stress on process' and 'an emphasis on becoming is implicitly embodied in its understanding of Tao, the ultimate concept in Chinese tradition' (p.178). Wang describes Tao as 'the creative advance of the world' (p.178), and notes that although Tao is translated into English as "way" or "path", i.e. as a noun, in Chinese the word serves as both noun and verb – it is the following of a path as much as it is a path to follow. Thus, it seems that Tao is best thought of not as analogous to Schaffer's 'history' (2007, p.83; see chapter 3, section 3.1.1) which lacks the dynamism essential to the Tao concept, and more similar to my concept of a highly determinable process (c.f. chapter 32 of the *Tao Te Ching*).

a certain kind). So, processes and events depend for their existence on substances, each in different ways, but not the other way around.[30]

Engagement and *instancing* connect entities which fall into distinct ontological categories, but it would be wrong to consider *engagement* and *instancing* as themselves 'elements of being', as Lowe puts it (2005b, p.35). Entities of different ontological categories never come by themselves: reality is substances engaging in processes and thereby bringing about events. Thus, entities of different ontological categories exist together. The concepts of *engagement* and *instancing* are used to explain how the togetherness of entities of different sorts is asymmetrically structured, but *engagement* and *instancing* are not themselves entities of any kind.[31] Lowe explains this point by appealing to a distinction between 'ontological content' and 'ontological form':

> Beings, or entities, we may say, provide ontological content. But all beings also have an ontological form. The ontological form of an entity is provided by its place in the system of categories, for it is in virtue of a being's category that it is suited or unsuited to combine in various ways with other beings of the same or different categories. (2005b, p.49)

Ontological content is exhausted by the entities which belong in the various ontological categories. Inter-category 'relations' like engagement and instancing help us explain how the ontological form of entities in these categories is

[30] In fact, I do not think there are any substances which are not, at some time, engaged in at least one process, hence I do not think there are any substances which have not, at any time, supporting any processes or events. But I take this to be a contingent truth. Worlds where substances exist without engaging in any processes at all seem metaphysically possible.

[31] Denying that *engagement* and *instancing* are elements of being allows me to avoid an objection akin to Bradley's regress (1893), namely: does a substance S need to engage in engagement in order to engage in a process P, and if it needs to engage in engagement, does it need to engage in engagement* to engage in engagement, and so on? The answer is: no, a substance S does not need to engage in engagement in order to engage in a process P, because engagement is not a process (or an entity of any kind). A similar answer is offered to the question of whether an event needs to be an instance of instancing in order to be an instance of a process. I intend this sort of response to be parallel to responses to Bradley's regress offered by Armstrong (1989, pp.109-110) and Lowe (2005, ch.3).

different. This means that to call engagement and instancing 'ties' or 'relations' or 'connections' is misleading – it suggests that engagement and instancing are entities, or elements of being, when they are not. Frank Ramsey said, of the 'tie' between a quality and the thing which possesses the quality, 'I cannot understand what sort of thing it could be, and prefer Wittgenstein's view that in the atomic fact the objects are connected without the help of any mediator' (1925, p.29). This point could equally be made about the tie between a process and a substance, or a process and an event. The correct response, I contend, is to deny that the tie is any sort of *thing* at all (and hence not a mediator).

Although I am claiming that processes are universals, it is important to distinguish processes from *properties*, which are also thought to be universals by some philosophers (including Armstrong 1978a, 1978b, 1989). The distinction between processes and properties can be drawn in the following way: properties concern the static nature of things – they are qualities, 'ways for things to be' – whereas processes are dynamic, that is, they are connected with how a thing is changing over time. My proposal is that processes are ways for a substance to be changing, to be resisting change, or to be effecting change (this last sub-group of processes are picked out by the concept 'activity').

There are similarities between my definition of a process and Lawrence Lombard's (1986) definition of a 'dynamic property'. Lombard defines a dynamic property as a 'property of moving from having the one to having the other static property' (1986, p.172). So, for example, when a leaf changes colour, from green to red, being green and being red are two static properties, the change in colour is an event, and chang*ing* from green to red is a dynamic property. All Lombard's dynamic properties would count as processes on my view. However, because I hold that processes can be ways for a substance to resist change, a substance can be engaging in a process even if no overall change is occurring – which is not possible for Lombard's dynamic properties. For example, on my definition of process, *keeping still* would be a process a person could engage in, and *thermoregulation* is a process many mammalian bodies engage in – when these processes are going on, the substance engaging in them is not undergoing

any overall change, or 'moving from having one to having the other static property'. The function of these processes is precisely to avoid overall change.

The bare bones of my theory are summarised in figure 1. In sections 4.2.1 and 4.2.2 I will explain in more detail what it is for a substance to engage in a process and for an event to be an instance of a process.

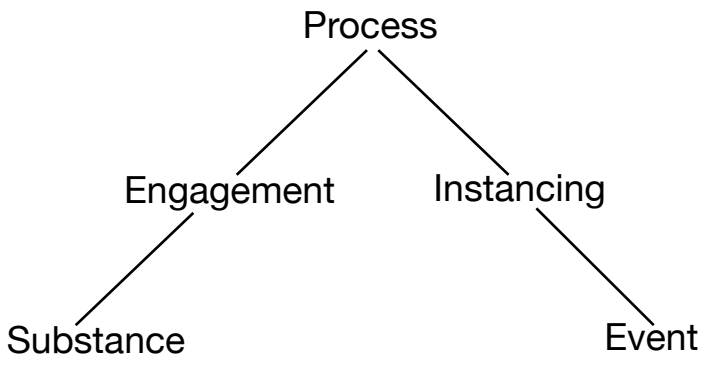

Fig.1

4.2.1 Engagement

The existence condition for processes is logically equivalent to the following principle:

> **Engagement Principle**: A process cannot go on without some substance engaging in the process.

The Engagement Principle gives us a better understanding of what it is for a process to go on: to go on is to be engaged in (this contrasts with what it is for an event to be happening: for an event to be happening a temporal part of the event must presently exist).

The Engagement Principle is analogous to David Armstrong's 'Instantiation Principle' for properties (conceived of as universals). Armstrong claims that there are no properties which are not instantiated by substances. In other words, properties must be instantiated by something at some point in time to exist; eternally uninstantiated properties don't exist (Armstrong 1978a, p.113; 1989, pp.75-82). I do not know whether the instantiation principle is true of properties, but something like it is true of processes, or so I claim. (I will

consider potential counterexamples to my Engagement Principle in section 4.3.2) In fact, I see engagement as similar, in a number of ways, to Armstrong's conception of instantiation. Both are supposed to connect a substance to a universal, and both concepts are supposed to be primitive. Furthermore, just as Armstrong claims that when a substance is metaphysically bound to a universal by instantiation, this unity of substance and universal is a 'state of affairs' (Armstrong, 1989, p.88), so I want to claim that a substance's engaging in a process is a state of affairs. To distinguish states of affairs constituted by a substance's instantiating a *property*, and states of affairs constituted by a substance's engaging in a *process*, I will call the former *static* states of affairs, and the latter *dynamic* states of affairs. It is dynamic states of affairs which are reported by Mourelatos's process predications.

4.2.2 Instancing

In Armstrong's ontology, everything that exists is either a state of affairs or one or other element of a state of affairs, and for Armstrong there are only two such elements: particular and universal. On my scheme, there is not one, but *two* distinct ways in which particulars and universals are connected: *engagement* and *instancing*. *Instancing* is the manner in which events and processes are connected; it should not be confused with either *engagement* or Armstrong's *instantiation*.

I take instancing to be a primitive concept which helps explain the ontological form of the entities it relates. Lowe (2005b) also takes the relation between a universal and its instance to be primitive. Lowe contends that this relation holds between 'non-substantial universals' or 'attributes', like the property redness, and 'modes' or 'property-instances', which are particularised properties like the redness of this particular rose. It also holds between 'substantial universals' or 'kinds', like doghood, and 'objects' or 'individual substances' like my dog Fido. Whether or not attributes and modes, and kinds and individual substances, really are connected by (what I call) instancing depends on whether Lowe's four-category ontology is coherent, plausible and justified. I will not attempt to settle this question. However, I will assume that

the following conditional is true: if attributes, modes, kinds and individual substances all exist, then modes are instances of attributes and individual substances are instances of kinds. Making this assumption enables me to use the relationships between attributes and modes, and kinds and individual substances, to explain, by analogy, the nature of the relationship between processes and events.

Lowe holds that the relationship between a universal and its instance gives rise to a distinctive sort of existential dependence which Lowe calls 'non-rigid existential dependence' (2005b, p.34). When two entities are related by an instancing relation the instanced entity non-rigidly existentially depends on the instance, according to Lowe. Taking as his example the kind doghood, Lowe explains non-rigid existential dependence in the following way:

> The thought here is that, necessarily, this universal exists only if some individual dogs exist, even though the universal does not depend rigidly for its existence upon any individual dog 'in particular' (as we say). Supposing, for instance, that Fido and Rover are the only existing dogs, we do not want to imply that doghood would not have existed if neither Fido nor Rover had existed, but only that it would not have existed if no individual dog whatever had existed. None the less, it seems that there is a perfectly good sense in which doghood does actually depend for its existence upon Fido and Rover, because it depends for its existence on there being some individual dogs and it turns out that Fido and Rover are (as we suppose) all the individual dogs that there actually are. (2005b, p.37)

A similar sort of dependence relation can be said to hold between processes and events. For example, it is possible to argue that the process *running* cannot exist unless some running event or other also exists. The Engagement Principle entails that processes cannot exist unless there is/was some substance engaging in the processes, and once some substance has engaged in a process and completed it, there exists an event. If we assume that all processes are eventually completed, then if the Engagement Principle is true, it is hard to see how a process could exist without there being events that are

instances of that process.[32] To this extent, processes are existentially dependent on events which are instances of those processes. But this existential dependence is not rigid. For example, suppose the only run that has ever or will ever take place is the four-minute run Roger took. The running process does not depend on this actual four-minute event 'in particular' but, paraphrasing Lowe, because the running process depends for its existence on there being some individual running event and it turns out that Roger's four-minute run is (as we suppose) the only running event that actually exists there is a perfectly good sense in which the running process does actually depend for its existence upon Roger's four-minute run.

This argument depends on the assumption that all processes are eventually completed. There are two difficulties associated with this assumption. First, it is not clear what it means to say that processes are the sorts of thing which can be completed or why events come into existence only when processes are completed. Second, it is not clear what entitles me to assume that all processes are eventually completed – for example, it seems possible that someone could have been building a house and yet never managed to get the house built.

Taking these three difficulties in turn, what does it mean to say that processes are completed? One might reasonably suppose that only processes like running a mile, building a house, and painting a picture can be completed. These processes are all telic: they have a natural end-point. They are what Vendler would call 'accomplishments'. They all involve a '"climax" which has to be reached if the action is to be what it is claimed to be' (Vendler, 1957, p.145), or as Mourelatos puts it, accomplishments 'involve a product, upshot, or outcome' (1978, p.417). There is an obvious sense in which telic processes can be completed: telic processes are completed when the relevant product, upshot,

[32] Hornsby (2012, p.237) acknowledges the truth of this claim.

or outcome has been attained. For other sorts of processes however, processes which do not involve a product, upshot, or outcome, it is more difficult to make sense of the idea that they can be completed. Running-a-mile is completed when a mile is run – running, in contrast does not have an obvious end-point, so what sense can be made of the idea that running can be completed?

The way to make sense of the idea is to recognise that, as Anthony Galton and Riichiro Mizoguchi point out, 'at sufficiently fine temporal resolution, the running process is seen as a succession of alternating leg movements, none of which on its own would constitute running' (2009, p.75). This means that no stretch or episode of running can last for only a millisecond: to run, one must make the right sort of leg movements – one needs to raise one leg, lift off from the other, land on the first, transfer weight, and so on – it is impossible to accomplish this in a millisecond. There is, therefore, *something* which one must complete before a running event can be said to exist. And if this something had never been completed, then we would say that no running process was ever engaged in. For example, if someone made the first movement of running, i.e. raised one leg, but got no further than this, then we would deny that that person was ever running. However, even though there cannot be a millisecond-long run, if there is a running event, then the runner was running, i.e. engaged in the process, for that millisecond and a running event partially occupies that first millisecond of time. So, even non-telic processes can, in a sense, be completed.

A key difference between telic and non-telic processes is that in the case of non-telic processes, if the process was started but never completed, then it was never really engaged in. In the case of telic processes, this is not true: a substance could have been engaging in a telic process even if they never accomplished the process's product, upshot or outcome. For example, it can still be true to say that Roger was running a mile, even if he never managed the full mile. This is the imperfective paradox: if Roger was running a mile, this does not imply that he ran a mile. (The imperfective paradox does not apply in the case of non-telic processes: if Roger was running, then this does imply that Roger ran.)

The idea that a running-of-a-mile event comes into existence only when a substance has run a whole mile seems right to me. However, acknowledging this creates a problem: it now seems that the running-a-mile process does not non-rigidly existentially depend on its instances, because a person can be running a mile even in cases where they never complete this process, and hence where running-of-a-mile events never come into existence. This point relates to the second difficulty raised above: what entitles me to assume that all processes are eventually completed? This difficulty might mean that the analogy I have tried to draw between events and processes on the one hand, and individual substances and kinds on the other, cannot be drawn, which would mean that this analogy cannot help explain what instancing is.

The best way to respond to this difficulty is, I think, to claim that not *every* process predication implicitly quantifies over a process. Just as Armstrong denied that every property-predicate corresponds to a property-universal, so I can deny that every process-predicate corresponds to a process-universal. I can claim that processes to which the imperfective paradox applies are non-natural: predicates like 'is running a mile' do not correspond to universals. What makes it the case that, for example, Roger is running a mile, is not that there is a distinctive running-a-mile universal. This predication may be true just in case Roger is engaged in a running process, and some other facts hold true.[33] The point here is that non-telic processes do non-rigidly existentially depend on their instances, just like kinds – telic processes do not, but perhaps non-telic processes are the only processes there really are.

I think the relationship between events and processes is analogous to the relationship between individual substances and kinds. However, it is also analogous to the relationship between attributes and modes. Modes are

[33] I do not know what these other facts would be. See Wolfson (2012) for a discussion of what the truth conditions for process predications involving imperfective aspect should be.

'unstructured'. That is, modes are not constructed from entities of other categories in the way that I said states of affairs are. An important aspect of the thesis that events are instances of processes is therefore that events are unstructured particulars.

Some philosophers have suggested that events just are modes, or property instances. For example, Donald Williams (1953, p.172) writes that 'generally speaking any event is a trope'. And Lowe suggests at one point in *The Four Category Ontology* (2005b, pp.80-81) that events are instances of relations. The problem with taking events to be modes is that it cannot account for the dynamic nature of events. Essential to our concept of event is that events are things that happen. Frederick Schmitt (1983) has suggested that a necessary condition for something's being a happening is that it comes into being over its duration. So, for example, the relation 'is colliding with' is something that two objects can exemplify, and an instance of this relation (if there is such a thing) is something that exists from the moment the first object starts colliding with the second to the moment it stops, but crucially, it doesn't come into being over this duration – it exists complete from the start. For this reason, instances of relations are not things that happen. So, an instance of the relation 'is colliding with' is not an event. Processes, unlike properties and relations, are dynamic, that is, they are concerned with how a substance is changing, or resisting change, or effecting change – they are entities which essentially unfold over time. This means that an instance of a process is precisely the sort of entity which comes into existence gradually, temporal part by temporal part, as a substance engages in the process.

Furthermore, it is intuitively *more* plausible to say that there are instances of processes than it is to say there are instances of properties. Jonathan Bennett nicely captures this:

> I used to object to the notion of a quality instance or trope, along these lines: "This stone is a particular substance; its shape is a universal property, flatness. The friends of tropes are trying to introduce a third item that is particular rather than universal but is a property rather than a substance,

> namely the flatness of this stone. This is a gratuitous, pointless addition." (1988, p.90)

Of course, most trope theorists are not trying to introduce an *additional* third item, they are instead trying to do away with substances and universals altogether, claiming that both are constructed from tropes. However, if one already endorses a metaphysics that includes substances and properties as fundamental, then also including property-instances as fundamental might seem, as Bennett suggests, 'gratuitous' and 'pointless'. But Bennett continues:

> Try the objection again, but this time taking a property that is more naturally thought of as constituting an event: "The stone is a particular substance; its way of moving is a universal property *falling*. The friends of tropes are trying to introduce a third item that is particular rather than universal, but is a property rather than a substance, namely the fall of this stone…" – at which point the objection peters out. One cannot confidently continue "…and this is a gratuitous, pointless addition". (1988, p.90)

What's right here is the idea that positing a 'third item', which is a particular, in addition to the stone and the falling (which Bennett calls a property, but which I say is a process), does not seem like a 'gratuitous, pointless addition'. This might be because we are already in less doubt about the existence of events than about the existence of property-instances. There are several reasons why this might be: first, we might think events must exist in order to be the relata of difference-making relations; second, we might think events must exist because we have singular terms that seem to refer to them; third, we might think events must exist to be truth-makers of claims about what has happened. We may also be in less doubt about the existence of events than about the existence of property instances because events are extended in time: there is a dimension of space-time that they fill or occupy, whereas it is unclear whether property-instances occupy space-time or not.

4.3 Objections and Replies

4.3.1 The Temporal Profile of Processes

An important consequence of the view of processes just outlined is that it entails that processes are not temporally extended. It is fairly uncontroversial that whatever the correct characterisation of universals is, universals do not occupy or fill space or time – they are not spatially or temporally extended. So, given that I take processes to be universals I am committed to the claim that processes are not spatially or temporally extended. This is contrary to the views of many philosophers writing about processes; most view processes as a species of occurrence and take temporal extension as part of the definition of occurrence. Thus, a potential objection to my theory is that it wrongly denies that processes are temporally extended.

One reason to think processes are temporally extended is because processes 'take time' and surely only entities which are temporally extended can take time. That processes are temporally extended is certainly one way of accommodating the intuition that they take time, but it is not the only way. If processes are ways for a substance to be changing, resisting change or effecting change, then it follows that nothing can be going on for only an instant (although, of course, it can be true *at* an instant that something is going on). It seems to me that the intuition that processes take time can be saved by acknowledging that no dynamic states of affairs can obtain for only an instant – the intuition that processes take time need not commit us to anything more than this.

Interestingly, not every philosopher willing to countenance process as a distinctive ontological category thinks processes are temporally extended. For instance, Rowland Stout thinks that processes are 'occurrent continuants' i.e. 'things which simultaneously occur and continue or endure' (2016, p.42). Thus, Stout thinks that processes persist by enduring. Stout (1997) writes:

> There is something absurd about saying that at any one time while something is happening only part of what is happening is present. What is

> happening at any moment during a process is the whole process, not just part of it. The claim that what is present at any moment is not the whole process but a process part is every bit as bad as the parallel claim that an object as a whole is not present at any one moment, but all that is present is a time-slice of an object. For it is also a distinctive feature of our conceptual scheme of processes that we suppose a process to be both present on one occasion and literally identical with a process present as a whole on another. The phrase 'What is happening now' is naturally taken to denote a whole process; and we do want to claim that what is happening now is literally identical with what is happening at some other time —the very same process. (1997, pp. 25–26)

Helen Steward (2013a) rightly points out that nothing said here supports the conclusion that processes do not have temporal parts. We can say of events (entities which do have temporal parts) that while they are happening, what is happening at any moment is the whole event, not just part of it (although at least one part of the event is also happening). This clearly does not licence the claim that the whole event, rather than a part of it, is *present* at any moment during which it is happening. The claim that the whole event is *happening* at some time during which the event is happening can be true, while the claim that the whole event is *present* at some time during which the event is happening is false. I agree with Stout that there is something absurd about saying that while a process is going on, only part of the process is present at any moment during which it is going on. However, merely noting what we can say about things while they are happening isn't enough to show this. Indeed, even though I agree with Stout that 'part of our conceptual scheme of processes is that we suppose a process to be both present on one occasion and literally identical with a process present as a whole on another', I think he holds this view for the wrong reasons.

Stout makes two mistakes. The first is one that Steward (2013a) points out. Steward argues that upholding both of the following claims about processes is an unstable position:

(a) Processes are occurrences (things that happen or occur).

(b) Processes are not temporally extended and do not have temporal parts.

The reason for this is: part of what it means for an entity to be an occurrence is that it is temporally extended and has temporal parts. Denying this leaves us with no clear way of drawing a distinction between things that happen or occur and things that exist at more than one time by enduring. The second mistake is that Stout assumes that because we think processes can be wholly present on multiple occasions this makes them comparable only with objects that persist over time by enduring. He doesn't consider that this *also* makes processes comparable with universals which are instantiated at more than one place and at more than one time without being *partly* present at one place/time and *partly* present at another.

The charge Stout faces, that his view is unstable because he upholds both (a) and (b) is potentially a charge that could be directed at me. If processes are universals they are not temporally extended and do not have temporal parts, so I accept (b). This means that, on pain of instability, I cannot accept (a). However, I am happy to reject (a). I agree that what it is to be an occurrence – to be something that happens or occurs – is to be the sort of entity which is temporally extended and has temporal parts. Denying that processes are the sorts of thing that happen or occur might seem very unintuitive. However, even if we deny that processes happen or occur, we can still say that processes are the sorts of things that *go on*, where 'to go on' means 'to be engaged in'. Furthermore, the difference between events and processes becomes much clearer if we deny that processes are occurrences: events are things that happen, which are temporally extended and which have temporal parts; processes are things that go on, i.e. things that are engaged in, they are universals, are not temporally extended and do not have temporal parts.

4.3.2 Subject-less Processes

The Engagement Principle states that a process cannot exist unless there is some substance engaging in it. However, some processes appear to be subject-less, that is, some processes appear to go on even though there are no substances which engage in them. My response to this objection will be, to put it as bluntly as possible, to deny that there are such processes.

Some putative counterexamples to my view include processes reported by predications such as “the water is boiling,” “the forest is dying,” “the country is celebrating,” “the university is failing,” “inflation is falling,” “the magnetic field is fluctuating,” “his health is deteriorating,” and “it’s raining,”. None of these examples seem to report processes engaged in by something which fits our intuitive conception of a substance: “the water” refers to a portion of stuff, a forest is an ecosystem, “the country” refers to a group of people who share a national identity, inflation is the *increase* in the average prices of goods and services, a magnetic field is an *area* within which magnetic forces act, someone’s health is a *condition* of them, and there is nothing which rains. Although the water, inflation, the magnetic field etc are being treated as subjects in these sentences, not everything that can be treated as a grammatical subject is, metaphysically speaking, a substance (consider Ramsey’s famous ‘Wisdom is a property of Socrates’).

There are two strategies of response available here which I think are perfectly reasonable. The first is to reduce the problematic sentence. This strategy works best in cases like “it’s raining” or “his health is deteriorating”. In these cases, one can deny that raining and deteriorating are really going on. Raining is perhaps a useful way to talk about *many* changes and processes engaged in by genuine substances (e.g. water droplets) which would be too difficult to directly refer to. And although we may speak of conditions changing, one can argue that it is not the condition *itself* which is engaging in a process of change, it is the *man* who is engaged in a process of change, in this case a process of health-deterioration; his health ‘changes’ only in the sense that earlier periods of it were good and later periods of it were poor.

The second strategy is to adopt a liberal conception of *substance* so that things like an ecosystem and inflation count as substances after all, perhaps by stipulating that having causal power is a sufficient condition for being a substance, i.e. the sort of thing that can engage in a process. I think there are advantages to being liberal about what the substances are. I certainly think it is true that we can, and do, make mistakes about what the substances are. With

medium-sized things, it is easy to distinguish the substances from the processes: the plant is the substance, photosynthesising is the process; the cat is the substance, purring is the process. But in other domains I suspect it is much harder to determine what are the substances and what are the processes. For example, are electrons substances? Or are they ways for areas of space to be changing over time? I am also open-minded about the possibility that, in some domains or in some cases, there may be no objective answer regarding what are the substances and what are the processes. Electrons – and perhaps inflation too – are theoretical entities, that is, they are entities which are posited to serve an explanatory purpose. It may be that in the case of theoretical entities, whether we regard these entities as substances or processes depends on the theoretical work we want to put them to. Liberalism about what the substances are allows for these sorts of indeterminacy.

The ontology I am proposing in this chapter involves weighty claims about the nature of substances. As well as claiming that *substance* is a fundamental ontological category, and that events and processes depend for their existence on substances, my view also entails that substances persist through time by enduring. That substances must persist through time by enduring follows from two claims: first, that a process exists only when a substance is engaging in the process, and second, that no process can be going on only for an instant, or equivalently that engaging in a process takes time. The first claim entails that substances are what engage in processes, and the second claim entails that what engages in a process at each instant during the time a process is going on must be the very same entity – if a distinct entity engaged in the process at each instant during the time the process was going on, e.g. a temporal part of a substance, then this would contravene my claim that engaging in a process takes time. Unfortunately, I do not have space to defend this view of substances here (although the advantages this ontological scheme as a whole confers when it comes to giving a theory of agency constitutes indirect evidence that the conception of substances here proposed is worthwhile).

Fortunately, an endurantist view of substances has been defended in detail elsewhere.[34]

4.3.3 Instantaneous Events

A potential objection to my proposal that events are instances of processes is that it would seem to imply that there are no instantaneous events. Processes, I have suggested, essentially take time. No process can be going on for only an instant. Events, I have suggested, come into existence when a substance engages in a process and then finishes or stops the process, which seems to entail that events must be temporally extended. However, some events don't take any time at all. For example, William's winning of the race is an instantaneous change of state. I do not wish to deny the existence of instantaneous events. However, I think they can be accommodated within my ontological scheme. The Engagement Principle entails that it is a contradiction to say that a substance engaged in a process and the process did not go on for any time at all, but there is no lower limit on how long a process has to go on for. Some processes are engaged in and completed in a very short amount of time. For example, when one lights a match, for a period of time the match is igniting – but this process probably goes on for only milliseconds before it is completed. The most extreme value of the variable length of time a process can go on for is zero – in this case, the process doesn't really go on at all, and hence this case is an exceptional one. I take instantaneous events to be, to borrow a term from mathematics, degenerate limiting cases. They are what happens when the variable length of time the process can go on for takes its most extreme value.

[34] See Lowe (1998) especially chapters 4 and 5. See Sattig (2002) and Rea (1998) for defences of endurantism against the problem of temporal intrinsics; Gilmore (2007) defends endurantism against issues deriving from the possibility of coinciding objects; Lowe (2005a) defends endurantism against issues to do with vagueness. See also Haslanger (1994) for a discussion of the relationship between the endurance/perdurance debate and the metaphysics of Humean Supervenience.

4.3.4 Event Individuation

The ontological scheme I have proposed includes substantive novel claims about what events are. One last potential issue with my account is that it is unclear what it entails about how events are individuated. Proposing identity conditions for events, that is, proposing conditions which stipulate when *x* is the same event as *y* (where '*x*' and '*y*' are event designators), which adequately capture our intuitions regarding event identity has proven difficult. Indeed, a strong reason for rejecting the property exemplification view of events championed by Kim (1976) is that this theory entails that events are more finely individuated than we would expect happenings to be.

On Kim's view, an event is the exemplification of a property by an object at a time. Kim's theory of events yields the following identity condition for events: event *x* and event *y* are one and the same if and only if *x* makes the same property attribution to the same object at the same time as *y* does. According to this identity condition, an event e_1, which is the exemplification of property *F*, and an event e_2, which is the exemplification of property *G*, could be the very same event, even if *F* and *G* are distinct, if e_1 were also an exemplification of *G* and e_2 were also an exemplification of *F*. But, as Lombard (1986, p.55) argues, this cannot be the case on Kim's view. This is because Kim thinks events are things which can explain and be explained, and if e_1 and e_2 really are identical, then one should be able to substitute one for the other in an explanation without affecting the truth of the explanation. If *F* and *G* are distinct then the exemplification of F at a time and the exemplification *G* at a time are not necessarily intersubstitutable in explanations. So, Kim is committed to the view that events are constructed entities made up of three constituents, a property, an object and a time, and an event which is the exemplification of property P_1 by object o_1 and time t_1 is identical with an event which is the exemplification of P_2 by object o_2 and time t_2 only if $P_1=P_2$, $o_1=o_2$ and $t_1=t_2$. This results in some unacceptable event individuations.

Bennett (1988) offers the following counterexample. Suppose that at noon David kisses Eva just once, doing it tenderly, on her cheek. According to Kim's theory, the following name three events:

a The tender kiss that David gave Eva at noon

b The kiss that David planted on Eva's cheek at noon

c The tender kiss that David planted on Eva's cheek at noon.

Each has the form [(David, Eva), P, noon] with different values of P – "kissed tenderly", "kissed on the cheek", "kissed tenderly on the cheek" – and because the Ps are different the named events are different (1988, p.79).

Bennett asks: was event b – the kiss David planted on Eva's cheek at noon – tender? And was event a – the tender kiss David gave Eva at noon – planted on Eva's cheek? It would be absurd to say no to either question, and Kim doesn't. But now it seems that there were many tender kisses planted on Eva's cheek at noon. It is no affront to common sense to say that there are many different facts about David and his kiss that are true, e.g. that it was tender, that it was planted on Eva's cheek, that it occurred at noon etc. But it is unintuitive to say that there was more than one tender kiss occurring at noon if David kissed Eva only once.

My theory would involve the very same error if it entailed that no event can be an instance of more than one process. Fortunately, nothing in the idea that events are instances of processes prevents us from accepting that the same event can be an instance of more than one process. As I said in section 4.2.2, the instancing relationship between processes and events is analogous to the relationship between kinds and individual substances, so because individual substances can be instances of more than one kind, it seems natural to suppose that events can be instances of more than one process. Put more abstractly, on my view, an event *x*, which owes its existence to substance *S* engaging in process *P* over period of time Δt, and event *y*, which owes its existence to substance *S'* engaging in process *P'*, which is distinct from *P*, over period of time Δt, can be the very same event. This principle allows the single tender kiss David planted on Eva's cheek at noon to be an instance of at least three distinct processes: kissing, tender-kissing and kissing-on-the-cheek. It also allows for the thesis that

when an agent φs by ψing, the event of the agent's φing is the same event as that of her ψing, as long as the activities φing are ψing were engaged in for the same length of time.[35] So, to borrow an example from Maria Alvarez and John Hyman (1998, p.234), when Jean gave the signal to her companions by opening the window, Jean's opening the window and Jean's giving the signal are one and the same event. This principle also allows us to identify an event *x*, which owes its existence to substance *S* engaging in process *P* over period of time *Δt*, with an event *y*, which owes its existence to substance *S'* engaging in process *P'* over period of time *Δt*, even when *S* and *S'* are distinct substances. So, for example, my view would allow us to identify an event of an agent's φing a patient with the event of the patient's being φed, even though φing and being φed are distinct processes (one is active one is passive), and even where the agent is distinct from the patient.[36]

However, although my view *allows* that a single event can be an instance of many processes (even of processes which are engaged in by different substances), my view does not entail any general principles which tell us how to decide which processes a single event is and is not an instance of. For example, my view *allows* us to identify an event of an agent's φing a patient with the event of the patient's being φed, but it does not entail that we *must*.

An alternative theory of events, which takes events to be spatiotemporal particulars, i.e. entities which occupy both time and space, also suffers problems to do with event individuation. In keeping with this theory, Edward Lemmon (1967) suggests that an event *x* and an event *y* are one and the same if and only if event *x* and event *y* occupy exactly the same spatiotemporal region. Willard van Orman Quine's theory of event individuation is very similar:

[35] The view that that when an agent φs by ψing, the event of the agent's φing is the same event as that of her ψing was perhaps defended by Anscombe (1963, pp.37-47) and has been endorsed by Hornsby (1980) and Davidson (1969).
[36] I will discuss this somewhat controversial idea again in chapter 5.

> Physical objects, conceived thus four-dimensionally in space-time [i.e. conceived of as entities stretched out in time as well as space, and whose earlier and later periods of existence count as temporal parts], are not to be distinguished from events or, in the concrete sense of the term, processes. Each comprises simply the content, however heterogenous, of some portion of space-time, however disconnected and gerrymandered. (Quine, 1960, p.171)

As chunks of space-time, events *x* and *y*, on the Quinian view, are one and the same if and only if *x* and *y* have the same spatiotemporal location.

However, as Davidson (1967) objects, this view entails that there cannot be more than one event happening in the same place at the same time, and intuitively there can be many events happening in the same place at the same time. Davidson (1967) provides an example: consider a metal sphere which is simultaneously heating and rotating, Lemmon's identity condition would have it that the sphere's rotation and its heating are one single event, because the rotation and the heating occur at the same time in the same place, but intuitively we want to regard the rotation and the heating as two distinct events which occur at the same time in the same place.

My view does not forbid multiple events from occupying a single spatio-temporal location. Put more naturally, my view allows that many things can happen at the same time and place. However, although my view *allows* that many events can happen at the same time, my view does entail any general principles which tell us when we *should* differentiate simultaneous events. In fact, the only constraint on event identity which is *entailed* by the idea that events are instances of processes is that a single event cannot occupy multiple temporal locations except by having parts at each location. For example, I made tea yesterday, and I made tea today. These are *two distinct* tea-making events, not one; they may be the same type of event, but they are distinct tokens of that type. Roderick Chisholm (1970) disagrees with me on this point: he thinks that one and the same event can recur. Chisholm's view is incompatible with what my process ontology entails about what events are. This is because, on my theory, events are particulars and as such cannot be repeated, and so cannot

recur. My theory entails that an event *x* is one and the same as an event *y* only if *x* and *y* occupy the very same temporal location. Furthermore, on my theory, events owe their existence to substances engaging in processes and then completing those processes, so the temporal extent of an event is determined by how long the substance that ontologically supports the event engaged in the process the event is an instance of.

My theory entails that a single event cannot occupy multiple temporal locations except by having parts at each location, but it leaves open whether a single event can be an instance of many processes, and whether many events can all happen in the same place at the same time. My theory does not say that our intuitions regarding these questions – that a single event *can* be an instance of many processes, and that many events *can* all happen in the same place at the same time – are wrong, unlike Kim's (1976) events as exemplifications view, and Quine's (1960) view of events as spatiotemporal particulars. But my theory offers no explanation why these intuitions are right either. In short, my theory as it stands simply does not tell us enough about the nature of events to allow us to put together necessary and sufficient conditions for event identity.

However, I think this problem demonstrates that my account is incomplete, not that my account is implausible. It may be possible to say more about the nature of events, without denying that they are instances of processes. For example, my causal pluralism holds that events are difference-makers. It may be that being the sort of entity that can be a difference-maker entails restrictions on what processes a single event can be an instance of, and provides a general principle which tells us when we ought to distinguish simultaneous events. Perhaps, in line with Davidson's (1969) suggestion, instances of distinct processes should be identified when and only when they occupy the very same position in a causal nexus. Davidson (1969, p.179) suggested that event *x* and event *y* are one and the same if and only if event *x* and event *y* have the same causes and the same effects. This criterion has been criticised for being

circular.[37] The problem with Davidson's proposal is that it individuates events by quantifying over causes and effects which are themselves events, therefore Davidson's identity conditions for events individuate events only if there are some events already individuated.[38] This problem may not be a problem for a view which synthesises Davidson's view with the theory that events are instances of processes. This is because if events are instances of processes as well as causal relata, then some events *are* already individuated before we consider the causal nexus, namely those events which occupy different temporal locations. To conclude, although my account does not offer definitive answers regarding how events are individuated, my account does not entail implausible identity conditions for events, and I think the idea that events are instances of processes could be developed in such a way as to yield identity conditions which are in line with our intuitions.

[37] See Beardsley (1975) and Wilson (1974).
[38] C.f. Bennett (1988, p.98).

5

A Neo-Aristotelian Theory of Agency

In chapter 3, I sketched a non-Humean theory of causation. According to this theory, causation is not always and everywhere a relation, but can be a process which substances engage in. In the previous chapter, I presented a novel metaphysical framework, which includes processes, conceived of as universals, in its ontology. This metaphysical framework gave content to the claim that causation can be something substances engage in, rather than merely an external relation holding between events (or any other particulars). In this chapter, I will argue that this alternative way of thinking about causation, and the ontology that permits it, allows us to put together a more successful theory of agency, one that avoids the problems facing the causal theory of action. Section 5.1 will contain my positive view. Drawing on the process ontology outlined in the previous chapter, I will present my view on the metaphysics of action, and explain how the agency concept should be analysed. In sections 5.2-5.4, I will compare my metaphysics of action with some other alternatives to the causal theory of action, namely agent-causation-based theories of action, and argue that my account has advantages over these theories. In section 5.5 I will defend my analysis of the agency concept.

5.1 A Neo-Aristotelian Theory of Agency

The proposal I have advanced in the last two chapters is that causation can be a determinable process. Causation is going on when a substance is engaging in it, when a certain sort of dynamic state of affairs obtains. On this proposal, if any process is a determination of causation, then it is causal *intrinsically*, just as if a colour is a determination of red (as scarlet is), then that colour is red

intrinsically. As to *which* processes are determinations of causation and which aren't, my answer is that the distinction is not absolute, and can be difficult to determine.

In the previous chapter, I said that processes are ways for substances to be changing, to be effecting change or to be resisting change. This means that some processes are active, i.e. those which are ways for substances to be effecting change, and some processes are passive, i.e. those which are ways for substances to undergo change (resisting change, I think, can be both active and passive). Ways for substances to be effecting change are causal processes. However, as stated in chapter 2, the active–passive distinction is not binary. It is therefore more accurate to say that some processes are more active that others, and some are more passive than others, but (probably) no process is completely active, and no process is completely passive. For example, the process of crushing something is mostly active: in crushing something, a substance is effecting more change than it is undergoing. The process of dying on the other hand is mostly passive: in dying, a substance is undergoing more change than it is effecting. And many processes involve ostensibly equal degrees of activity and passivity. For example, processes by which we move ourselves about, like walking, and running, seem to involve a mix of activity and passivity: when we move ourselves about, we effect change on ourselves, so we are both agent and patient with respect to those changes. Processes which result in no overall change, like thermoregulation or keeping still, also seem to involve elements of activity and passivity. When one stands still, for example, one must exert some degree of force in opposition to the forces which would cause one to fall to the ground (e.g. gravity), but not so much force that one ends up moving. Thus, standing still seems to involve a roughly equal mix of activity and passivity.

I mentioned in chapter 2 that this view of the active–passive distinction seems to entail that this distinction is less than fully objective. This is because whether what a substance is doing is activity or passivity is relative to the degree of change it is wreaking and/or undergoing, and assessing how much change a substance is wreaking and/or undergoing may not be a fully objective matter.

For example, when some salt dissolves in water, how much change one thinks the water undergoes when salt dissolves into it may depend on one's views about the nature of water. Now that I am using this distinction to distinguish between processes which are determinations of causation and processes which are not, the full scale of this subjectivity problem is apparent. If the distinction between activity and passivity is partly a subjective matter, and this distinction is key to distinguishing processes which are determinations of causation from processes which are not, then it seems that what is and is not causation is itself partially a subjective matter. I think that this reasoning is sound, so I accept that what is and is not causation is partially a subjective matter. However, I do not consider this to be problematic because, while it may be true that how we classify the processes being engaged in by substances is partly dependent on our own perspective, the *existence* of dynamic states of affairs, i.e. substances engaging in processes, is not mind-dependent.

The mostly active processes I will call *activities*. What it is for a substance to be causing something is for there to be an activity which the substance is engaging in. A substance engaging in an activity is an agent, and the event that results once the substance has completed the activity it has been engaging in is an action. Actions are thus events of a special kind: they are events which are instances of activities, and as engaging in an activity is what it is for an agent to be causing something, actions can also be said to be instances of substance or agent causation. This metaphysics is summarised in figure 2 (I have also represented the agency relation in this figure).

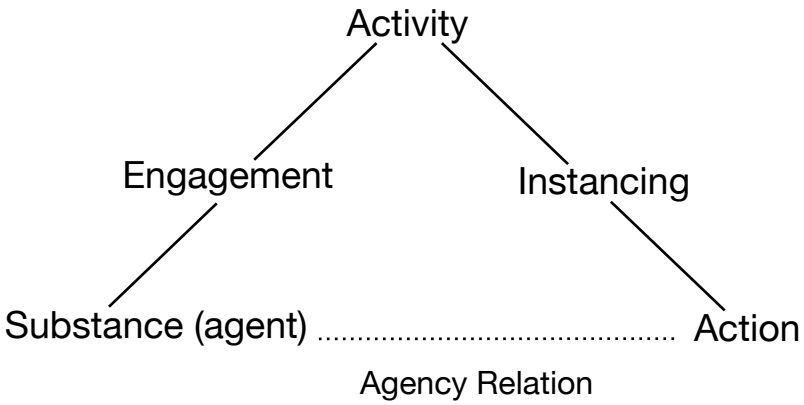

Fig.2

This metaphysics helps us see why the causality of action is something that essentially involves the agent. On my theory, the causation exemplified by actions is the activity the agent engages in; it is something which goes on, but only insofar as it is engaged in by an agent. Furthermore, the dynamic state of affairs which is an activity's going on is something that is partially constituted by the agent. A dynamic states of affairs is, as I proposed in chapter 4, a substance's engaging in a process – it is a complex entity composed of a substance and a process. So, if we take the causality of action to be a dynamic state of affairs, then the agent herself *partially constitutes* the causality of action – she cannot, therefore, be merely the arena within which the causality of action takes place.

My metaphysics of action also entails a plausible account of the relationship between agents and their actions. I proposed in the previous chapter that events depend for their existence on substances engaging in processes and then completing them. As actions are events on my view, they are subject to the same existence conditions. Actions depend for their existence on agents' engaging in activities and completing them. This means that actions come into existence because of agents engaging in activities – but this 'because' indicates existential rather than causal dependence. I have said that the agency relation, i.e. the relation that obtains between a substance exercising a power and the event the substance produces in exercising that power, is not a cause-effect relation or a 'difference-making' relation, and that being the agent of an event is not the same as being a cause of it. I can now state in positive terms what the agency relation is: it is a relation of ontological dependence.

The theory I have just proposed tells us what sort of entity an action is (an event, i.e. an instance of activity). My theory also tells us what sort of entity the exercise of power is: the exercise of power by a substance is a dynamic state of affairs, i.e. a substance's engaging in a process. However, providing a *metaphysics of action* is not all that is required for a complete and adequate theory of *agency*. It takes more to provide an adequate theory of agency than simply to describe the ontological structure of the worldly entities which are picked out by the concepts *action*, *agent* and *activity*. To provide a complete

theory of agency one must consider the concept of agency, and provide some sort of dissection of this concept. In chapter 2, I claimed that there are two distinctions crucial to our concept of agency: the distinction between activity and passivity, and the distinction between one-way and two-way powers. Agency does not reduce to the exercise of active power, because some substances can manifest their agency by remaining passive, and therefore by not engaging in activity. Neither does agency reduce to the exercise of two-way power, because not all substances which cause things to happen do so by exercising two-way powers, but all substances which cause things to happen are agents. My view is that agency is a complex, highly abstract concept which incorporates both distinctions. Some substances' agential powers are one-way; these substances manifest their agency when they are active, but not when they are passive. For these substances, exercising their agential power is to engage in an activity. Other substances' agential powers are two-way; these substances manifest their agency when they are active, but also sometimes when they are passive. For these substances, in *some* cases exercising their agential power is to engage in an activity, but in other cases exercising their agential power is to allow other substances to act upon them.

This is, in a nutshell, my positive account of agency. I will attempt to make this account clearer in the following sections by comparing my metaphysics of action with some other 'agent-causation-based' theories of action. Like my own view, these theories are presented as alternatives to the causal theory of action. Also like my view, these theories are 'non-Humean' insofar as they presuppose that substance or agent causation cannot be reduced to a special kind of event-causation. However, I shall argue that these alternative non-Humean, agent-causation-based theories suffer problems that my account avoids.

5.2 Agent Causationism

The first kind of agent-causation-based theory of action I shall compare my account to is a kind of theory I shall call 'traditional agent causationsim'.

Traditional agent-causationist theories of action, the kinds of theories put forward by Taylor (1966) and Chisholm (1976), hold that an irreducible notion of agent causation is essential for understanding agency.[39] Traditional agent causationists hold that agency is the obtaining of a causal relation between an agent, *qua* substance, and a change which is the action of the agent. For example, Chisholm states that 'whenever an agent performs such an act [as raising his arm or stealing the money], he contributes causally to the fact that he performs that act' (1976, p.71). Taylor claims that 'nothing can be represented as a simple act of mine unless I am the initiator or originator of it', where 'initiator' and 'originator' are supposed to be synonymous with 'that which brings about' or 'that which has the power to produce' (1966, p.112).

Although some agent causationists maintain their view because they think it is the only theory that can do justice to the causal role played by the agent in action (e.g. Franklin, 2016), traditionally, agent causationism is adopted because it is thought to be essential for an adequate treatment of free will. Traditional agent causationists subscribe to a libertarian view of freedom, which is the combination of two theses: that an action cannot be free if it is deterministically caused to happen by a prior event (incompatibilism), and that we do indeed act freely. However, agent causationists are also sympathetic to an argument which appears to show that simply injecting indeterminism into the causal chain leading up to an action cannot secure freedom (Chisholm, 1976, pp.58-59).[40] The solution, agent causationists claim, is to hold that an action is free just in case it is caused to happen by the agent. Now the agent, rather than any prior event, is the causal determiner of the action.

One major problem with agent causationism is that it makes irreducible substance causation seem like something *unnatural*. This is because, on this

[39] Clarke (2003) is a contemporary defender of this kind of theory.
[40] See also Mele (2006, pp.6-9) and Pereboom (2001, pp.38-59).

view, *most* causation in the world is really causation of one event by another, and it is only in the case of things done freely by human beings where we find irreducible substance causation. On traditional agent-causationist views, most causation, including interactions between animals and inanimate objects, is nothing over and above causation of one event by another. Only in the case of things done freely by human agents, is there something extra – causation by *a substance*. Traditional agent causationism thus entails that animals and inanimate objects are not, as I have put it previously, *causers* – active powers are exercised only by human beings. This view entails that there is no continuity between causation in the non-human and human world. Active power, and hence agency, is made to seem like 'a relatively rare and exotic exception to the rules governing the world's normal causal functioning' (Steward, 2012, p.198). I have mentioned already that, on my view, substance causation is not unique to human doing – it is found wherever there is causation. Animals and inanimate objects also exercise active powers, and cause things to happen. I believe such a view avoids making substance causation seem unnatural, because on such a view there is no discontinuity between causation in the non-human and human world.

The argument for agent causationism can be challenged in several ways. For example, one might challenge the agent causationist's incompatibilism and argue that causal determination by prior events is not on a par with the kind of compulsion which negates freedom.[41] One might also challenge the argument that freedom is not secured by making the event-causal chain leading up to the action indeterminisitic (Kane, 1999, 2016). However, I think there is something to the intuition that there is a kind of causal determination by the agent demonstrated in cases where an agent *chooses* what to do, which is missing in

[41] See McKenna and Coates's (2016) Stanford Encyclopaedia entry on 'Compatibilism' for a useful survey of contemporary compatibilism.

cases where the agent does not choose what to do. For example, I think Taylor (1966) is correct when he claims that an inanimate object, like a match, is such that it cannot 'wreak changes in itself', unlike a man who '*can* bring about such a change as a motion of his arm quite by himself' (p.122). However, traditional agent causationists, like Taylor, confuse being able to wreak changes in oneself with agency as such.

In chapter 2, I claimed that inanimate objects do not possess two-way powers. What inanimate objects do is always a response to the circumstances they are in and what's acting upon them. What inanimate objects do is not up to them, they are moved-movers not self-movers. Human beings (and many animals I think), in contrast, are capable of self-movement; human beings have control over some of what they do, and this control consists in the exercise of two-way power. I think the notion of 'wreaking changes in oneself' is parallel to the notion of self-movement. Taylor is quite right that inanimate objects like matches are not self-movers – he is right that inanimate objects cannot bring about change by themselves. But this does not mean that inanimate objects are never truly *active* and are never agents. Taylor, it seems to me, is confusing the exercise of a two-way power with activity and the exercise of a one-way power with passivity.

Another unappealing aspect of traditional agent causationism relates to its claim that agents stand in causal relations to their own actions. On the agent-causationist view, the event of my raising my arm, which is my causing my arm to rise, is an action because it is an event which I, *qua* substance, caused to happen. However, there is a well-known problem with this view. If my action is an event of which I am the cause, then we can ask of the causing of my action whether *this* is an action of mine or not. If it is, then, on the agent-causationist theory, it is also an event of which I am the cause, but now we seem to have opened an infinite regress: is the causing of my causing of my action another action? However, if we deny that the causing of my action is an action, then it seems we have two sorts of 'causings', some of which are actions and some of which are not. For example, my causing my arm to rise is an action, but the

causing of my causing my arm to rise is not – what makes this difference? It is unclear what the agent causationist can or should say.[42]

On my view, agents do not stand in causal relations to their own actions. Rather, actions are the events that come into being once an agent has completed the activity she has been engaging in. On my view, actions are instances of agent causation, where what it is for an agent to be causing something is for an agent to be engaging in an activity. So, actions can be thought of as events of an agent causing something, but an agent's causing something is not an event which the agent stands in a causal relation to.

In summary, there are two key similarities between my view and traditional agent causationsim: first, both my view and traditional agent causationsim hold that an irreducible notion of agent causation is essential for understanding agency; and second, my view, like agent causationsim, entails that there is a form of causal determination which cannot be demonstrated by inanimate objects. However, there are important differences. Firstly, my view does not make substance causation something which is only demonstrated in the case of free human action, because animals and inanimate objects as well as humans exercise active powers – the form of causal determination which cannot be demonstrated by inanimate objects is self-movement, or settling, not agency as such. Secondly, my view does not hold that agents stand in causal relations to their own actions. In these two ways, my view avoids the problems that beset traditional agent causationism.

5.3 Actions as Causings

Other agent-causation-based rivals to the causal theory of action do not contend that agents cause their own actions. Like traditional agent causationism, these alternative agent-causation-based theories maintain that in examples of agency

[42] See Alvarez and Hyman (1998, p.222); Davidson (1971/2001, p.52); Hornsby (1980, p.101).

the agent, *qua* substance, causes something. However, what the agent causes is not her own action – an agent's action *is* her causing of something, it is not what is caused. Most contemporary agent-causation-based theories of agency take it that an action is the causing of an event, or sometimes a state of affairs. For example, Mayr argues that human agency is an instance of substance causation (2011, p.219), and substance causation, for Mayr, should be understood in terms of a causal relation obtaining between a substance exercising an active power and the effect produced when the substance exercises active power: 'when such an "active power" is exercised, the cause of the resulting event is the substance which possess the power itself' (pp.145-146). Similarly, Lowe describes agent causation as a species of causation 'in which the cause of some event or state of affairs is not (or not only) some other event or state of affairs, but is, rather, an agent of some kind' (2008, p.121). And even though Alvarez denies that the power of agency is one and the same as the power to cause events to happen, she admits that 'the paradigmatic case of agency is the exercise of our power to move our body and thus cause change' (2013, p.107) and 'causing these changes involves causing events' (p.103). Alvarez and Hyman also claim that 'an action is a causing of an event by an agent' (1998, p.224), where the event caused by the agent is 'intrinsic' to the agent's action, an event they call the 'result' of the action (p.233). Often, in the case of human action at least, the 'result' of an action is a bodily movement.

It is important to remember that the thesis being endorsed by Alvarez and Hyman is not simply that agents can cause events to happen by acting, a thesis no-one seriously questions. The thesis being endorsed is the stronger thesis that action consists in an agent coming to stand in a causal relation to an event. To bring out the difference, consider someone who denies the existence of irreducible substance causation. Such a person can readily agree that agents cause events to happen – they do so when their actions stand in causal relations to other events. For example, I caused the explosion to happen when I lit the fuse, because my lighting of the fuse, my action, stands in a causal relation (an event-event causal relation) to the explosion. The thesis being endorsed is that

even in the case of basic action, where I do something not by doing something else, my action consists in my causing an event to happen. So, my arm raising consists in my causing the rising of my arm. My action is not the cause of my arm rising; *I* am the cause of my arm rising, and my so being the cause of my arm-rising is what my action consists in.

According to Alvarez and Hyman, actions are *causings* of the bodily movements one's body makes when one acts, so for example the action of raising my arm is my causing the rising of my arm. This entails that actions cannot be identical with the bodily movements one's body makes when one acts, so my action of raising my arm cannot be identical with the rising of my arm. Alvarez and Hyman (1998) provide an argument demonstrating this entailment:

> Davidson is one philosopher who claims that, in some cases, 'my raising my arm and my arm rising are one and the same event'. But my raising my arm is my causing my arm to rise. Hence, if my raising my arm is an event, it is the same event as my causing my arm to rise. And hence, if my raising my arm and my arm's rising are one and the same event, then my causing my arm to rise and my arm's rising are one and the same event. But it cannot be plausible that causing an event to occur is not merely an event itself, but the very same event as the event caused. (1998, p.229)

Spelled out, the argument runs as follows:

Assume for *reductio*:

1. My raising my arm is one and the same event as my arm's rising.

Now assume the very plausible:

2. My raising my arm is my causing my arm to rise.

And:

3. If my raising my arm is one and the same event as my arm's rising, then my causing my arm to rise is my arm's rising.

Together these premises entail:

4. My causing my arm to rise is my arm's rising.

A conclusion which, when generalised, is revealed to be absurd:

5. My causing an event is the event caused.

In response to this argument, Alvarez and Hyman, and many writers sympathetic to agent-causation-based theories of action, have rejected the thesis that one's action is identical with the bodily movements one's body makes when one acts. Alec Hinshelwood calls this 'the Separation Thesis' (2013, p.626).

To explain how the Separation Thesis is compatible with the plausible claim that many actions are bodily movements, Alverez and Hyman (1998) make use of an ambiguity associated with the word 'movement' noted by Hornsby (1980). Many verbs can be transitive (i.e. used with a grammatical object) or intransitive (i.e. used without a grammatical object). The verb 'move' is also *ergative*, which means that it can be transitive or intransitive and that the direct object of the verb when transitive becomes the subject of the verb when intransitive. For example, 'move' is transitive in the sentence "I moved my arm," but intransitive in "My arm moved," and the object of the transitive 'move' is the subject of the intransitive 'move'. This feature of the verb 'move' renders the nominalisation of 'move', 'movement', ambiguous. When we speak of, for example, my arm movement there are two movements we might be talking about. There is one that corresponds to the transitive use of move as in, "I moved my arm," which can be otherwise picked out by the expression 'my moving of my arm' and the one that corresponds to the intransitive use of move as in, "My arm moved," which can be otherwise picked out by the expression 'the motion of my arm'. To help keep the two senses of 'movement' separate I will follow Hornsby's notation and use 'movement$_T$' for the first sense, and 'movement$_I$' for the second sense. Alvarez and Hyman (1998) hold that many actions are bodily movements$_T$, which they claim are *causings* of bodily movements$_I$, and hence cannot be identical with bodily movments$_I$.

Alvarez and Hyman (1998) also argue that actions, i.e. *causings* of bodily movements$_I$, are not events of any kind. To establish this conclusion Alvarez and Hyman assume that there are only two possible sorts of event actions could be:

i) bodily movements$_I$; or

ii) events which are causes of bodily movements$_I$.

Alvarez and Hyman take the first possibility to have been ruled out already by the argument outline above. To show that bodily movments$_T$ are not events which cause bodily movements$_I$ Alvarez and Hyman argue as follows:

> […] if bodily movements$_T$ are events which cause bodily movements$_I$, then either bodily movements$_T$ are events, perhaps neural events, which occur inside the agent's body, as for example Hornsby maintains in her book *Actions* or they are events of another sort, which do not – presumably events which have no location at all, if there are such events. The first alternative implies that bodily movements$_T$, unlike their effects, are not normally perceptible without a special apparatus. The second implies that bodily movements$_I$ are caused both by neural events and by events of another sort, and therefore raises the difficult question of how these two sorts of events are related. It also implies that bodily movements$_T$ can never be perceived, whatever sort of apparatus we are equipped with. But we can and do see people and animals moving their limbs without making use of any sort of apparatus; and seeing a person or an animal moving its limbs is seeing a bodily movement$_T$. Hence neither alternative is tenable; and it follows that bodily movements$_T$ are not events which cause bodily movements$_I$. (1998, pp.229-230)

If Alvarez and Hyman's argument succeeds, then bodily movements$_T$ are not events, so the causing of an event by an agent is some *other* sort of entity. If Alvarez and Hyman's view is the natural position for anyone who believes that in acting an agent causes something, then this theory of agency would seem to involve ontological commitment to a novel kind of entity, which is the engendering of a causal relation between an agent and an event. To give these novel entities a name let's call them 'causings'. The proposal that actions are causings naturally raises the question 'what are causings?'. Desire to answer this question is, I believe, part of the reason why philosophers sympathetic to agent-causation-based theories of action have begun to think about whether processes are a unique sort of entity. However, the view of processes which has become popular is quite different to the view I espoused in chapter 4. I will discuss, and criticise, this alternative theory of processes in section 5.4.

As mentioned, I think that actions are events. They are instances of processes. I also agree with Davidson that 'at least where basic actions are concerned "my raising my arm and my arm rising are one and the same event"' (1987, p.37; see also Davidson 1967/2001, p.128 and Anscombe 2000, pp.52-53). That is, I reject the Separation Thesis. On my view, there is the dynamic universal *raising one's arm*, which an agent engages in, and which is distinct from any event, including the event of the agent's arm going up, but I deny that there is any particular arm-raising occurrence distinct from the event of arm-rising. The agent's action, which is an instance of an arm-raising process, *is* the event of the agent's arm going up. The main advantage of this view is that it avoids proliferating the number of individual occurrences – we do not have to posit 'causings'. However, I also think that actions are instances of agent causation, which means that when an agent acts, the agent, *qua* substance, causes something. One might wonder how it is possible for me to maintain that actions are instances of agent causation *and* reject the Separation Thesis given that Alvarez and Hyman's (1998) argument seems to show that such a position is incoherent.

I believe that Alvarez and Hyman's (1998) argument is invalid, because it wrongly assumes that the expression 'caused to rise' means 'caused an arm-rising event to happen'. Alvarez and Hyman assume that when an agent raises her arm, a relation of causation comes to obtain between the agent and an arm-rising event. My own theory of agency is premised on the idea that causation is not always, everywhere a relation. I deny, whereas many contemporary agent-causation-based views of agency accept, that the Agency Relation, i.e. the relation between an agent and the event she is agent of, is a causal relation. As I explained in chapter 3, the agency relation is not a difference-making relation, and agents are not difference-makers. This is because difference-makers must be dated entities: in looking for that which makes the difference to an effect's occurring, we are looking for a part of history which stands in relation to another part of the history of the world *atemporally*; agents are not dated entities and as such do not stand in relations atemporally.

I accept the platitude that raising one's arm is causing one's arm to rise, but if agent causation is a dynamic state of affairs as I contend, then an agent can be causing her arm to rise without thereby coming to stand in a causal relation to an arm-rising event. This means that Alvarez and Hyman's (1998) argument can be resisted. It is implausible 'that causing an event to occur is not merely an event itself, but the very same event as the event caused' (Alvarez and Hyman, 1998, p.229). However, this only falsifies lemma 4 if 'causing my arm to rise' is taken to mean 'causing an arm-rising event to occur'. And why should we 'Relationalise' the infinitival phrase 'causing my arm to rise'? Why should we assume that what claims like 'the agent caused her arm to rise' mean is that an agent is the cause of an arm-rising event? As Stout suggests 'The phrase "your arm to rise" is not really a noun phrase at all and certainly does not encode some implicit reference to an entity which is the event of your arm's rising' (2010, p.104).

Ursula Coope (2007) outlines a similar response to Alvarez and Hyman's (1998) argument which is available to Aristotle, who also thought that my arm's going up, the arm-rising event, was identical with my action of raising my arm. Coope suggests that Aristotle would deny that his view commits him to the implausible idea that the causing of an event is one and the same as the event caused, because Aristotle would deny that an action is a causing of an event to happen. According to Coope's Aristotle, an action is the causing of a state, or the causing of a state of affairs to obtain:

> Aristotle's view, I shall argue, is that the power that is exercised in an action of moving X is a power to produce the end of X's movement: a power to produce a state, rather than a movement. In this sense, what I am causing when I move X is the state that X's movement is directed towards. For example, when I raise my arm, what I am causing is my arm's being up, rather than my arm's going up. More generally, the action of changing something towards being F is, for Aristotle, a particular kind of causing of the state being F. (Coope, 2007, pp.113-114)

My suggestion is, in some ways, more controversial. I am suggesting that 'causing my arm to rise' should not be given a relational interpretation at all.

That is, I am suggesting that an agent's causing her arm to rise does not entail that the agent stands in a causal relation to any particular. So not only does an agent raising her arm not stand in a causal relation to an arm-rising event, she also does not stand in a causal relation to the state of her arm being up.

There is one respect in which I think my response to Alvarez and Hyman's argument fares better than the response offered by Coope's Aristotle. Coope points out a problem facing the theory of action she attributes to Aristotle.

> This difficulty stems from his view that the action of changing something is the causing of the end state of the change: raising one's arm is causing one's arm to be up; walking to the pier is causing oneself to be at the pier. This raises an obvious question. There are, surely, different types of action that result in one's arm being up, and also different types of action that result in one's being at the pier. How is Aristotle to distinguish between these? How, for instance, is he to distinguish between walking to the pier and swimming to the pier? (Coope, 2007, p.132)

Aristotle's view seems to suggest that walking to the pier and swimming to the pier are exercises of the same agential power (i.e. a power to cause oneself to be at the pier). But, intuitively, these actions are exercise of different powers. The agency demonstrated in walking to the pier is not the same as the agency demonstrated in swimming to the pier. Coope suggested that Aristotle could avoid this problem by stipulating that there are different *ways* to cause the end state of the change, and the manner in which an end state is caused determines what power a given action is an exercise of. On my theory, an agent's causing something is their engaging in an activity, and an activity is a way for a substance to be effecting a certain kind of change, so what powers are exercised by an agent when she acts is determined by the activities she engages in. My theory thus avoids the problem facing Coope's Aristotle, but does so without having to make an *additional* stipulation, instead holding that an action of changing something is an instance of activity, and not the causing of the end state of the change.

A key difference between my neo-Aristotelian theory of agency and agent-causation-based theories of agency of the kind championed by Mayr (2011),

Lowe (2008), and Alvarez and Hyman (1998) is that these theories accept, whereas I deny, that actions are 'causings'. That is, I deny that action consists in an agent's coming to stand in a causal relation to an event. On my view, when an agent engages in an activity and then completes that activity, an event (the agent's action) comes into existence – but the agent does not stand in a causal relation to this event.

5.4 The Temporal Stuff View

Partly due to Alvarez and Hyman's argument that actions are not events, the idea that actions are a special sort of occurrence, 'causings', has gained some traction. I briefly mentioned in the previous section that the idea of actions as 'causings' has prompted some philosophers to examine whether recognising *process* as a distinctive ontological category is important for understanding action and agency. However, the theory of processes which has become popular is different from the theory of processes I advanced in chapter 4. The theory of processes which some have argued will be useful to philosophy of action is a theory I call the 'temporal stuff view' of processes. According to the temporal stuff view, processes are the 'temporal stuffs' from which events are composed. For example, Hornsby suggests that 'the relation between the stuff of the spatial world and the particulars therein is analogous to the relation between the activity [a kind of process] of the temporal world and the particulars there' (2012, p.238). And Thomas Crowther maintains that 'What things are doing throughout periods of time and substance stuff are constituents of the same basic ontological category; they could be thought of as temporal and spatial masses' and 'Both substance-stuffs and time-occupying stuffs, respectively, fill out space and time in the same way' (2011, p.17).

Steward (2013a) also argues that the spatial and temporal realms have analogous ontologies. Steward argues that, in the spatial realm, we can distinguish two sorts of particular: 'substances' and mere 'lumps of stuff'. Substances, Steward argues, are entities which can survive the loss or replacement of their spatial parts; they have 'a certain distinctive form by means

of which they are singled out in thought and which underwrites their relative independence from the actual parts of which they consist in any particular instant' (p.487). Lumps, in contrast, are such that 'the merest addition or subtraction [of spatial parts], however tiny, makes for a different lump' (p.804). Steward's view is that we can make similar distinctions between different types of occurrences in the temporal realm. She argues that there are analogues of substances in the temporal realm, entities which Steward calls 'individual processes' – these entities have a distinctive form which determines what intrusions, shortenings and lengthenings they could and could not have survived. The walk to the newsagent's I took this morning is an example of an individual process: my walk to the newsagent's could have taken a little longer than it actually did, as long as this lengthening didn't change the distinctive 'walk-to-the-newsagent's' form of my walk (e.g. my walk didn't take longer because I stopped walking and started crawling, or because I stopped walking to the newsagent's and started walking to the gym instead). According to Steward, there are also analogues of 'lumps of stuff' which Steward calls 'stretches of activity' – these are occurrences which do not have temporal boundaries indicative of any distinctive form; consequently, stretches of activity have their temporal parts essentially. In addition to these two sorts of temporal occurrence there is also 'massy' temporal stuff from which the occurrences are made.

The temporal stuff view of processes is justified on the grounds that it helps resolve certain problems in philosophy of action. For example, Crowther suggests that 'distinctions in the way that things occupy periods of time may help us to explain important concepts within the philosophy of mind and action' (2011, p.6). Steward also suggests that the concept of process, as distinct from event, 'promises to do important philosophical work – especially, as others have also hoped, in the philosophy of action' (2013a, p.782). The problem which the temporal stuff view is supposed to solve is the very problem I have suggested my non-Humean theory of causation and universals theory of processes will help solve, namely the problem of providing a theory of agency which casts the agent

as a causal player, rather than merely the setting for events to cause other events.

As previously mentioned, Hornsby has argued that to recognise the essential role of the agent in the causality of action, we need to think of causation as something other than a relation between events. According to Hornsby, what's needed to recognise the essential role of the agent in the causality of action is a metaphysical framework which provides intellectual space for thinking of causation as something other than a relation between events. Without such a framework, it is impossible to see how the causality of action might be something other than a causal relation between mental event and action, and instead something that casts the agent as a causal player rather than merely the setting for events to cause other events. Thus far, Hornsby and I agree.

However, Hornsby's own suggestion is that adopting a metaphysical framework where activity – a kind of process – is *the stuff from which actions are composed*, is what allows us to think of the causality of action as something other than a causal relation between mental event and action. Hornsby suggests that 'the agent is given her due only when it is acknowledged that she engages in activity, where no activity is any particular' (2012, p.233). She claims that:

> In order to recognise causality as present on an occasion of someone's raising her arm, one needs to think of a person's raising her arm as a type of causal activity in which she engages. (2012, p.234.)

The idea here is that once we accept that an agent's causing something (for example, an agent's causing her arm to go up) is an activity or process, i.e. the temporal stuff from which her action is composed, we are permitted to think of the causality of action as something which essentially involves the agent herself. The causality of action is thus thought of as the activity which composes the agent's action, and once we acknowledge this, we are no longer at risk of failing to include the agent in an account of the causality of her action – or so Hornsby proposes.

However, I do not think that the temporal stuff view, and its associated notion of temporal composition, successfully explains how the agent of an action is necessarily involved in the causality of her action. This is because I do not think the relationship between the agent and the causality of her action is made any clearer by positing the existence of temporal entities which are distinct from, but compose, events, in the same way that matter composes substances. The aim, in proposing a process ontology, is to enable an account of agency which casts the agent as a causal player, rather than merely the setting for events to cause other events. Proponents of the temporal stuff view try to achieve this goal by claiming that the agent's playing her causal role is a unique sort of entity, which is related to the agent's action by something analogous to composition. However, why is an entity which is related to an agent's action by 'temporal composition' uniquely well-suited to be *the agent's* causing of something? What makes the entity that composes an agent's action *her* causing of something? How is the agent attached to, or involved in, the temporal stuff which composes her action? What is it to 'engage in' a temporal stuff? It is not clear to me that answers to these questions are contained within the temporal stuff ontology and its associated concept of temporal composition. In short, as long as processes/activities are thought of as the temporal stuff which composes an agent's action, the connection between the agent and the process/activity she engages in will remain obscure – and this is precisely the relationship we need to understand, if we are to understand the agent's place in the causality of her action. Drawing analogies between processes and matter, and invoking the notion of 'temporal composition', may help with the construction of a metaphysical framework where processes and events are clearly distinguished – but it seems to me that the notion of 'temporal composition' does not help us understand the agent's place in the causality of action.

My view abides by Hornsby's suggestion that 'the agent is given her due only when it is acknowledged that she engages in activity, where no activity is any particular' (2012, p.233), but it solves the problem of recognising the essential role of the agent in the causality of action without positing temporal

stuff or invoking the concept of temporal composition. On my view, the causality of action is an agent's engaging in activity – it is a dynamic state of affairs, and as such, is an entity which has the agent as a constituent (just as the state of affairs of the rose's being red has the rose as a constituent). The agent herself partially constitutes the causality of her action – she is, therefore, essentially involved. For example, the dynamic state of affairs reported by the sentence 'Sally is raising her arm', is *Sally's* causing of something because Sally partially constitutes the dynamic state of affairs reported; she is the particular element of this dynamic state of affairs.

Steward offers a slightly different account of how the temporal stuff view helps us understand agency and action:

> […] it is natural to think of the cause of an event simply as another event, or perhaps a collection of pre-existing events and states which give being to the caused event simply by triggering it off. But what is the cause of a token process, conceived of as an entity with a robust form, a normal course of development, something which can obtain extra temporal parts without detriment to its continued identity? Not only what triggers it, but also what sustains it, what keeps it on course, what prevents it from ceasing or disintegrating. In this observation lurks the promise of an account of activity from which the agent does not suffer the disappearance so often complained of in event-based views, precisely because she is needed in order to ensure the continuation in the right direction of the process which constitutes her activity. (2013a, p.810)

Steward suggests that because individual processes have a distinctive form, to cause such an occurrence, there needs to be both a trigger and something which can ensure that the individual process continues and develops in accordance with its distinctive form. In the case of individual processes which are actions, this latter role can only be played by the agent. In other words, because of the distinctive forms actions possess, for actions to exist they must be 'sustained' by agents. On Steward's view, to understand agency we need to acknowledge that agents ensure that the (massy) process which composes their actions takes on a specific form.

Steward's view avoids the problem of failing to explain how the agent is necessarily involved in the causality of her action, because on Steward's view the agent is required to, as it were, mould the temporal stuff that composes her action into its distinctive shape. However, Steward's view faces another problem. The problem with holding that agents cause their own actions, outlined above, is also a problem for Steward's proposal that agents sustain their actions. Supposing that agents sustain their actions, keep their actions on course and prevent their actions from ceasing or disintegrating, is sustaining one's action itself an action? If it is then it is also presumably the sort of entity that needs to be sustained by an agent, in which case we are in danger of requiring that whenever an agent acts there is an infinity of 'sustaining' actions they must perform. However, if sustaining one's action is *not* itself an action, then what is it? What is it for an agent to ensure than her activity continues and develops in accordance with the distinctive form of her action?

My view avoids this problem as on my view agents are not causally related to their actions or to their activities. On my view, agents engage in their activities, but *engagement* is not an 'element of being', it is rather the concept by which we understand the ontological form of substances and processes. This means that the relationship between the agent and the causation exemplified by her action, i.e. the activity she engages in, is not a relation at all. The agent does not stand in any relation to her activity. Similarly, engaging in an activity is not itself an activity, and there are no actions which are instances of engagement. It is not the case therefore that whenever an agent acts there is an infinity of actions they must perform. Furthermore, as mentioned, on my view, agents are not causally related to their actions (which are instances of their activities). Rather, actions *existentially depend* on agents engaging in activities and then completing them.

5.5 Agential Power

I have compared my metaphysics of action to three different kinds of agent-causation-based theory: traditional agent causationism, agent-causation-based

theories which take actions to be causings, and agent-causation-based theories which posit a 'temporal stuff' metaphysics to help explain what action is. These comparisons focused on issues related to what an action is, what it is for an agent to be causing something, and how we ought to think of the relationship between an agent, her activity and her action. However, as I mentioned above, providing a *metaphysics of action* is not all that is required for a complete and adequate theory of *agency*. To provide a complete theory of agency one must provide a dissection of the concept of agency. In this section, I will defend my understanding of this concept.

On my view, agency cannot be identified with either the exercise of active power or with the exercise of two-way power. Instead, *both* concepts are key to understanding agency. Among the agential powers things have are powers like the power to bend, the power to crush, the power to fold, the power to build. What all these various agential powers have in common is that they are all powers to control what transpires, in one way or another. But 'control' is an ambiguous concept, which is arguably no easier to understand than the agency concept. My proposal is that competence with the agency concept demands that one have some grasp of at least two distinctions: between activity and passivity, and between settling and non-settling. These distinctions capture two important ways in which an agent can be said to be controlling what transpires. In this section, I will outline my reasons for adopting this view.

Hyman claims that 'to act is to intervene, to make a difference, to make something happen, to cause some kind of change' (2015, p.33). On Hyman's view, the connection between agency and substance causation is very tight. Agency is a causal concept on his view; the power of agency is the power of a substance to be active or to cause change. Agents *cause* change and should be contrasted with patients, who *undergo* or *suffer* change (Hyman, 2015, p.34). This conception of agency is compatible with acknowledging that 'action occurs throughout the world, some of it voluntary and some of it not, some of it by human being and some of it not, some of it by living beings and some of it not' (p.29). According to Hyman, the idea that this concept does not apply in non-

human, or non-voluntary, contexts is the result of combining the false doctrine that matter is inert with the idea that 'human beings, or in some philosophical systems living beings, are not constrained in their behaviour by the same conservation laws as the rest of the natural world' (p.29). Hyman acknowledges that human agency is distinctive in many ways, but agency is still a broad and highly abstract concept, of which human agency is just a special sub-species.

Steward (2012) argues that agency is the exercise of two-way power. For Steward, a key part of our concept of action is that agents settle heretofore open questions (like whether or not they will φ) when they act. What it is to act is to settle some matter. On Steward's view, 'agents are entities that things can be up to' (p.26). Steward, like me, thinks that the power to settle a question should be understood as a two-way power. So, because Steward thinks that agency is essentially settling, she thinks agency is essentially the exercise of two-way power.[43] Not all examples of substance causation count as examples of agency on Steward's view, as not all examples of substance causation are exercises of two-way power: inanimate substances, for example, cannot exercise two-way powers. Such substances are not able both to cause some change and refrain from causing such a change. Furthermore, not all examples of agency are examples of substance causation – sometimes when substances demonstrate agency they do so by refraining, by *not* acting, by not causing a change which they could have caused.

I think both views have got something right and something wrong. Hyman's view wrongly entails that refraining from acting is never a demonstration of agency (unless it is something which is achieved by performing some positive action). Steward's view wrongly entails that substances which are incapable of settling (such as inanimate objects) cannot demonstrate agency. My

[43] Steward's conception of two-way power is somewhat different to my own, but for the purposes of this chapter it is safe to ignore this disagreement. See chapter 2, section 2.1.2.

preferred view holds that agential powers are sometimes one-way and sometimes two-way. Substances which possess only one-way powers to act are agents only when they are active; for such agents their exercising agential power is their exercising active power. However, substances which possess two-way powers can be agents not only when they are active, i.e. when their exercising agential power is their exercising active power, but also when they are passive.

5.5.1 Agential Power is not Active Power

Alvarez (2013) acknowledges that there is 'a pre-theoretical notion of agency according to which an agent is something or someone that makes things happen, something or someone with the power to cause things. This pre-theoretical notion of agency extends to animals and plants, and also to inanimate things' (2013, p.102). However, she argues that *human* agency should be understood in terms of the exercise of a two-way power. Alvarez proposes that, for human beings, their agential powers are two-way powers, and every instance of human agency is an instance of the exercise of a two-way power, or something done by exercising a two-way power.[44] Alvarez motivates this view by providing a diverse selection of examples of human agency and arguing that what these diverse examples have in common is that they are all exercises of two-way power.

First, Alvarez cites examples where human beings cause an event to occur by not doing something: 'we can cause offence by not greeting someone, cause a

[44] The clause 'or something done by exercising a two-way power' is added to cover cases such as the following: I press a button, which unbeknownst to me issues an order to launch nuclear missiles, and thereby start a war. I did not know pressing the button was a way of starting a war, so I did not possess or exercise a two-way power to start a war. Nevertheless, starting the war was an action of mine. Alvarez allows for this kind of case by pointing out that in this case starting a war was done by exercising a two-way power to press a button.

death by not feeding someone, and so on' (2013, p.104). These are examples of *allowing* something to happen, and according to Alvarez they 'suggest that not doing something itself can be an instance of agency' (p.104). Alvarez considers and rejects the objection that when we causally attribute an event to someone's inaction this is not a genuine causal claim as being unduly restrictive.[45] Importantly, in cases of allowing, the putative agent does not exert any active power. For example, in failing to water my plant, I do not *actively* cause the death of the plant. Substances in the vicinity which might have actively caused the death of the plant probably include parts of the plant itself (e.g. the plant's chloroplasts may have actively caused the death of the plant by using up what water was stored in the plant, thereby causing the plan to wilt, which in turn prevented the plant from capturing light etc). In cases of allowing, like the case where I allow my plant to die by not watering it, I demonstrate agency by letting the active powers of *other* substances to manifest themselves, rather than by exercising any active powers myself.

Alvarez also cites examples of preventing something from happening. One can prevent something from happening either by doing something – 'for instance, I can prevent a paper from flying away by holding it down; doing that is not causing but is rather preventing a change by doing something: holding the paper down' (2013, p.106) – or by not doing something – 'for example, if I stand motionless in front of a laser-beam mechanism that controls a door, and thus keep it open' or prevent it from closing, or 'an ambassador may prevent a diplomatic incident by keeping quiet when provoked' (p.106). This latter kind of example, preventing-by-not-doing, is particularly interesting as this kind of example is more obviously a case where the putative agent does not exercise any active power.

[45] See also Alvarez's (2001) argument for the conclusion that an agent who lets an event happen causes it.

Alvarez thinks that both allowing and preventing-by-not-doing are manifestations of agency. I agree: I think these cases count as demonstrations of agency, even though they do not involve the agent exercising active power. It should be granted, however, that these examples are not descriptions of actions. Refraining from acting, or failing to act, is not a kind of action. I agree with Hyman that 'ensuring that something doesn't change by not changing something may be as consequential as an act, but it is not an act, even though not doing something can be deliberate and voluntary, and one can try not to do something such as bite one's nails, and either succeed in the attempt or fail' (2015, p.14). Actions are what come into existence when an agent engages in an activity and completes that activity. If an agent refrains from, or fails to, engage in an activity, then no action will come into being. Some philosophers think that there are such things as negative actions (e.g. Vermazen, 1985), but to posit not-doings in addition to doings, is to proliferate the number of occurrences beyond necessity. However, even though refraining from acting, or failing to act, is not a kind of action, this does not mean that such cases are not manifestations of agency. *Agency* and *action* are distinct concepts, even though they are closely related. Something can be an instance of agency, or a manifestation of agential power, even though it is not an action. I think the agency concept extends to cases of allowing and preventing-by-not-doing, in part because such cases are ethically important: people can be held morally responsible for refraining from acting, and failing to act, and it would be unjust to hold people morally responsible for refraining from acting, and failing to act, if these weren't demonstrations of agential power.

It might be argued that the ethical significance of allowing and preventing-by-not-doing does not hinge on whether such cases can be counted as an instance of agency, but on whether such cases can be counted as instances of *voluntariness* or *intentionality*. One might argue that it is *voluntariness* or *intentionality* which is important for moral responsibility, not agency *per se*. I think there are two problems with this line of argument. First, I am not sure that cases of allowing and preventing-by-not-doing can really be examples of

voluntariness and intentionality *without* being examples of agency. More specifically, I doubt that something can be counted as an instance of intentionality or voluntariness if it cannot be counted as a demonstration of the kind of control over the course of events we would want to call agency. (Of course, I accept that there can be examples of voluntariness and intentionality which are not examples of *activity* or substance causation – but that is not what is at issue.)

Second, I think that there are some cases of allowing and preventing-by-not-doing which fall short of being intentional, and hence fall short of being voluntary, but where we would still assign moral responsibility.[46] For example, suppose I am in the kitchen while some pasta is cooking, but I am distracted, so I allow the pasta to boil over by not turning the gas down. In this case, I did not intentionally let the pasta boil over. However, I think I would be held responsible for letting the pasta boil over. A case of non-intentional preventing-by-not-doing might be not moving out of someone's way on a busy train. I may not have noticed that I am blocking someone's path, so it would be incorrect to say I am blocking their path intentionally. However, I think I can be justly held responsible for impeding their progress. In these cases, moral responsibility does not depend on the intentionality or voluntariness of the passivity – so what does it depend on? Determining a precise answer to this question requires much more detailed examination of the concept responsibility than I can undertake here. However, I think it is plausible that responsibility for unintentional omissions will depend, in some way, on whether agential powers were operational.[47]

[46] I am taking it for granted that being intentional is a necessary, but insufficient condition for being voluntary.

[47] Raz suggests that we are responsible for outcomes 'which result from failure, due to the malfunctioning of our capacities of agency, to complete as intended an action within our domain of secure competence' (2010, p.21). In other words, we are responsible for unintentional omissions when our conduct would have been guided by our rational capacities were it not for the fact that these capacities were malfunctioning. The concept of agential power thus plays a part in Raz's account.

5.5.2 Agential Power is not Two-Way Power

Alvarez is right that most of the time when human beings exercise their agential powers, the power they are exercising is a two-way power. For human beings, their agential powers are, in most cases, two-way. (I also think this holds for many animals.) Recognising this means that we can explain how agency can be demonstrated in passivity as well as in activity: when one's agential power is two-way, one can demonstrate this power by *not* performing the action one's agential power is a power to do. However, this doesn't prove that *all* agential powers are two-way powers, or even that all *human* agential powers are two-way powers.

Agency can be found where the power to settle is lacking; that is, there can be agency even in cases where what happens is not up to the agent. In particular, inanimate objects can be agents even though inanimate objects do not possess two-way powers. To claim that our agency concept does not really apply to inanimate objects seems inconsistent with the fact that we use the same causative verbs to report human actions as to report what inanimate objects have done.[48] I also think that, in some circumstances, human beings' agential powers are one-way rather than two-way. For example, when I am unconscious, my power to crush a crushable object (e.g. by rolling onto it) is one-way. So, unlike Alvarez, I do not think that all human agential powers are always two-way. Human agency is *sometimes* a matter of two-way powers being exercised and *sometimes* a matter of one-way powers being exercised.

An example which seems to speak against my proposal that human agential powers can be one-way or two-way is Davidson's example where I spill my coffee because you jiggle my hand (1971/2001 p.45). In this case, it was not up to me whether my coffee spilled. It was not me who settled whether my

[48] See chapter 2, section 2.2.5.

coffee would spill: it was you. In this example, I am a helpless victim. It is even legitimate to say that in this example, *I* did not spill my coffee, you did. I was just the conduit for *your* action. If all human agency is the exercise of two-way power, as Alvarez proposes, we can explain why there is no agency in this example by appealing to the fact that it was not up to me whether my coffee spilled: I did not possess or exercise a power to spill my coffee or not.

Hyman provides an alternative way to accommodate the intuition that I am not an agent in Davidson's example, which does not depend on the idea that all human agency is the exercise of two-way power. Hyman contends that when a substance is complex, i.e. has parts, the substance's status as an agent depends on the integrated operation of the substance's parts. Hyman proposes that 'all animal agency is really collective agency', and so 'individual human agency is always really collective agency, since a human being, like every other multicellular organism with specialised tissues, is in reality a highly integrated colony of functionally differentiated but genetically similar cells' (2015, p.48). Therefore, an act can be considered a demonstration of a human being's agency, rather than a demonstration of the agency of some proper part of that human being, when it involves the integrated operation of 'metabolic, motor, and cognitive systems' (p.50). In Davidson's example, my spilling of my coffee doesn't involve the integrated operation of all my functionally differentiated parts: my arm is certainly involved, and perhaps my metabolic and motor systems too (or else my arm would not be gripping my coffee cup to start with), but my cognitive systems are not involved, therefore, I am not the agent.

A third alternative is to say that in Davidson's example, despite intuitions to the contrary, I am an agent, it is just that in this example a) I am not exercising a two-way power and b) I am also a patient. The case is confusing because human beings' agential powers can be one-way or two-way. The agency of human beings is *usually* a matter of exercising two-way powers – we are usually settlers – so on the odd occasion where we cause things to happen without settling, we might be inclined to say our agency has disappeared – but it hasn't, it has just been down-graded. Although this last way of dealing with

Davidson's example might seem unappealing, I prefer it. I have considered three alternative analyses of the agency concept in this section and all three require that we give up an intuition. Identifying agential power with active power means that we have to deny that cases of allowing and preventing-by-not-doing are demonstrations of agency; identifying agential power with two-way power means that we have to deny that inanimate objects are capable of exercising agential power. My preferred view, which holds that agential powers are sometimes one-way and sometimes two-way, saves both the intuition that cases of allowing and preventing-by-not-doing can be demonstrations of agency, and the intuition that inanimate objects are capable of exercising agential power, but it conflicts with the intuition that in cases like Davidson's I am not an agent. Of the three intuitions mentioned here, I am least attached to the third.

In this chapter, I outlined a neo-Aristotelian metaphysics of action, which draws upon the process ontology outlined in chapter 4, and my analysis of the agency concept. I contended that what it is for a substance to be causing something is for there to be an activity which the substance is engaging in. A substance engaging in an activity is an agent, and the event that results once the substance has completed the activity it has been engaging in is an action. Actions are thus instances of activities, and as engaging in an activity is what it is for an agent to be causing something, actions can also be said to be instances of agent causation. In sections 5.2-5.4, I compared this metaphysics of action with three prominent agent-causation-based theories of action, and argued that my account has advantages over each. I also suggested that the agency concept should be understood in terms of two distinctions: the distinction between activity and passivity, and the distinction between one-way and two-way powers, and I defended this analysis in section 5.5.

6

Action Explanation

As stated in chapter 1, the problem of mental causation is the problem of how to reconcile the idea that our mental states have some sort of causal influence over how we act, with the principle of causal closure, which entails that every bodily movement has a completely physical mode of generation. Few philosophers writing on the problem of mental causation are willing to resolve this tension by giving up the idea that there is mental causation. There is good reason for this reluctance. It is natural to think that some form of mental causation is indispensable to our conception of ourselves as agents who act intentionally and bear moral responsibility. Our conception of ourselves as agents who sometimes act intentionally presupposes 'the reality of causal processes involving cognitive phenomena', as Menzies (2013, p.58) puts it. Given that we sometimes act intentionally, I don't think it can be doubted that mentality is causally relevant in the physical world – but this claim is very vague. What exactly is it for mentality to be causally relevant in the physical world?

In chapter 1, I contended that philosophers writing on the problem of mental causation often assume what I called the relational understanding of mental causation:

> **Relational understanding of mental causation:** mental causation is a matter of mental items (events, processes or states which are conceived of as particulars) standing in causal relations to physical events, e.g. bodily movements.

When philosophers writing on the problem of mental causation claim that the possibility of intentional action presupposes the existence of mental causation, it is often the relational understanding that they have in mind. For example, in

Physicalism, or Something Near Enough, Kim gives the following reason why it is important that mental causation is real:

> First and foremost, the possibility of human agency, and hence our moral practice, evidently requires that our mental states have causal effects in the physical world. In voluntary actions our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged. (2005, p.9)

What Kim thinks is required by the possibility of human agency – and specifically voluntary action – is that 'our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move' which either means that mental states are themselves the agents of bodily movements, or that mental states stand in causal relations to bodily movements. Later in the same work, Kim claims that 'mental phenomena must be capable of functioning as indispensable links in causal chains leading to physical behaviour, like movements of the limbs and vibrations of the vocal cord' (2005, p.10). Kim assumes mental causation is a matter of mental items (states or events) standing in causal relations to physical events.

The question of whether our concept of intentional action presupposes mental causation is related to the question of what an intentional action is, or what it is to act intentionally. Many have tried to provide an account of intentional action by examining the way we typically explain intentional actions. It is commonly held that intentional actions can be explained by the reasons for which they were performed, or by citing the motive or aim of the agent. I will call such explanations 'rationalising explanations'. The assumption made by philosophers who theorise about what intentional actions are is that we can achieve an adequate account of what it is to act intentionally by examining the distinctive sort of explanation with which intentional actions are associated.

Davidson (1963) argues that explanations of intentional actions which cite the agent's reasons are causal explanations. They are true if a mental event suitably related to the reason stands in a causal relation to the action. Hornsby suggests that many (including Davidson himself) have assumed, on the basis of

Davidson's argument, that reason-giving explanations of intentional actions are causal, that 'beliefs and desires stand to actions as causes to effects (with decisions or intentions maybe intervening)' (2015, p.134). Hornsby's point is that Davidson's argument is often taken to justify the claim that mental states or events stand in causal relations to actions. Davidson's argument seems to be the source of the common view that our conception of ourselves as intentional agents presupposes that mentality is causally relevant in the physical world *and* that this mental causation should be conceived of in relational terms.

In section 6.1, I will look at explanations of intentional action which cite the agent's reasons. I will clarify what I take reasons to be, and how they relate to two important sorts of mental state: states of believing and states of desiring. In section 6.2, I will outline Davidson's argument for thinking that explanations of intentional action which cite the agent's reasons are causal. In section 6.3, I will consider objections to the idea that rationalising explanations are made true by mental events or states standing in causal relations to the actions explained. I will conclude that while Davidson's opponents, whom I shall call 'non-causalists', make several correct claims about how rationalising explanations work, their objections do not successfully refute Davidson's position. This is because, firstly, the idea that rationalising explanations are causal is very appealing, and secondly, the availability of a metaphysical theory like Anomalous Monism allows the Davidsonian to accept the correct claims made by non-causalists even while insisting that rationalising explanations are made true by causal relations between events. In other words, I will conclude that the debate between Davidson and non-causalists is at something of an impasse. In chapter 7, I will explain how my non-Relationalist view of causation, and neo-Aristotelian view of agency, provides a way of resolving this impasse.

6.1 Rationalising Explanation

We do something intentionally when we do it because, in a very broad sense, we want to. 'Want to' needs to be understood broadly so that mental states as diverse as craving a cup of coffee, having a life-long ambition to climb Everest,

and seeing that some otherwise detestable action is necessary in order to achieve something valuable all count as wanting to perform the action in question. This is because making a cup of coffee to satisfy a craving, climbing Everest to fulfil a life-long ambition, and performing a detestable action in order to achieve something valuable are all examples of intentional action.

Often (but not always), when we explain an intentional action, that is, say why the agent acted (intentionally) as she did, we do so by giving the agent's reason for acting as she did.[49] I will call such explanations 'rationalising explanations'. Rationalising explanations are explanations which work by showing why what the agent did seemed, to the agent, like a rational or sensible or good thing for the agent to do, given certain facts (e.g. what the agent's preferences or values are). Rationalising explanations explain why an agent acted as she did (this is the explanandum) by telling us why, in the agent's eyes, what they did was a rational thing for them to do (this is the explanans).

Rationalising explanations should be distinguished from what I will call 'mere rationalisations'. Mere rationalisations are similar to rationalising explanations in that they also tell us why the course of action taken by the agent seemed, to the agent, to be a rational course of action to take. However, what mere rationalisations explain – i.e. the explanandum of a mere rationalisation – is slightly different. Mere rationalisations do not tell us *why an agent acted as she did* – they *only* tell us why what the agent did seemed, to the agent, to be a rational thing for them to do. For example, imagine that Diana is deciding whether or not to speak at a conference. She knows that speaking at a conference will be good for her career, but in the end, she decides to speak at the conference because it will draw praise from her friends, and not because it will be good for her career (perhaps she does not really care about her career).

[49] Two examples of an explanation of an intentional action which does *not* cite the agent's reasons or motives are: 'Sally bit the policeman because she was drunk' (Hyman, 2015, p.105) and 'She threw the water at him because she was angry at him'.

Diana actually spoke at the conference because she would get praise from her friends, not because it would be good for her career. In this context, the following statement would be a mere rationalisation of Diana's action:

(a) Speaking at the conference seemed rational to Diana because it would be good for her career.

This is a mere rationalisation because it explains why speaking at the conference seemed to Diana to be a rational thing for her to do – but it does not explain why Diana actually spoke at the conference. Consequently, *sometimes* citing a consideration which, in the agent's eyes, renders a course of action a rational course of action for the agent to take will *not* explain why the agent took that course of action.

A key question, one which has a bearing on whether rationalising explanations are causal or not, is *how* rationalising explanations explain. How does a statement telling us why what an agent did seemed to them to be a rational thing to do explain why the agent did as she did? How does the explanans of a rationalising explanation illuminate the explanandum? This question is especially pertinent given that *sometimes* a statement telling us why what an agent did seemed to them to be a rational thing to do does *not* explain why the agent did as she did. However, before attempting to address this question, a prior question needs to be addressed, namely, what exactly is the explanans of a rationalising explanation? I have said that it is a statement which tells us why what an agent did seemed to them to be a rational thing to do – but what sorts of statement are able to do this? I will assume throughout that the explanans of any explanation is always a fact. That is, I assume that explanation is a relation between facts and only facts can explain other facts.[50] So, the first question to address is: what sort of facts can tell us why a certain sort of action

[50] Van Fraassen (1980, p.134-153) proposes a theory of explanations as answers to why-questions where both the answer and the topic of the why-question are true propositions. Raley (2007) has also defended the view that all explanation is factive.

seemed to an agent to be a rational thing for them to do, and by doing this also tell us why the agent performed that sort of action?

To this question, I think Davidson has essentially the right answer. Davidson claims that the explanantia of rationalising explanations are facts about what the agent wants to do (or what the agent has an urge to do, or what the agent has an ambition to do) and what the agent believes about how to do it. Facts about what an agent wants and believes can tell us why a certain sort of action seemed to the agent to be a rational thing for them to do and why the agent performed that sort of action. Davidson calls the dual possession of a desire to perform some type of action and a belief about how performance of that action may be achieved 'the primary reason why the agent performed the action' (1963/2001, p.4). Rationalising explanations don't typically take the form 'agent A φed because A wanted to φ and believed that ψing was a way to φ'. Sometimes this is because it suffices to explain why someone acted as they did to only mention what the agent wanted to do. For example, in (b) Beth's action is explained in terms of her desire only:

(b) Beth is buying flour because she wants to make bread.

We don't need to be told that Beth believes or knows that buying flour is an essential preparatory action for making bread. We take it for granted that Beth possesses this knowledge. Other times it is because it suffices to explain why someone acted as they did only to mention what the agent believes, or knows, about how to achieve what they want to do. For example, in (c), John's action is explained in terms of his belief only:

(c) John is adding rosemary to the sauce because he believes it will make it taste better.

We don't need to be told that John wants to make the sauce taste better – we take it for granted that he wants this. Davidson's point is not that all rationalising explanations *explicitly* give the primary reason why the agent acts, his point is rather that for the explanans of a rationalising explanation to illuminate the explanandum, 'it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason' (1963/2001, p.4). That is,

the explanatory power of rationalising explanations rests on our ability to construct a primary reason from any rationalising explanation.

Although I think Davidson is broadly correct in thinking that the explanantia of rationalising explanations are facts about what the agent wants and believes, there are two complications. Firstly, when an agent acts for a reason, the reason for which they act is not usually a fact about the agent's own mental states. For example:

(d) David took the A-road because the motorway was shut.

In (d) David's reason is 'that the motorway was shut', not 'that David believed or knew that the motorway was shut'. At least, that is how things seem.

How does this square with Davidson's claim that the primary reason why an agent acts is a belief-desire pair? Is the appearance that David's reason is 'that the motorway was shut' and not 'that David believed or knew that the motorway was shut' an illusion? Or are there different senses to the term 'reason'? The question of what reasons are or can be has a bearing on the question of whether acting intentionally is to act for a reason. Davidson thinks that intentional action can be defined as an action done for a reason. However, sometimes we do things intentionally but 'for no reason' – such as when one hums a tune to oneself, or spontaneously decides to take the scenic route home. Davidson responds by claiming that 'for no reason' means 'for no *further* reason', that is no reason besides wanting to do it (1963/2001, p.6). But this response assumes an understanding of 'reason' where reasons are always facts about the agent's mental states. It ignores the possibility that there may be a sense of 'reason' where 'for no reason' can be taken literally, and consequently where acting intentionally and acting for a reason are not the same thing.

The second complication is that we typically explain an agent's intentional action in terms of what the agent believes *only* when they are acting on a false belief. When the agent is acting on something they know, we typically give a rationalising explanation like (d), which does not explicitly mention any facts about the agent's mental state at all. In other words, the form rationalising explanations typically take varies systematically in accordance with whether the

agent was acting on something they knew or something they falsely believed. Rationalising explanations of the form 'A φed because A believed that P' are typically only given when P is false.

6.1.1 Reasons-for-acting, Reasons-why and Aims

The best way to tackle the first complication is, I think, to acknowledge that the word 'reason' can be used in more than one way. Firstly, the term can be used to denote an agent's *reason for acting*. I follow Alvarez (2010) in thinking that an agent's reason-for-acting is that which makes the action a sensible or rational or good thing to do. As Alvarez puts it, an agent's reason-for-acting is 'the desirability characterisation' the action has for the agent. As such, reasons-for-acting are not usually facts about an agent's mental states. Strictly speaking, Beth's reason for buying flour is not that she *wants* to make bread and *believes* that buying flour is a means of doing so. Similarly, David's reason for taking the A-road is not that he *wants* to get some place and *believes* that, because the motorway is shut, taking the A-road is the only means of getting there. Alvarez (2010) explains why one ought not to think that the mental state of wanting to φ and believing that ψing is a means of φing is itself the reason for which the agent acted:

> This suggestion seems plausible but, on examination, it is unconvincing. Remember that my reason for acting is the desirability characterisation that an action has for me. But, if I want to φ and believe that ψing is a means of φing, what makes ψing desirable to me is precisely that: namely, that my ψing is a means of φing (which is something I want). The desirability characterisation that my ψing has is that it is a means of φing. (2010, pp.109-110)

The good Beth sees in buying flour is that buying flour is a necessary preparatory action for making bread (something that she wants). The good David sees in taking the A-road is that, given that the motorway is shut, taking the A-road is the only way he can get to where he wants to go. Of course, *sometimes* wanting to do something may be an agent's reason for acting. For example, someone who wants to eat coal may take her wanting to eat coal as a

reason to take a pregnancy test. Similarly, *sometimes* believing something is an agent's reason-for-acting. For example, someone who believes she is being pursued by the security services may take her so believing as a reason to see a psychiatrist. But such cases are exceptional.[51] That you want something and/or believe something is not *usually* what makes an action seem sensible to you. Reasons-for-acting are not usually facts about an agent's own mental states.

As well as being used to denote the desirability characterisation an action has for an agent, the word 'reason' can also be used as a synonym for 'explanans'. When we give *the reason why* such and such is the case, we are providing an explanans. Reasons-why are explanans of explanations. I think Davidson's claim that primary reasons given by rationalising explanations are belief-desire pairs is plausible only if 'primary reason' is taken to mean 'primary reason-why' or 'primary explanans', because reasons-for-acting are not usually facts about the agent's own mental states. However, I believe that primary reasons-why, i.e. the primary explanantia, of rationalising explanations *are* facts about what the agent wants and believes. That is, I believe that the explanatory power of rationalising explanations rests on our ability to construct a belief-desire pair from any rationalising explanation.[52]

If reasons-for-acting are desirability characterisations, then this means that (d) explicitly gives David's reason-for-acting, but (b) does not give Beth's reason-for-acting. Instead of giving Beth's reason-for-acting, I think that (b) gives Beth's *aim*, or *purpose*, or *goal*, or *motive*, or *intention*[53]: what Beth wants to do,

[51] This point has been clearly made by Dancy (2000, p.125) and Setiya (2011, p.132).
[52] Or a knowledge-desire pair – see section 6.1.2.
[53] I take these five terms to be synonyms. In some contexts, it may be useful to draw distinctions between aims, purposes, goals, motives and intentions. Anscombe, for example, suggests that

i.e. *to make bread*, is Beth's aim, it is not her reason. It is controversial to distinguish between reasons and aims – some would say that 'to make bread' is Beth's reason. However, if Alvarez (2010) is correct in thinking that an agent's reason-for-acting is that which makes the action a sensible or rational or good thing to do, then the infinitival phrase 'to make bread' does not express a reason because this infinitival phrase does not state the good Beth sees in either buying flour or making bread – it is not a desirability characterisation of either activity.

Like reasons-for-acting, aims are not facts about mental states of the agent. Specifically, aims are not facts about desires. Beth's aim is 'to make bread', it is not the fact that Beth wants to make bread. Indeed, aims are not facts at all. Aims are expressed by an infinitival phrase, for example, Beth's aim is 'to make bread'. Facts, on the other hand, are expressed by that-clauses. Perhaps one might argue that the infinitival phrases which express aims can be translated, without any change in meaning, into that-clauses which express facts. For example, perhaps 'to make bread' could be translated into 'that Beth will make bread' or 'that Beth may make bread'. To this suggestion, I would reply that neither 'that Beth will make bread' nor 'that Beth may make bread', as they are normally understood, capture the meaning of 'to make bread' when the latter is used to express an aim. This is because 'that Beth will make bread' and 'that Beth may make bread' are usually taken to be assertions, the first makes an assertion about how things will be in the future, the second makes an assertion about how things might be in the future.

there is 'popularly' a distinction between 'motive' and 'intention'. A motive is 'the spirit in which' someone does something; so, when one acts out of pity, pity would be one's motive (2000, p.18) Whereas an intention, in the popular sense, is the end to which one's action is a means; it is 'a future state of affairs to be produced' (pp.18-19). Such distinctions are useful, but they do not matter for my purposes here. What matters for my purposes here is that rationalising explanations like (b) do not give explicitly the agent's reason-for-acting. Instead they explain the agent's actions in terms of something else, something which is expressed by an infinitival phrase like 'to make bread'.

Kenny (1975) suggested that assertions contrast with things like commands and wishes. The contrast can be captured in the following way. Grant that both assertions and commands can be said to have a content which describes a state of affairs. So, suppose the content of 'that Beth will make bread' describes a future state of affairs where Beth is making bread, and suppose the content of the command 'Beth! Make bread!' also describes a future state of affairs where Beth is making bread. Now suppose this state of affairs does not obtain: in the future Beth does not make bread. In the case of the assertion, if in the future Beth does not make bread, this means that there's something wrong with the assertion, the assertion is false. In the case of the command however, if in the future Beth does not make bread, this means there's something wrong with the world, the world has failed to conform with the command. Aims are like commands. They are not statements about what the future is like, or even statements about what the future might be like; they are more like private decrees one makes to oneself. They are measured against the world in a different way to assertions. When the facts fail to conform with them, the fault lies with the world.[54]

Distinguishing reasons-for-acting from reasons-why also lets us see how 'I acted intentionally but for no reason' can be literally true. This sentence is literally true when 'reason' is taken to mean 'reason-for-acting'. When I hum to myself intentionally but for no reason, I hum to myself because I want to, but not because humming has a desirability characterisation for me. When Davidson claimed that intentional action could be defined in terms of acting for a reason, it is likely the first sense of 'reason' that he had in mind. But, if it really is the first sense of 'reason' that Davidson had in mind, then his claim is false. Acting intentionally and acting for a reason are not the same thing, because I can act intentionally without acting for a reason. Claiming that intentional action is

[54] See Kenny (1975, pp.29-45).

action for which there is a ‘reason-why’ is true, but it is no definition of intentional action. ‘Reason-why’ just means ‘explanans’, and plenty of actions which are not intentional have an explanans (e.g. I ducked because there was a ball was coming my way). Claiming that intentional action is action for which there is *a special kind* of ‘reason-why’, namely the kind of ‘reason-why’ given by a rationalising explanation, is true but also fails as a definition of intentional action. This is because there is likely no way of distinguishing the special kind of reason-why from other sorts of reason-why except by saying that the special kind are those which explain intentional actions. In other words, such a definition would be circular.

6.1.2 Acting on a False Belief and Acting on What You Know

The second complication is that rationalising explanations of the form ‘A φed because A believed that P’, i.e. those which explicitly reference the agent’s beliefs, tend to be given only when it is false that P. We typically only give a rationalising explanation of an agent’s intentional action which explicitly mentions their belief when the agent is acting on a false belief. So, (c) would normally be read as implying that adding rosemary to the sauce will *not* make it taste better (or at the very least that it is not certain whether adding rosemary will make the sauce taste better or not).

(c) John is adding rosemary to the sauce because he believes it will make it taste better.

In cases where an agent acts on a false belief we can still give a rationalising explanation of their action, but it is not clear whether such explanations give the agent’s reason for acting. For example, (e) is a rationalising explanation, because it tells us why Columbus acted as he did by telling us why sailing west across the Atlantic seemed, to Columbus, like a rational thing for him to do.

(e) Columbus sailed west across the Atlantic because he thought that was the way to India.

However, does this rationalising explanation give Columbus's reason for acting? Alvarez (2010) thinks it does not. This is because Alvarez thinks that reasons are facts. That India can be directly reached by sailing west across the Atlantic is not a fact, hence it could not have been a reason for Columbus to act. Although Columbus doesn't know it, he doesn't really have a reason to sail west across the Atlantic. The false proposition that India can be reached by sailing west across the Atlantic was *mistakenly judged* to be a reason by Columbus. If this is right then (e) does not attribute to Columbus a reason-for-acting, instead it tells us what he *thought* his reason-for-acting was. Dancy (2000) in contrast thinks that a rationalising explanation like (e) does give Columbus's reason-for-acting. Dancy thinks that reasons do not have to be true. According to Dancy, Columbus's reason is the (false) content of his belief.

I do not know whether Alvarez or Dancy is right on this point (although, I lean towards Alvarez's position). However, either way, what does the explaining in (e) is, I contend, the fact that Columbus *believed* that India can be reached by sailing west across the Atlantic. In other words, the explanans in (e) is a fact about Columbus's mental state, a fact about what he believes. Dancy disagrees: he thinks that the explanans of (e) is the false content of Columbus's belief, which is what Dancy takes to be Columbus's reason-for-acting. This is because Dancy thinks that the explanans of a rationalising explanation is always the reason-for-acting provided by the rationalising explanation. However, this position commits Dancy to the view that explanans do not have to be facts, which goes against the plausible assumption that explanation is a relation between facts, and that only facts can explain.[55] For this reason, Dancy's claim that the explanans in (e) is the false content of Columbus's belief should be rejected. Regardless of whether we think Columbus really has a reason-for-

[55] Dancy goes on to use this as an argument against the idea that reason-giving action explanations are causal, as *causal* explanations, according to Dancy, must be factive.

acting or not, what explains why Columbus acted as he did is the fact that he had a certain (false) belief.

When an agent acts intentionally on a false belief, the fact that they believe some false proposition is the explanans for their action. When an agent acts intentionally on something they know, the rationalising explanation we give of their action will not typically explicitly mention any belief of the agent. We are far more likely to give a rationalising explanation like (d) which does not explicitly mention any facts about David's mental states.

(d) David took the A-road because the motorway was shut.

The explanans of (d) appears to be a fact about the world external to David, i.e. a fact about the motorway – which is David's reason for acting – and not a fact about David's mental states. What does this mean for Davidson's proposal, which I have endorsed, that the primary explanantia of rationalising explanations are facts about the agent's mental states, specifically facts about what the agent believes and desires?

The idea that the primary explanantia of rationalising explanations are facts about the agent's mental state is not automatically falsified by the observation that, when an agent is acting on what they know, the fact explicitly mentioned as explanans is often a fact about the world external to the agent. After all, Davidson concedes that rationalising explanations do not always explicitly mention the primary reason. In (d)'s case, had David not been *aware* that the motorway was shut, then even though the fact that the motorway was shut renders taking the A-road a sensible course of action, mentioning this fact would not explain why David took the A-road. This appears to suggest that offering a statement like (d) as an explanation of David's action is only *explanatory* when we take it as read that David *knew* the motorway was shut. So perhaps the fact that's really doing the explanatory work in (d) is the fact that David *knew* the motorway was shut.

However, it is unclear what sort of state a state of knowledge is. Perhaps knowing that P is just a special sort of belief-state. In this case, the Davidsonian idea that the primary explanantia of rationalising explanations are facts about

what the agent wants to do and what the agent believes about how to do it is correct as it is. However, perhaps knowing that P is a distinctive kind of mental state, possession of which is incompatible with believing that P. In this case, the Davidsonian idea needs to be amended. We would have to say that the primary explanantia of rationalising explanations are facts about what the agent wants to do and what the agent believes *or knows* about how to do it.

The significance of adding this disjunction is unclear. How important is the difference between rationalising explanations which have *beliefs* as part of their primary explanatia and rationalising explanations which have *knowledge* as part of their primary explanatia? Should we treat the two types of rationalising explanation differently? Furthermore, because I have assumed, along with many philosophers writing on intentional action, that rationalising explanations are a guide to the nature of intentional actions, if there are distinct kinds of rationalising explanation, then we have to ask if that means anything for our understanding of intentional action. Does it mean, for example, that intentional action is not a homogeneous phenomenon? Does it mean that there are different sorts of intentional action? These questions are beyond the scope of this dissertation. However, when I come to put forward a view of intentional action in chapter 7, I will try not to pre-empt answers to these important questions.

To summarise, I believe Davidson is broadly correct in thinking that for the explanans of a rationalising explanation to illuminate the explanandum, 'it is necessary and sufficient that we see, at least in essential outline, how to contrast a primary reason' (1963/2001, p.4). Davidson's idea needs to be modified slightly in light of two complications. Firstly, the primary reason-why should be distinguished from the agent's reason-for-acting. The former is a fact about the agent's mental states which must at least be implicitly gestured to by the rationalising explanation in order for that rationalising explanation to explain. The latter is the good the agent saw in the action she performed, it is a desirability characterisation of the action. Usually, reasons-for-acting are not facts about the agent's mental states but facts about the world external to the agent. Secondly, although Davidson thought that the primary reason-why an

agent acts was always a fact about what the agent wants to do and what the agent *believes* about how to do it, given that we only seem to offer rationalising explanations which mention beliefs when the agent lacks knowledge, it is possible that when an agent is acting on what she knows, the primary reason-why is a desire-knowledge pair and not a desire-belief pair. In other words, it is possible that there are two different kinds of rationalising explanation. It is unclear how significant this difference is.

There is much more to be said on the topic of rationalising explanations. However, I hope that I have said enough to show that the explanatory power of rationalising explanations seems to rest on their suggesting the existence of a connection between the agent's mental states and their action. So, when we explain why an agent acted as she did by saying why what the agent did seemed, to the agent, like a rational thing for the agent to do, the fact that the agent acted as she did is explained by facts about her mental states.

6.2 Are Rationalising Explanations Causal?

How do rationalising explanations explain? How is it that a statement telling us why an agent's action seemed to them to be rational is able to explain why the agent did as she did? One answer is that rationalising explanations explain by giving a *causal* account of the agent's action. That is, a statement telling us why what an agent did seemed to them to be a rational thing to do explains why the agent did as she did by giving us causal information. Davidson (1963) favoured this answer. Davidson's (1963) argument is best thought of as a challenge to anyone who thinks that rationalising explanations are not causal, as Davidson does not offer any *positive* reason to think that they are.

In brief, Davidsons's argument as follows: *Some* statements which tell us why what an agent did seemed to them to be rational do *not* explain why the agent did as she did. This phenomenon was demonstrated in the case of Diana described above. Diana believes that speaking at a conference will win her praise from her friends and will help her career. But she ends up speaking at a conference because of the praise she will receive from her friends rather than

because speaking at the conference will help her career. To say only that Diana spoke at the conference and believed that doing so would help her career is not to have explained why Diana spoke at the conference. This is because even though Diana's action is justified by the fact that speaking at the conference would help her career, this was not why Diana spoke at the conference. It is not true that Diana spoke at the conference *because* she thought it would help her career. On the other hand, it is true that Diana spoke at the conference because she would receive praise from her friends. That Diana would receive praise from her friends if she spoke at the conference *does* explain why Diana acted as she did. (Diana's case focuses on beliefs – but a similar case can be constructed using desires.) Because some statements which tell us why what an agent did seemed to them to be a rational do not explain why the agent did as she did, those statements which do both must achieve this by doing more than simply revealing why what an agent did seemed to them to be a rational thing to do. And if the extra thing rationalising explanations do is not revealing causal information – what is it? Davidson thinks there is no satisfactory answer to this question.

Dancy (2000) denies that successful rationalising explanations do more than reveal why what an agent did seemed to them to be a rational thing to do. The difference between statements that rationalise but do not explain and statements which rationalise *and* explain is simply that, in the former, the belief/desire mentioned is not the belief/desire the agent acted in the light of, and in the latter the belief/desire mentioned *is* the belief/desire the agent acted in the light of. Davidson insists that the explanatory connection between beliefs/desires an agent acts in light of and the agent's action cannot be brute – it has to hold in virtue of some other connection between the agent's beliefs/desires and their action. But, Dancy objects, Davidson provides no argument against the following view:

> [...] the difference between those reasons for which the agent did in fact act and those for which he might have acted but did not is not a difference in causal role at all. It is just the difference between the considerations in the light of which he acted and other considerations he took to favour acting as

> he did but which were not in fact ones in the light of which he decided to do what he did (2000, p163).

In other words, Dancy doesn't think that Davidson provides any argument against taking 'acting in the light of' as primitive.

On Dancy's view, 'acted in the light of' performs the function in the case of rationalising explanations that truth plays in the case of other sorts of explanation. Like truth, 'acted in the light of' is a status capable of belonging to statements given as explanans, which is a necessary condition for their explanatoriness. For example, compare 'George is the first born of William and Kate' with 'George is the first born of Elizabeth and Philip' as putative explanans of the following explanandum: why is George heir to the throne? Both statements posit the kind of relationship which would guarantee George's being the heir to the throne, but only the first statement can genuinely explain why George is heir to the throne because only the first statement is true. There is nothing perplexing about the fact that truth can make the difference between two statements which both posit something that would make sense of the explanans. That only true statements can explain is plausibly a brute fact.

However, there is something perplexing about the fact that 'acted in the light of' also seems to be able to perform this function. That 'acted in the light of' can perform this function seems like something that needs accounting for – it does not seem like a brute fact. There must be something about statements which tell us the reason the agent acted in the light of which grounds their explanatoriness. The question Davidson's challenge raises is: *why* does learning that Diana's reason for acting was that she would receive praise *explain* why Diana spoke at the conference? Why does 'acted in the light of' bestow explanatory power? Tanney (2009) expresses the puzzle well:

> Davidson claims that it would be a mistake to conclude from the fact that placing the action in a larger pattern explains it, we now understand the sort of explanation involved, and that 'cause and effect form the sort of pattern that explain the effect in the sense of "explain" that we understand as well as any' [(1963/2001, p.10)]. Davidson challenges the opponents of the

> causal view to identify what other pattern of explanation illustrates the relation between reason and action if they wish to sustain the claim that the pattern is not one of cause and effect. (2009, p.96)

The task is to spell out what 'pattern of explanation' is demonstrated by rationalising explanations.

I have said that Davidson thought that the pattern of explanation demonstrated by rationalising explanations is a causal one. That is, that rationalising explanations explain by giving a *causal* account of the agent's action. But what is the nature of the causal information rationalising explanations are supposed to provide? Davidson's answer is that 'the primary reason for an action is its cause' (1963/2001, p.4); Davidson's view is that rationalising explanations are true if and only if the belief or desire which explains the action (or some mental event suitably related to the belief or desire) stands in a causal relation to the action explained. So, for example, the truth of (b) depends on there being a causal relation between Beth's desire to make bread (or, at least, the onset of her desire to make bread) and her buying flour. If (b) is true, then it necessarily implies the existence of a causal relation between Beth's desire to make bread and her action, or so the thought goes.

Opponents to Davidson, whom I will call 'non-causalists', deny that rationalising explanations are causal. Non-causalists deny that true rationalising explanations necessarily imply the existence of causal relations between mental states/events and actions. Non-causalists object to the idea that the mental concepts employed in rationalising explanations point to mental items which stand in causal relations to the actions explained. They also question the soundness of Davidson's argument that the pattern of explanation demonstrated by rationalising explanations must be causal. According to non-causalists, what grounds the explanatory power of statements which tell us the reason the agent acted in the light of is not that such statements pick out the cause of the action explained, but something else.

6.3 Objections to Davidson

6.3.1 Mental Concepts

Opponents of Davidson argue that rationalising explanations display features which set them apart from typical causal explanations, like 'the patient developed cancer because he was exposed to radiation', and which challenge the plausibility of the idea that mental concepts employed in rationalising explanations designate the causes of the actions they explain.

Firstly, believing (or knowing) that something is the case and desiring to do something are not events but states. For example, the explanans given in (a), i.e. 'she wants to make bread', would be classified by Mourelatos as a state predication, not an event predication. On the assumption that causal explanations are typically explanations which tell us what event stands in a causal relation to the event whose occurrence we want to explain, the fact that the explanans of most rationalising explanations is a state predication seems to speak against classifying these explanations as causal.

This first difference does not carry much force. Although it is true that believing (or knowing) that something is the case and desiring to do something are not events but states, the onset of belief and the onset of desire are events and the Davidsonian could argue that it is causal relations between *these* mental events and the action explained which the truth of rationalising explanations depends on. As Davidson puts it, 'In many cases it is not difficult at all to find events very closely associated with the primary reason. States and dispositions are not events, but the onslaught of a state or disposition is' (1963/2001, p.12).

A second difference between rationalising explanations and causal explanations is that when we causally attribute one event to another, this is usually taken to imply the existence of a law which states that there is an event-kind F, of which the cause event is a token, and an event-kind G, of which the effect event is a token, such that F events always cause G events. However, when we say that an agent acted as she did because of the beliefs and desires she had, there is no implication that other agents with the same beliefs and desires will

(or are likely to) do the same thing, or that the same agent will act in the same way when she has the same beliefs and desires on another occasion (Hart and Honoré 1985, p.55).

To deal with this objection Davidson proposes that the laws which cover the causal relation rationalising explanations necessarily imply cannot be stated in the language of beliefs and desires. This is Davidson's Anomalous Monism. Davidson proposes that when a mental event and an action are causally related, these two events fall under event-kinds which feature in a causal law. This follows from Davidson's Principle of the Nomological Character of Causality: all causal relations are covered by strict deterministic laws. However, the event-kinds which feature in the causal law, which the mental event and action fall under, are *physical* kinds, not mental kinds. Furthermore, the law that covers the causal relation can *only* be stated in a language of physical kinds. So, it is true that when we say that an agent acted as she did because of the beliefs and desires she had, there is no implication that other agents with the same beliefs and desires will (or are likely to) do the same thing. But this is because in giving a rationalising explanation we are picking out the cause of an action using *mental* kinds, and these mental kinds do not feature in any universal regularity, not even the universal regularity which covers the causal relation which the rationalising explanation owes its success to. As Davidson puts it:

> The laws whose existence is required if reasons are causes of actions do not, we may be sure, deal in the concepts in which rationalisations must deal. If the causes of a class of events (actions) fall in a certain class (reasons) and there is a law to back each singular causal statement, it does not follow that there is any law connecting events classified as reasons with events classified as actions – the classifications may even be neurological, chemical, or physical. (1963/2001, p.17)

Thirdly, Anscombe pointed out that when 'one says what desire an act was meant to satisfy, one does not identify a feeling, image or idea that precedes the act the desire explains: one does not answer the question "what did you see or hear or feel, or what ideas or images cropped up in your mind and led up to it?"' (2000 p.17). The desire which an act satisfies is not the 'mental cause' of

the act in the same way that, to use Anscombe's example, noticing a face appearing at the window might be the mental cause of one's jumping. Anscombe defines a 'mental cause' as 'what someone would describe if he were asked the specific question: what produced this action or thought or feeling on your part: what did you see or hear or feel, or what ideas or images cropped up in your mind, and led up to it?' (2000 pp.17-18). Giving a 'mental cause' of something, in the special sense of 'mental cause' that Anscombe has isolated, is thus to say what prior mental event *triggered* the effect. We sometimes do explain actions by giving their 'mental causes'. Anscombe provides one example. Other examples include explaining why a football fan cheered by saying it is because he saw his team score a goal, or explaining why someone is going outside by saying it is because she has remembered she needs to put the bins out. But, Anscombe's point is that rationalising explanations are not usually like this – or they do not seem to be. Explaining why a football fan cheered by saying it is because he saw his team score a goal could be seen as a rationalising explanation because as well as being the 'mental cause' of his action, *that his team scored a goal* can also be seen as the football fan's reason for acting. In this case, the distinction between reason and 'mental cause' is hard to discern. But, Anscombe's point is that many, if not most, rationalising explanations do not seem to be explanations which provide a 'mental cause'. When we explain our actions in terms of our beliefs and desires, we are usually not identifying something that occurred at a particular time which triggered our action, or which moved us from a state of inaction to a state of action. Furthermore, for a rationalising explanation to be explanatory it is not necessary that it suggest, or imply, an Anscombian 'mental cause' attribution. This is not to say that considering what 'mental causes' there might be for a particular action is completely *irrelevant* for judging the truth or explanatoriness of a rationalising explanation. It is rather to say that considering what 'mental causes' there might be for a particular action is no *more relevant* than a host of other factors, and the concepts employed in a rationalising explanation can be employed correctly even when there is no 'mental cause'.

Tanney makes a similar point. Tanney argues that the concepts that are at work in rationalising explanations perform their explanatory role 'without designating anything; let alone causally efficacious states or events; let alone causally efficacious states or events whose nature awaits discovery' (2009, p.100). Tanney supports this view by arguing that mental concepts cannot be treated as theoretical terms, that is, to terms which purport to refer to or designate an event, property, state, fact or condition, whose intrinsic nature is up for discovery, which causes the phenomena to be explained. An example of such a theoretical term would be 'gene': genes are entities we posit on the grounds that their existence would explain some observable phenomena; 'gene' is a term that purports to refer to a hidden, but causally efficacious entity. Tanney argues:

> The problem in assuming that the motive, intention or reason is (in principle describable as) a logically independent, temporally antecedent, causally efficacious event (perhaps identified with its alleged 'onset') is that it mis-assigns the explanatory function of these concepts. The position commits us to postulating an event, unobservable to others and possibly even to the agent herself that would, if known, provide the sought-after reason-explanation for the agent's action. In such cases, as Ryle insisted, an epistemological puzzle arises as to how anyone could ever know whether a person acts for reasons or what, if she does, her reasons are, since the hypothesis is not even in principle testable. Not only do we not, in every-day situations have access to these hidden events, but even if we were, say, to monitor the neural activity of someone's brain or access their stream of consciousness, we would never be able to set up the kinds of correlations that would establish a particular occurrence as an instance of a particular reason without already having a way of deciding whether someone acted for a particular reason in order to make the correlation. (2009, p.100)

Tanney has, I think, hit upon an important truth here, which is that construing rationalising explanations as explanations which posit an entity which is logically independent from, temporally antecedent to and causally related to the action which is explained encourages us to think that concepts like *reason*, *aim*, *belief* and *desire* refer to mental *items* (perhaps mental events,

perhaps mental states). This view, I believe, legitimises a metaphysics of mind wherein our status as minded creatures depends on the existence of mental events and states whose nature we have yet to discover and whose existence must, one way or another, be reconciled with the fact that the world is physical in all its fundamental aspects. The view also seems to mischaracterise rationalising explanations: when we say that an agent acted as she did because of the reasons she had, we do not take ourselves to be making existential claims about what obtained or occurred (or is obtaining or occurring) which may turn out to be false, if there is no way to reconcile the truth of these existential claims with the fact that the world is fundamentally physical.

According to Tanney, we should resist characterising rationalising explanations as explanations which identify 'mental causes', because in doing so we *distort* mental concepts like *reason*, *aim*, *belief* and *desire*; we end up taking these concepts as designating *items* whose nature is up for discovery. However, there is an issue regarding how we can be sure that taking mental concepts as concepts which designate items whose nature is up for discovery represents a *distortion* of those concepts rather than a *correction*. Anscombe and Tanney may be right that the concepts employed in rationalising explanations do not seem to function by postulating an event which stands in a causal relation to the explanandum. But perhaps the Davidsonian could argue that there is good reason to believe that things are not as they appear. Davidson's Anomalous Monism allows one to maintain that the explanatory power of rationalising explanations depends on their implying the existence of causal relations between mental items and actions, despite the fact that rationalising explanations do not *appear* to have this implication. The anomalousness of mental concepts means that the causal nature of mental states and events is not revealed when these entities are picked out by mental concepts. So, one can grant that mental concepts do not *seem* to perform their explanatory function by designating causes – as *mental* concepts one would not expect them to – but at the same time insist that rationalising explanations would not be true if mental concepts did not somehow pick out events which stand in causal relations to actions. After

all, what else could the explanatory pattern displayed by rationalising explanations be?

6.3.2 Context-placing Explanations

Non-causalists have argued that there is no reason to think that the pattern of explanation demonstrated by rationalising explanations must be causal. For example, Tanney argues that, despite the challenge proposed in Davidson's *Actions, Reasons and Causes*, 'there is no obligation to construe the deployment of [concepts like reason, aim, belief and desire] as the identification of events or standing states' (2009, p.100). Tanney argues that Davidson's argument does not force us think of rationalising explanations as implying the existence of causal relations between mental items and actions because there is another, equally acceptable way of understanding how rationalising explanations explain. Tanney suggests rationalising explanations ought to be understood as 'context-placing' explanations. The explanans of a rationalising explanation explains the action by placing it in a context that makes it intelligible. According to Tanney, rationalising explanations are explanations which work by giving us more information about what is going on. Tanney provides an example of a 'context-placing explanation' which helps us see why this sort of explanation is not the sort of explanation in which one event follows another:

> (f) The teacher has written 'CAT' on the board because she is writing 'CATALYST' on the board.

Tanney says that the explanans in (f) 'serves to re-characterise what happened so that it – as newly described – is no longer puzzling' (2009, p.98). The explanans does not illuminate 'any mysterious connection between the occurrences of two contingently related events – the writing of 'c', 'a', and 't', on

the one hand and the writing of 'catalyst', on the other' (2009, p.98). Tanney argues that rationalising explanations are all, essentially, of this kind.[56]

Anscombe makes a similar suggestion. Anscombe suggests that rationalising explanations 'interpret' the action explained:

> To give a motive […] is to say something like "See the action in this light". To explain one's own actions by an account indicating a motive is to put them in a certain light. (2000, p.21)

George Schueler (2009) also offers an account of rationalising explanations which takes them to be 'interpretive explanations'; they are explanations which tell us how to interpret the physical, causally related, events we are able to perceive. Mayr (2011, p.269) also endorses the idea that rationalising explanations 'explain actions by making them intelligible' and not by positing an event-causal link between the agent's action and an appropriate mental event. For Mayr, rationalising explanations explain by providing us with a way of framing the agent's actions – a way of seeing the agent's actions as manifesting a certain pattern.

Tanney argues that Davidson's point about how some rationalisations explain why the agent did as she did and some do not does not provide sufficient reason for thinking that rationalisations that do explain the agent's action *must* function by designating events causally related to the action. Even if successful rationalising explanations must 'do more' than mere rationalisations, the 'more' they do need not be designating a cause of the action. When re-characterising an agent's action as an action rationalised by certain reasons fails to explain the agent's action, we may simply need to 'probe further for a different or more far-

[56] Thompson (2008) suggests that 'the explanation of action as it appears most frequently in human thought and speech is *the explanation of one action in terms of another*' (p.85, emphasis in original). Thompson calls this form of action explanation 'naïve action explanation', with action explanations that reference reasons, motives, beliefs and desires qualifying as 'sophisticated' (pp.85-89). Thompson argues that a better understanding of intentional action is achieved if we take naïve action explanation as the most basic form of action explanation.

reaching context-placing explanation that will succeed or give up the initial expectation that the action can be explained by reasons' (2009, p.100). Other writers have also suggested that there is no obligation to think that the difference between a rationalisation that succeeds in explaining an agent's action and one that does not is that the former but not the latter designates the cause of the action. It may be that the successful rationalising explanation is the one that fits better with the agent's general character, or moral code; or perhaps the successful rationalising explanation is one which accords better with how the agent has herself weighted the various considerations she takes to favour her action (Owens, 1992, pp.164-165; Tanney, 1995, p.110). The point here is that, if we assume that rationalising explanations are context-placing or interpretative then we have at our disposal several ways to say why some statements which tell us why the agent's action seemed rational to them fail to explain the agent's action without supposing that successful rationalising explanations designate events causally related to the action.

Another observation that speaks in favour of taking rationalising explanations to be context-placing or interpretative is that sometimes when an agent has more than one reason for performing some action it is genuinely indeterminate which of the reasons was *the* reason she acted for. As Mayr puts it, there is not always a fact of the matter about which reason an agent acted for:

> Consider cases where the agent has a bundle of strong motives to do X, but it is not clear – even after thorough examination of his action, its circumstances, and his general character – on which of these motives he has acted. We do not have then to assume that our inability to decide this question rests on merely practical grounds – that is, that there is a fact of the matter which we are unable to establish only because we lack further evidence; for it may well be that we would not even know what kind of further evidence would decide the question. Instead, we should accept that in such cases our inability may stem from the fact that these cases are truly indeterminate, because the criteria for judging whether the agent acted on a particular reason have 'run out', without unequivocally determining an answer. (2011, p.261)

The idea that rationalising explanations explain by identifying the cause of the action is inconsistent with allowing for this kind of indeterminacy. On the Davidsonian view, any indeterminacy regarding what belief or desire the acted in the light of is epistemic – this is because an agent *acts in the light of* a belief or desire if and only if the onset of that belief or desire is the cause of the action, and the latter relation cannot be indeterminate. Of course, it could be that when an agent has many reasons favouring a course of action, their action is causally overdetermined by these many reasons. However, it seems possible that an agent could have many reasons favouring a course of action, where none of these reasons is *the* reason the agent acted, and where the agent would not have acted if the case for acting was not overwhelming. For example, imagine Diana is again deciding whether or not to speak at a conference, and because the conference is quite far away, Diana vows only to speak at the conference if the case for doing so seems overwhelming, where overwhelming for her means that there are at least *n* strong reasons favouring the action (where *n* is more than one). Then suppose Diana discovers *n* reasons for speaking at the conference, and so goes on to speak at the conference, but none of Diana's reasons stands out as *the* reason for which Diana spoke at the conference. In this case, it does not seem like Diana acts in the light of just one of the many reasons favouring speaking at the conference, but it is also not plausible to describe this as a case of overdetermination by her *n* reasons, because it is not the case that Diana would have acted in the same way had any one of her *n* reasons been missing.

The idea that rationalising explanations are context-placing explanations is plausible. Furthermore, that there is a difference between mere rationalisations and rationalising explanations does not seem to *force* us to accept that rationalising explanations necessarily imply the existence of a causal relation between a mental item and the action explained. There seem to be other ways to account for the fact that some rationalisations explain an agents action whereas others don't. Davidson's argument is thus not as strong as it first appears.

However, the plausibility of the context-placing account of rationalising explanations may not be enough to offset the intuition, brought out by Davidson's argument in *Actions, Reasons and Causes,* that rationalising explanation are causal. I think the intuition that rationalising explanations are causal explanations is hard to resist. This means that, there is a strong motivation to accommodate valid points made by the non-causalists without giving up the idea that rationalising explanations are causal. Davidson's Anomalous Monism lets one do this. Anomalous Monism allows us to acknowledge the validity of the points made by non-causalists – including the observation that we distinguish rationalising explanations from mere rationalisations by considering facts like how far-reaching the rationalisation is, or how it fits with the agent's general character – without giving up the intuition that rationalising explanations are causal. On Anomalous Monism, the facts that we consider when we work out why an agent did as she did are not constitutive of what it is to act for a reason. As Mayr puts it:

> For Davidson, the epistemological criteria that we use for determining for which reason an agent has acted are the considerations of rationality and overall coherence among his mental states that are generally relevant for the interpretative enterprise of 'making sense of the agent'. What makes the reasons-explanation true, however, is something completely different: the obtaining of an event-causal link between reason and action, which for Davidson must be based on a strict causal law. (2011, pp.269-270)

So, if Anomalous Monism is true, we might expect there to be a gap between the facts we consider when judging the veracity of a rationalising explanation and the facts that make that rationalising explanation genuinely explanatory. Of course, Anomalous Monism might not be correct, but I think that the opposition between Davidson and non-causalists on the matter of rationalising explanations is at something of an impasse because Anomalous Monism is an available position.

Perhaps there is another way to accept the non-causalist's view that rationalising explanations are context-placing, and that the mental concepts appealed to in rationalising explanations do not discharge their explanatory role

by designating the causes of the action they explain, without giving up the intuition that rationalising explanations are causal. Non-causalists assume that if rationalising explanations reveal causal information, the causal information they reveal is that there are mental items, which the mental concepts employed in rationalising explanations pick out, which stand in causal relations to actions. Tanney is explicit about this assumption:

> […] the position I wish to bring back into focus says that what it is for an action to be in execution of an intention or for it to be explicable by reasons is not a matter of there being a causal relation [understood as 'a relation between two logically and temporally distinguishable events'] between intention or reasons and action. If causation is to be thus understood the pattern in virtue of which a person's intentions, motives or reasons explain her action is not *eo ipso* causal. (Tanney, 2009, p.95)

However, is it right to assume that a rationalising explanation is causal only if it posits a causal relation between an item somehow picked out by the mental concept employed in the explanation and the action explained?

In the next chapter, I will argue that explanations can be causal even when they do not necessarily imply the existence of causal relations between certain particulars. If this argument succeeds, then it is possible that rationalising explanations could be causal even though the mental concepts cited in the rationalising explanation do not pick out items which stand in causal relations to the action explained. If so, then it may be possible to grant that rationalising explanations are context-placing, and at the same time accept that they are causal, without turning to Anomalous Monism, and thus without having to accept the physicalist metaphysics Anomalous Monism entails.

7

Causal Explanations and Intentional Action

In the previous chapter, I examined the debate concerning whether explanations of intentional actions which cite the agent's reasons or aims are causal explanations. Davidson (1963) argues that they are. On Davidson's view, rationalising explanations depend for their truth on causal relations between mental items, which are picked out by the mental concepts employed in rationalising explanations, and the actions which are explained. Assuming an action is intentional if and only if what it takes for a rationalising explanation of it to be true obtains, if Davidson is right, then the existence of an intentional action conceptually entails that there is mental causation *and* that this mental causation should be conceived of in relational terms. However, in section 6.3 of the previous chapter, I described objections and arguments which aim to show that, as Tanney puts it, the mental concepts in rationalising explanations do not 'discharge their explanatory role' by designating causes of the actions they explain (2009, p.100). Instead, rationalising explanations are 'context-placing' or 'interpretative': they place the action in a context that makes it intelligible. However, I suggested that this position could be resisted by a Davidsonian. This is because the Davidsonian can claim that because of the anomalousness of the mental, we can judge the appropriateness or correctness of rationalising explanations according to criteria like 'what makes best sense of the agent', even while these rationalising explanations are made true by causal relations between events.

I suggested that the debate between Davidsonians and non-causalists is at something of an impasse because both positions have intuitive appeal. On the one hand, the idea that when we explain someone's actions in terms of their

beliefs and desires we are giving causal information is very appealing. On the other hand, to say that when we attribute beliefs and desires to an agent we are positing internal causes of their observable behaviour does seem to misconstrue the explanatory function of these mental concepts. In this chapter, I will present a view which tries to save both intuitions.

Non-causalists assume that to show that the mental concepts employed in rationalising explanations do not discharge their explanatory role by signifying items which are the causes of the actions explained *just is* to show that rationalising explanations are not causal. In other words, non-causalists assume that causal explanations are precisely those explanations whose explanandum designates an effect and whose explanans designates an item which is the cause of that effect. On this view of what makes an explanation causal, the 'because' of a causal explanation signifies the obtaining of a causal relation. I will call this view the Davidsonian view, because it is likely that Davidson held it.[57]

In this chapter, I will question the truth of the Davidsonian view. In section 7.1, I will discuss counterexamples to the Davidsonian view. I will suggest that some causal explanations may depend for their truth not on causal relations between events but on facts about dynamic states of affairs. In section 7.2, I will examine an alternative view of what makes an explanation causal, which allows some explanations to count as causal even though they are not made true by the obtaining of causal relations between events. I will then reconsider the question of whether rationalising explanations are causal in light of these more relaxed criteria for an explanation to be causal. In section 7.3, I will present a view on intentional action which aims to resolve the impasse between Davidsonians and non-causalists by granting that rationalising explanations are a unique kind of disposition-citing explanation.

[57] See Evnine (1991 pp. 49-52).

7.1 Causal Explanations

As mentioned, Davidsonians and non-causalists alike assume that causal explanations are precisely those explanations whose explanandum designates an effect and whose explanans designates an item which is the cause of that effect. As William Child describes the Davidsonian view:

> The general idea, then, is that the truth (or acceptability) of a causal explanation rests on the presence of appropriate relations of causation. And a natural thought would be to put the point in the following way: a causal explanation is one whose explanatory power depends on the assumption that there are events mentioned, or pointed to, in the explanans and explanandum sentences, between which the natural relation of causation obtains; and whose truth (or acceptability) requires that the relation does indeed obtain. (Child, 1994, p.102)

This view assumes that a causal *explanation* is the statement of a non-natural, intensional relationship which holds between true propositions. The causal *relation* in contrast is a natural, extensional relation which 'holds in the natural world between *particular events or circumstances*, just as the relation of temporal succession does or that of spatial proximity' (Strawson,1985, p.115). This theory does not demand that the events, whose causal connectedness grounds the truth of a causal explanation, should be *explicitly* referred to or mentioned by the sentences which form the explanandum and explanans of the causal explanation, or that the explanandum and explanans sentences can be transformed into sentences which involve explicit quantification over events.[58] As Child notes, 'the fact that, in some (or even most) cases, reference to causally related events is concealed is compatible with the idea that the truth of an explanation depends on the presence of appropriate relations to causality

[58] The process of transforming a sentence like 'Roger ran a mile' into a sentence that explicitly quantifies over an event ('Roger's running of a mile'), is a process Mourelatos calls 'nominalisation transcription'. Nominalisation transcription is discussed in chapter 4.

between particular events' (1994, p.102). In this section, I will outline four kinds of counterexample to the Davidsonian view of what makes explanations causal.

7.1.1 Negative Causal Explanations

The first sort of counterexample to the Davidsonian view is negative causal explanations, i.e. causal explanations where either the explanans, or the explanandum, or both, is a fact about an event failing to occur.

(i) Don did not die because his rope did not break. (Child, p.106)

(ii) The water swept away the fish because the sluice gate did not shut.

(iii) The policeman wasn't hurt because the bullet got stuck in his Kevlar vest.

On the Davidsonian view, these explanations are causal explanations if and only if they are made true by a causally related pair of events. But in (i), it seems like no events are mentioned or pointed to by the explanation. In (ii), the explanans clause does not seem to mention an event. And in (iii) the explanandum clause does not seem to mention an event. One might posit so-called 'negative events' – so that something's *not* happening is an event. This allows one to argue that in fact the explanans clauses and the explanandum clauses of (i)-(iii) do all explicitly mention events whose causal connections serve as truth-makers for the explanations. On some theories of events, the existence of negative events might be plausible.[59] But on any theory which takes seriously the idea that events are happenings, this proposal is implausible: something's not happening is not a thing that happens.[60]

A more plausible response to negative causal explanations is suggested by Child. Child suggests that the Davidsonian could potentially accommodate negative causal explanations within his account of causal explanations by

[59] For example, on certain theories of events as property exemplifications it might be possible for there to be negative events. Philosophers who have argued for the reality of negative events include: Vermazen 1985; De Swart 1996; Higginbotham 2000.
[60] See Mele (2005) for further reasons to reject negative events.

allowing the relation between a causal explanation and the causally related events which make the explanation true to be opaque (1994, p.106). The Davidsonian position is safe if the truth of negative causal explanations depends on there being causal relations between events; it is not necessary that the negative causal explanation itself mention the causally related pairs of events which make it true. The idea would be that 'Don did not die because his rope did not break' succeeds as an explanation only because rope-breakings are causally related to deaths when they occur in circumstances similar to Don's – the explanation depends for its truth on causal relations between rope-breakings and deaths. Another way of putting this point is to say that negative causal explanations are true when they are backed by a causal law – i.e. a generalisation which says that events of one type always (or usually) cause events of another type to occur.[61]

There is nothing wrong with the idea that the relation between an explanation and what makes the explanation true can be opaque. As Kevin Mulligan, Peter Simons and Barry Smith put it, it is 'perfectly normal for us to know *that* a sentence is true, and yet not know completely *what* makes it true' (1984, p.299). However, it seems odd to me to suggest that the truth of a negative causal explanation should depend on causal relations between events that take place somewhere else (perhaps even on causal relations between events that take place in non-actual possible worlds, because even if no rope-breakings had ever occurred, and so no-one had ever died as a result of one, 'Don did not die because his rope did not break' could still be true, and a Davidsonian might say this is because if some rope-breakings had occurred, these events *would* have caused deaths). It seems to me that the truth of negative causal explanations should depend on something within the causal system the causal explanation concerns. So, for example, 'Don did not die

[61] C.f. Beebee (2005).

because his rope did not break' should depend, for its truth, on Don, or something about Don – or the rope, or something about the rope. This not a decisive objection against the response Child gives on behalf of the Davidsonian, but it does highlight a cost of the Davidsonian view: on the Davidsonian view some causal explanations are made true by causally related events which occur outside the circumstances the causal explanation specifically concerns.

7.1.2 Process-citing Explanations

A second kind of counterexample to the Davidsonian view is causal explanations which cite the continuous operation of causal processes, such as:

(iv) The snow is melting because the sun is shining.

Are causal explanations like (iv) made true by causally related pairs of events? As Mourelatos (1978) argued, and as I outlined in chapter 4, process predications, of which 'the snow is melting' and 'the sun is shining' are examples, do not implicitly quantify over events. So, (iv) does not say that some melting event was caused by some shining event. The tense of (iv) indicates that melting and shining are still going on, so it is not completed events, but ongoing processes which the explanation references. Nevertheless, it may well be true that whenever the sun melts some snow by shining on it causal relations between events always obtain. For example, it might be that whenever the sun melts some snow by shining on it a series of causally related chemical events involving light particles and ice molecules occur. Perhaps it is *these* causally connected events on which the truth of (iv) depends.

In most cases, when we say some causal process is in operation, we can find pairs of causally related events occurring at a finer temporal resolution. However, the vocabulary which we use to express the original causal explanation does not indicate what pairs of causally related events we should expect to find. For example, it is not part of the *meaning* of 'shining' or 'melting' that instances of shining or melting involve causally related pairs of events of

certain types.[62] It might be necessary that whenever the sun melts some snow by shining on it a series of causally related chemical events involving light particles and ice molecules occur, but this is an *a posteriori* necessity. The idea that an explanation must be made true by causally related events falling under types which have no connection to the *meaning* of the predications featuring in the explanation seems contrary to the reasonable principle that whatever makes some sentence true should be what the sentence is about. The notion of what a sentence is about is imprecise. Possibly, a Davidsonian could argue that on a loose enough definition of 'aboutness' (iv) is about events involving light particles and ice molecules. However, for this response to work the Davidsonian would have to convince us to adopt his loose definition of 'aboutness'.

If one thought, as seems reasonable, that explanations are causal if and only if they answer to causal reality, and that all there is to causal reality is events standing in causal relations to other events, then it would be natural to suppose that (iv) *must* depend for its truth on causally related pairs of events, if it is a causal explanation at all. However, as I argued in chapters 3 and 4, one need not think of causation as always, everywhere a relation between events. Causation can be a determinable process engaged in by substances. If this view of causation is plausible, then facts about what events are causally related to what others are not the only causal facts which causal explanations could answer to. Some causal explanations may answer to facts about dynamic states of affairs. Furthermore, the idea that (iv) is made true by facts about a dynamic state of affairs has intuitive appeal. What seems to matter for the truth of (iv) is that it is the sun which is causing what the snow is suffering.

[62] C.f. Child (1994, p.108).

7.1.3 Stative Causal Explanations

A third sort of counterexample to the Davidsonian view is stative causal explanations. Here are three examples:

(v) The bridge collapsed because the bolt was weak. (Child, 1994, p.106)

(vi) The floor is dirty because Mary's dog was here.

(vii) My leg is broken because I fell off my bike. (Child, 1994, p.105)

These examples are problematic for the Davidsonian view because in each of them either the explanans clause, or the explanandum clause, or both, seems to reference a state, not an event. In (v) that an event occurred is explained by the fact that a state obtains, in (vi) that one state obtains is explained by the fact that another state obtained, and in (vii) that a state obtains is explained by the fact that an event occurred.

Once again, the Davidsonian can respond by stressing that reference to the events, whose causal connectedness grounds the truth of the causal explanation, can be concealed. The reply would go like this: when we talk of a state as the cause of some event, 'there is a causal relation between events; the state [is] part of the circumstances in which the cause occurred; and mentioning that state can help to explain why the cause had the effect it did' (Child, 1994, p.106). So, in the case of (v), something happened to cause the collapse of the bridge (e.g. a train went over the bridge), the bolt's being weak was part of the circumstances in which this event occurred and helps explain why the event caused the collapse of the bridge. Similarly, when someone offers 'the floor is dirty because Mary's dog was here' as a causal explanation, we can suppose events occurred which stand in causal relations to each other (e.g. Mary's dog arrived, then ran around the room with muddy feet, and this latter event caused the floor to become dirty) and these causally related events are what makes the stative causal explanation true. And in (vii), the causal explanation is made true by the causal relation obtaining between my falling off my bike, and my leg breaking.

However, to suppose that whenever we offer a stative causal explanation, there *must* be appropriate pairs of causally related events to serve as the grounds

for the stative causal explanation seems to me to be metaphysically suspect. Events are not included in our ontology for the sole reason that they serve as truth-makers for causal explanations. Whether or not certain events exist and stand in causal relations, and whether or not a certain stative causal explanation is true, can therefore be determined independently. 'Was there an event which triggered the collapse of the bridge?' and 'did the bridge collapse because it the bolt was weak?' seem like independent questions, in the sense that an answer to the first need not impact an answer to the second and vice versa. Confidence in the truth of the stative causal explanation, should not, therefore, govern the truth of a claim about what events exist. Steward (1997, pp.173-174) also questions the assumption that appropriate pairs of causally related events can always be found to serve as the grounds for a stative causal explanation. In the bridge case, for example, what if the bridge just collapsed, apparently spontaneously? Are we always entitled to assume that there *must* have been a triggering event which stands to the event explained as cause to effect?

7.1.4 Disposition-citing Explanations

Stative causal explanations for which Steward's point seems particularly pertinent are stative explanations which seem to cite powers or dispositions. Indeed, (v) probably counts as a disposition-citing explanation. Other examples of disposition-citing explanations include:

(viii) Peter sneezed because he is allergic to flowers.

(ix) The cat died after eating the lilies because they are poisonous to cats.

(x) The aspirin relieved Joe's pain because it is a cyclo-oxygenese inhibitor.

It is possible that all stative causal explanations are disposition-citing explanations. For example, if it could be argued that (1) all stative predications attribute properties, and (2) all properties are really powers or dispositions, then it would follow that all stative causal explanations are really disposition-citing

causal explanations. However, both of these premises are controversial.[63] I will not attempt to establish that all stative causal explanations are really disposition-citing explanations, but I will assume that *some* stative causal explanations are disposition-citing explanations. I will also assume that disposition-citing explanations are causal explanations. As Hyman puts it:

> […] explanations that refer to disposition are *echt* causal explanations, whatever kind of disposition they refer to. *How* they explain, exactly what part of a causal story they tell, and whether a disposition is the cause, or part of the cause, of its manifestation – these are contentious questions. But *that* explanations that refer to dispositions are causal explanations should be beyond doubt. (2015, p.121)

Do disposition-citing explanations depend for their truth on the obtaining of causal relations between events? One might think that disposition-citing explanations are causal because they report causal relations between the triggering or stimulus event of the manifestation and the manifestation event. So, for example, perhaps (vii) 'Peter sneezed because he is allergic to flowers' reports a causal relation between Peter moving near to a flower (the trigger event) and Peter's sneeze (the manifestation event). For many dispositions, when they are manifested, causal relations between trigger and manifestation exist. Indeed, if they did not we might wonder whether the disposition has really been manifested at all. If there was no causal relation between Peter's moving near a flower and his sneeze, we might doubt that his sneezing was really a manifestation of his allergy. This is because to have an allergy is to be liable to exhibiting an immune reaction in the presence of an allergen – it is part of the meaning of 'allergy' that allergic reactions have specific triggers.

[63] Shoemaker (1980), Mumford (2004) and Whittle (2008) are three philosophers who have defended (2); Armstrong (1997, pp.69-84) has argued against it.

However, there are two problems with this suggestion. First, some dispositions do not seem to have triggers at all, either because they are always manifested (e.g. the disposition of a massive body to deform space-time) or because their manifestation is spontaneous (e.g. radioactive decay). Explanations which make reference to these sorts of dispositions therefore will not be made true by causal relations between triggers and manifestations, and on the assumption that all disposition-citing explanations to have the same sort of truth-maker, this makes the idea that disposition-citing explanations are made true by trigger-manifestation causal relations doubtful. Second, it is possible for there to be a causal relation between two events, the first of which is of the same type as the trigger of a disposition's manifestation, and the second of which is of the same type as a disposition's manifestation, without the disposition being manifested at all. For example, suppose the flower Peter moves near is bright white in colour, and the bright light reflected off the flower induces a photic sneeze reflex in Peter and he sneezes. In this example, moving near the flower caused Peter to sneeze, but his disposition to exhibit an immune response to flowers wasn't manifested. For all dispositions where the manifestation of a disposition involves a series of causally related events starting with a triggering event and ending with a manifestation event, it is possible for this type of causal chain to obtain without the disposition being manifested because the causal chain is 'deviant' in some way.[64] This throws into doubt the idea that causal relations between trigger-events and manifestation-events are what disposition-citing explanations report.

One might think that disposition-citing explanations are made true by causal relations holding between the dispositions themselves and the events explained. However, I reject this suggestion because I do not think that dispositions or powers can be causal relata.

[64] C.f. Hyman (2015, pp.121-127).

A number of philosophers have doubted that dispositions or powers themselves can be causally efficacious. Debate about the causal efficacy or causal relevance of dispositions mirrors the debate about the causal efficacy or causal relevance of mental states. Frank Jackson (1995, p.257) argued that because part of what it is for a substance to possess a disposition, like 'fragility', is for that substance to be prone to exhibit the manifestation behaviour, this entails that the disposition is non-contingently connected to the manifestation behaviour. And because the connection between cause and effect is contingent, this entails that the connection between disposition and manifestation cannot be causal. This parallels Melden's (1961, p.52) objection to the idea that desires are causes of actions: desires are non-contingently related to actions which satisfy the desire. Elizabeth Prior, Robert Pargetter, and Frank Jackson (1982) argued that dispositions lack causal efficacy because there is always a 'causal basis' of the disposition – i.e. there is always a 'property or property-complex of the object that, together with the [triggering or stimulus event] is the causally operative sufficient condition for the manifestation in the case of "surefire" dispositions, and in the case of probabilistic dispositions is causally sufficient for relevant chance of the manifestation' (p.251). According to Prior et al. this means that there's no 'causal work' left for the disposition to do (unless the manifestation event is overdetermined). This argument parallels Kim's causal exclusion argument discussed in chapter 1. And, just as philosophers have responded to Kim by questioning assumptions about what it means for a mental property or state to be causally relevant, philosophers have responded to Prior et al. (1982) by questioning assumptions about what it means for a disposition be causally relevant (e.g. McKitrick, 2005).

However, I think that the debate about the causal efficacy or causal relevance of powers/dispositions is often misconceived. In chapter 3, I expressed support for the Rylean view that powers are not *things*; they are not 'elements of being', to borrow a phrase from Lowe. In Ryle's view, to attribute a power to an entity is not to report a state of affairs, it is not to say that the entity has some attribute or stands in some relation. For an entity to have a power is for an open-

ended set of facts about what that substance can do, or can be relied upon to do – what processes it can engage in – to be true of it. Powers are ways of thinking about how substances are connected to the processes they engage in. In this respect, *power* is akin to concepts like *engagement* and *instancing*: it is a concept which does not name any kind of being, but instead helps us explain the ontological form of entities belonging to the categories the concept concerns. If this view is correct, and for a substance to have a power is not for it to have a certain attribute or stand in a certain relation, then powers (or the state of having a power) cannot be relata of *any* relation, let alone a causal relation. Arguments like Prior et al.'s only have bite if one assumes that powers are the sorts of entities that even could 'do causal work' – and I do not think powers or dispositions are the sorts of entities that even could 'do causal work' because I do not think they are any sort of entity at all.

If one thought that causal reality were nothing but events standing in causal relations, then explanations which make reference to dispositions, if they are causal at all, would have to depend for their truth on the obtaining of certain types of causal relations. However, as I mentioned in section 7.1.2, if the non-relationalist view of causation put forward in this dissertation is plausible, then causal reality is more than events standing in causal relations to other events. Causal reality is also a matter of substances engaging in processes. The idea that it is something about this latter aspect of causal reality which disposition-citing explanations answer to is plausible. On the non-Humean theory of causation I outlined in chapters 3 and 4, what it is for a substance to be exercising a power, or manifesting a disposition, is for that substance to be engaging in a process. Therefore, the obvious candidate for what a disposition-citing explanation reports is the fact that some dynamic state of affairs is a manifestation of the disposition cited. In other words, disposition-citing explanations depend for their truth on the relationship between the disposition cited and the dynamic state of affairs that is the manifestation of that disposition. This is a *relationship* rather than a relation of any kind (and hence not a causal relation), because the disposition is not an element of being, and hence cannot be a relatum of any

relation. This means that the 'because' in disposition-citing explanations does not indicate a causal relation (or indeed a relation of any kind).

7.2 Causal Explanations and Manipulation

We have seen that some causal explanations, namely negative causal explanations, causal explanations that cite the operation of causal processes, stative causal explanations, and disposition-citing causal explanations, do not explicitly mention events whose causal connectedness could ground their truth. In the face of causal explanations like this, the Davidsonian is forced to maintain that reference to the causally related events which make true a causal explanation can be opaque. This suggestion is not implausible itself, but in the case of negative causal explanations and causal explanations that cite the operation of causal processes, it threatens to contravene the reasonable assumption that what makes a sentence true must be what the sentence is about. Furthermore, even this response seems insufficient in the case of stative causal explanations and disposition-citing explanations. This is because, for at least some stative causal explanations and disposition-citing explanations, it is not obvious that causally related pairs of events can be found to serve as implicit referents of explanandum and explanans.

Child suggests that in the face of counterexamples like those discussed in section 7.1, we could 'give up the idea that what makes an explanation a causal explanation is its dependence on the presence of causal relations between events' (1994, p.109). There is more than one way to 'give up the idea that what makes an explanation a causal explanation is its dependence on the presence of causal relations between events'. First, we can give up this idea without giving up the idea that what makes an explanation causal is its dependence on the presence of causal relations of some other kind (perhaps between states). Second, we can deny that what makes an explanation a causal explanation is its dependence on the presence of causal relations of *any* kind – what unites causal explanations into a single category is something else, perhaps a fact about the sort of information they provide.

Some remarks of Child suggest that he has sympathy for the second option. Child describes the alternative to the Davidsonian account as a view where 'causal explanations are not united by their dependence on a natural relation of causality, but rather by the fact that they are all explanations of the occurrence or persistence of particular events or circumstances, or of general types of event or circumstance' (p.100). In any case, it should be obvious that I prefer the second option. I concede that causal explanations depend for their truth on an underlying causal reality, but this underlying reality need not involve any causal *relations* – some causal explanations are not grounded by the presence of any causal relation at all. Instead, I think that explanations are causal because of the sort of information they provide.

In chapter 3, I discussed an objection to my view that we think of causation in two distinct ways, as a process and separately as a cause-effect relation. According to this objection, the idea that we think of causation in two different ways is inconsistent with the idea that causation is a single phenomenon. I responded to this objection by maintaining that the concept of causation as a cause-effect relation is derived from our concept of causation as a process which substances engage in. I noted that if substances possess and exercise causal powers, then substances with certain powers must behave in certain ways when the conditions for the manifestation of the power arise, provided there is nothing interfering. The exercise of powers will therefore be the source of regular and stable relations between trigger-events and manifestation-events. We can use knowledge of these relations to change how powerful substances behave. For example, if one knows that being near flowers triggers an allergic reaction, then one can prevent the allergic reaction by avoiding flowers. From this we get the idea that events, particularly (but not exclusively) trigger-events, can be *devices* for manipulating later events. Events are not literally devices, but even though talk of events as devices is metaphorical there are still conditions under which use of this metaphor is correct and conditions under which use of this metaphor is incorrect. This metaphor is thus the source of the idea that there is a special sort of relation

between events which is causation. So, the causation concept can cover ontologically diverse phenomena because from the concept of causing as something substances engage in, we can derive the idea that some relations between events are causal, via the intermediary notion of using knowledge of stable relations between trigger-events and manifestation-events to manipulate powerful substances. The notion of manipulation thus ties the concepts of causation-as-a-process and causation-as-a-relation together. I suggest that the notion of manipulation is also what explains how many diverse explanations can all count as causal. Causal explanations are explanations which provide information relevant to the manipulation of an effect. They are explanations which provide us with information about how to stop something from happening, or how to get something to happen again, or how to get it to happen in a different way. Or at least, they are explanations which provide us with information about how to make such outcomes more likely.

One might argue that my proposal gives conditions which are unnecessary for an explanation to be causal, because there are some causal explanations where the named causal factor cannot be manipulated even in principle. For example, one might think that 'Fido is warm-blooded because he's a dog' and 'Sarah didn't get promoted because she's a woman,' are causal explanations.[65] It is impossible to consider whether or not Fido would have been coldblooded had he not been a dog, because any possible being which is not a dog is not Fido; similarly, it is impossible to consider whether or not Sarah would have got promoted had she not been a woman, because any possible being who is not a woman is not Sarah, or so the thought goes. For this reason, these cannot be examples of explanations which give information relevant to the manipulation or control of an effect.

[65] Holland (1986) considers examples of this kind, arguing that if these really are causal claims then they are causal claims which lack a clear meaning (pp.954-956).

In response to the first example, it is not obvious to me that this explanation is a causal explanation at all. Fido's being warm-blooded is not causally explained by his being a dog – being warm-blooded is part of what it is to be a dog. The second example, in contrast, does seem to me to be a causal explanation. However, it is not obvious that Sarah's gender is an essential property of her, so it is not obvious that any possible being who is not a woman is not Sarah. Furthermore, even if Sarah's gender were an essential property of her, I would argue that social categories like gender, race and class (and perhaps also categories like criminal, employee, preacher, grandmother etc) are peculiar in that the dispositional properties one enjoys or suffers as a result of being placed into one or other of these categories only exist because of certain cultural practices and behaviour. Sarah's being a woman is a causal factor in the explanation of her not getting promoted, but only because, as a society, we are liable to treat people differently when they fall into different social categories. So, even granting that Sarah's gender is not, even in principle, something we can manipulate, the cultural practices and behaviours which turn being a woman into a causal factor in the first place certainly are things we can manipulate. In other words, 'Sarah didn't get promoted because she's a woman,' is an explanation that provides information relevant to manipulation of an effect after all because of the peculiar connection between social categories and changeable cultural practices. Of course, exactly how social categories function is a debated topic, but this only emphasises the point that 'Sarah didn't get promoted because she's a woman,' is not an uncontroversial counterexample to my proposal.[66]

The assumption made by both Davidsonians and non-causalists, that an explanation is causal only if it depends for its truth on the obtaining of a causal relation, looks questionable. What does this mean for the question of whether rationalising explanations are causal? That some causal explanations are causal

[66] See Woodward (2003, pp.114-117) for a good discussion of this issue.

even though that are not made true by the obtaining of a causal relation allows for the possibility that rationalising explanations are causal even though, as Tanney argues, the mental concepts cited in the rationalising explanation do not pick out items which stand in causal relations to the action explained. Rationalising explanations could be the kind of causal explanation which answers to the non-relational aspect of causal reality.

This thesis, that rationalising explanations are causal explanations which are made true by the non-relational aspect of causal reality, is attractive for at least two reasons. Firstly, it allows us to save the intuition that explaining someone's actions in terms of their beliefs and desires is to give causal information, while at the same time accepting that the mental concepts appealed to in rationalising explanations do not refer to items which stand to the action explained as cause to effect. In other words, the thesis that rationalising explanations are causal, but made true by the non-relational aspect of causal reality, allows us to acknowledge what's intuitive about both the Davidsonian and the non-causalist views, without accepting anomalous monism and the physicalist metaphysics that theory entails.

Secondly, there are similarities between rationalising explanations on the one hand, and process-citing and disposition-citing explanations on the other, which lends support to the idea these three kinds of explanation belong in the same general category. Some rationalising explanations appear to be very similar to causal explanations which cite the continuous operation of causal processes. Causal explanations which cite the continuous operation of causal processes are roughly of the form: some effect occurred or is occurring, or obtained or obtains, because substance S is or was engaging in causal process P. Michael Thompson (2008) outlines a class of rationalising explanations which take this form. These rationalising explanations, which Thompson calls 'naïve action explanations', explain one action in terms of another. (g) would be an example of such a rationalising explanation:

(g) Tom is drilling a hole in the wall because he is hanging a picture.

Tanney's paradigm context-placing explanation also explains why an agent engaged in some activity in terms of something else the agent is doing:

(f) The teacher has written 'CAT' on the board because she is writing 'CATALYST' on the board.

Many rationalising explanations are not of this form. For example, none of the rationalising explanations given in chapter 6 have this form:

(b) Beth is buying flour because she wants to make bread.
(c) John is adding rosemary to the sauce because he believes it will make it taste better.
(d) David took the A road because the motorway was shut.
(e) Columbus sailed west across the Atlantic because he thought that was the way to India.

However, rationalising explanations like (b) – (e) are similar in form to stative causal explanations. Indeed, some have suggested that rationalising explanations are a kind of disposition-citing explanation. Hyman (2015, pp.103-132) argues that explanations of intentional actions which cite desires are disposition-citing explanations because desires are dispositions. Mayr also suggests that that explanatory function of rationalising explanations is 'quite similar to the function of explanations in terms of dispositions or tendencies' (2011, p.295).

However, even if mental states like desiring, believing and knowing are dispositions, they are not *ordinary* dispositions. Most dispositions are dispositions to engage in or undergo a certain specific activity or process. In contrast, having a desire to do something or achieve something (for example) disposes one to undertake whatever activities are deemed, by the agent, to be acceptably good means of achieving what one wants; to deliberately refrain from acting should that turn out to be an acceptably good means of achieving what one wants; to feel happy or pleased if one's desire gets satisfied or disappointed if it is frustrated; and to use one's desire as a premise in practical deliberation about what to do. Desires are not dispositions to do any one specific thing (or even any two specific things) – they are rather dispositions for one's activities to instantiate a certain pattern or goal-directedness, which is made sense of by the

content of the desire. Similar claims can be made about other mental concepts. As Ryle suggested, it would be wrong to think that just because the verbs 'know' and 'believe' are 'ordinarily used dispositionally', that 'there must therefore exist one-pattern intellectual processes in which these cognitive dispositions are actualised' (1949, p.44). Rather, states of believing and states of knowing, if they are dispositions at all, are 'dispositions the exercise of which are indefinitely heterogeneous' (p.44). So, while there are some similarities between rationalising explanations on the one hand and process-citing and disposition-citing explanations on the other, it is important not to forget that rationalising explanations are unique: they are very variable in form, and even if we suppose that the mental states cited in rationalising explanations are dispositions, they are not, by any means, *ordinary* dispositions.

There is one more impediment to concluding that rationalising explanations are causal explanations, even the kind of causal explanations which depend for their truth on something about the non-relational aspect of causal reality. When you learn that some agent's activity is a manifestation of her desire or an output of her rational capabilities, you learn that you might be able to alter her activity by altering what she believes about the world, or by changing her desires, perhaps by changing her environment, but more usually by reasoning with her, talking to her, or persuading her. However, learning this information only makes it the case that you *might* be able to alter the agent's activity. This is because reasoning with an agent in an attempt to get them to φ, for example, does not guarantee that the agent will φ – it does not even ensure that it is more likely that the agent will φ. This is because the agent can ignore you, or remain unconvinced, or even just act against her better judgement. In short, rationalising explanations do not seem to be the sort of explanations which provide us with information about how to stop something from happening, or how to get something to happen again, or how to get it to happen in a different way, or even how to make such outcomes more likely. They seem only to provide information about how we *might* stop something from happening, or get something to happen again, get it to happen in a different way, or make such

outcomes more likely. Rationalising explanations do not seem to provide information that enables the kind of manipulation or control of an effect which I have said causal explanations provide. Thus, it is still unclear what sort of causal information rationalising explanations provide.

As mentioned, it is commonly held that we can achieve an adequate account of what it is to act intentionally by examining the distinctive sort of explanation with which intentional actions are associated. If rationalising explanations are causal explanations which do not designate mental items which stand to the action explained as cause to effect, but instead answer to non-relational causal reality, then the case for thinking intentional actions are distinguished from non-intentional actions by their mental causes is significantly weakened. However, without getting clearer on exactly what facts about dynamic states of affairs rationalising explanations could plausibly be said to answer to, it is difficult to offer a positive account of what the distinguishing mark of intentional action is. I do not have a fully worked out answer to this question. However, in the next section, I will present a view on intentional action which grants that rationalising explanations are a form of disposition-citing explanation, but which respects the two key ways in which rationalising explanations are unique.

7.3 Intentional Action

7.3.1 Mayr's Theory of Intentional Action

Mayr (2011) offers a theory of intentional action which takes seriously the idea that intentional action is the manifestation of a special sort of power. According to Mayr, 'intentional behaviour displays a certain characteristic structure of "purposefulness"' (2011, p.271). Mayr proposes that to act for a reason is for one's behaviour to display a particular kind of structure, i.e. 'the characteristic structure of taking something as one's "standard of success and failure", or "of correctness and incorrectness"' (p.271). Mayr takes this proposal to be supported by the fact that when searching for a rationalising explanation of someone's action the facts we consider relevant are facts about whether the agent's

behaviour, feelings and reasoning display – or would display – a certain pattern. For example, when we wonder if Beth is buying flour because she wants to make bread, we seek to find out things like 'will Beth also buy yeast?', 'if Beth got home and found out her bread tin was missing, would she feel disappointed?', 'would Beth make use of her desire to make bread in a practical deliberation?'. For Mayr these facts do not merely constitute the epistemic criteria for determining what reason an agent acted in light of, they are also the facts that *make it the case* than an agent acted for a specific reason. There's nothing more to acting for a reason than for this welter of facts concerning the agent's actual or hypothetical behaviour and thinking to obtain.

What are the facts, the obtaining of which, makes it the case that an agent acted for a specific reason? According to Mayr's theory, there are three sorts:

1. Facts concerning the teleological structure or 'plasticity'[67] of the agent's actual or hypothetical behaviour. Mayr claims that when an agent has a certain goal, they will 'react sensitively to changes in the environment which threaten the attainment of that goal or make it otherwise necessary to adopt different means for attaining his goal' (p.271) – or *would* if such environmental changes occurred. Agents with a goal will take 'corrective measures' and perform actions 'conducive to overcoming obstacles' should such mistakes or obstacles occur (p.271). These 'corrective movements' indicate to an observer that the agent has a 'standard by which – at least implicitly – he assesses his behaviour and considers himself – in cases of non-conformity of his behaviour to this standard – to have "made a mistake"' (p.273). When an agent does not encounter any obstacles or make any mistakes, the agent's actions may not display plasticity. Mayr insists that, in this case, 'our ascriptions of aims rely on our confidence that certain counterfactual conditionals about what the agent would do if obstacles arose are true, and that the hypothetical behaviour he would display

[67] Mayr takes 'plasticity' to be an alternative term, used by Woodfield (1976), for this pattern in an agent's activity.

would have an adequate teleological structure' (p.274). In other words, the plasticity of hypothetical as well as actual behaviour is important.

2. Facts concerning the agent's actual and hypothetical success and failure feelings. Achieving one's aim is often accompanied by feelings of satisfaction or joy, and failing to achieve one's aim is associated with feelings of disappointment or frustration. For Mayr, what occurrences trigger (or would trigger) feelings of satisfaction or disappointment are important for determining what the agent is aiming at, or what the agent considers to be a success and what he considers to be a failure. Of course, success is not always accompanied by feelings of joy, and failure is not always accompanied by feelings of frustration. For example, when one achieves something one considers a 'necessary evil', one may feel bitter and unhappy upon achieving it. In such cases, Mayr thinks that 'the only success feeling of the agent may be a half-hearted or even bitter feeling of "having done it" or "being finished"' (p.277).
3. Facts concerning whether the agent makes use of their purported aim as a premise in the practical deliberation leading to the action, or at least would if practical deliberation were called for. According to Mayr, when an agent is guided by the requirements he takes to be placed on him by his aims, this guidance will express itself in 'individual or joint practical deliberation about what to do, before or during the action, and in *ex post* justifications of his actions. In practical deliberation, the purpose provides the premise in the agent's deliberation, from which he proceeds to the conclusion that he should act in this way; and after the action it is to this aim that he appeals in justifying his action (as far as he is sincere)' (p.279).

According to Mayr, an agent's behaviour displays the structure characteristic of 'purposefulness' when facts of these three sorts obtain. Mayr claims that it is not necessary that facts of *all* three sorts obtain for an agent to act for a reason. Mayr thinks that sometimes an agent may not deliberate about what to do before acting, may be at a loss when asked later why he acted as he did, have no success and failure feelings, and yet still act for a reason. For example, someone who has an unconscious (or sub-conscious) desire to sabotage a rival might give them bad advice. In this case, the agent has an aim (to sabotage his rival), but does not deliberate, would not be able to give an *ex*

post justification for his action, and might not feel satisfied once the sabotage has been achieved. According to Mayr, 'what is present in such cases is only the (actual or hypothetical) teleological structure of the agent's behaviour' (p.282). Mayr thinks this indicates that facts of type 1 are privileged in the sense that where an agent is acting with an aim, facts of type 1 *must* obtain – something which doesn't hold true for facts of type 2 or 3.

7.3.2 Expanding on Mayr's Theory

There are two issues with Mayr's account I would like to discuss. Firstly, not all intentional activities display a pattern as sophisticated as the one Mayr describes. Some intentional actions are not done for reasons. For example, when I skip just for the fun of it, I have no aim I want to achieve by skipping. In such cases, because I have no aim I want to achieve, I have no aim to use in practical deliberation. Similarly, because there's nothing I want to achieve by skipping, there are no success or failure feelings.[68] It is also unclear that I would engage in actions which are conducive to overcoming obstacles when I skip just for the fun of it. When I skip just for fun, it is more than likely that should some obstacle to skipping occur – e.g. my path becomes blocked or dangerously slippy – I would just stop skipping. I'm doing it just for fun after all, not to achieve anything, so I have no motivation to continue skipping when doing so becomes difficult. Similarly, some animal behaviour seems to be intentional, in a minimal sense, even though it does not display anything as sophisticated as Mayr's 'plasticity'. For example, it seems to me that when a cat grooms itself, the grooming is intentional, but it doesn't seem that, had the cat's environment presented an obstacle to grooming – e.g. had it started to rain – the cat would try to overcome this obstacle and continue grooming itself. In such circumstances, the cat is as

[68] If I go to skip and suddenly find myself unable, this will no doubt incur negative feelings, but they are not obviously 'failure feelings' – I am more likely to feel surprised and possibly concerned that a skill I thought I had has suddenly disappeared!

likely to run off and hunt for mice as it is to go inside and continue grooming itself there. Many animal actions are, I think, intentional, but few have as sophisticated a teleological structure as Mayr describes.

Secondly, Mayr seems to endorse the context-placing or interpretative view of rationalising explanations:

> When we understand acting for a reason as following a standard of success […] it must be the function of reasons-explanations to locate the action within the structure constituted by the agent's behaviour, emotional responses, thoughts, and practical reasoning which is constitutive for following the relevant standard of success. (2011, p.292)

But, as mentioned above, Mayr also thinks that rationalising explanations are a kind of disposition-citing explanation (2011, p.295). Mayr claims that when a rationalising explanation is offered a 'certain item of behaviour is explained *as the manifestation of one of the dispositions* connected with the welter of material and counterfactual conditionals which are responsible for the characteristic structure of intentional agency' (p.294, emphasis added). Mayr claims that the power manifested in intentional action is a 'complex power to act in certain ways in specific situations'; it is a power *of the agent* to structure her own activities (which are exercises of her abilities to act), a power which is 'superimposed on the pre-existing active powers of the agent' (p.295). So, on Mayr's view, rationalising explanations do two things: (1) they place the action explained within a specific structure and (2) they explain an action as the manifestation of a special sort of power to structure one's own activities, a power which is 'superimposed' on the pre-existing active powers of the agent. The second issue with Mayr's account I want to draw attention to concerns how rationalising explanations can perform both roles, and where this special power of an agent to structure her own activities comes from.

In response to the first issue, one might simply insist that actions like skipping for the fun of it and animal actions are not intentional because they do not meet the criteria Mayr sets out. However, even though actions like skipping for the fun of it and animal actions do not display a teleological structure as

complex as the one Mayr describes, it is not true that they display no teleological structure at all. Anyone who can skip is able to make all sorts of small adjustments to their movements to maintain balance, or to ensure that the steps and hops that constitute skipping are executed with the required co-ordination. Skipping still involves some 'corrective measures' albeit on a smaller scale than the kind of corrective measures Mayr talks about. Similarly, when a cat grooms itself, it must co-ordinate the movements of its body so that its tongue catches its fur in just the right way. Again, there is a form of teleological structure demonstrated. In both cases, there is a pattern demonstrated by the agent's actions – a pattern which makes sense once one learns what the agent is trying to do. I think that it is more in keeping with Mayr's core claim, that what makes an activity intentional is its characteristic structure of 'purposefulness', to grant that actions like skipping for the fun of it and animal actions are intentional in virtue of the teleological structure they display, than to insist that such actions do not count as intentional because they fail to demonstrate a teleological structure of the right level of sophistication. If we are content to depart from traditional theories of intentional action and instead adopt a theory which ties the intentionality of some activity to the plasticity of that activity, then why not also accept the phenomenon of intentionality itself is not a homogenous phenomenon, but instead something that can be more or less sophisticated?

The difficulty with weakening Mayr's view so that all activities that display some degree of plasticity count as intentional is that plasticity can be displayed in the behaviour of things that do not have mental states and which do not really act intentionally, for example, machines and robots. This difficulty parallels issues surrounding Daniel Dennett's (1987) intentional stance theory. Dennett proposed that treating objects as rational agents with beliefs and desires helps us understand and predict the behaviour of those objects. Treating objects as rational agents with beliefs and desires is to take an intentional stance with respect to that object. According to Dennett, 'any object – or as I shall say, any *system* – whose behaviour is well predicted by this strategy is in the fullest sense of the word a believer' (1987, p.15). Dennett goes on: '*What it is* to be a true

believer is to be an *intentional system*, a system whose behaviour is reliably and voluminously predictable via the intentional strategy' (p.15). The problem with Dennett's theory is that we can take the intentional stance to objects which do not really have beliefs and desires, like machines and robots.

It is commonly thought that there is a difference between *really* believing something, and behaving *as if* you believed something, and that the difference lies in there being something extra, something hidden, in the case of genuine belief. I think this is the wrong way to capture the difference. True, machines and robots do not really have beliefs and desires, but this is not because believing something is a peculiar kind of property, or involves engaging in a peculiar kind of process. Rather it is because machines and robots do not possess and exercise *two-way powers*. Their behaviour is not up to them. There is a real difference between behaviour of machines which seems to instantiate a pattern which can be made sense of by attributing mental states and genuine intentional action, but the difference does not consist in there being something *extra* present in the latter case. The difference is that machines are not capable of intentional action, because they do not possess two-way powers, and possessing and exercising a two-way power is a necessary condition for acting intentionally.

A consideration which supports the idea that intentional agency always involves the exercise of two-way power is the fact that when an agent is constrained so that they only have the opportunity to φ, and lack the opportunity to not φ, if the agent φs in this situation we wouldn't want to say they φed intentionally.[69] For example, suppose Ben's hands have been

[69] Frankfurt cases (Frankfurt 1969) are thought to demonstrate that this claim is false, that an agent can intentionally φ, and indeed be morally responsible for φing, even when they could not have done otherwise. However, I would argue that even in Frankfurt cases the agents in question do, in fact, have the ability and opportunity not to φ. The presence of neuroscientists with fancy machinery, who could take control over an agent's body just in case they start to look like they might not φ by themselves, may foreclose the physical possibility that a φing won't happen, but, as I suggested in footnote 13, these facts are not relevant to what is an open agential possibility for the agent.

temporarily paralysed so that he is denied the opportunity to move his hands. Whether Ben moves his hands or not is not up to him. Is it possible for Ben, in this situation, to intentionally refrain from moving his hands? Suppose someone unaware of Ben's situation said to him: "if you keep your hands perfectly still I'll give you £10". Ben may want to comply, but even if not moving is what Ben wants, it does not seem like he is remaining still intentionally when his hands are paralysed. It seems like being able to both move and not move your hands is a precondition for doing one or the other intentionally, and lacking this two-way power renders intentionally doing one or the other action impossible.

Another consideration that may speak in favour of the view that exercising a two-way power is a necessary condition for intentional action are cases of deviant causation. As mentioned in chapter 2, deviant causal chain cases are a well-known problem for causal analyses of intentional action, i.e. analyses which attempt to reduce intentional action to causation of bodily movements by appropriate mental states and/or events. The most famous deviant causal chain case is Davidson's nervous climber example:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. (Davidson, 1973/2001, p.79)

There are two potential explanations available for why there is no intentional action in this case:

(1) The causal chain does not follow the sort of causal path that counts as 'the "right" way in which beliefs and desires must yield behaviour for genuine intentional action to occur' (Bishop, 1989, p.135), the 'right way' being: …, where the '…' has to be filled in without reference to intentional action.

(2) The agent's reasons or intentions operate via causal chains which rob the agent of the relevant two-way power, most probably by robbing the agent of the opportunity to both φ and not φ. For example, in Davidson's nervous climber case, the climber's nervousness robs the climber of the opportunity not to let go

> of the rope. Just as extreme grief can render a person incapable of not crying out, the climber's control over his body has been hijacked by the conditions responsible for his nervous state.

The success of (1) as an explanation depends on how the '…' is filled in. One promising strategy is the 'sensitivity approach' (e.g. Peacocke 1979, Mele 1992, 2003). This approach suggests that a necessary condition for intentional action is that the bodily movement is 'responsive' or 'sensitive' to the content (or some other variable feature) of the mental state that caused it. One way of spelling out this sensitivity requirement is in terms of counterfactuals: a bodily movement is sensitive to the mental state that caused it if and only if had the mental state had a slightly different content, then a slightly different bodily movement – one that conformed to the slightly different mental state – would have occurred.[70] Smith (2010) gives a clear example: suppose a pianist wants to appear nervous to his audience and believes he can achieve this end by playing a C# instead of a C during his piece. The pianist's pressing C# is sensitive to this belief-desire pair if only if had the pianist thought that pressing B would achieve his goal, then the pianist would have pressed B. Cases of deviant causation are thought not to satisfy this sensitivity requirement.

[70] The counterfactual version of the sensitivity approach isn't the only version available. Peacocke (1979) offers an alternative version. Peacocke argues that there is an intentional action if and only if the bodily movement is caused by an intention and that the intention differentially explains the occurrence of the bodily movement (p.69). A state or event differentially explains another when there is a law backing the explanation, according to which changes in the intensity or value of the explanandum are correlated (one-to-one) with changes in the intensity or value of the explanans. For the sake of brevity, I won't discuss Peacocke's version of the sensitivity approach in this essay. See Sehon (1997) for a convincing argument that Peacocke's proposed criterion for intentional action is neither necessary nor sufficient.

However, this proposal faces a counterexample. Consider Amy and her device which can manipulate my brain and nervous system to make me get up out of my seat at 4pm. Suppose that I do in fact form the intention to get up at 4pm because I want to make tea. Suppose further that Amy would *only* use her device if I formed the intention to get up; had I formed a different intention Amy wouldn't have used her device *to make me get up,* she would have used her device instead to make sure my body moved in conformity with my alternative intention. Thus, the bodily movement that results from my intention to get up is sensitive to the content of that intention. However, where Amy uses her device to manipulate my brain and nervous system, I am not performing an intentional action: *I* am not in control over what is going on with my body, *Amy* is.[71] Bishop (1989) calls cases like this, where the causal path from intention to bodily movement passes through a benevolent second agent, *heteromesial* causal chain cases.[72]

One response to this counterexample may be to stipulate that the causal chain cannot be heteromesial if intentional action is to occur. However, as Bishop (1989, pp.158-159) points out, this cannot be right, as not *every* heteromesial causal chain is such that it blocks intentional action. Imagine machinery like Amy's is used to make sure that an agent's damaged neural pathways carry on functioning as normal (e.g. suppose some synapse isn't functioning properly, Amy's machinery might work by stimulating the second neurone when the first is in the right electrochemical state, just as the first

[71] This counterexample is adapted from an example given by Peacocke (1979, p.87).
[72] A more recent suggested solution to the causal deviance problem, suggested by McDonnell (2015), also cannot deal with this counterexample. McDonnell suggests that there is an intentional action if and only if the mental cause of the bodily movement is 'proportional', in Yablo's (1992) sense, to the bodily movement. My intention to get up is a proportional cause of my subsequent getting up if and only if the following counterfactual conditionals are true:

1. Had my intention to get up been absent, then I would not have gotten up.
2. Had my intention to get up been absent, then had I intended to get I would have gotten up.

These are both true even in the hetereomesial case.

neurone *would* if it were working properly). Now suppose I'm the one with the damaged neural pathways and when I decide to get up at 4pm to make tea, Amy has to switch her machine on to make sure that neural signals are transmitted as normal thus allowing me to get up. In this case, the causal chain from my intention to my bodily movement has to go via an action of Amy's – were it not for Amy, my damaged neural pathways would thwart any bodily movement, thus rendering my intention inefficacious. However, it does not seem, in this case, that my getting up at 4pm is not intentional. In this case Amy is *helping* me carry out my intention to get up by helping my nervous system remain in working order, she's an essential component of the causal chain that lets me carry out my intention, but my action doesn't fail to be intentional.

The project of specifying what it is for a causal chain from beliefs and desires or intentions to bodily movement to be non-deviant may suffer a similar plight to that faced by the project of specifying necessary and sufficient conditions for knowledge, namely, that every new proposal faces new counterexamples and the project seems nowhere near an end.[73] I would like to suggest that adequately solving the problem of deviant causal chain cases has proven difficult because the correct explanation for why there is no intentional action in deviant causal chain cases is because in these cases is the agent lacks the relevant two-way power. In other words, I endorse explanation 2 above. This is why some heteromesial cases are such that intentional action is blocked, and others do not block intentional action: not every heteromesial case is such that the agent is stripped of either the ability to φ or not φ or the opportunity to φ or not φ. Where Amy is using her machine to keep my nervous system in working order, she hasn't robbed me of the ability or opportunity to not get up. Whereas where she uses her machine to control the movements of my body, she has

[73] See Zagzebski (1994) for an argument that Gettier-style counterexamples are inescapable for almost every analysis of knowledge.

robbed me of the opportunity not to get up. If 2 really is the better explanation, then we have further support for the idea that there can be no intentional action that is not an exercise of a two-way power. Thus, deviant causal chain cases provide conditional support for the idea that possessing and exercising a two-way power is a necessary condition for acting intentionally.

The idea that possessing and exercising a two-way power is a necessary condition for acting intentionally suggests a possible answer to the second problem facing Mayr's account. It is because we have two-way powers that our activities can demonstrate patterns of the kind Mayr describes. When we have two-way powers, it is up to us whether we perform the activites these two-way powers are powers to do. In virtue of this, the pattern our actions display is also up to us. This is where, I think, the special power of an agent to structure her own activities, the power which Mayr says is 'superimposed' on the pre-existing active powers of the agent, comes from. Because we have many two-way powers, we also have an extra power to organise our actions in such a way so as to meet our aims. The power to act intentionally is thus an emergent power – a power that emerges from our possessing two-way powers to act. Having such a power does not mean we will always use it – many exercises of two-way powers are not intentional, for example absent-minded fiddling. The power may also come in degrees: creatures whose powers are mostly two-way will be able to organise their activities into a greater variety of patterns than creatures whose powers are mostly one-way. For example, it is up to the cat whether it grooms itself now or later, but it is probably not up to the cat whether it grooms itself at all. For human beings, however, refraining from taking care of one's hygiene is, perhaps unfortunately, under our personal control.

This view has interesting consequences for the question of what causal information rationalising explanations provide. First, the view grants that rationalising explanations are a form of disposition-citing explanation. Intentional actions are manifestations of a special sort of power, namely a power to organise one's activities in accordance with a certain form (a power which depends on having two-way powers to act), and the function of rationalising

explanations is to tell us which form the agent was disposed to structure her activities in accordance with. Second, the view allows that rationalising explanations are context-placing or interpretive: the mental concepts cited in rationalising explanations make the form of an agent's intentional activity intelligible. Third, the view can explain why determining whether rationalising explanations provide information relevant to the manipulation or control of an effect, and hence whether rationalising explantions are causal, is difficult.

As I mentioned in section 7.2, when you learn that some agent's activity is a manifestation of her desire or an output of her rational capabilities, you learn that you might be able to alter her activity by altering what she believes about the world, or by changing her desires, perhaps by changing her environment, but more usually by reasoning with her, talking to her, or persuading her. However, learning this information only makes it the case that you *might* be able to alter the agent's activity. The view of intentional action sketched in this section allows us to explain why this is: reasoning with an agent in an attempt to prevent them from φing (or get them to φ) doesn't take away the agent's two-way power to φ. Because her power to φ is two-way, it is up to her whether she φs or not. Of course, we can always control someone else's φing by removing their two-way power to φ, for example by tying them down so that they no longer have the opportunity to φ. But, learning about the reasons and motives behind an agent's activity is not relevant for our exercising *this* kind of control over the agent. If learning about the reasons and motives behind an agent's activity is relevant for the manipulation or control of their behaviour at all, then it is relevant for a kind of control that leaves the agent's two-way powers intact.

Determining whether rationalising explanations provide information relevant to the manipulation or control of an effect is difficult because it is unclear whether this latter sort of control is a form of *causal* control. Is convincing someone to behave in some way to exercise a *causal* power? Is it to cause something to happen? These questions matter if, as I have proposed, an explanation is causal if and only if it provides information relevant to manipulation and control, where manipulation and control are causal activities

that powerful particulars, such as ourselves, can undertake. I do not know if the answer to these questions should be yes. I do not think that the causal concept sits comfortably with concepts like *convincing*, *persuading* and *reasoning with*. On the other hand, the concept does not feel wholly inappropriate either. In short, because the disposition manifested when an agent acts intentionally is one which is dependent on their having and exercising *two-way* powers, learning about the reasons and motives behind an agent's activity does not provide us with information that enables us to *ensure* that the activity is (or is not) engaged in. However, it is not obvious that exercising causal control over a situation is always a matter of *ensuring* certain outcomes. The causal status of rationalising explanations is unclear. But if something like the account of intentional action I have sketched in this section is true, then the dubious causal nature of rationalising explanations isn't an anomaly, it is instead something that should be expected given the nature of the agential powers demonstrated in intentional action.

In this chapter, I have sought to show that it is not obviously true that an explanation is causal only if its explanandum designates an effect and its explanans designates an item which is the cause of that effect. My non-relational theory of causation allows that some causal explanations may depend for their truth on facts about dynamic states of affairs. Furthermore, it is quite plausible that process-citing explanations and disposition-citing explanations are the kinds of causal explanation which answer to the non-relational aspect of causal reality. In other words, it is plausible that the 'because' of these causal explanations does not signify the obtaining of a causal relation. There is some reason to think that rationalising explanations are also causal explanations which are not made true by a pair of causally related events. This view has consequences for how we ought to think about the nature of intentional action. It casts doubt over the view that intentional actions are distinguished from non-intentional actions by their causes. I proposed an alternative view of intentional actions, inspired by Mayr (2011), which takes intentional actions to be manifestations of a special power of agents to organise their activities into a pattern of determinate form

(an emergent power that depends on one possessing two-way powers to act). Rationalising explanations reveal this form by attributing mental states with certain contents to the agent.

8

Conclusion

This dissertation challenges the orthodox understanding of the mental causation that is on display when an agent acts intentionally. In discussions of the so-called problem of mental causation, mental causation is typically presented as a cause-effect relation between mental and physical entities. As I put it in chapter 1, mentality and physicality are presented as two sides of a causal exchange. I called this understanding of mental causation the relational understanding of mental causation:

> **Relational understanding of mental causation**: mental causation is a matter of mental items (events, processes or states which are conceived of as particulars) standing in causal relations to physical events, e.g. bodily movements.

The clearest endorsement of this understanding of mental causation can be found in Kim's remarks on why it is important that mental causation is real:

> First and foremost, the possibility of human agency, and hence our moral practice, evidently requires that our mental states have causal effects in the physical world. In voluntary actions our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged. (2005, p.9)

The aim of this dissertation was to show that Kim is wrong to claim that 'in voluntary actions our beliefs and desires, or intentions and decisions, must somehow cause our limbs to move in appropriate ways' (2005, p.9). When we say that someone acted intentionally because of what she believed, desired, intended or decided, the concepts *belief*, *desire*, *intention*, perhaps even *decision*, do not refer to items which stand in causal relations to limb movements. In other

words, in this dissertation I sought to defend the thesis that even if there is mental causation on display when we act intentionally, this mental causation should not be relationally understood.

To defend this thesis, I began by arguing that the canonical theory of action, the causal theory of action, fails as a theory of agency. I argued that, contrary to the claims of the causal theory of action, agency is misconstrued when it is taken to be something that can be reduced to a special kind of event-causation. I endorsed Hornsby's (2004a, 2012) suggestion that what is needed to properly understand agency is a radical departure from the Humean approach to causation which the causal theory of action presupposes. Specifically, to understand agency we need a metaphysical framework that allows us to think of causation as something other than a relation between events. Only then is it possible to see how the causality of action might be something other than a causal relation between mental event and action, and instead something that casts the agent as a causal player, rather than merely the setting for events to cause other events.

In chapter 3, I outlined a non-Humean approach to causation which involved rejecting the three core theses of Humean theories of causation: 1) reductionism 2) denial of powers and 3) relationalism. According to this approach, causation is not always and everywhere a relation, and giving a full account of causation is not merely a matter of explaining what a relation must be like to be a causal relation. Put positively, I maintain that causation can be a process rather than a relation, of which processes like breaking, crushing, bending etc are more determinate species. In chapter 4, I outlined and defended the process ontology which this non-Humean theory of causation depends on. My process ontology maintains that processes are universals which substances engage in, and events are instances of processes – they are particular occurrences which come into being when a substance has engaged in a process and completed it.

I argued in chapter 5 that this non-Humean approach to causation, and the process ontology that accompanies it, allows us to put together a more

successful understanding of agency. On my view, agents are substances which exercise agential powers, where to exercise a power is for a substance to engage in a process, i.e. for a dynamic state of affairs to obtain. Actions are the events which come into existence when agents exercise their agential powers – i.e. engage in processes – and then complete those processes. I argued in chapter 5 that there are two distinctions crucial to our concept of agential power: the distinction between activity and passivity, and the distinction between one-way and two-way powers. Agency does not reduce to the exercise of active power, because some substances can manifest their agency by remaining passive, and therefore by not engaging in activity. Neither does agency reduce to the exercise of two-way power, because not all substances which cause things to happen do so by exercising two-way powers, but all substances which cause things to happen are agents. My view is that agency is a complex concept which incorporates both distinctions. Some substances' agential powers are one-way; these substances manifest their agency when they are active, but not when they are passive. Other substances' agential powers are two-way; these substances manifest their agency when they are active, but also sometimes when they are passive.

In chapters 6 and 7, I turned my attention to intentional action. Many philosophers have tried to provide an account of intentional action by examining the distinctive sort of explanation with which intentional actions are associated, i.e. rationalising explanations. Davidson (1963) argues that rationalising explanations are causal explanations. They are true if a mental event suitably related to the mental concept cited in the rationalising explanation stands in a causal relation to the action explained. Davidson's argument that rationalising explanations are causal is often taken to justify the claim that mental states or events stand in causal relations to intentional actions. Thus, Davidson's argument is the source of the common view that our conception of ourselves as intentional agents presupposes that mentality is causally relevant in the physical world and that this mental causation should be conceived of in relational terms.

In chapter 6, I outlined a number of considerations which appear to suggest that rationalising explanations do not explain by causally attributing one event to another. As Tanney puts it, the mental concepts in rationalising explanations do not discharge their explanatory role by designating causes of the actions they explain (2009, p.100). Instead, rationalising explanations are 'context-placing' explanations, explanations which place the action explained in a context that makes it intelligible. To say that when we attribute beliefs and desires to an agent we are positing internal causes of their observable behaviour misconstrues the explanatory function of these mental concepts. However, I acknowledged that, despite Tanney's arguments, the idea that when we explain someone's actions in terms of their beliefs and desires we are giving causal information is very appealing.

In chapter 7, I argued to say that a rationalising explanation is causal is not thereby to commit oneself to the view that mental items stand in causal relations to actions, and that maybe rationalising explanations are causal even though they do not function by designating mental items which are the causes of the actions they explain. I did this by first examining what makes an explanation causal. I argued against the view that causal explanations are precisely those explanations whose explanandum designates an effect and whose explanans designates an item which is the cause of that effect. My non-Humean theory of causation implies that facts about what events are causally related to what others are not the only causal facts which causal explanations could answer to; some causal explanations may answer to facts about dynamic states of affairs. I suggested that explanations that cite the operation of causal processes and disposition-citing explanations are amongst those causal explanations which depend for their truth on facts about dynamic states of affairs.

I proposed that rationalising explanations are also causal explanations which are not made true by causally related events. The most important consideration favouring this view is that is saves two strong intuitions: 1) that rationalising explanations are causal, and 2) that the mental states cited in rationalising explanations do not denote items which stand in causal relations to

the actions they explain. The idea that rationalising explanations are causal explanations which answer to the non-relational aspect of causal reality is also supported by the fact that rationalising explanations bear some similarities to both process-citing and disposition-citing explanations.

The view that rationalising explanations are causal explanations which do not designate mental items that stand to the action explained as cause to effect has consequences for how we ought to think about the nature of intentional action. Most importantly, it casts doubt on the view that intentional actions are distinguished from non-intentional actions by their causes. I proposed an alternative view of intentional actions, inspired by Mayr (2011), which takes intentional actions to be manifestations of a special power to organise one's activities into a pattern of determinate form. Rationalising explanations reveal this form by attributing mental states with certain contents to the agent – in this way, rationalising explanations are both context-placing and disposition-citing. In other words, intentional actions are manifestations of a power to organise one's activities so that they instantiate a certain structure, and rationalising explanations make this structure comprehensible.

The theory of intentional action I propose relies heavily on two ideas: the notion of plasticity, or teleological structure, whose defining characteristic is its 'corrective behaviour'; and the concept of a two-way power. Following Mayr, I proposed that what makes activity intentional is the fact that it displays the characteristic structure of 'purposefulness' or 'of taking something as one's "standard of success and failure", or "of correctness and incorrectness"' (Mayr, 2011, p.271). However, I recommended that we be more liberal regarding what sorts of structure counts as a structure of 'purposefulness' than Mayr proposes, so that actions which are done simply because one wants to do them and not for any reason and some animal actions count as intentional. To avoid the consequence that all activities that display some degree of plasticity count as intentional, even activities engaged in by substances which do not really act intentionally, like machines and robots, I suggested that only activities engaged in by substances which possess two-way powers can be intentional. The activities

of machines and robots are not even candidate intentional actions because they are exercises of one-way powers. This account requires further defence. One key problem with it is that I have said nothing about how we *know* which substances possess two-way powers and which don't. I mentioned in chapter 2 that I doubt that exercises of two-way power come with distinctive perceptual markers. One might worry whether an account of intentional action which relies so heavily on a concept which suffers this kind of epistemological problem can be successful. However, even though the account I have proposed is underdeveloped, I hope it is enough to show that what's distinctive about intentional action need not be its cause.

I think the considerations put forward by non-causalists indicate that our concept of intentional action does not presuppose that mental items (states or events) stand in causal relations to actions, or physical events (e.g. bodily movements). However, I hope that the argument of chapter 7 shows that we can deny that the concepts employed in rationalising explanations refer to items which stand in causal relations to actions even if we take rationalising explanations to be causal. The fact that we causally explain people's intentional actions in terms of their mental states does not justify the contention that necessarily, whenever there is intentional action there is a causal relation between a mental item and an action or a bodily movement. The causal nature of rationalising explanations does not give us any reason to think that necessarily, relational mental causation is on display whenever we act intentionally. This is not to say that causal relations between mental events and physical events do not exist. The conclusion I have argued for is weaker than that. My conclusion is that the existence of intentional action does not presuppose the existence of mental causation relationally understood. The existence of causal relations between mental events and physical events is not *conceptually entailed* by the existence of intentional action.

Is there anything worthy of the name 'mental causation' necessarily on display whenever an agent acts intentionally? I believe we can, and should, answer this question positively. I have, at various points in this dissertation,

mentioned that it is natural to think that some form of mental causation, or 'the reality of causal processes involving cognitive phenomena' as Menzies (2013, p.58) puts it, is indispensable to our conception of ourselves as agents who act intentionally and bear moral responsibility. A positive answer to this question is possible once we acknowledge that we need not, and should not in this case, understand 'mental' in 'mental causation' as a 'transferred epithet' as Crane (1995, p.219) puts it. Understanding 'mental' in 'mental causation' as qualifying the cause relatum of a causal relation, rather than causation itself, is a prescription of the relational understanding of mental causation.

In chapter 5, I proposed a theory of agency where to act is to engage in an activity. If this theory is correct, then acting is to engage in a process of causation (or, more accurately, a determinate form of the process of causation). And if, as I suggested in chapter 7, acting intentionally is to manifest a special power to organise one's activities into a pattern that can be made sense of by appeal to mental concepts, why not say that the causation that is engaged in when an agent acts intentionally is 'mental' in virtue of this fact? In other words, perhaps the causation on display when an agent acts intentionally is mental in virtue of the fact that it is part of a teleological structure whose form is revealed by attributing knowledge, beliefs, desires or aims to the agent. As Hornsby (2015, p.135) points out, it is wrong to think of instances of human agency as occurrences which interrupt idleness. Human beings are not things which just lie there until something comes along and prods them into action. This is true regardless of whether we think of the thing that is doing the prodding as another substance, or as a mental item. Human beings, as Hornsby puts it, are 'rational creature[s], leading a life, equipped with powers of thought and self-movement' (p.135). Why not think of the mentality of the causal processes human beings engage in when they act intentionally as consisting in the fact that these processes are part of a larger pattern of meaningful, or interpretable, activity?

This is not the only way in which the truth of the idea that some form of mental causation is indispensable to our conception of ourselves as agents who act intentionally and bear moral responsibility can be borne out. Human beings

are capable of performing activities which we would naturally describe as ‘mental’, such as imagining and reasoning, and persuading and convincing. Exactly what these activities amount to is a difficult philosophical question, one which I cannot answer here. However, it seems to me that these activities are ways to deliberate (individually or in groups) about what beliefs and desires it is best to have, and can be means by which we can alter what beliefs or desires an agent has. That we have such capacities is relevant to our bearing moral responsibility.

How it is that we have such capacities is, I think, a very difficult question. How are we able to engage in activities like imagining and reasoning? How does our capacity to imagine, reason, persuade, or convince relate to the physical capacities of our bodies? How is it possible that we can change the action-plans and projects an agent is disposed to enact by imagining or reasoning or persuading or convincing? I have no idea how to answer these questions. But it is *these* questions – and not questions about how mental items can stand in causal relations to physical events – which constitute the real problem of mental causation. The real mystery is not how mental items can stand in causal relations to physical events, but how it is that we can perform mental activities.

Works Cited

Árnadóttir, S. T. & Crane, T., 2013. There is No Exclusion Problem. In: R. Ingthorsson, S. C. Gibb & E. J. Lowe, eds. *Mental Causation and Ontology*. Oxford: Oxford University Press.

Alvarez, M., 2001. Letting Happen, Omissions and Causation. *Grazer Philosophische Studien,* 61(1), pp. 63-81.

Alvarez, M., 2010. *Kinds of Reasons: An Essay in the Philosophy of Action.* Oxford: Oxford University Press.

Alvarez, M., 2013. Agency and Two-Way Powers. *Proceedings of the Aristotelian Society,* Volume 113, pp. 101-121.

Alvarez, M. & Hyman, J., 1998. Agents and Their Actions. *Philosophy,* 73(284), pp. 219-245.

American Honda Motor Co. Inc., 2017. *Inside ASIMO.* [Online]
Available at: http://asimo.honda.com/inside-asimo/ [Accessed 30 December 2017].

Anscombe, G. E. M., 1963. Critical notices. *Mind,* 72(286), pp. 288-293.

Anscombe, G. E. M., 1971. *Causality and Determination: an Inugural Lecture.* Cambridge: Cambrdige University Press.

Anscombe, G. E. M., 2000. *Intention.* Cambridge(Massachusetts): Harvard University Press. Originally Published in England in 1957 by Basil Blackwell.

Armstrong, D. M., 1968. *A Materialist Theory of the Mind.* London: Routledge.

Armstrong, D. M., 1978a. *Universals and Scientific Realism: Nominalism and Realism.* Cambridge: Cambridge University Press.

Armstrong, D. M., 1978b. *Universals and Scientific Realism: A Theory of Universals.* Cambridge: Cambridge University Press.

Armstrong, D. M., 1989. *Universals: An Opinionated Introduction.* Boulder: Westview Press.

Armstrong, D. M., 1997. *A World of States of Affairs.* Cambridge: Cambridge University Press.

Armstrong, D. M., 1999. The Open Door: Counterfactual Versus Singularist Theories of Causation. In: H. Sankey, ed. *Causation and Laws of Nature.* Dordrecht: Kluwer Academic Publishers.

Baumgartner, M., 2008. Regularity Theories Reassessed. *Philosophia,* Volume 36, pp. 327-354.

Beardsley, M. C., 1975. Actions and Events: The Problem of Individuation. *American Philosophical Quarterly,* 12(4), pp. 263-276.

Beebee, H., 2005. Causing and Nothingness. In: L. A. Paul, E. J. Hall & J. Collins, eds. *Causation and Counterfactuals.* Cambridge(Massachusetts): MIT Press, pp. 291-308.

Beebee, H., 2006. Does Anything Hold the Universe Together? *Synthese,* 149(3), pp. 509-533.

Beebee, H., 2007. Humes Old and New: Peter Millican and Helen Beebee: The Two Definitions and the Doctrine of Necessity. *Proceedings of the Aristotelian Society,* Volume 107, pp. 413-431.

Bennett, J., 1988. *Events and Their Names.* Indianapolis: Hackett Publishing Company.

Bennett, K., 2003. Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract it. *Noûs,* 37(3), pp. 471-497.

Berofsky, B., 2011. Compatibilism Without Frankfurt: Dispositional Analyses of Free Will. In: R. Kane, ed. *Handbook of Free Will.* 2nd ed. New York: Oxford University Press, pp. 153-175.

Bishop, J., 1989. *Natural Agency: An Essay on the Causal Theory of Action.* Cambridge: Cambridge University Press.

Bradley, F. H., 1893. *Appearance and Reality: A Metaphysical Essay.* Oxford: Oxford University Press.

Bratman, M. E., 2001. Two Problems About Human Agency. *Proceedings of the Aristotelian Society,* Volume 101, pp. 309-326.

Campbell, J. K., 2005. Compatibilist Alternatives. *Canadian Journal of Philosophy,* 35(3), pp. 387-406.

Cartwright, N., 2009. Causal Laws, Policy Predictions, and the Need for Genuine Powers. In: T. Hanfield, ed. *Dispositions and Causes.* Oxford: Oxford University Press.

Child, W., 1994. *Causality, Interpretation, and the Mind.* New York: Oxford University Press.

Chisholm, R., 1970. Events and Propositions. *Noûs,* 4(1), pp. 15-24.

Chisholm, R., 1976. *Person and Object: A Metaphysical Study.* London: Routledge.

Clarke, R., 2003. *Libertarian Accounts of Free Will.* New York: Oxford University Press.

Coates, J. D. & McKenna, M., 2016. *Compatibilism.* [Online]
Available at: https://plato.stanford.edu/archives/win2016/entries/compatibilism/
[Accessed 30 December 2017]

Coope, U., 2007. Aristotle on Action. *Proceedings of the Aristotelian Society, Supplementary Volumes,* Volume 81, pp. 109-138.

Crane, T., 1995. The Mental Causation Debate. *Proceedings of the Aristotelian Society,* Volume 69, pp. 211-36.

Crowther, T., 2011. The Matter of Events. *Review of Metaphysics,* 65(1), pp. 3-39.

Dancy, J., 2000. *Practical Reality.* New York: Oxford University Press.

Davidson, D., 1963. Actions, Reasons, and Causes. *Journal of Philosophy,* 60(23), pp. 685-700. Reprinted in Davidson 2001.

Davidson, D., 1967a. The Logical Form of Action Sentences. In: N. Rescher, ed. *The Logic of Decision and Action.* Pittsburgh: University of Pittsburgh Press. Reprinted in Davidson 2001.

Davidson, D., 1967b. Causal Relations. *Journal of Philosophy,* 64(21), pp. 691-703. Reprinted in Davidson 2001.

Davidson, D., 1969. The Individuation of Events. In: N. Rescher, ed. *Essays in Honor of Carl G. Hempel.* Dordrecht: D. Reidel Publishing Company, pp. 216-234. Reprinted in Davidson 2001.

Davidson, D., 1970. Mental Events. In: L. Foster & J. W. Swanson, eds. *Experience and Theory.* 2nd ed. Cambridge: The University of Massachusetts Press. Reprinted in Davidson 2001.

Davidson, D., 1971. Agency. In: R. Binkley, R. Bronaugh & A. Marras, eds. *Agent, Action, and Reason.* Toronto: University of Toronto Press. Reprinted in Davidson 2001.

Davidson, D., 1973. Freedom to Act. In: T. Honderich, ed. *Essays on Freedom of Action.* New York: Routledge and Kegan Paul. Reprinted in Davidson 2001.

Davidson, D., 1987. Problems in the Explanation of Action. *Metaphysics and Morality: Essays in honour of J.J.C. Smart*.

Davidson, D., 2001. *Essays on Actions and Events.* 2nd ed. Oxford: Oxford University Press.

de Swart, H., 1996. Quantification Over Time. In: J. van der Does & J. van Eijck, eds. *Quantifiers, Logic, and Language.* Cambridge: Cambridge University Press.

de Swart, H., 2012. Verbal Aspect. In: R. I. Binnick, ed. *The Oxford Handbook of Tense and Aspect.* New York: Oxford University Press, pp. 752-781.

Dennett, D. C., 1987. *The Intentional Stance.* Cambridge(Massachusetts): MIT Press.

Evnine, S., 1991. *Donald Davidson.* Stanford: Stanford University Press.

Fales, E., 1990. *Causation and Universals.* London: Routledge.

Filip, H., 2012. Lexical Aspect. In: R. I. Binnick, ed. *The Oxford Handbook to Tense and Aspect.* New York: Oxford University Press, pp. 721-752.

Frankfurt, H. G., 1969. Alternate Possibilities and Moral Responsibility. *Journal of Philosophy,* 66(23), pp. 829-839.

Franklin, C. E., 2016. If Anyone Should Be an Agent-Causalist, then Everyone Should Be an Agent-Causalist. *Mind,* 125(500), pp. 1101-1131.

Frost, K., 2013. Action as the Exercise of a Two-Way Power. *Inquiry: An Interdisciplinary Journal of Philosophy,* 56(6), pp. 611-624.

Galton, A. & Mizoguchi, R., 2009. The Water Falls but the Waterfall does not Fall: New Perspectives on Objects, Processes and Events. *Applied Ontology,* 4(2), pp. 71-107.

Ganeri, J., Noordhof, P. & Ramachandran, M., 1996. Counterfactuals and Preemptive Causation. *Analysis,* 56(4), pp. 219-225.

Gibb, S., 2013. Introduction to Mental Causation and Ontology. In: S. C. Gibb, E. J. Lowe & R. D. Ingthorsson, eds. *Mental Causation and Ontology.* Oxford: Oxford University Press, pp. 1-17.

Gilmore, C. S., 2007. Time-Travel, Coinciding Objects and Persistence. In: D. W. Zimmerman, ed. *Oxford Studies in Metaphysics.* New York: Oxford University Press, pp. 177-198.

Gvozdanović, J., 2012. Perfect and Imprefect Aspect. In: R. I. Binnick, ed. *The Oxford Handbook to Tense and Aspect.* New York: Oxford University Press, pp. 781-803.

Hall, N., 2004. Two Concepts of Causation. In: J. Collins, N. Hall & L. Paul, eds. *Causation and Counterfactuals.* Cambridge(Massachusetts): MIT Press, pp. 225-276.

Harré, R. & Madden, E. H., 1975. *Causal Powers.* Oxford: Blackwell.

Hart, H. L. & Honoré, A. M., 1985. *Causation in the Law.* 2nd ed. Oxford: Oxford University Press.

Haslanger, S., 1994. Humean Supervenience and Enduring Things. *Australasian Journal of Philosophy,* 72(3), pp. 339-359.

Haugeland, J., 1982. Weak Supervenience. *American Philosophical Quarterly,* 19(1), pp. 93-103.

Heil, J., 2013. Mental Causation. *Mental Causation and Ontology*, pp. 18-35.

Hellman, G. P. & Thompson, F. W., 1975. Physicalism: Ontology, Determination, and Reduction. *Journal of Philosophy,* 72(17), pp. 551-564.

Higginbotham, J., 2000. On Events in Linguistic Semantics. In: J. Higginbotham, F. Pianesi & A. Varzi, eds. *Speaking of Events.* New York: Oxford University Press, pp. 49-81.

Hinshelwood, A., 2013. The Metaphysics and Epistemology of Settling: Some Anscombean Reservations. *Inquiry: An Interdisciplinary Journal of Philosophy,* 56(6), pp. 625-638.

Hitchcock, C., 1993. A Generalized Probabilistic Theory of Causal Relevance. *Synthese,* Volume 97, pp. 335-364.

Holland, P. W., 1986. Statistics and Causal Inference. *Journal of the American Statistical Association,* 81(396), pp. 945-960.

Hopkins, J., 1978. Mental States, Natural Kinds and Psychophysical Laws. *Proceedings of the Aristotelian Society,* Volume 221, pp. 221-236.

Horgan, T. E., 1993. From Supervenience to Superdupervenience: Meeting the demands of a Material World. *Mind,* 102(408), pp. 555-586.

Hornsby, J., 1980. *Actions.* London: Routledge.

Hornsby, J., 2004a. Agency and Actions. In: H. Steward & J. Hyman, eds. *Agency and Action.* Cambridge: Cambridge University Press, pp. 1-23.

Hornsby, J., 2004b. Agency and Alienation. In: D. Macarthur & M. De Caro, eds. *Naturalism in Question.* Cambridge(Massachusetts): Harvard University Press, pp. 173-187.

Hornsby, J., 2012. Actions and Activity. *Philosophical Issues,* pp. 233-245.

Hornsby, J., 2015. Causality and "the Mental". *Humana Mente,* Volume 29, pp. 125-140.

Hume, D., 1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals.* 3rd ed. Oxford: Clarendon Press.

Hursthouse, R., 1991. Arational Actions. *Journal of Philosophy,* 88(2), pp. 57-68.

Hyman, J., 2015. *Action, Knowledge, and Will.* New York: Oxford University Press.

Jackson, F., 1995. Essentialism, Mental Properties and Causation. *Proceedings of the Aristotelian Society,* Volume 95, pp. 253-268.

Kane, R., 1999. Responsibility, luck, and chance: Reflections on free will and determinism. *Journal of Philosophy,* 96(5), pp. 217-240.

Kane, R., 2016. Moral Responsibility, Reactive Attitudes and Freedom of Will. *Journal of Ethics,* 20(1), pp. 229-246.

Kapitan, T., 2011. Free Will as an Open Scientific Problem, by Mark Balaguer. *Mind,* 120(479), pp. 848-852.

Kenny, A., 1963. *Action, Emotion and Will.* New York: Routledge and Kegan Paul.

Kenny, A., 1975. *Will, Freedom and Power.* Oxford: Basil Blackwell.

Kim, J., 1976. Events as Property Exemplifications. In: M. Brand & D. Walton, eds. *Action Theory.* Dordrecht: D. Reidel Publishing Company, pp. 310-326.

Kim, J., 1989. The Myth of Non-Reductive Materialism. *Proceedings and Addresses of the American Philosophical Association,* Volume 63, pp. 31-47.

Kim, J., 1993. *Supervenience and The Mind: Selected Essays.* Cambridge: Cambridge University Press.

Kim, J., 1998. *Mind in a Physical World.* Cambridge: Cambridge University Press.

Kim, J., 2001. Mental Causation and Consciousness: The Two Mind-Body Problems for the Physicalist. In: G. Carl & B. Loewer, eds. *Physicalism and its Discontents.* Cambridge: Cambridge University Press, pp. 271-283.

Kim, J., 2005. *Physicalism, or Something Near Enough.* Princeton: Princeton University Press.

Lau-Tzu, 1993. *Tao Te Ching.* Indianapolis: Hackett Publishing Company.

Lemmon, E., 1967. Comments on The Logical Form of Action Sentences. In: N. Rescher, ed. *The Logic of Decision and Action.* Pittsburgh: University of Pittsburgh Press.

Lewis, D. K., 1973a. Causation. *The Journal of Philsophy,* Oct, 70(17), pp. 556-567.

Lewis, D. K., 1973b. *Counterfactuals.* Oxford: Basil Blackwell.

Lewis, D. K., 1986. *Philosophical Papers.* New York: Oxford University Press.

Lewis, D. K., 1994. Humean Supervenience Debugged. *Mind,* 103(412), pp. 473-490.

Lewis, D. K., 2000. Causation as Influence. *Journal of Philosophy,* 97(4), pp. 182-197.

List, C., 2014. Free Will, Determinism, and the Possibility of Doing Otherwise. *Noûs,* 48(1), pp. 156-178.

List, C. & Menzies, P., 2009. Nonreductive Physicalism and the Limits of the Exclusion Problem. *Journal of Philosophy,* 106(9), pp. 475-502.

Locke, J., 1975. *An Essay Concerning Human Understanding.* Oxford: Clarendon Press.

Lombard, L. B., 1986. *Events: A Metaphysical Study.* London: Routledge & Kegan Paul.

Lowe, E. J., 1998. *The Possibility of Metaphysics.* Oxford: Oxford University Press.

Lowe, E. J., 2000. Causal Closure Principles and Emergentism. *Philosophy,* 75(294), pp. 571-586.

Lowe, E. J., 2005a. Vagueness and Endurance. *Analysis,* 65(2), pp. 104-112.

Lowe, E. J., 2005b. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science.* Oxford: Oxford University Press.

Lowe, E. J., 2008. *Personal Agency: The Metaphysics of Mind and Action.* New York: Oxford University Press.

Lowe, E. J., 2013. Substance Causation, Powers and Humean Agency. In: S. C. Gibb, E. J. Lowe & R. D. Ingthorsson, eds. *Mental Causation and Ontology.* Oxford: Oxford University Press, pp. 153-173.

Mackie, J. L., 1974. *The Cement of the Universe.* Oxford: Clarendon Press.

Mackie, J. L., 1976. Causality and Determinism. *The Journal of Philosophy,* 73(8), pp. 213-218.

Marmodoro, A., 2013. Causes as Powers: Book Symposium on Stephen Mumford and Rani Lill Anjum: Getting Causes from Powers. *Metascience,* 22(3), pp. 545-559.

Mayr, E., 2011. *Understanding Human Agency.* New York: Oxford University Press.

McDermott, M., 2002. Causation: Influence versus Sufficiency. *The Journal of Philosophy,* 99(2), pp. 84-101.

McDonnell, N., 2015. The Deviance in Deviant Causal Chains. *Thought: A Journal of Philosophy,* 4(2), pp. 162-170.

Mckitrick, J., 2005. Are Dispositions Causally Relevant? *Synthese,* 144(3), pp. 357-371.

McKitrick, J., 2013. Getting Causes from Powers By Stephen Mumford and Rani Lill Anjum. *Analysis,* 73(2), pp. 402-404.

Melden, A. I., 1961. *Free Action: Studies in Philosophical Psychology*. London: Routledge & Kegan Paul.

Mele, A., 2003. *Motivation and Agency*. Oxford: Oxford University Press.

Mele, A. R., 1992. *Springs of Action: Understanding Intentional Behavior*. New York: Oxford University Press.

Mele, A. R., 2005. Action. In: F. Jackson & M. Smith, eds. *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press, pp. 78-88.

Mele, A. R., 2006. *Free Will and Luck*. Oxford: Oxford University Press.

Menzies, P., 2013. Mental Causation in the Physical World. In: S. C. Gibb, E. J. Lowe & R. D. Ingthorsson, eds. *Mental Causation and Ontology*. Oxford: Oxford University Press, pp. 58-88.

Menzies, P. & Price, H., 1993. Causation as a Secondary Quality. *British Journal for the Philosophy of Science,* 44(2), pp. 187-203.

Millican, P., 2007. Against the New Hume. In: R. Read & K. Richman, eds. *The New Hume Debate: Revised Edition*. London: Routledge, pp. 211-252.

Mill, J. S., 1843. *A System of Logic, Ratiocinative and Inductive Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Longmans, Green, Reader, and Dyer.

Mourelatos, A., 1978. Events, Processes and States. *Linguistics and Philosophy,* Volume 2, pp. 415-434.

Mulligan, K., Simons, P. & Smith, B., 1984. Truth-Makers. *Philosophy and Phenomenological Research,* 44(3), pp. 287-321.

Mumford, S., 2004. *Laws in Nature*. London: Routledge.

Mumford, S., 2009. Passing Powers Around. *The Monist,* 92(1), pp. 94-111.

Mumford, S. & Anjum, R. L., 2010. A powerful theory of causation. In: A. Marmodoro, ed. *The Metaphysics of Powers: Their Grounding and Their Manifestations*. New York: Routledge, pp. 143-159.

Mumford, S. & Anjum, R. L., 2011. *Getting Causes From Powers.* New York: Oxford University Press.

Mumford, S. & Anjum, R. L., 2013. Causes as Powers: Book Symposium on Stephen Mumford and Rani Lill Anjum: Getting Causes from Powers. *Metascience,* 22(3), pp. 545-559.

Nagel, T., 1986. *The View From Nowhere.* New York: Oxford University Press.

Noordhof, P., 1999. Micro-based properties and the supervenience argument: A response to Kim. *Proceedings of the Aristotelian Society,* 99(1), pp. 115-118.

O'Shaughnessy, B., 1980. *The Will.* Cambridge: Cambridge University Press.

Owens, D., 1992. *Causes and Coincidences.* Cambridge: Cambridge University Press.

Papineau, D., 1993. *Philosophical Naturalism.* Oxford: Blackwell.

Papineau, D., 2001. The Rise of Physicalism. In: C. Gillett & L. Barry, eds. *Physicalism and its Discontents.* Cambridge: Cambridge University Press. pp. 3-36.

Papineau, D., 2002. *Thinking About Consciousness.* New York: Oxford University Press.

Peacocke, C., 1979. Deviant Causal Chains. *Midwest Studies in Philosophy,* 4(1), pp. 123-155.

Pearl, J., 2000. *Causality: Models, Reasoning, and Inference.* New York: Cambridge University Press.

Pereboom, D., 2001. *Living Without Free Will.* Cambridge: Cambridge University Press.

Pettit, P., 1993. A Definition of Physicalism. *Analysis ,* 53(4), pp. 213-223.

Price, H., 2017. Causation, Interventionand Agency: Woodward on Menzies and Price. In: H. Beebee, C. Hitchcock & H. Price, eds. *Making a Difference: Essays on the Philosophy of Causation.* New York: Oxford University Press, pp. 73-98.

Prior, E., Pargetter, R. & Jackson, F., 1982. Three theses about dispositions. *American Philosophical Quarterly,* 19(3), pp. 251-257.

Psillos, S., 2002. *Causation and Explanation.* London: Routledge.

Quine, W. v. O., 1948. On What There Is. *Review of Metaphysics,* 2(1), pp. 21-38.

Quine, W. v. O., 1960. *Word & Object.* Cambridge(Massachusetts): MIT Press.

Raley, Y., 2007. The Facticity of Explanation and its Consequences. *International Studies in the Philosophy of Science,* 21(2), pp. 123-135.

Ramsey, F. P., 1925. Universals. *Mind,* 34(136), pp. 401-417.

Raz, J., 2010. Responsibility and the Negligence Standard. *Oxford Journal of Legal Studies,* 30(1), pp. 1-18.

Rea, M., 1998. Temporal Parts Unmotivated. *Philosophical Review,* 107(2), pp. 225-260.

Reid, T., 1788. *Essays on the Active Powers of Man.* Edinburgh: John Bell, Parliament-Square, and G. G. J. & J. Robinson, London.

Russell, B., 1912. On the Notion of Cause. *Proceedings of the Aristotelian Society,* Volume 7, pp. 1-26.

Ryle, G., 1949. *The Concept of Mind.* London: Hutchinson's University Library.

Salmon, W. C., 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

Sartorio, C., 2005. Causes as Difference-Makers. *Philosophical Studies,* 123(1/2), pp. 71-96.

Sattig, T., 2002. Temporal Parts and Complex Predicates. *Proceedings of the Aristotelian Society,* Volume 102, pp. 279-286.

Schaffer, J., 2007. Causation and Laws of Nature: Reductionism. In: T. Sider, J. Hawthorne & D. W. Zimmerman, eds. *Contemporary Debates in Metaphysics.* Malden: Blackwell, pp. 82-107.

Schmitt, F. F., 1983. Events. *Erkenntnis,* 20(3), pp. 281-193.

Schueler, G. F., 2009. Interpretative explanations. In: C. Sandis, ed. *New Essays on the Explanation of Action.* Basingstoke: Palgrave-Macmillan, pp. 112-132.

Sehon, S. R., 1997. Deviant Causal Chains and the Irreducibility of Teleological Explanation. *Pacific Philosophical Quarterly,* 78(2), pp. 195-213.

Setiya, K., 2011. Reasons and Causes. *European Journal of Philosophy,* Volume 19, p. 129-157.

Shoemaker, S., 1980. Causality and Properties. In: P. v. Inwagen, ed. *Time and Cause.* Dordrecht: D. Reidel Publishing Company, pp. 109-135.

Shoemaker, S., 2013. Physical Realization Without Preemption. In: S. C. Gibb, E. J. Lowe & R. D. Ingthorsson, eds. *Mental Causation and Ontology.* Oxford: Oxford University Press, pp. 35-57.

Skillen, A., 1984. Mind and Matter: A Problem which Refuses Dissolution. *Mind,* 93(372), pp. 514-526.

Skyrms, B., 1984. EPR: Lessons for Metaphysics. *Midwest Studies in Philosophy,* 9(1), pp. 245-255.

Smith, M., 2004. The Structure of Orthonomy. *Royal Institute of Philosophy Supplement,* Volume 55, pp. 165-193.

Smith, M., 2010. The Standard Story of Action: An Exchange 1. In: A. A. Buckareff & J. H. Aguilar, eds. *Actions: New Perspectives on the Causal Theory of Action.* Cambridge(Massachusetts): MIT Press, pp. 45-56.

Smith, M., 2012. Four Objections to the Standard Story of Action (and Four Replies). *Philosophical Issues,* 22(1), pp. 387-401.

Sober, E., 1984. Two Concepts of Cause. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,* Volume 1984, pp. 405-424.

Steward, H., 1997. *The Ontology of Mind: Events, Processes and States.* Oxford: Oxford University Press.

Steward, H., 2009. Sub-Intentional Actions and the Over-mentalization of Agency. In: C. Sandis, ed. *New Essays on the Explanation of Action.* Basingstoke: Palgrave Macmillan.

Steward, H., 2012. *A Metaphysics For Freedom.* Oxford: Oxford University Press.

Steward, H., 2013a. Processes, Continuants, and Individuals. *Mind,* 122(487), pp. 781-812.

Steward, H., 2013b. Responses. *Inquiry: An Interdisciplinary Journal of Philosophy,* 56(6), pp. 681-706.

Stout, R., 1997. Processes. *Philosophy,* 72(279), pp. 19-27.

Stout, R., 2010. What Are You Causing in Acting? In: J. H. Aguilar & A. A. Buckareff, eds. *Causing Human Actions: New Perspectives on the Causal Theory of Action.* Cambridge(Massachusetts): MIT Press.

Stout, R., 2016. The Category of Occurrent Continuants. *Mind,* 125 (497), pp. 41-62.

Strawson, G., 1987. Realism and Causation. *The Philosophical Quarterly,* 37(148), pp. 253-277.

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

Strawson, P. F., 1985. Causation and Explanation. In: B. Vermazen & M. B. Hintikka, eds. *Essays on Davidson: Actions and Events.* Oxford: Clarendon Press, pp. 115-135.

Sturgeon, S., 1998. Physicalism and Overdetermination. *Mind,* 107(426), pp. 411-432.

Tännsjö, T., 2009. On Deviant Causal Chains – no Need for a General Criterion. *Analysis,* 69(3), pp. 469-473.

Tanney, J., 1995. Why Reasons may not be Causes. *Mind and Language,* 10(1-2), pp. 103-126.

Tanney, J., 2009. Reasons as Non-Causal, Context-Placing Explanations.. In: C. Sandis, ed. *New Essays on the Explanation of Action.* Basingstoke: Palgrave-Macmillan, pp. 94-111.

Taylor, R., 1966. *Action and Purpose.* Englewood Cliffs(New Jersey): Prentice-Hall, Inc.

Thompson, M., 2008. *Life and Action: Elementary Structures of Practice and Practical Thought.* Cambridge(Massachusetts): Harvard University Press.

Tooley, M., 1977. The Nature of Laws. *Canadian Journal of Philosophy,* 7(4), pp. 667-698.

Tooley, M., 1990a. Causation: Reductionism versus realism. *Philosophy and Phenomenological Research,* Volume 50, pp. 215-236.

Tooley, M., 1990b. The Nature of Causation: A Singularist Account. *Canadian Journal of Philosophy,* 20(1), pp. 271-322.

van Fraassen, B. C., 1980. *The Scientific Image.* Oxford: Clarendon Press.

Velleman, D. J., 1992. What Happens When Someone Acts? *Mind,* 101(403), pp. 461-481.

Vendler, Z., 1957. Verbs and Times. *Philosophical Review,* Volume 66, pp. 143-160.

Vermazen, B., 1985. Negative Acts. In: B. Vermazen & M. B. Hintikka, eds. *Essays on Davidson: Actions and Events.* Oxford: Clarendon Press. pp.93-104.

von Wright, G. H., 1971. *Explanation and Understanding.* Ithaca: Cornell University Press.

von Wright, G. H., 1974. *Causality and Determinism.* New York: Columbia University Press.

Wang, Z., 2013. *Process and Pluralism: Chinese Thought on the Harmony of Diversity.* Berlin: Walter de Gruyter.

Whittle, A., 2008. A Functionalist Theory of Properties. *Philosophy and Phenomenological Research,* 77(1), pp. 59-82.

Williams, D. C., 1953. The Elements of Being: II. *Review of Metaphysics,* 7(2), pp. 171-192.

Wilson, J., 1999. How Superduper Does a Physicalist Need to Be? *Philosophical Quarterly,* 50(194), pp. 33-52.

Wilson, J., 2005. Supervenience-based Formulations of Physicalism. *Noûs,* 39(3), pp. 426-459.

Wilson, N. L., 1974. Facts, Events and their Identity Conditions. *Philosophical Studies,* 25(5), pp. 303-321.

Wittgenstein, L., 1953. *Philosophical Investigations.* Oxford: Wiley-Blackwell.

Wittgenstein, L., 1958. *The Blue and Brown Books.* Oxford: Blackwell.

Wolfson, B., 2012. Agential Knowledge, Action and Process. *Theoria,* Volume 78, pp. 326-357.

Woodfield, A., 1976. *Teleology.* Cambridge: Cambridge University Press.

Woodward, J., 2003. *Making Things Happen: A Theory of Causal Explanation.* New York: Oxford University Press.

Yablo, S., 1992. Mental Causation. *Philosophical Review,* 101(2), pp. 245-280.

Zagzebski, L., 1994. The Inescapability of Gettier Problems. *Philosophical Quarterly,* 44(174), pp. 65-73.