

Kent Academic Repository

Full text document (pdf)

Citation for published version

Van Aken, Erik (2018) Causal Interactionism: A New Agency Approach to Causation. Doctor of Philosophy (PhD) thesis, University of Kent, University of Kent.

DOI

Link to record in KAR

<https://kar.kent.ac.uk/69390/>

Document Version

UNSPECIFIED

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

Causal Interactionism

A New Agency Approach to Causation

Erik Van Aken

A thesis presented for the degree of
Doctor of Philosophy

University of
Kent

2018

Contents

1	Introduction	4
1.1	Background	4
1.2	Philosophical Claims	8
I	The Agency Approach	12
2	A Historical Sketch	13
2.1	Introduction	13
2.2	On The Notion of Cause	14
2.3	Collingwood, Gasking, and von Wright	18
2.4	Contemporary Agency and Manipulationist Theories of Causality	25
2.5	Interventionism	29
3	Prospects and Problems	38
3.1	Introduction	38
3.2	Prospects	39
3.2.1	Spurious Causes	40
3.2.2	Causal Asymmetry	43
3.3	Problems	45
3.3.1	Subjectivity	45
3.3.2	Scope	48
3.4	A Problem for Woodward Too?	49

II	Defending the Agency Approach	52
4	Overcoming Objectivism	53
4.1	Introduction	53
4.2	Two Kinds of Objectivity	55
4.3	Objectivism About Causation: An Example	60
4.4	The <i>Either-Or</i> Mentality	64
4.5	Objectivity in Modern Science	66
4.6	The Debate Between Woodward and Price	71
4.7	Price’s Perspectivism	72
4.8	Woodward’s Interventionism	78
4.9	Collingwood and Objectivity	83
5	Overcoming Foundationalism	87
5.1	Introduction	87
5.2	Foundationalism About Causation	88
5.3	Ney’s Foundationalism	90
5.4	Physical Facts and Difference Making	92
5.5	The Fate of Ney’s Foundationalism	95
5.6	Foundationalism and Interactionism	98
5.6.1	A Note on Fundamentality	103
6	Agency and Anthropocentrism	106
6.1	Introduction	106
6.2	Agency and Anthropocentrism	108
6.3	Anthropocentrism in Collingwood, Woodward, and Price	111
6.3.1	Collingwood’s Anthropocentrism	111
6.3.2	Anthropocentrism in Woodward	114
6.3.3	Price’s Anthropocentrism	115
6.4	How to Understand the Concept of Agency	119
6.4.1	Hypothetical Agency	123
6.4.2	Who’s an Agent?	128
6.5	Do Interventions Involve Agency?	130
6.5.1	Interventionism and Agency	130

III	Building a Coherent Agency Approach	136
7	Causal Interactionism	137
7.1	Introduction	137
7.2	How do Human Agents Interact with the World <i>via</i> Causal Models?	139
7.3	My Theory of Causal Interactionism	142
7.3.1	How does Causal Interactionism Relate to Other Manipulation Approaches?	143
7.3.2	How Does Causal Interactionism Differ From Other Manipulation Approaches?	146
7.4	The Metaphysics of Causal Interactionism	149
8	Prospects for Causal Interactionism	154
8.1	Introduction	154
8.2	Overcoming the (Analytic) Problem of Scope	156
8.2.1	A Solution to the Problem of Scope	164
8.3	The Experiential Problem of Scope	166
8.3.1	Gasking and Price	167
8.3.2	What Does Causal Interactionism Say about the Experience of Agency?	172
8.4	Causal Variation and Ascription	173
8.4.1	Moral Judgment	183
9	Conclusion	186

Chapter 1

Introduction

1.1 Background

Is there a constitutive relationship between agency and causation? This question seeks to determine whether or not causation needs to be analysed in terms of agency; thus, the relationship between agency and causation should be read in a particular direction—that is, agency is a constitutive element of causation. Many readers might think the answer to this question is obviously “no”. There are many cases where a causal relationship (i.e., a relationship between a cause and an effect) is entirely unassociated with human agency. Therefore, in order to refute the claim that causation is constituted by agency, it appears that one must only provide a simple counterexample. For example, we now know that the dinosaurs went extinct 65 million years ago because a large asteroid struck the earth near the Mexican peninsula. The dinosaurs’ unfortunate extinction is a clear example of a relationship between cause and effect, which appears to have nothing to do with human agency. Other examples include the 1989 San Francisco earthquake, the expansion of the universe, or tidal movements on Earth, to name a few. In other words, the claim that there is a constitutive relationship between agency and causation seems to be easily refuted by a mere counterexample.

It could also be argued that the claim that agency is a constitutive element of causation is outdated. The book of Genesis states that ‘God created the heavens and the earth’. In other words, God (i.e., an agent) *did something* that resulted in the creation of the heavens and the earth. Historical Christianity claims that God created by *fiat*—that is, God spoke and thereby

created. However, a scientific perspective might suggest that this claim is antiquated and instead cite the Big Bang as another example of an agent-independent causal relationship.

Yet, there are cases where a connection between agency and causation is rather clear. Tabloids are filled with stories about relationships that end because one of the partners was unfaithful. In these cases, we can reasonably claim that the cause was an action of the unfaithful partner. As another example, when I was a young boy, I once broke my arm playing tennis. My opponent hit a hard shot down the baseline, and I dove (rather carelessly) to return the shot. To this day, I consider the cause of my injury to be a direct result of my own action (i.e., diving).

While there are many other cases that could be used to support the claim that causation is related to agency, there are just as many cases—indeed many more—where a clear connection between agency and causation appears to be absent. The best one might do, in light of these examples, is claim that agency *can* be connected to causation but acknowledge that the relationship is of little or no consequence to a plausible theory of causation.

Interestingly, a so-called *agency theory* of causation, which originated in the middle of the twentieth century, was developed in the early 1990s. The agency theory is generally associated with the work of Huw Price. In a 1993 paper with Peter Menzies, ‘Causation as a Secondary Quality’, Price argues that causation is contingent on human agency, much in the same way that colour is contingent on the human visual system. Thus, Menzies and Price conclude that *there is* a constitutive relationship between causation and agency. Yet, with counterexamples so readily available, why do Menzies and Price defend the idea that agency is essential to causation?

The answer, as you will soon discover in this thesis, is rather complicated. However, there is something of a historical precedent for the idea that causation is related to agency—indeed, Menzies and Price did not conjure the idea from nothing.

There is a line in the philosophy of causation, traced through such thinkers as Hume, Mach, and Russell, which says that causation is not something that *really* exists or something that exists in terms of a mind-independent reality. For Hume, the causal relation was an idea (or habit) formed by the mind. A cause as a natural relation is defined as

...an object precedent and contiguous to another, and so united with it that the idea of the one determines the mind to form the

idea of the other, and the impression of the one to form a more lively idea of the other. (Hume, 2015, p. 28).

For Mach, when we speak of cause and effect,

...then we arbitrarily emphasize those aspects, the connections among which are the ones on which we have to focus, when we represent a fact from a certain perspective that is important to us. In nature there are no cause nor an effect. (Mach, 1901, p. 513)

A decade later, Russell would repeat Mach's claim, saying, 'there is no cause, nor an effect', because physicists had ceased to look for causes. According to Russell, physicists no longer looked for causes because fundamental physical laws were incompatible with the ordinary notion of cause. Russell concludes that causation is merely a 'relic' and that the notion of cause should be eliminated. In relation to the so-called law of causation, Russell famously remarks that

The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm (Russell, 1913, p. 1).

While Mach's view gives some minimal credence to causal notions in relation *to us*, his view that causal notions are in some sense perspectival would become an idea that later philosophers would appraise, as we will soon see. Moreover, the notion that causation is incompatible with physics is a hypothesis that contemporary philosophers of science have taken seriously. Yet, in light of the apparent incompatibility, philosophers have overwhelmingly rejected Russell's conclusion that the notion of cause should be eliminated. Why?

The answer can be traced to a new line in the philosophy of causation; this line of thought states that, while causation may *not* be located in the natural world (or reducible to fundamental laws), it *can* be located in human practice or human agency.

In her influential essay, 'Causal Laws and Effective Strategies' (1979), Nancy Cartwright argues that while Russell is correct regarding the so-called law of causation, Russell is incorrect to conclude that we can eliminate the

notion of cause. According to Cartwright, we need the notion of cause because it enables us to think about and pursue *effective strategies*. Effective strategies are methods that allow us to achieve certain *ends*. For example, one could prevent further cases of malaria by gassing swamps. While Cartwright does not explicitly defend the idea that causation is linked to human agency, her idea that causation is useful for pursuing effective strategies certainly makes the connection perceptible.

Indeed, Cartwright was not the first to notice the connection. In his book *An Essay on Metaphysics* (1940), R.G. Collingwood describes three “senses” of the word cause.¹ In the third sense, the *theoretical* sense, causation is taken to be an absolute relationship, which is made up of basic forces or powers that exist in nature independent of human agents. Much like Russell (though for different reasons), Collingwood rejects the theoretical sense of cause for being nonsensical. Yet, following Collingwood, one could reject the theoretical sense of cause but naturally maintain that causation was still important because, for example, one still relies on the notion of cause in sense one or sense two. Indeed, sense two, which Collingwood called the *practical* sense of cause, is still important for sciences like medicine and engineering, while sense one (i.e., the *historical* sense) is still important for historical enquiry.²

Notably, Collingwood describes the practical sense of cause as agent-relative. Although causes, as Mach pointed out, are perspectival, they are non-arbitrary, according to Collingwood, because they provide us with control. For Collingwood, causation is explicitly defined as agent-relative because a cause is something that allows an agent to manipulate nature and ultimately to achieve certain ends. For example, I could *prevent* scurvy by consuming vitamin C or I could *produce* fire by manipulating two pieces of wood.

Here, we find a plausible grounding for the agency theory of cause. Causation is (i) not something that exists at a fundamental level independent of us as human beings, and (ii) causation is important for us (or from our perspective) as human agents.³

¹Prior to Collingwood, Ramsey (1929) articulated the idea that the distinction between cause and effect could be grounded in deliberate action, i.e., human agency.

²In his influential book, *What is History?* (1961), the historian E.H. Carr actually suggests that the practical sense of cause is the important sense for historical enquiry.

³(i) and (ii) are similar to what Price and Corry (2007) call *causal republicanism*. A causal republican is someone who believes that, while causation is not a natural kind or

Grounding the agency theory in (i) and (ii) may provide the theory with some credibility. However, two obvious problems remain. First, there needs to be a way to handle the numerous counterexamples which indicate that causation is not connected to human agency. Second, relating causation to human agents or to the perspective of human beings with distinct ends or goals suggests that causation is subjective. I take it to be self evident and incontrovertible that a viable theory of causation should be able to capture both the scope of well-established scientific claims, such as the extinction of the dinosaurs, as well as a clear sense of objectivity. Indeed, these are both problems that threaten the agency theory of causation.

Lastly, even if one were to accept (i) and (ii) above, it is not evident that this entails a constitutive relationship between agency and causation. For example, in his book, *Making Things Happen* (2003), the philosopher James Woodward argues (like Russell) that the notion of cause seems to be incompatible with physics. Moreover, Woodward argues (like Collingwood) that causes are handles that allow us to pursue certain ends. Yet, Woodward explicitly disagrees with Price that causation is constituted by human agency, precisely on the grounds that it makes causation unduly subjective and limited in scope.

1.2 Philosophical Claims

The philosophical position defended in this thesis, Causal Interactionism, is meant to provide a consistent position on causality, whereby the causal relation is neither a physical, mind-independent feature of the world, nor is it arbitrary, subjective, or something that varies from individual to individual.

I claim that there *is* a constitutive relationship between agency and causation. I develop a novel approach, *Causal Interactionism*, whereby we can say that the causal relation involves an interaction between human agents and the physical world and thus that causation is constituted by agency. This approach allows us to overcome numerous problems associated with standard agency theories of causation (e.g., Collingwood [1940], Menzies and Price

something that is discoverable at the level of fundamental physics, it is, nonetheless, ineliminable. ‘Causal republicanism’, Price and Corry claim, ‘is thus the view that although the notion of causation is useful, perhaps indispensable, in our dealings with the world, it is a category provided neither by God nor by physics, but rather constructed by us’ (p. 2).

[1993]) and to clarify the precise relationship between agency and causation.

This thesis is organised around a specific debate, and the debate can be described in general terms—it is a disagreement between those who think that causation is agent-independent (i.e., unrelated to human agents) and those who believe that causation somehow depends on us as agents. The thesis follows an ongoing debate between two *manipulationists*, James Woodward and Huw Price. I also bring Collingwood into the debate because Collingwood created the approach, and much of what he has to say is still relevant (in the debate and my work).

While Woodward and Price agree that causes are handles that allow us to pursue certain ends, they have been debating the precise relationship between causation and agency for more than a decade. The debate between Woodward and Price can be summarised quite easily. On one hand, Price argues that causation is constituted by human agency.⁴ On the other, Woodward (2003, 2013) has argued that Price is mistaken—he claims that causation is not constituted by human agency. For Woodward, the causal relation is objective and agent-independent.

I claim to have resolved the debate between Woodward and Price. The details will emerge throughout this work. The resolution to the debate does not show a clear winner: *both* Woodward and Price are slightly mistaken. Although there are important differences between Woodward and Price’s accounts, in most places, I have tried to show that there is, in fact, a consensus between Woodward and Price and indeed a consensus between Woodward, Price, and Collingwood.

In Part 1, I will introduce the agency approach under its various forms and highlight the prospects and problems associated with it. I will situate the agency approach by investigating Russell’s (1913) arguments against the notion of cause, and argue that the agency approach should be viewed in light of Russell’s concerns (Chapter 2). In Chapter 3, I will argue that the agency approach is able to overcome two serious challenges associated with standard notions of causation, viz., the problem of causal asymmetry and the problem of spurious causes. However, I will also argue that the agency approach is open to two serious concerns, notably, the problem of scope and the problem of subjectivity.

Part 2 of the thesis offers a defence for the agency approach. In Chapter

⁴I’ll specifically focus on the work of Price, and not Menzies, because Price has continued to develop and defend the agency theory, whereas Menzies has not.

4, I argue against Michael Strevens, who holds that the causal relation is absolute and mind-independent. Against Strevens, I argue that the causal relation cannot be absolute because the causal relation is relative to what I call ‘frameworking’. The idea that the causal relation is non-absolute leads to a valuable discussion for the agency approach. If causation is relative, Collingwood’s claim that causation is relative to *human agency* will have increased plausibility. To show that the agency approach *can* and *should* be considered objective, I highlight the debate between Woodward and Price, which I take to be centred around the question of objectivity. By making a distinction between two kinds of objectivity—absolute objectivity and relative objectivity—I argue that the agency approach is objective in the relative sense. Indeed, I reveal a consensus between Woodward, Price, and Collingwood regarding causation and objectivity. They all agree that causation is objective in the relative sense and reject the idea that causation is objective in the absolute sense.

Chapter 5 includes an argument against Alyssa Ney, who defends foundationalism about causation. Foundationalism about causation is the thesis that the nature of causation is physical. Ney’s foundationalism allows for frameworking, but she maintains that what is important for the causal relation—that is, what is essential—is physical facts about the world. While Ney provides a fascinating insight into the relationship between the physical world and practices like frameworking, I argue that her account does not show what she intends. What Ney shows, contrary to her thesis, is that practices like frameworking are an *essential* element of the causal relation. Thus, we cannot claim, following Ney, that the nature of causation is physical.

In Chapter 6, I argue that Woodward’s interventionist theory relies on agency, and I investigate the relationship between agency and anthropocentrism. By creating a distinction between *overt* agency (e.g., deliberate bodily movements) and *mental* agency, which includes deliberate reasoning, I reveal that Woodward’s theory is distinct from Price’s agency theory in that it does not rely on the concept of overt agency. However, I further bring to light that interventionism does, indeed, rely on agency—notably, mental agency. Furthermore, by distinguishing between different forms of anthropocentrism, I exhibit a consensus regarding the manipulation approach and anthropocentrism while showing how the agency approach avoids an untoward anthropocentrism.

Part 3 of the thesis introduces my own version of the agency approach and explores its prospects. In Chapter 7, I bring together the work of Part

2 and develop my theory, Causal Interactionism. I defend the thesis that the causal relation involves an interaction between human agents and the physical world (and thus that the causal relation is agent-dependent). I further explore the metaphysical implications for Causal Interactionism and conclude that agency is a constitutive element of causation.

I use the final chapter of the thesis (Chapter 8) to highlight the leading prospects for Causal Interactionism. I argue that Causal Interactionism can overcome the problem of scope by taking into consideration facts about mental agency. I further uncover how Causal Interactionism has the tools to account for distinct concerns regarding the manipulation approach, e.g., how to explain causal variation (why the causal relation sometimes varies) and causal ascription (what we ascribe as causal).

Part I
The Agency Approach

Chapter 2

A Historical Sketch

2.1 Introduction

In this chapter, I will provide a historical overview of the agency approach. In the next chapter, I will explain the main prospects for the approach while also elucidating the most serious problems.

For now, the following questions are relevant:

1. What does an agent-dependent theory of causation look like?
2. Why does Price think that agency should be introduced into a theory of causation?
3. What have other philosophers said about the relationship between agency and causation?

My goal in this chapter is to answer these questions. In doing so, I will elucidate the historical precedent of the agency approach, which begins nearly fifty-five years before the work of Price. In §2, I will provide an overview of Russell's arguments for causal eliminativism. In §3, I will situate the agency approach by looking at portions of the philosophy of causation that followed Russell's (1913) essay. In many ways, the agency approach can be viewed as a rebuttal to Russell's conclusion that the notion of cause should be eliminated. In other words, I argue that the agency approach has a particular origin in an idea made famous by Bertrand Russell over a hundred years ago.

History will show us that the idea of an important relationship between agency and causation is better thought of as a general approach to causation

rather than a single theory. I will show, quite generally, that the motto for the agency approach can be paraphrased as follows: *agency is a vital component of causation*.¹ Many philosophers after Russell’s 1913 essay have explicitly accepted the proposition that *agency is a vital component of causation*, while many others have implicitly incorporated the idea into their work. In §4, I will highlight some of the most notable contemporary work on causation, which indicates that agency is related to causation. Of course, there are significant issues related to the idea that agency is related to causation, which I will explore in Chapter 3. Lastly, in §5, I will explain the interventionist approach to causation, highlighting some of the important similarities and differences with the agency approach.

But first things first—what does agency have to do with causation?

2.2 On The Notion of Cause

In 1913, Bertrand Russell published the essay ‘On the Notion of Cause’. In the article, Russell makes two significant claims: (1) the (ancillary) claim that the word “cause” had become so vague that philosophers should stop using it, and (2) that the “law of causality” cannot be grounded in our most advanced science, fundamental physics. Russell’s conclusion was that we should stop using (i.e., *eliminate*) the term ‘cause’ because fundamental science had shown it to be irrelevant, outdated, and non-existent.

The law of causality, which was Russell’s direct target in the essay, was thought (at the time) to be a deterministic law which stated that when a finite number of localised things happen at one time, some other localised thing would happen a short time later. To illustrate, if I strike a match, given that some limited number of conditions are satisfied, the law of causality states that there will be a flame just after the strike. What Russell intended to show by investigating the plausibility of a fundamental (physical) theory of causation was that the law of causation was false. Russell’s conclusion can be taken as the idea that a physical theory of causation cannot ground the relation between cause and effect and that, granting the physicalist assumption that our concepts should be reducible to physics, the notion of cause should be eliminated.

¹Going forward, I’ll use the term *approach* for the general theory. When I speak of the agency *theory*, it will be in line with a specific approach, e.g., Price (1993).

One of Russell’s secondary targets was causal realism. Russell wanted to show that from the perspective of our most advanced sciences, causation (by which Russell means the law of causality)—much like witches and the ether—*did not exist*. Advanced sciences, Russell says, had ‘ceased to look for causes’, because ‘in fact, there are no such things’ (p. 2). Thus, Russell’s 1913 position on causality was non-realist and eliminativist. Because the relation between cause and effect could not be grounded in physics, Russell thought we should eliminate the idea of causation much like we eliminated the concepts of witches and ether.

The Directionality Argument

One of Russell’s arguments concerns causal *asymmetry*, the so-called directionality of causation. The relationship between cause and effect is said to be *temporally* asymmetric—cause events (e.g., striking the match) always precede effect events (e.g., the flame) in time. In everyday life, we experience many temporal asymmetries that rely on the notion of cause; how we act and how we explain individual events lead us to say that causes and effects can be *contiguous* through time. We say that one event, the cause event, leads directly to another event, the effect event. Causation is asymmetric because causes precede effects, and effects do not precede causes.

Fundamental physical laws are symmetric. From the perspective of physics, the future explains the past just as well as the past explains the future. In other words, fundamental physical laws have the same character in both forward and backward temporal directions. The obvious problem for the “law of causation” is that causal asymmetry cannot be grounded by fundamental physical laws, which are temporally symmetric. So, given the symmetrical character of physical laws, Russell concludes that our notion of cause has no grounding in fundamental physics.

Russell’s argument works as follows. First, Russell makes a conceptual claim, that is, he assumes that the notion of cause must have a certain feature. In terms of the directionality argument, the feature in question is temporal asymmetry. Second, Russell attempts to show that fundamental physics cannot ground the feature in question. Russell concludes that fundamental physics cannot ground temporal asymmetry, and therefore fundamental physics cannot ground causal relations. Alyssa Ney (2009) correctly summarises Russell’s directionality argument as follows:

1. Causal relations are temporally asymmetric.

2. However, fundamental laws have the same character in both forward and backward temporal directions.²
3. Therefore, there is nothing in fundamental physics to ground temporal asymmetry.
4. Therefore, there is nothing in fundamental physics to ground causal relations.

Russell also points out that fundamental physics cannot capture the ordinary distinction between cause and effect, i.e., what counts as *the* cause of an event and what counts as *the* effect. For Russell, this is because our ordinary notion that two events are distinguishable within a time sequence makes no sense—or, at the very least, it becomes incredibly vague—from the perspective of fundamental physics. From the perspective of fundamental physics, our intuition that one event, the cause (e.g., striking the match), is clearly distinguishable from another event (e.g., the flame) is simply false. Fundamental physics can show that, in some sense, the two events are actually simultaneous (that is, they are indistinguishable within a certain time sequence).

The Localisation Argument

The principle of the same cause/same effect, which is articulated by philosophers such as Bergson and Mill, says that when a finite number of localised things happen at one time, some other localised thing will happen a short time later. The principle leads to a general *law of causation*: there are observable laws that tell us that given a cause A (e.g., striking a match), an effect B (e.g., a flame) will occur. Regarding the law of causation, here is a line from Mill:

The Law of Causation, the recognition of which is the main pillar of inductive science, is but the familiar truth, that invariability of succession is found by observation to obtain between every fact in nature and some other fact which has preceded it (quoted from Russell, 1913, p. 6).

²Hartry Field (2003) points out that there is no need to assume determinism for the argument to be sound. It could be the case that the laws of physics are indeterministic but have the same indeterministic character in both temporal directions.

And Bergson states:

Now it is argued that this [the law of causality] means that every phenomenon is determined by its conditions, or, in other words, that the same causes produce the same effects (quoted from Russell, 1913, p. 6).

Russell's objection to the principle of the same cause/same effect concerns our definition of 'events'. If we assume that the law of causation is true, if A nomologically produces B, Russell points out that we cannot define an event as localised to a limited number of conditions (as we ordinarily do). Rather, we must describe the event in such a way, as Russell says, to 'include the environment'. By 'the environment', Russell means any physical facts that influence a particular phenomenon, e.g., whatever physical facts affect the match prior to the strike. The problem with 'including the environment' is that the physical facts that affect the match prior to the strike amounts to a ridiculously large range of influence. Essentially, physics can describe how anything impacts an event, assuming that event is within a distance that can be reached by light in a finite period of time; this is often referred to as the *past light cone of an event*. What Russell notes is that when we include the environment into A, event A will become so large that it will never reoccur. If event A is unique, the law of causation is false. It cannot be true that (in ordinary language) "all As cause Bs" when A will never reoccur (A will never reoccur because the exact physical specifications of A are so vast that they are essentially unrepeatable).

For the law of causation to be true, Russell points out, we must define the event A by noting invariable uniformities, i.e., we must *localise* the event by abstracting away most of the physical influences on A. Russell calls this type of localisation 'an essential vagueness' (ibid, 8). Russell's contention with vagueness was based on a serious problem. When we abstract away from the multitude of physical influences on an event, we can never be sure that we have not abstracted away something *causal*, i.e., something that makes a difference to A. If we were to abstract a causal influence on A, the law of causation would fail (we cannot say that 'A causes B' in terms of the law of causation when some causal influence on A has been abstracted away). So, the only way to *guarantee* that we are not leaving out a genuine causal influence is to include the environment back into A, but we have already seen that this solution fails.

Thus, Russell presents a dilemma for the law of causation: either we include the environment into A, thereby making the event so large that it will never reoccur, or we abstract the seemingly irrelevant influences on A, making it uncertain whether B will occur. Either way, the law of causation is shown to be false. Russell concludes that the existence of causes of the kind that our ordinary concept requires cannot be grounded in fundamental physics and subsequently that we should eliminate the notion of cause.

Russell's conclusion gives rise to an important question. If causation is not a physical phenomenon, something that exists in nature or something grounded by our most advanced physics, then what is it? Russell's answer in 1913 was that causation is 'a relic of a bygone age,' which is ultimately outdated and obsolete.

Whether or not fundamental physics has made causation outdated and obsolete is a question that will be addressed throughout this thesis. To begin, let us consider some of the philosophy of causation that was done after Russell's critical essay, focusing on what I believe to be the inception of the agency approach to causation.

2.3 Collingwood, Gasking, and von Wright

The agency theory originated in 1940, with the work of R.G. Collingwood. As I mentioned in the previous chapter, prior to Collingwood, Ramsey had argued that the distinction between cause and effect was agency based. Ramsey's idea is one that agency theorists have embraced, as we will soon see. Generally speaking, Ramsey was a pragmatist about causation, or as Price (2017b) now argues, an expressivist with links to Hume (and in the contemporary landscape of expressivism, Simon Blackburn). The basic idea behind this brand of pragmatism is that concepts such as 'cause' must start with, and remain linked to human experience and inquiry. Nevertheless, Collingwood was the first philosopher to offer a systematic *theory* of causation that was based on human agency.

In 1940, Collingwood wrote what has become a classic text in the philosophical literature on causation, his controversial *An Essay on Metaphysics*. The book primarily addresses Collingwood's views on metaphysics—an aspect of philosophy that Collingwood believed was in need of serious reform.

Philosophy, for Collingwood, is a reflective activity. The task of philosophy, Collingwood thought, was to reflect on the fundamental principles or

absolute presuppositions on which knowledge rests. For example, an absolute presupposition of medicine is that understanding causes allows us to produce or prevent certain conditions. In fact, the subject matter of philosophy for Collingwood *was* the fundamental principles that underlie certain areas of knowledge and experience, such as medicine, natural science, and history—principles implicitly followed by individual practitioners, such as medical doctors.

At the time that Collingwood was writing *An Essay on Metaphysics*, A.J. Ayers' *Language, Truth and Logic* (1936) had been published, and neo-empiricism had become popular in the anti-metaphysical school of logical positivism. Logical positivists believed that statements were meaningful only if they could be verified empirically. Since the absolute presuppositions on which knowledge rests could not be tested empirically, Collingwood uses *An Essay on Metaphysics* to defend his conception of philosophy. To do so successfully, Collingwood would need to show that there are important propositions that are empirically unverifiable.

Collingwood thought that we could reform metaphysics by eliminating ontology and by making metaphysics the search for the fundamental principles, or presuppositions, on which knowledge rests. Collingwood calls fundamental presuppositions 'absolute' presuppositions. He defines absolute presuppositions as, '[a presupposition] which stands, relatively to all questions to which it is related, as a presupposition, never as an answer' (1940, p. 31). He goes on to explain what he means with the following example

Thus if you were talking to a pathologist about a certain disease and asked him 'What is the cause of the event E which you say sometimes happens in this disease?' he will reply 'The cause of E is C'...You might go on to ask: 'I suppose before so-and-so found out what the cause of E was, he was quite sure it had a cause?' The answer would be 'Quite sure, of course.' If you now say 'Why?' he will probably answer 'Because everything that happens has a cause.' If you are importunate enough to ask 'But how do you know that everything that happens has a cause?' he will probably blow up right in your face, because you have put your finger on his absolute presupposition, and people are apt to be ticklish in their absolute presuppositions. But if he keeps his temper and gives you a civil and candid answer, it will be to the following effect. 'That is a thing we take for granted in my

job. We don't question it. We don't try to verify it. It isn't a thing anybody has discovered like microbes or the circulation of the blood. It is a thing we just take for granted.'

According to Collingwood, we get 'ticklish' about our absolute presuppositions because they do not stand as answers to questions. As the above quote suggests, if you were to ask a pathologist why she believes that everything that happens has a cause, there is no sensible answer which she could give. As Collingwood says, it's simply something taken for granted. Collingwood argues that metaphysics should be the search for just these kinds of ticklish things—unverifiable assumptions held by certain individuals working within the context of a particular field of knowledge, e.g., pathology, history, or natural science.

Collingwood's idea is that to be a legitimate science, a *scientia* or body of knowledge requires a unique subject matter. Collingwood argues that metaphysics, understood as the science of ontology or pure being, lacked a unique subject matter and thus that metaphysics failed to be a science at all. Understood as the science of pure being, metaphysics possesses no subject matter of its own because, according to Collingwood, it attempts to study what exists without asking specific questions and without making any presuppositions. Legitimate sciences, such as the science of mind (or history; see Collingwood [1946]), and the science of matter (or natural science; see Collingwood [1945]), had their own unique subject matter. Because each science had its own unique subject matter, each science would bring different questions, presuppositions, and interests to bear on their individual fields. The historian is interested in the motives that guide our actions because the goal of the historian is to understand. The natural scientist is interested in constant conjunctions because the goal of the natural scientist is to predict. In order for either science to be possible, historians and natural scientists have to make different presuppositions about the nature of reality. Historians assume that man is rational; otherwise, they could not explain the past in terms of rational processes. The natural scientist assumes that nature is uniform; otherwise, they would not be able to make sense of the inductive generalisations on which their predictions rely.

After Collingwood describes his theory of reform, he uses the idea of causation as an example to demonstrate the concept of metaphysics as the investigation of absolute presuppositions.

Like Russell, Collingwood notes an inherent ambiguity in our notion of

cause. To resolve the ambiguity, Collingwood offers a description of causation based on three different ideas or ‘senses’ of the word ‘cause’; the *historical sense*, the *practical sense*, and the *theoretical sense*. Each sense of the word ‘cause’ corresponds to a different idea, where ‘idea’ refers to specific presuppositions that determined how certain practitioners think about cause and effect.

The Historical Sense of Cause

Collingwood attempts to show that the practical and theoretical senses of cause are derived from the historical sense, which is itself derived from the Latin *causa*, which meant ‘guilt’ or ‘blame’ (1940, 291). In the historical sense, ‘cause’ means ‘a free and deliberate act of a conscious and responsible agent’ (ibid, 290) and is typically intended to imply that A affords someone a motive for doing B. A clear example of this is found in the text of Genesis. When Adam discovers his nakedness, God asks Adam, ‘Who told you that you were naked?’ Adam replies, ‘The women whom you gave to be with me, she gave me from the tree, and I ate’ (Genesis 3: 11-12, New Standard American Bible). We discover the cause of Adam’s nakedness (according to Adam) to be the persuasion of Eve—Eve is the cause in the sense that she “made” Adam do it.³ In its original sense, a cause is essentially the action of an agent (human or supernatural) that motivates and compels another agent into action. The word ‘cause’ becomes analogous to words like ‘making’, ‘inducing’, ‘persuading’, ‘urging’, ‘forcing’, and ‘compelling’.

The Practical Sense of Cause

The practical sense of cause, which is the most widely discussed in the philosophical literature, concerns events in nature. In the practical sense, the word ‘cause’ means ‘an event or state of things which it is in our power to produce or prevent’ (ibid). This idea of causation is practical because it presupposes that we can control things in nature by manipulating their causes. Collingwood says that causation in the practical sense represents the Baconian idea that knowledge is power: if we know how to manipulate the cause in the right way, we will know how to control the effect. Collingwood is quick to point

³Ironically, the same reasoning is seen again in the next line. Genesis 3:13 reads: ‘Then the Lord God said to the women, What is this you have done?’ And the women said, ‘The serpent deceived me, and I ate’.

out that the practical idea of causation is relativistic. In the practical sense, a ‘cause’ is relative to an agent’s knowledge of ‘producing’ and ‘preventing’. The so-called principle of the relativity of causes (p. 304) says that, ‘for any given person the cause in sense II [the practical sense] of a given thing is that one of its conditions which he is able to produce or prevent’ (ibid).

To illustrate the principle of relativity, Collingwood describes the case of an automobile accident, where different agents with different skill sets and expertise try to determine the cause of the crash. Imagine, for example

[A] car skids while cornering at certain point, strikes the curb, and turns turtle. From the car-driver’s point of view, the cause of the accident was cornering too fast, and the lesson is that one must drive more carefully. From the county surveyor’s point of view the cause was a defect in the surface or camber of the road, and the lesson is that greater care must be taken to make roads skid-proof. From the motor-manufacturer’s point of view the cause was defective design in the car, and the lesson is that one must place the centre of gravity lower. (ibid)

In the above example, causation is relative to individual expertise. However, in many other cases, e.g., hunger, thirst, and fatigue, human agents will share the same ‘point of view’ about the causes of these events, and thus they will use the same manipulation strategies for producing and preventing. That is, most human agents produce or prevent things like hunger, thirst, and fatigue using similar manipulation strategies. So, for Collingwood, in some cases, causation is relative to human agents in general, while in other cases, causation is relative to a particular field of expertise (e.g., engineering, medicine).

A cause in the second sense represents what Collingwood calls a *handle* or a switch. A handle is something that gives the agent *power*. A cause is a handle because if we know how to “work the cause” in the right way, we can control the effect in some desirable form. A cause is related to human agency because, as we would manipulate a door handle, we manipulate a cause to ‘produce and prevent’. So, a cause in the second sense necessarily relies on human agency. In the second sense, Collingwood notes, ‘for a mere spectator, there are no causes’ (ibid, 307).

The Theoretical Sense of Cause

The theoretical sense of cause concerns an ‘unconditional’ cause or a cause as it would exist in nature apart from any human perspective. The theoretical sense of cause is meant to be *agent-independent*. In this sense, there can be no relativity of causes, and causes are not handles. Unlike the generality of cause in the second sense, the theoretical sense of cause is “tight”, meaning the cause will be indistinguishable from the effect. Collingwood makes reference to Russell’s ‘On the Notion of Cause’ and likewise has difficulty analysing the (theoretical) idea of cause within a time sequence. When we light a fuse and see a corresponding explosion, there is no handle, so to speak, as there was in the second sense that we deem “the” cause (e.g., lighting the fuse). How the actual cause then persists from one event to the other is a problem for the theoretical sense of cause.

Collingwood suggests that ‘cause’ in the third sense is empty. We assimilate talk of causes in the third sense by assuming nature to have a unique causal “power” or “force” by which we can describe the relationship between natural events. Collingwood suggests that this is merely an antiquated belief in animism or an inability to understand causality in non-anthropocentric terms. Like Russell, Collingwood thinks that modern scientific laws (e.g., Newton’s laws of motion) have made the idea of (theoretical) causation a relic. For Collingwood, the relic is not the law of causality but the animistic metaphor. Yet, unlike Russell, Collingwood does not allow the demise of causality in its theoretical sense to affect the notion of cause in the historical and practical sense. Since the idea of causality in the third sense is a derivative from the first and second senses, its elimination cannot jeopardise its earlier functions.

Interestingly, Collingwood agrees with only half of Russell’s conclusion: the idea of physical (agent-independent) causation is incoherent and false; however, given Collingwood’s conception of philosophy, Collingwood is happy to leave things as they are. We would not eliminate the idea that causes are handles, for example, because it is an absolute presupposition of medical science that causes allow us to produce and prevent events. Nor would we eliminate the idea that a cause can be a ‘free and deliberate act’ because it is a presupposition in law.

Causal Recipes

Another philosopher who argued that causation is related to human agency was Douglas Gasking. At the beginning of his essay ‘Causation and Recipes’ (1955), Gasking admits that we often speak of one thing causing another. But, he asks the important question: ‘in what circumstances do we do so’ (p. 479)? Gasking’s answer is that we can speak of one thing causing another in circumstances where we have learned *manipulation strategies* or what he calls ‘causal recipes’. Bodily movements can lead to the discovery that manipulating certain things in certain ways tends to make things happen. If I walloped the table with my knee, I might learn how to produce a large black spot on the carpet, e.g., by spilling the inkwell. Gasking explains, ‘men found out how to produce certain effects by manipulating things a certain way’ (ibid, 482). For Gasking, we speak of event A as causing event B when we have learned to produce B as a result of doing A. Gasking calls this a ‘producing by means of’ relation. Gasking says:

We learn by experience that whenever in certain conditions we manipulate objects in a certain way a certain change, A, occurs. Performing this manipulation is then called: “producing A”. We learn also that in certain individual cases, or when certain additional conditions are also present, the manipulation in question also results in another sort of change, B. In these cases the manipulation is also called “producing B”, and since it is, in general, the manipulation of producing A, in this instance, it is called “producing B by producing A”. For example, one makes iron glow by making it hot (ibid, 486).

According to Gasking’s definition, a cause is something that a human agent can overtly manipulate. Strictly speaking, causation is related to human agency because when we say that A causes B, A will be something that, by overtly manipulating it, allows us to produce B.

Explaining Causal Asymmetry

In, ‘On the Logic and Epistemology of the Causal Relation’ (1993), another agency theorist, von Wright, asks the following question: ‘What distinguishes cause from effects?’ He answers, ‘the fact that by manipulating A, we can bring about changes in B’. The idea is that if we think about the causal

relation from a *human perspective*, viz., the perspective of human agents, the problem of causal asymmetry does not arise. Asymmetry, von Wright explains, is simply “built in” to our conceptual experience of agency and time. Von Wright suggests from this fact that there is an implicit dependency (what he also called a ‘limitation’) of the notion of cause on the concept of agency and action (p. 123).

As we will see in later chapters (viz., chapter 3), the idea that we should think about causation from a human point of view, which is common to all of the agency theorists we have seen thus far, allows us to explain the asymmetry of cause and effect by grounding causal asymmetry in the asymmetry of human action or agency. Indeed, we will come to see that an essential component of Price’s agency theory is that the concept of cause is ‘situated’ in terms of human agency. In von Wright’s language, when we act, the effects of our actions lead to changes in the *future*. Thus, when we think of causation in terms of *acting* on A to produce changes in B, the asymmetry of action entails the asymmetry of the causal relation. The important point to note here is that we can allegedly overcome the problem of causal asymmetry, which is one of the two key problems that led to Russell’s eliminativism, by relating causation to human agency.

2.4 Contemporary Agency and Manipulationist Theories of Causality

Effective Strategies

In considering contemporary agency and manipulationist theories of causality, a good place to start is with Nancy Cartwright’s influential theory of *effective strategies*. While Cartwright herself is not an agency theorist, her ideas regarding effective strategies sheds important light on the relationship between agency and causation.

In her influential essay ‘Causal Laws and Effective Strategies’, Cartwright distinguishes two kinds of laws: laws of physics, which she calls ‘laws of association’, and ‘causal laws’:

Laws of association are typically causally neutral because they provide no account of what makes things happen. Causal laws, by contrast, have the word ‘cause’—or some causal surrogate—right in them (1979, p. 21).

Cartwright agrees with Russell's localisation argument, that is, causal laws cannot be derived from the laws of physics. Cartwright wishes to support the claim that causal laws cannot be derived from the laws of physics, but she wants to argue against Russell's claim that causal laws can be eliminated (much like Collingwood). For Cartwright, we need causal laws to distinguish between effective and ineffective strategies. An effective strategy is defined as a situation where the cause increases the probability of its effect. Effective strategies, which Cartwright thinks of as causal laws, can be discovered by 'partitioning' (i.e., localising) the space of possible situations to include the factors that are probabilistically relevant to the given strategy. Once we have partitioned our situations accurately, we will be able to determine effective strategies. Cartwright is subsequently able to claim that there is a 'natural connection between causes and strategies that should be maintained' (Ibid, 36).

Although implicitly, Cartwright is suggesting that causation is related to human agency, much like the agency theorists we have just considered, by relating causation to manipulation. An effective strategy is simply a manipulation strategy that allows us to produce or prevent some desired end or goal. As I said in the previous chapter, if I wanted to reduce the spread of malaria, an effective strategy would be to gas swamps. If I wanted to prevent scurvy, an effective strategy would be to consume vitamin C.

Price's Agency Theory

As explained in the previous chapter, the so-called *agency theory* of causality is primarily associated with the work of Huw Price. In his 1993 paper with Peter Menzies, 'Causation as a Secondary Quality', Price defends the thesis that causation is essentially related to human agency. For Price, A causes B means that an agent can 'bring about' B by bringing about (or doing) A. For example, one might say that throwing rocks (A) causes windows to break (B) because one can bring about B by bringing about A. According to this definition, Price is claiming that for a causal relationship to exist, there must be *intentional action*, or what he takes to be an occurrence of overt human agency.⁴

⁴I will explain the standard concept of agency in greater detail in Chapter 6. For now, it is important to note that an overt action is an action that involves *bodily movement*. Overt actions include raising an arm, bending a knee, throwing, catching, running, driving, etc.

Regarding the agency approach, it is also common for philosophers to reference two earlier papers by Price: ‘Agency and Causal Asymmetry’ (1992) and ‘Agency and Probabilistic Causality’ (1991). In his 1992 article, Price attempts to explain the asymmetry of causation—causes precede effects, and effects never or almost never precede causes—by introducing the notion of agency into a theory of causation. Because our actions affect the future and not the past, the asymmetry of causation, according to Price, can be reduced to the asymmetry of manipulation or human agency. In his 1991 paper, Price argues that intentional actions raise the probabilities of certain events—for example, by throwing a rock at your window, I increase the probability that it breaks. Thus, when we read that A causes B means bringing about B by doing A, we should interpret this in terms of probabilities: doing A raises the probability that we can bring about B.

In the contemporary literature on causation, the tendency to relate the agency approach with Price’s works (e.g., Woodward, 2003; 2013; Hausman, 1998) has led to the supposition that the agency approach relates to the work of Price alone; the only serious advancement and defence of the agency approach being seen as a selection of Price’s earlier works and some modifications over the last two decades (viz., Price 1991, 1992, 1992b, 1993, 1996, 2007, 2017). The sentiment in contemporary philosophy of science is that the agency approach states that the causal relation depends on overt actions or an occurrence of human agency (in other words, the agency approach is as Price defines it).

In ‘Causation as a Secondary Quality’, Price and Menzies (hereafter MP) define the relation between cause and effect in terms of a ‘means/end relationship’. A causal relationship exists if and only if (iff) bringing about A is an effective means of bringing about some end, B. Significantly, agency is explicitly introduced into MP’s theory of causation so that the causal relation can be defined in terms of agent probabilities: ‘Agent probabilities are to be thought of as conditional probabilities, assessed from the agent’s perspective under the supposition that the antecedent condition is realised *ab initio*, as a free act of the agent concerned’ (1993, pp. 187–190).

According to MP, causation is a ‘secondary quality’—that is, it is an external relation among events that is explained by its connection to the experience of agency. In much the same way that an apple looks red to an ordinary observer under standard conditions, ‘an event A is a cause of a distinct event B just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B’

(ibid). The benefits and problems associated with MP's notion of causation will be explored in the next chapter.

Causal Perspectivalism

Price's later work (e.g., 2007, 2017) is a slightly altered defence of the agency theory that was developed in the early 1990s. In his recent work, Price develops a much closer relationship with Ramsey (1929) and describes causation as *perspectival*.

Price has us imagine a world where time runs in the opposite direction. Price (2007) says that, 'the hypothesis [of a time-reversed universe] simply gives us an easy way to imagine the possibility that there might be creatures, elsewhere in the actual universe, whose time-sense is a mirror-image of ours' (p. 273). From our perspective, time for these creatures would run backwards from future to past. The time-reversed world would be *retrocausal*, i.e., causes would be preceded by effects. The important point to note is that agents in the time-reversed location would disagree with agents from our location about the causal relation: agents in our part of the universe would say that causes preceded effects, and agents in the time-reversed part of the universe would say that effects preceded causes. In the same way that our notion of foreignness is perspectival, or relative to a specific location—from Y's location, X's are foreigners, and from X's location Y's are foreigners—Price argues that our notion of cause is perspectival, or relative to our situation in time. In regards to the creatures who live in a world that we would call retrocausal, Price notes, 'their perspective would be as valid as ours' (ibid).

Like Ramsey (and von Wright), Price thinks he can explain the asymmetry of cause and effect by relying on facts about *the agents themselves*. The idea is that causal facts, whatever they are, partly depend on facts about human agents. Speaking of the connection with Ramsey, Price states:

Ramsey is famous as a pioneer of pragmatic subjectivism about probability. In one of his last papers, [Ramsey] extends this subjectivist viewpoint to laws and causation. He links the asymmetry of cause and effect explicitly to the perspective we have as agents, saying that, 'the general difference of cause and effect' seems to arise 'from the situation which we are deliberating'. He then goes on to identify what he seems to take to be the crux of the agent's perspective, namely the fact that from the agent's point of view, contemplated actions are always taken to be *sui generis*, uncaused

by external factors. As [Ramsey] puts it, ‘my present action is an ultimate and the ultimate contingency (2007, p. 281).

Price argues that causal chains begin *from the agent’s perspective* with a free act. He attempts to explain the direction of causation by appealing to the process of *deliberation*. Deliberation is the asymmetric process of reasoning about *future* events with the specific caveat that when *we* reason about manipulations, we assume things like the direction of time. Of course, agents in a time-reversed universe would argue that, from their perspective, we experience retro-causation. The right way to think about causation, they would say, is that effects always precede their causes. Causal perspectivalism is the idea that neither our intuitions nor the intuitions of the agents in the time-reversed universe are, technically speaking, false—the direction of causation is simply a matter of perspective.

2.5 Interventionism

The last philosopher that I want to highlight is James Woodward. Woodward defends a so-called interventionist theory of causation. The interventionist theory of causation is remarkably similar to the agency approach. One of the reasons that the two ideas are similar is that interventionists and agency theorists are motivated by a key purpose—the discovery of causal relationships. Another reason is that both theories are correctly regarded as manipulation theories, where causes are regarded as handles. The notable difference is that interventionism is claimed to be agent-independent. For example, Woodward says that

an “interventionist” approach [to causation] avoids a classical problem besetting manipulability theories—that of anthropocentrism and a privileged status for human action (2013, p. 14).

Woodward claims that ‘interventions’ (correctly construed) do not make reference to human agency and that interventionism, ‘is quite different from traditional agency theories (such as those of von Wright & Price)’ (2003, p. 103).

While Woodward has continually criticised Price’s theory for being unacceptably anthropocentric and subjective, I will argue that much of Woodward’s criticism of Price is misplaced and misguided (though not all of it).

Indeed, there are significant differences between Woodward’s theory of interventionism and Price’s agency theory. For now, I want to explain Woodward’s approach and situate it within the context of the earlier work done by Judea Pearl. Although I do argue (Chapter 6) that Woodward’s account represents an agency approach to causation, here one should simply note the details of interventionism, perhaps with a keen eye on the notion of ‘cause’ and ‘manipulation’ within or regarding the interventionist’s model. A helpful question to bear in mind as you read through the details of interventionism is the following: *Is Woodward’s non-commitment to human agency consistent with the manipulation approach that he prefers, viz., interventionism?* The answer to this question is quite difficult to discern when being provided with merely a surface reading of the relevant material. However, recall that for Price, Collingwood, Gasking, von Wright, and indeed even Cartwright, agency was (albeit implicitly) taken to mean *overt* action. We agents overtly *do* A and produce or prevent B. A key aspect of Woodward’s theory—and his non-commitment to human agency—relies on the supposition that agency means overt human action. Thus, what Woodward intends to show is that interventionism does not rely on *overt* agency.

More on this later. For now, let us move on to consider interventionism.

Causal Modeling

Judea Pearl’s (2000) book *Causality* is considered to be the inception of the interventionist approach. Therein, Pearl notes that the notion of ‘cause’ is a man-made concept. Indeed, Pearl believes that the concept began as an anthropocentric explanatory tool. Much like Collingwood’s historical sense, only God, people, and possibly animals could cause things to happen (since ‘cause’ is connected to something like ‘motivation’ and ‘will’). Natural events would only later enter into causal explanation. Pearl writes

The agents of causal forces in the ancient world were either deities, who cause things to happen for a purpose, or human beings and animals, who possess free will, for which they are punished and rewarded...this notion of causation was naïve, but clear and unproblematic (2000, p. 333).

According to Pearl, a problem for the concept of cause begins with the science of engineering. As humans began to build complex systems, physical objects began to acquire causal characteristics. When a system like a

catapult or a waterwheel broke down, ‘it was futile to blame God or the operator—instead, a broken rope or a rusty pulley were more useful explanations, only because these could be replaced quickly and make the system work again’ (ibid). Causal explanation was still regarded in terms of purpose (e.g., Aristotle’s notion of teleology), and engineering seemed to merely split the concept of cause into two directions: one being the old sense of credit and blame, and the new being a sense of flow or control. The split fits nicely within the work of Collingwood. The science of engineering moves us into ‘cause’ in the second sense.

Pearl sees the real trouble for the notion of cause beginning with the work of Galileo. Galileo wanted to put explanation aside and focus solely on *description*. Also, he believed that the fundamental nature of description was not going to be qualitative. Galileo thought we could describe nature with incredible accuracy using only mathematical equations. Galileo’s insight carries us into Collingwood’s theoretical idea of causation.

Pearl notes a similarity between Galileo and Hume. Hume was not concerned with explaining the nature of cause and effect. Hume, like Galileo, was concerned with description. Hume would only describe *how* we come to think and speak in causal terms. Pearl says that Hume was not involved with the *why* but with the *how*. Interestingly, Pearl notes a concern with Hume’s treatment of the *how*. If experience leads us to observe individual correlations, Pearl asks, ‘how do people ever acquire knowledge of causation?’ (ibid, p. 336). This question is what Pearl calls the ‘first riddle of causation.’ The ‘second riddle of causation’ asks, ‘what difference does it make if [we discover] that a certain connection is causal?’ (ibid, 337). Pearl’s somewhat obvious but no less important answer to the second riddle is that it makes a difference to how we *act*. Pearl’s theory will be set against the background of these two riddles.

Pearl’s Two Riddles of Causation

1. How do people acquire knowledge of causation?
2. What difference does it make if we know that a connection is causal?

Pearl is not a philosopher by trade. Pearl is an engineer, and he has a practical question in mind when thinking about the two riddles of causation. Pearl wonders:

How should a robot acquire causal information through interaction with its environment? How should a robot process causal information received from its creator-programmer? (ibid, 343).

Pearl argues that *causal modelling* can solve the second riddle of causation and that this strategy has the double advantage of making the first riddle much less formidable. Pearl would say that the motivation for the interventionist is the discovery of causal knowledge and that this knowledge is essential regarding human (or robotic) behaviour. Essentially, Pearl argues that causal knowledge is discoverable *via* interventionism and that we can learn about causation by causal modelling.

This solution includes three basic ideas:

- Treating causation as a summary of behaviour under interventions.
- Using equations and graphs as a mathematical language within which causal thoughts can be represented and manipulated.
- Treating interventions as a *surgery* over equations. (ibid, 344)

This sounds rather abstruse. Indeed, Pearl develops a method of calculus that he uses as a formal structure for representing the results of targeted interventions. While Pearl's so-called 'do calculus' is rather sophisticated, the intuitive idea behind interventionism is quite simple. At its core, interventionism is a system that allows scientists to model specific 'small world' systems (i.e., limited aspects of the world that we include into a model called 'variables'). Once we have a 'small world', investigators can make surgical changes to any variables within that system (e.g., by severing one functional link and replacing it with another or changing the value of individual variables) and observe the results. Once manipulated, a variable that produces changes in other variables can be thought of as causally related. Variables that produce no such changes can be thought of as not causally related (this is why Pearl thought we could treat causation as a 'summary of behaviour' under interventions). Consider an alleged causal relationship between the rooster and the sun. If we model the variables (where the arrow represents an alleged causal pathway) as

rooster \longrightarrow *rising sun*

we can “intervene” on the model by manipulating [rooster] (i.e., surgically cut out [rooster], leaving [rising sun] completely untouched). The model then becomes the following:

→ *rising sun*

In the case that we manipulated (cut out) the variable [rooster] and we observed no corresponding changes to the variable [rising sun], we can say that there is no causal connection between roosters and the rising sun. Admittedly, this model is overly simplistic. A better model would include more data, e.g., variables related to Kepler’s laws of planetary motion, the weight of the rooster, and the inverse square law for sound. This way, we could test whether it was somehow the physical presence of the rooster that affected the sun (e.g., if our rooster was [variable X] an 8 pound bird, what would happen if we changed variable X to 6 pounds?), or whether it was the rooster’s crow (e.g., take the sound intensity, I , of the birds crow Ix , where x equals the decibels ratio of a given intensity I relative to the threshold of hearing intensity, and see what happens if we lower x). On noticing the lack of change to variable [rising sun], we could be confident that there is no causal relationship to [rooster]. The results of the model might indicate that the two variables are effects of a common cause, e.g., the rotation of Earth.

Pearl claims that diagrams such as these capture ‘the very essence of causation’, which is the ability to ‘predict the consequences of abnormal eventualities and new manipulations’ (ibid p. 345). This leads us to what is likely the most significant outcome of Pearl’s theory—*deep understanding*. Deep understanding comes from ‘knowing how things will behave under new circumstances’ and gives us a feeling of being ‘in control’ (ibid).

There are more interesting details to Pearl’s theory, some of which we will consider in later chapters. For now, I want to highlight the work of Woodward.

Woodward’s Interventionism

When thinking about causation, Woodward believes that a useful heuristic is to ask what the notion is intended to contrast with. Not surprisingly, Woodward claims that when we employ causal terms in the special sciences and the context of ordinary life, ‘the relevant contrast is very often with mere correlations or associations’ (2007, p. 72). As investigators, we want

to determine genuine causal relationships so that these ‘might be exploited for purposes of manipulation and control’ (ibid). According to Woodward, the strategy that best allows us to achieve this is interventionism. Interventionism allows an investigator to discover effective strategies or what Woodward sometimes refers to as *difference-making* strategies. The idea behind difference-making strategies is that causal claims are understood as claims about what would happen (or what would be different) ‘under interventions on...one or more variables’ (ibid, p. 73). For Woodward, X is said to be the cause of Y iff under an appropriate intervention that changes the value of X , there is an associated change in the value of Y .

Woodward has four basic criteria that capture the idea of an intervention I on a given variable X :

1. I should cause (variation in) X (that is, I should have total control over X and determine its new value).
2. I acts as a ‘switch’ for all the other variables that cause X (that is, I cuts off X from any previous causal pathways).
3. Any directed path from I to Y must go through X (X will be ‘independent’ of any other causal influence or pathway).
4. I is independent of any variable Z that causes Y and is on a directed path from I to Y that does not go through X .

An intervention involves isolating the variable X (making X ‘independent’) from other ways that X and Y might be correlated. By turning off or ‘breaking’ the influences of certain variables within a complex system, we are essentially solving an inference problem—that is, we are discovering whether or not two variables are causally connected or whether they are merely correlated (like we did with [rooster] and [rising sun]). An intervention represents

An idealised experimental manipulation carried out on some variable X for the purpose of ascertaining whether changes in X are causally related to changes in Y ...[T]he idea we want to capture is roughly this: an intervention on some variable X with respect to some second variable Y is a causal process that changes the value of X in an appropriate exogenous way, so that if a change in the value of Y occurs, it occurs only in virtue of the change in

the value of X and not through some other causal route (2003, p. 94).

More simply:

Interventionist accounts take as their point of departure the idea that causes are potentially means for manipulating their effects: if it is possible to manipulate a cause in the right way, there would be an associated change in its effect (Woodward, 2007, p. 20).

Woodward describes interventionism as a counterfactual theory and claims that once we have discovered a causal pathway from X to Y , we can explain Y by citing X (in counterfactual terms, $\neg X \rightarrow \neg Y$). Indeed, interventionism is fundamentally a theory of causal explanation.

In his essay, ‘Causation With a Human Face’ (2007), Woodward describes interventionism as both *coarse grained* and *weak*. Interventionism is *coarse grained* because Woodward believes that causal explanations apply to high-level, macroscopic events. Thinking in terms of macroscopic (coarse grained) events, according to Woodward, is useful or relevant to sciences like economics, law, medicine, psychology (see Gopnik and Schulz, 2007), and to the context of ordinary life, and it is less useful (though not irrelevant) to microscopic, *fine-grained* sciences like fundamental physics; this idea is consistent with Collingwood’s endorsement of the practical sense of cause and his rejection of the theoretical sense of cause. Similarly, interventionism is *weak* because causal explanations do not amount to, nor do they specify, nomologically sufficient conditions. Causal explanations will include the kind of invariant generalisations that we described above, which Russell described as ‘vague’. So, what Woodward is attempting to offer is an explication and a justification of the types of (causal) explanations that we ordinarily use outside of physics.

The motivation behind the interventionist approach is causal discovery and causal explanation. Significantly, once we are in a position to determine a causal relationship, Woodward tells us, we are in a position to utilise causal explanation. Causal explanations support, what Woodward calls, ‘what if things had been different’ questions, that rely on a basic counterfactual theory of causal explanation.

This chapter explained several difficulties that plague the notion of causation if we investigate the notion from the perspective of fundamental physics.

As Russell pointed out, localisation and asymmetry pose serious, if not fatal, challenges to our common notion of cause. I have argued that the agency approach attempts to avoid the issues posed by investigating causation from the perspective of fundamental physics by instead investigating the notion of causation from the perspective of human agents.

We should note one important factor that agency accounts have in common and one that they do not. The most significant similarity is that they all describe what appears to be an agent-dependent approach of causation. Collingwood, Gasking, von Wright, Cartwright, Price, Pearl, and Woodward all attempt to avoid Russell's causal eliminativism by focusing on causation from the perspective of human agents (or at least, in Woodward's case, a macroscopic view that aligns with a human perspective). For Collingwood, causation cannot be eliminated because it is an absolute presupposition on which some practical knowledge rests. The idea, which is now orthodox to the manipulation approach to causation, is that causes act as handles because they allow us to do things in the world: doctors can cure diseases, and engineers can repair broken machinery.

An important difference in these accounts has to do with what can be called the *commitment to agency*, meaning whether or not an account of causation relies on human agency. Collingwood, Gasking, von Wright, and Price are all explicitly committed to human agency; Woodward and Cartwright are not. Instead, Woodward and Cartwright seem to rely on an *implicit* form of human agency. In Cartwright's case, it is clear that the idea of an 'effective strategy' involves some reference to human agency, i.e., an effective strategy is a manipulation technique that allows us to *do* something. In Woodward's case, it is much harder to locate a clear reference to agency. Often, this is because Woodward is openly critical of Price's use of agency, and he is vocal about the fact that a successful manipulation approach must be agent-independent. However, it may be the case that a manipulationist is unable to avoid *any* commitment to agency since causation is being described from the agent's perspective and causes are regarded as handles.

Later, I will argue (Chapter 6) that much of the confusion stems from ambiguity regarding the meaning of agency. As I explained earlier, most agency theorists implicitly take agency to mean *overt* agency—that is, some kind of bodily movement. However, it is not clear that agency is, in fact, limited to actions of this kind. Indeed, Woodward may be correct that his account of interventionism avoids a specific type of agency, e.g., overt actions. Yet, and I will argue in Chapter 6 that this is true, it may be

the case that interventionism relies on a different type of agency. In other words, Woodward's non-commitment to agency may turn out to be a non-commitment to one *type* of agency, where the non-commitment is still friendly to some other type.

Before all of the details can emerge, we must first consider in greater detail both the leading prospects for the agency approach and the most serious challenges associated with the idea that causation is agent-dependent. Having explained and situated the agency approach, we can now move on to discuss the prospects and problems associated with it.

Chapter 3

Prospects and Problems

3.1 Introduction

In this chapter, I will show that Price introduces the notion of agency into his theory of causation because he believes that, by doing so, we can improve standard evidential and probabilistic theories of causation by (i) distinguishing causes from correlations¹ and (ii) explaining causal asymmetry. By contrasting Price's approach with standard evidential and probabilistic theories, I will locate the benefits and prospects for Price's use of agency in a theory of causation.

Despite these alleged benefits, I will further argue that the introduction of agency into a theory of causation leads to two significant problems: the problem of subjectivity and the problem of scope. I will locate these problems in relation to Price's use of agency and further reveal how they are related to both the use of agency in Collingwood (1940) and the (implicit) use of agency in Woodward (2003, 2007).

My arguments will be structured around three questions:

1. How does the introduction of agency into a theory of causation help us improve standard accounts?
2. How does agency lead to the problem of subjectivity?
3. How does agency lead to the problem of scope?

¹This benefit relates specifically to causal *epistemology*. An important aspect of Price's theory, indeed an important aspect of all the manipulation theories thus far considered, is the ability to *know* that a relationship is causal.

In §2, I will highlight the leading prospects for Price’s use of agency, that is, the reasons why he thinks the notion of agency should be introduced into a theory of causation. §3 will be used to explain the most significant problems—subjectivity and scope—which seem to result from relating agency to causation by investigating the use of ‘agency’ in the work of Price and Collingwood. Lastly, I will use §4 to investigate whether or not the problems of subjectivity and scope arise in Woodward’s interventionist theory.

3.2 Prospects

Before I discuss the problems with Price’s agency theory, I want to further explain two of the reasons why Price thinks agency is vital to any plausible theory of causation.

When thinking about the work of Price, it is important to bear in mind that Price is attracted to the notion of agency because he is a pragmatist who thinks that we should investigate causation by investigating what we do and what we say as human creatures. Yet, Price’s argument, in his 1991 paper ‘Agency and Probabilistic Causality’ and in his 1992 paper ‘Agency and Causal Asymmetry’, is that the introduction of agency into a formal theory of causation can help us overcome long-standing problems associated with the notion of cause. Price claims that by introducing the notion of agency into a formal theory of causation, we can accomplish the following:

1. Overcome the problem of spurious causes and
2. Explain the asymmetry of cause and effect.

To put these problems in context, consider the highly influential theory of causation known as the *regularity theory*.² The Regularity theory is based on the work of Hume. For Hume, causation was nothing more than a habit of the mind. The ‘habit’ or ‘custom’, as Hume refers to it, is the process of believing ‘like events’ to have ‘like causes’. After we have observed two events to be constantly conjoined, we infer, by a habit of the mind, that they are

²There are various interpretations of the regularity theory [see Beebe, 2009; Noonan, H.W. 1999] which I do not intend to support or analyse in this brief overview of Hume. Rather, I intend only to express an important concern which was critical for Hume. The concern is epistemological—with what can and cannot be know with certainty, and, more critically, with what, if anything, justifies causal claims?

causally related. Constantly conjoined events are events that regularly occur in succession, that is, one regularly follows the other. These events include day and night, fire and heat, and rain and wet. After we have observed two events, e.g., rain and wet, as constantly conjoined, we come to call the first event ‘cause’ and the subsequent event ‘effect’; we say that rain causes wetness.

As I explained in the last chapter, Hume does not provide, nor is he concerned with providing, an *explanation* for what causation really is, apart from our psychological understanding of it. Hume offers a *descriptive* account of *how* we come to think and speak in causal terms. In everyday life, Hume thinks, we form beliefs on the basis of experience—in the case of causation, the experience of observing ‘constantly conjoined events’, where experience leads us to form the idea that two events, e.g., the flame and the sensation of heat, are causally related.

While being extremely influential to modern theories of causality, Hume’s account is not perfect. For example, Hume’s account does not address the problem of *spurious causes* or the problem of *causal asymmetry*. Spurious causes are two events that are in constant conjunction without one being the cause of the other. The crow of the rooster and the sunrise are constantly conjoined, and yet the rooster has no effect on the sun.

The problem of causal asymmetry concerns probability and regularity relationships. When two events are constantly conjoined or in regular conjunction, they are evidentially and probabilistically symmetric; two events can be causally related when they are regularly or probabilistically conjoined, and it does not matter which event comes first. For example, the evidential and probabilistic relationships are approximate for wet/rain and rain/wet and heat/fire and fire/heat. In other words, these relationships appear to be evidentially and probabilistically symmetric. The specific problem relates to the fact that we think causes and effects are asymmetric—generally, we think that causes raise the probability of later events, viz., effects.³

3.2.1 Spurious Causes

Regarding the case of the rooster and the sunrise, on a regularity account of causation, we have (experiential) evidence that the rooster and the sunrise are

³Admittedly, Hume’s account does not attempt to *explain* the asymmetry of cause and effect because Hume thought it was simply a matter of convention to call the earlier event ‘cause’ and the latter ‘effect’.

causally related—the past evidence or *causal history* of roosters and sunrises leads us to infer that they are causally related. We do not *know* that the rooster and the sunrise are causally related, we simply infer from their causal history (our experience of their conjunction) that they are.

Allegedly, one benefit of Price’s agency theory is that it gives us the ability to know whether or not two things are causally related. How might this be done? Price’s answer is that we can discover whether or not a relationship is *causal* by creating ‘independent causal histories.’ Notably, Price thinks that we can create independent causal histories by performing an action. To illustrate, we can act on the rooster by giving him sleeping pills. By observing the subsequent sunrise without the rooster, we have created an independent causal history, where one deprives roosters of their usual evidential bearing on rising suns (e.g., the rooster would no longer be in conjunction with the rising sun). The same can be done with many other cases, e.g., nicotine-stained fingers. Having nicotine-stained fingers is evidentially and probabilistically related to having lung cancer (likewise, in some cases the two events may be in conjunction). Yet, we could have smokers act, say by washing their hands after every cigarette or putting on gloves before every smoke, so they no longer had nicotine stains on their fingers. These actions would have no effect on the relationship between smoking and lung cancer. Significantly, our *actions* would be creating new causal histories, where one deprives the original events of spurious causes.

If we translate the case of the rooster and the sunrise into a case of so-called joint effects of a common cause, we can further reveal a benefit for Price’s use of agency. To illustrate, we know that the rooster and the rising sun are merely correlated with the rotation of the earth; they are, so to speak, joint effects of a common cause—joint effects of the rotation of the earth⁴ It just so happens that roosters typically awaken when the sun is about to appear over the eastern horizon.⁵ There is no causal relationship between the rooster and the sun—the rooster and the rising sun are merely correlated. Price’s idea is that conditional on the realisation of X , manipulation *via* a free act, e.g., causing the rooster to sleep in, will cause the correlation with Y to disappear. The correlation disappears because we *cannot* manipulate Y by acting *via* correlations, e.g., we could not manipulate the sun by getting,

⁴One might think that the *sunrise* causes the rooster to crow. But we know that the sun does not actually rise. It appears to rise because of the rotation of the earth.

⁵This is because the earth’s day/night cycle sets the roosters circadian rhythm.

or getting rid of, a pet rooster. This is why Price claims that the

common idea to agency accounts of causation is that an event A is a cause of a distinct event B just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B (1993, p. 3).

A stock example can further reveal the idea. A low-pressure system arrives at the weather station. First, the pressure system causes the barometer's needle to point to a low value. Later, it causes a rainstorm. There is a higher probability that there will be a rainstorm given that the barometer's needle is pointing to a low value. But barometers do not cause rainstorms. The *evidential* probability of a rainstorm is higher given that the barometer has a low value because the low barometer reading is typically a good indicator that a low-pressure system has arrived, which does *causally* raise the probability of rain.

The question that Price wants to address is this: How are we to distinguish between evidential-probability raising and causal-probability raising?

The introduction of 'agent probabilities' (Price, 1993, p. 190–91) is meant to distinguish between evidential relationships—which contain possible spurious causes—and causal relationships. Agent probabilities are 'the probability that should enter into calculations of a rational agent whose abilities consist in the capacity to realise or to prevent C, and whose goal turns entirely on E...the agent probability that one should ascribe to B conditional on A (which we symbolize as ' $P_{\underline{A}}(\underline{B})$ '), is the probability that B would hold were one to choose to realize A' (1993, p. 190–1). Agent probabilities thus measure the degree to which cause C is an effective means to bringing about effect E. Price says that, 'to say that A is an effective means by which an agent could achieve B is to say that if the agent were to have an overriding desire that B should obtain, then an adequate rational decision theory would prescribe realizing A rather than $\neg\underline{A}$...In other words, A constitutes a means for achieving some end B just in case $P_{\underline{A}}(\underline{B})$ is greater than $P_{\neg\underline{A}}(\underline{B})$ ' (ibid, p. 191). Price believes that agent probabilities are helpful because they 'abstract away' from the evidential import of an event. Significantly, by abstracting away from the evidential import of an event, we have the ability to break spurious causes.

3.2.2 Causal Asymmetry

In ‘Agency and Causal Asymmetry’, Price asks:

What feature of our perspective could it be that manifests itself in the cause-effect distinction? (1992, p. 513)

Price’s response is that,

The answer may lie in the agency or manipulation theory of causation in general. The core of this theory is the view that we acquire the notion of causation in virtue of our experience as agents. Roughly, to think of A as a cause of B is to think of A as a potential *means* for achieving (or making more likely) B as an *end* (1992, p. 514).

Following Price here, we can describe the asymmetry of cause and effect as agency based. As agents, we know that our actions can only effect future events, and it is through this fundamental experience—of doing one thing *now* (in the present) and effecting something later on (in the future)—that we come to think and speak in causal terms. This is to say, causal reasoning comes to reflect the asymmetry of action and manipulation.

So how might ‘our experience as agents’ help us to make sense of causal asymmetry? Consider the problem in terms of probability and regularity. There are many cases where C is the cause of an effect E, and E is not the cause of C. Alternatively, there are no cases where C is the cause of E, and E is the cause of C. These facts reflect our intuition that causation is asymmetric. However, standard probability and regularity approaches to causation suggest that there *will* be cases of symmetric causation (here I will only speak of probability, but the same results will follow for regularity approaches as well).

Consider, for example, that turning the ignition on my car raises the probability of the engine starting. This is not problematic. However, it is also true that starting my engine is more likely preceded by turning the ignition than by using jumper cables or hitting the engine with a hammer. According to a standard probability theory, when the engine is running, there is a high probability that the ignition has turned. This means that we can switch the places of C and E. Probability raising runs equally from C to E and from E to C. In terms of probability, the relationship appears symmetric.

The common “solution” to the problem of causal asymmetry is not to *explain* the source of asymmetry but to simply stipulate that causes must precede their effects. This quasi-solution makes the asymmetry of causation a matter of convention. Price considers the alternative possibility of explaining asymmetry in terms of some further modal notion, for example, in terms of counterfactuals. But this just results in what Price calls ‘conceptual buck-passing’ (1992b, p. 254). Conceptual buck-passing is ‘the mistake of appealing to some notion whose own temporal asymmetry and orientation is no less problematic than that of causation itself’ (ibid). For Price, conventionalism and conceptual buck passing is unhelpful because it does not *explain* the problem of asymmetry, and this is precisely what Price wants to do.

Price thinks that human agency can provide an explanation for causal asymmetry. Ramsey, remember, argued that the asymmetry of cause and effect was linked to the fact that we are agents. Ramsey says, ‘from the situation when we are deliberating seems to...arise the general difference between cause and effect’ (Ramsey, 1929). Following Ramsey, Price argues that when we emphasise the agent’s perspective, the unique perspective of ‘deliberation’, we can explain or ground causal asymmetry. The asymmetry of our modal notions (here causality) can be explained by something actual, viz., ‘*our own constitutions*’, as Price says (1992b, p. 263). Thus, Price argues the distinction between cause and effect turns out to be anthropocentric.

To illustrate, consider the notion of deliberation. When we deliberate, we reason about the effects of an action, ϕ . Given our epistemic constitutions—our knowledge of the past, our limited knowledge of the future, and our location in a future-directed world—when we deliberate about an action, we consider our action to be the *means* to a later and desired *end*. Thus, a free act, which always occurs before some desired effect, is considered to be a cause. What results from the free act is considered to be the effect. For Price, this is why the asymmetry of cause and effect is grounded by our perspective as deliberative agents.

According to Price, there is no *de re* sense of the causal distinction—that is, nothing “in the world” that aligns to our talk of causes and effects. Causal asymmetry is simply something that reflects the agent’s perspective. As agents, we take our actions to be statistically independent of everything but their effects. Given that this ‘experience of agency’ includes *our* temporal arrow, how we experience time, we know that our actions raise the probability of later events *and* that our actions cannot raise the probability of earlier

ones. It is not just that we *say* causes precede their effects. The distinction naturally arises for us as situated agents.

So, given that we take into consideration certain facts about human agency, e.g., our epistemic constitutions and our situation in time, we can explain the cause-effect distinction: a cause, as a free act, always occurs upstream of some desired effect. The same is true for the problem of spurious causes. *Given* that we consider certain facts about human agency, e.g., that free actions have the ability to create new causal histories, we can discover whether or not a certain relationship is causal (and thus solve the problem of spurious causes). (Von Wright agrees. Von Wright (1975) claims that the asymmetry of causation consists in the general truth that an event can be manipulated by its causes, but never by its effects. I can, for example, ‘imagine’ controlling floods by controlling rain, but I cannot, however, imagine controlling rain by controlling floods.)

3.3 Problems

Despite the benefits of introducing agency into a formal theory of causation, relating causation to agency leads to serious worries. The most serious, those which make up the most common objections to the agency approach, are the claims that agency leads to an untoward subjectivity about causation and that it severely limits what we are allowed to call causes (or the scope of causation). I will begin with the problem of subjectivity.

3.3.1 Subjectivity

To understand the notion ‘subjective,’ one must understand the notion ‘objective.’ There are many ways to think about objectivity. Usually, something is taken to be objective if it is mind-independent—that is, something is objective if it is *viewed from nowhere*. Mind-independence means *aperspectival* and is often thought to be a hallmark of good science (in that objectivity, as mind independence, constitutes ‘accurate representations’ of the external world. See Daston and Galison [2007]). The products of good science, it is further claimed, should be *judgment free*, that is, free from human experience, beliefs, or opinions. So, generally speaking, something is said to be subjective if it is perspectival or if it relies on human judgment.⁶

⁶I will elaborate on this and other ideas of objectivity in Chapter 4.

It is not difficult to understand how the agency approach can be seen as subjective, given the idea of objectivity as a perspectival or mind-independence. We have already considered how manipulationists like Collingwood, Price, and Woodward are anthropocentric in the sense that causation relates to *our* perspective as human agents. Furthermore, I argued that the agency approach is best developed in light of Russell's *rejection* of mind-independent causes. As we saw in chapter 2, a hallmark of the agency approach to causation is the belief that there are no mind-independent or agent-independent causes. Indeed, Collingwood argues that causation is relative to what an individual can 'produce and prevent', and Price argues that causation is relative to what a human agent can 'bring about.'

To illustrate the relativity of causation once again, consider a remark from Collingwood. He says:

Suppose that the conditions of an event included three things, a, b, c; and suppose that there are three persons A, B, C, of whom A is able to produce or prevent a and only a; B is able to produce or prevent b and only b; and C is able to produce or prevent c and only c. Then if each of them asks, 'What was the cause of y?' each will have to give a different answer. For A, a is the cause; for B, b; and for C, c (1940, p. 304).

Following this example, we could say that causation is subjective in the sense that what causes what depends on what an individual agent is able to produce and prevent. Price is explicit that causation is subjective in a similar manner. He says:

Suppose that the world had developed in such a way that we had fewer manipulative abilities and skills than we actually possess but that we still apply our concept of causation roughly in conformity with the agency approach. In this case the reference of the expression 'relation between events such that an actual agent could manipulate one event as a means to bringing about the other' would have been fixed on different relations, even though our way of fixing the reference would have been the same (Menzies and Price, 1993, p. 99).

According to Price, causation is relative to the manipulative capacities of agents. Whereas Collingwood seems to suggest that causes are relative to

an individual agent, Price thinks causation is relative to a group of *creatures* who share the same manipulation capacities.

When we think about causation being subjective in either of the two ways just described, it seems to go against our ordinary intuitions. Intuitively, what causes what seems to be objective and factual. Yet, if we follow Collingwood's principal of relativity, this cannot be the case. Recall the example of the automobile accident. From the car driver's point of view, speeding is the cause of the crash, because this is what the driver can control. However, from the county surveyor's point of view, the surface of the road is the cause of the crash, because surveyors have control over the condition of roads. And for the motor-manufacturer the design of the car is the cause, again, because this is what they can control.

Following Collingwood, we could say that the agency approach represents a form of epistemic relativism. You are justified to your belief B^* based on what you know how to produce and prevent; and I am justified in my belief B based on what I know how to produce and prevent.

Even if we can grant that theoretical, or "real", causes are nonsensical, it seems extremely counterintuitive (and highly undesirable) to accept this form of subjectivity. It is an essential desideratum of any viable approach to causation that it provides some sense of objectivity. In similar (and often more important) cases, e.g., the causes of cancer, autism, and the origins of the universe, we need to be able to distinguish between contradictory beliefs of the type B and B^* .

Furthermore, Price's claim that had our 'experience of agency' been different (that is, had we developed some alternative or fewer manipulative capacities) we would have different answers to questions about causation is counter-intuitive and unduly subjective. If this were true, it would mean that agents with different capacities to our own will rationally disagree about causal relationships and that there would be no objective basis to distinguish between them. As Jon Williamson points out:

This looks to be a problem not just across possible worlds but across agents in this world. Just as the capacities of a human, a robot, and a Venus Fly trap differ, so too would causality-for-a-human, causality-for-a-robot, and causality-for-a-Venus-Fly-trap (2007, p. 119).

The problem with agency accounts is that causation becomes subject to individual understanding (in Collingwood's analysis) or to what an individual

is capable of doing (in Price's analysis), and there appears to be no objective fact of the matter about what causes what.

3.3.2 Scope

The problem of scope says that the agency approach unduly limits what we are allowed to call causes. Indeed, Price's account makes causation subject to what an agent can 'bring about' or to our 'unique experience' of agency (1993). Woodward has been an open critic to these limitations:

If the only way we understand causation is by means of our prior grasp of the experience (or notion) of agency, then we face the obvious problem about the extension of causal notions to circumstances in which the relevant experience of agency is unavailable (2003, p. 123).

Again, we are unable to answer some essential questions. For example, it is not clear how we should assess the 'agent probabilities' for (what seem to be) uncomplicated facts. What is the 'agent probability' for the claim that a star's mass caused it to become a supernova? What is the 'agent probability' that a large asteroid caused the dinosaur's extinction? There does not seem to be any straightforward answers to these questions, and yet we know that explosions in space and extinction events have causes.

These difficulties translate to the problem of scope—that is, what are we allowed to call causes on the agency theory model? The problem for the agency theorist is that, by defining a cause as something we can overtly manipulate or bring about, there seems to be many causes where we know that C causes E but human agency (i.e., overt manipulation or bringing about) does not or cannot apply to C. Thus, we unduly limit the scope of what we are allowed to call causes. Here, again, is apt criticism from Woodward:

To use [Menzies and Price's example] what can it mean to say that, "the 1989 San Francisco earthquake was caused by friction between continental plates" if no one has (or, given the present state of human capabilities, could have) the direct personal experience of bringing about an earthquake by manipulating these plates? (2003, p. 124)

Thus, according to Woodward, limiting causality to the manipulative capacities of human agents unduly limits what we can call causes.

The agency approach, by defining causality in terms of manipulation and control, clearly seems to unduly limit the scope of our causal indicators. In other words, it does not seem practical to say that *all* causal claims will relate to human agency. Even if it is true that our concept of cause *developed* from our experience of manipulation, it appears to be much too rigid to demand that the only appropriate bearing on our causal judgments is agency. By gauging our causal judgments solely on the basis of agency, we risk the possibility of limiting the scope of causal indicators to things that we can manipulate and control, and this seems highly undesirable.

Indeed, there are many cases where we might appeal to further indicators, such as causal mechanisms or causal processes. Wesley Salmon (1998) and Phil Dowe (2000), for example, argue for a physical interpretation of causality, one that can identify causal relationships with causal processes like energy transfer or what Salmon calls ‘mark transmissions.’ Also, it would seem in cases like the expansion of the universe and the San Francisco earthquake that we *would* be better off explaining some facts in terms of physical processes, such as energy transfer, rather than facts about human agency.

3.4 A Problem for Woodward Too?

Woodward has been the most vocal critic of Price’s agency theory. According to Woodward, Price’s use of agency makes causation unduly subjective. But how much of the subjectivity does Woodward avoid? I have already suggested that Woodward may be incorrect that his theory of interventionism is agent-independent. For Woodward, causation is relative to what we include into a causal model, to our interests, and to assumptions about serious possibilities.

For Collingwood, we could say that causation co-varies with our manipulative know-how; for Price, causation co-varies with our manipulative capacities. Both kinds of variation lead to subjectivity. Interventionism is meant to avoid the subjectivity by ‘idealising’ causal models and experimental manipulations. For example, an idealised causal model would include what both of us know about car accidents. Note that this still makes causation relative to certain epistemic conditions. If *you* were to model the state of the accident (i.e., you were the “scientific investigator”), you might include things like the condition of the road and the speed of the vehicle. You might not include

any details of the car's design because you lacked the appropriate knowledge. Likewise, when scientists create models, idealised or not, they have no way of knowing, just like you and I, that they lack some type of relevant knowledge. In a way, this brings us back to Russell's localisation argument. When we limit what we include into the definition of an event or a causal model, we can never be *certain* that we have not left out something causal or, more to the point with causal models, that we have neglected some (unknown) variable that makes a difference to the effect.

Woodward understands these limitations. Indeed, interventionism is characterised as 'weak' and 'incomplete' because the best we can do, Woodward thinks, is describe coarse grained, invariant generalisations. Woodward says that

The variables involved in [coarse grained] generalizations, *shattering, non-shattering, window breaks, window does not break...* take their values across extended spatio-temporal regions with imprecise boundaries...In addition, coarse grained variables may fail to partition the full possibility space as seen from the point of view of an underlying fine-grained theory (2007, p. 81).

What Woodward is getting at, I presume, is that interventionism can avoid *some* of the relativism and subjectivity that we saw in Collingwood and Price but not all of it. For example, an idealised experimental manipulation could avoid the subjectivity of the automobile accident *given* that we have the appropriate knowledge. Yet, given that idealised experimental manipulations *are* relative to knowledge, human interests, and to judgments about serious possibilities, interventionism does not altogether avoid the problem of subjectivity. Simply stated, interventions do not qualify as objective in the sense of mind-independence.

A possible counter argument for Woodward would be that, while interventionism is friendly to a type of subjectivity, the subjectivity is not untoward. Indeed, Woodward's argument against Price and the agency approach, as I suggested in the last chapter, seems to specifically relate to the concept of overt action. In other words, Woodward would likely respond by claiming that interventionism is subjective according to a certain specification, e.g., mind-independence, but that interventionism is not subject to the specific limitations that arise for the agency approach, e.g., making sense of the causal

relation in terms of overt action.⁷

Indeed, in an extended debate with the philosopher Michael Strevens, Woodward (2008) explicitly argues that his theory of interventionism is *not* intended to specify the nature of causation, or describe the causal relation from an absolute, mind-independent perspective. When Strevens (2007) argues that Woodward's theory of interventionism fails because it is *relativistic* (e.g., interventionism is relative to variable sets and judgments regarding serious possibilities), Woodward argues that Strevens misrepresents his account by framing interventionism as a metaphysical analysis of causation. Woodward's response to Strevens is twofold: (i) interventionism is not meant to be a metaphysical or absolute theory of causation, and (ii) it does not follow from (i) that interventionism is radically subjective. Woodward (2007, 2008) is quick to point out that a rejection of absolutism (what is sometimes referred to as 'objectivism') does not entail a radical subjectivity. Hence, Woodward's debate with Strevens indicates that he does view the agency approach as *more* subjective, indeed radically subjective, in contrast with his own theory.

Happily, Woodward's argument and his conflict with Strevens opens a host of interesting questions for the agency approach. Is Woodward correct that his theory of interventionism escapes a radical and untoward subjectivity to which the agency approach succumbs? Is Strevens correct that *any* subjectivity is untoward and therefore preferably avoided?

In the following chapter, I will elaborate on Strevens' position and gauge the prospects for an absolute, i.e., non-relativist, approach to causation.

⁷I will argue for this position and attempt to show that Woodward supports it in Chapter 4.

Part II

**Defending the Agency
Approach**

Chapter 4

Overcoming Objectivism

4.1 Introduction

In this chapter, I will argue that objectivism about causation—the idea that causation is objective if it is frame-independent or absolute and radically subjective if it is not—fails by showing that it is impracticable, that it rests on an invalid assumption, and that it is inimical to modern science. I will achieve this by answering three questions:

1. What is objectivism about causation?
2. What is the *Either-Or* fallacy?
3. Does objectivism mean objectivity?

My arguments in this chapter will be focused on causation and will have a significant impact on the prospects of the agency approach to causation. If objectivism about causation is correct, the agency approach will fail for the reasons detailed in the previous chapter, viz., because it makes causation unduly subjective. However, if objectivism about causation is unsuccessful, we will be able to view the relativity of causation, which is a consequence of the agency approach, as non-problematic. Still, I will need to address critical questions about the agency approach—specifically, questions regarding whether or not (and if so, how) the problems of scope and subjectivity can be solved. Significantly, my arguments against objectivism, if they are successful, will allow us to investigate new possibilities. It is my belief that

there is a viable solution to the problem of scope and the problem of subjectivity. To support this claim, I develop a two-pronged approach. First, I reveal how many of the objections regarding agency, subjectivity, and scope rest on misguided intuitions, fears, and assumptions (Part II of this thesis). The second prong of my approach is to develop a viable agency approach that overcomes both problems of subjectivity and scope (Part III). My theory of Causal Interactionism will be presented as an agency approach that can overcome the problems of (i) subjectivity and (ii) scope.

In this chapter, §2 will be used to introduce two kinds of objectivity—absolute objectivity and relative objectivity. The distinction between the two ideas of objectivity is critical because I will argue that what is often taken as the default picture of objectivity regarding causation (i.e., absolute objectivity) is false, and I will provide what I take to be a feasible alternative (i.e., relative objectivity).

In §3, I will locate the objectivist position on causation by focusing on the recent work of Strevens. Strevens argues against a ‘*frameworked*’ notion of explanation and causation and in turn defends the concept of *deep, standalone* explanations as alternatives. I argue that Strevens is unable to eliminate the need for frameworking and suggest that the reason for this is that *deep* explanations are discordant with modern science. I will also elaborate on the idea of relative objectivity and the idea that explanations in science are better understood relative to frameworking.

§4 provides evidence to show that the objectivist position is based on a false dichotomy. According to objectivism, something is thought to be *either* frame-independent and objective *or* radically subjective. This dichotomy reveals itself in the work of Strevens (e.g., 2007, 2009) and Russell’s famous essay, ‘On the Notion of Cause’ (1913). I will argue that the *Either-Or* mentality allows us to clarify and make sense of the objectivist position. This will set the stage for §5, where I reveal how objectivism about causation fails.

§5 will be used to elaborate on the distinction between absolute and relative objectivity and, by relying on evidence from the history of objectivity in science, argue that the *Either-Or* dichotomy is false. It is my belief that the objectivist is working with an outdated sense of ‘objectivity’ and that the two concepts, ‘objectivism’ and ‘objectivity,’ are often mistakenly conflated. To the contrary, I will argue that objectivism does not necessarily mean objectivity. I argue that modern science reveals an alternative position, relative objectivity.

In the last four sections, I will use these results to highlight and ulti-

mately resolve the debate between Woodward and Price. In §6, I will argue that the debate between Woodward and Price rests on the distinction between absolute and relative objectivity. I will show that the debate is a matter of emphasis by highlighting Price’s advocacy of subjectivism as a rejection of absolute objectivity and Woodward’s support of objectivism as an acceptance of relative objectivity. Furthermore, I will clarify Price’s idea of subjectivity in §7 and locate a subtle form of objectivity that he implicitly defends. Likewise, I will locate Woodward’s idea of objectivity and a kind of subjectivity that he implicitly supports in §8. Finally, in §9, I will show how Collingwood shares the same stance on objectivity as both Woodward and Price.

4.2 Two Kinds of Objectivity

The influential anthropologist, Clifford Geertz, began his essay, ‘Anti Anti-Relativism’ (1984), by saying that, ‘I want not to defend relativism, which is a drained term anyway, yesterday’s battle cry, but to attack anti-relativism, which seems to me broadly on the rise and to represent a streamlined version of an antique mistake’ (p. 1). Like Geertz, I want to attack anti-relativism. Generally speaking, relativism claims that something, e.g., truth, goodness, or beauty, is relative to a frame of reference. A frame of reference can be thought of as a context, perspective, or a point of view.

It is useful to note what is meant to be contrasted with when using the term ‘relativism’. Generally speaking, when something is relative to a frame of reference, we say that it is non-absolute; there are no *absolute* overarching factual judgments about x , and there are no overarching standards to judge between competing reference frames. When philosophers such as Michael Krauz and Maria Baghramian claim that relativism contrasts with absolutism (see Krauz, 2010; Baghramian, 2010), there are two different varieties, so to speak, of what they mean by *absolutism*. One variety means objectivism, which Krauz refers to as ‘a matter of frame-independence’ (2010, p. 19). Objectivism affirms, Krauz says, ‘that sticks and stones exist as such irrespective of reference frames’ (ibid, 23). The second variety of absolutism is foundationalism, that is, the idea that whatever exists can be reduced to one fundamental, ultimate level. Note that the next two chapters represent an extended argument against absolutism regarding causation. In this chapter, I argue against objectivism.

When x is relative to a frame of reference, objectivists will claim that x is subjective. Significantly, objectivism entails that objectivity is an *absolute* term. According to objectivism, something is either frame- or mind-independent or it is subjective. So, an objectivist believes that (i) objectivity is an absolute term (i.e., they accept what I call *absolute objectivity*) and that (ii) something is either objective in the absolute sense, or it is subjective (when I label someone as an objectivist, e.g., Strevens, [2007, 2009] it will be in reference to (i) and (ii)). My aim in this chapter is to show that objectivism is false; it is not necessarily true that relativising x to a frame of reference means that x is radically subjective. The problem with objectivism is that it works as a *dichotomy*—something is objective if it is framework-independent and subjective if it is not. Regarding causation, the idea looks like this:

Absolute Objectivity: There is a fact of the matter as to whether A causes B irrespective of any context. The truth-makers of this fact do not include the subject. If two rational agents disagree about the causal relation, then at least one of them is wrong.

My claim is that one can reject absolute objectivity, as I will call it, and avoid subjectivism. I will argue that x can be relative *and* objective. Regarding causation, the idea looks like this:

Relative Objectivity: There is a fact of the matter as to whether A causes B given a specified context. If two rational agents occupying the same context disagree about the causal relation, then at least one of them is wrong.

One can reject objectivism about causation and still view causation as objective. What this shows, I argue, is that we can think of objectivity as a trichotomy. In one sense, x , e.g., the causal relation, is considered to be objective in the absolute sense— x is objective if it is frame independent. In another sense, x is seen as *relatively objective*. X is relatively objective if one rejects the absolute sense of objectivity (i.e., the idea that x is frame independent) but still sees x as objective given a specified context. According to relative objectivity, x is subjective in one sense because it rejects absolute objectivity—i.e., it relies on some type of context or frame (contexts or frames are limited, i.e. non-universal ‘points of view’, to use Collingwood’s

phrase). Yet, it is believed that we can still think about objectivity once the appropriate context has been specified. In another sense, x can be *absolutely subjective*. X is absolutely subjective if one rejects the absolute *and* relative senses of objectivity and sees x as subjective. Only by accepting this last sense does one qualify as a subjectivist. A subjectivist is someone who believes that x is relative to some type of context or frame and that the context is comprised of *individual agents*.¹

When considering relative objectivity, it is important to consider the frame of reference to which x is relative. For example, an objective Bayesian (E.T. Jaynes, 2003; Jon Williamson, 2010) who thinks that probabilities are relative to an ideal agent would not qualify as an absolute objectivist. This is because the *ideal agent*, which represents an ideal or *universal* frame—a frame that includes all human agents—still represents a specified context, albeit a large one. Thus, the objective Bayesian will accept relative objectivity, where the appropriate frame of reference is ideal and universal. Alternatively, a subjective Bayesian who thinks that probabilities are relative to individual agents would qualify as a subjectivist about probability.

Before I analyse objectivism in more detail, it will be useful to situate objectivism within the landscape of objectivity and frameworking that was laid out above and to locate a few agency approaches within this landscape. Table 4.1 plots theories of probability within the landscape of objectivity according to my analysis of objectivity, that is, objectivity as a trichotomy. For example, subjective Bayesians will fall under *subjectivism* because they believe that probability is relative to individual agents. Objective Bayesian’s fall under *relative objectivity* because they believe that probability is relative to a frame of reference, for example, an ideal agent. Chance theorists will fall under *absolute objectivity* because they believe that probability is non-relative and absolute.

Table 4.1: Objectivity and Probability

<i>Subjectivism</i>	<i>Relative Objectivity</i>	<i>Absolute Objectivity</i>
Subjective Bayes	Objective Bayes	Chance Theory

¹It’s true that the relative objectivist can think of the appropriate frame or context in terms of human agency. But a relative objectivist will not think that the context is individual agents, because she will still want to make objective claims given a specified context. And we would not be able to speak about objectivity if x was relative to individuals.

Where would agency theorists be placed on the graph? Agency theorists like Collingwood and Price think that causation is agent-dependent and thus reject the idea that the causal relation is objective in the absolute sense.

Table 4.2 *assumes objectivism*, where something is either frame-independent and objective, or it is subjective. Table 4.2 is useful because objectivism about causation (which results in a dichotomy) is often taken as the default structure of objectivity (e.g., it is assumed by Price, Woodward, Strevens, and probably Ney, as we will see in the next chapter). Given *this* structure of objectivity, I have plotted Collingwood under *subjectivism*, which is the only option besides *absolute objectivity*. Because Collingwood thinks that causes are handles and that causation is relative to manipulation (see Chapter 2§3), he must be placed to the left of *absolute objectivity*. So, if we take the philosophical literature on causation and objectivity, which often assumes objectivism, at face value, it appears that Collingwood must be placed under *subjectivism*.

For Price, causation is best thought of as a ‘situated’ concept (Price, 2017), where the *situation* in question refers to a creature’s place in time and a creature’s manipulation abilities. So, Price, like Collingwood, must be placed to the left of *absolute objectivity* because he takes the relationship between *c* and *e* to be non-universal. Again, given the structure of table 4.2, which reflects objectivism about causation, Price must be placed under *subjectivism* because it is the only option other than *absolute objectivity*, which Price clearly rejects. As we will see in §7, the structure of table 4.2 reflects Price’s intuition that the causal relation is subjective. In other words, *given* objectivism about causation, Price is correct, and he would be a subjectivist about causation.

I have also plotted Alyssa Ney, Phil Dowe, and Woodward according to the objectivist structure. Interestingly, the results in table 4.2 closely reflect the intuitions of the philosophers themselves, or the position that they would give themselves. I have placed Woodward under *absolute objectivity* because he has referred to his position as ‘objectivist’ (though it is not entirely clear what Woodward means by this) and other philosophers, e.g., Strevens and Price, also see Woodward’s project as some kind of defence for an objective position on causation.

Table 4.3 shows the same philosophers given *my* analysis of objectivity. Whereas objectivism works as a dichotomy, my analysis of objectivity results in a trichotomy. Below, I have structured the graph in such a way that, according to my arguments, causation and objectivity is shown *how it really*

Table 4.2: Causation According to Objectivism: How it Appears

<i>Subjectivism</i>	<i>Absolute Objectivity</i>
Collingwood Price	Strevens Ney Dowe Woodward

is rather than simply how it appears in the philosophical literature (where objectivism is often assumed). Notably, Collingwood, Woodward, and Price all fall under *relative objectivity* (I will argue that this is the correct placement later in this chapter). I have placed Hume under *subjectivism* because, according to a certain reading (see Beebe, 2007; Bernstein, 2017), one could argue that the causal relation, according to Hume, is simply a matter of projection, or something an *individual* projects based on their unique experience. In this chapter and the next, I will argue that *absolute objectivity*, a position held by Strevens, Ney, and Dowe, is untenable. The position is untenable because, even though a philosopher might endorse an absolute position, e.g., they may endorse the idea that the causal relation is in some way frame-independent, *what they actually say about the causal relation* contradicts an absolute position. For example, in this chapter, I will show that Strevens fails to meet the criteria for absolute objectivity about causation because he must rely on frameworking. In the next chapter, I will reveal that Ney fails to meet the criteria for absolute objectivity about causation because she relies on difference-making theories and ordinary pragmatics, which are frame-relative.

It should be noted that, given the structure of objectivity in table 4.3, one who falls under the middle category, *relative objectivity*, is not necessarily an agency theorist. For example, Jon Williamson (2007) would fall under *relative objectivity*, but Williamson would deny that his theory of causation amounts to an agency approach. Likewise, Cartwright (1979) would fall under *relative objectivity*, (recall that effective strategies can be discovered by ‘partitioning’ (i.e., localising) the space of possible situations to include the factors that are probabilistically relevant to the given strategy) yet she does not endorse an agency approach to causation. For Woodward, even though he denies an affiliation with agency theorists like Price, I will argue (in Chapter 6 §5) that his theory does amount to an agency approach to

causation because he relies on the notion of agency.

I should also point out that one can be a physicalist about causation (I will explore the physicalist position in the next chapter) and yet fall under *relative objectivity*. One could say that the causal relation is in some way relative to frameworking *and* that any causal relationship will be reducible to physical facts. (This is a position that likely results from a critical analysis of Ney, Strevens, and Dowe.)

Table 4.3: Causation and Objectivity: How it is

<i>Subjectivism</i>	<i>Relative Objectivity</i>	<i>Absolute Objectivity</i>
Hume	Collingwood Price Woodward	Strevens Ney Dowe

That table 4.3 accurately reflects the positions of Collingwood, Woodward, and Price will become more clear as we move ahead. For now, I would like to investigate Strevens’ argument for an objectivist position on causation and ultimately show that it’s unsuccessful.

4.3 Objectivism About Causation: An Example

In Chapter 3, I suggested that Strevens misrepresented Woodward’s approach by framing interventionism as a metaphysical analysis of causation. Strevens’ argument, which Woodward has attempted to refute, is that a successful approach to causation must be non-relativistic. While Strevens (2007) argued that Woodward’s approach failed due to interventionism’s accommodation to relativity, Strevens did not explicitly defend an objectivist position; instead, he simply argued that Woodward’s theory failed to meet its standards.

Interestingly, Strevens has recently provided an argument for his objectivist position. In *Depth* (2009), Strevens advocates ‘a two factor causal account of explanation.’ According to Strevens, the first factor of a satisfactory explanation is that the explanation of *e* is only made up of those causal influences of *e* that pass the second test for ‘explanatory relevance’ (see *Depth*, Chapter 3). The second factor is supposed to single out those causal influences that make a difference to the occurrence of *e*.

Strevens defends an *ontological* sense of explanation. He says that ‘an explanation [is] something out in the world, a set of facts to be discovered’ (p. 6). Additionally, Strevens claims that his project is ‘purely descriptive’ (p. 37). What Strevens intends to show is that *deep* or *standalone* explanations are preferable to frameworked explanations. Frameworked explanations are what Woodward calls ‘weak’ and ‘incomplete’ (see Woodward’s chapter on invariance, 2003, p. 239–314). They are relative to variable sets, which typically include high-level, coarse grained events and to assumptions about things like serious possibilities and background conditions (see Woodward [2007] for an overview). By contrast, standalone explanations are purportedly frame-independent. Strevens calls them ‘complete,’ i.e., sufficient for understanding the nature of a phenomena, because they are grounded in fundamental-level causal ‘realisers’ or influences (Strevens, 2009, p. 117).

To determine which fundamental-level influences bear on explanatory relevance, Strevens develops the so-called Kairetic account of difference-making (Καιροϛ is the ‘decisive moment’ or the ‘critical place’). A kairetic influence makes a difference; a non-Kairetic influence, i.e., a trivial one (which may bear some small influence), is readily dispensable. If a hurricane destroys my home in California, according to Strevens, what made a difference (to the state of my home) was the fundamental, physical properties that make up the movement and force of the storm (among other things). What failed to make a difference was the fundamental, physical properties of a ham sandwich being eaten in Barcelona. Speaking of Strevens’ account, Federica Russo, in her review of *Depth* (2011), says that, according to the kairetic account of difference making, ‘causes explain because, when you go *deep* down in the structure of the phenomenon to be explained, you will find the event-cause there, where it has to be, namely at the ‘right place,’ such that it makes a difference’.

According to Strevens, we can find kairetic difference makers by applying an optimising procedure to a causal model. For any given event, there is an optimised or idealised model that captures the fundamental level causal processes that led to the event.²

²The optimising procedure alters the model according to two specifications: the model is (i) made as abstract as possible while (ii) maintaining causal continuity among fundamental realisers. Continuity means that the trajectories or connections of the models’ many realisers remain unbroken. Strevens correctly acknowledges, ‘you may have to visit many other university departments, finishing a course with the physics department’ (2009, p. 161) to understand the depth that explanations require.

For Strevens, standalone explanations are preferable because they are purportedly sufficient for *full understanding*. Frameworked explanations are only partial and are thus weak candidates for explanations in the ontological sense. However, Angela Potochnik, in her article ‘Explanation and Understanding: An Alternative to Strevens’ *Depth*’ (2011), correctly points out that Strevens employs explanatory frameworks ‘to accommodate various features of explanatory practice that otherwise would not align with his account’ (p. 6). These features include ‘explanations that appeal to functional specifications, reference multiply realisable high-level properties, or sideline causal factors that qualify as difference-makers’ (p. 6).³ As an example of frameworking, Strevens uses the Lotka-Volterra equation, which can explain population dynamics across certain species, *given the proper explanatory framework*.⁴ Strevens insists that, while such frameworks are often practically indispensable, explanatory frameworks are inferior to standalone explanations that purportedly specify the nature of predation.

Strevens’ account raises two serious concerns. The first is that Strevens does not seem to provide a purely descriptive account of explanation in science, despite his claim to the contrary. It seems much more fitting, given Strevens’ strong ontological predispositions, that his account of explanation in science is *normative* rather than descriptive; he’s depicting how scientists *ought* to behave rather than how they do behave. Moreover, this raises the second concern, that is, the value that Strevens places on standalone explanations seems to rest too heavily on his supposition that only fundamental causal relationships are relevant to explanation.

To his credit, Strevens provides a compelling argument for an objectivist approach to causation. However, Strevens’ defence of objectivism about causation fails. Strevens’ defence of objectivism about causation fails because his account cannot eliminate frameworking. If we take his account at face value, that is, as a purely descriptive account of explanation in science, there is no way to eliminate the practical need for frameworking. Indeed, Strevens does not even try. The Lotka-Volterra model, which makes various preda-

³Strevens also relies on frameworking to accommodate explanations that cite omissions.

⁴Explanatory frameworks can be thought of as localised models. Models, such as the Lotka-Volterra equation, generate certain predictions and explanations via inputting data variables and manipulating the variables. If we are satisfied that the data is accurate, the results generated from the manipulation of the model provide us with an explanation of some target phenomenon. Thus the ‘explanatory framework’ can be thought of as a limited context based on data selection and manipulation procedures.

tion models relative to certain background conditions, is just one example in science where full, ontological understanding has nothing to do with the particular goal of scientific inquiry. Strevens uses this example to reveal that there are cases in science where something less than a full understanding is apparently useful. The problem for Strevens is that the example is far from unusual. There are many cases in science, including many cases in physics, where frameworking is used as an essential tool that enables understanding, not something that detracts from it.⁵ In his book *Causal Reasoning in Physics* (2014), Mathias Frisch forcefully argues that frameworking (Frisch says creating ‘boundary conditions’) is an essential practice in physics. Physicists use localised or frame-relative models (i.e., what interventionists would call small world systems) because it enables them to answer questions and make useful predictions. To illustrate the role of causal representation in physics, Frisch uses the example of the Large Hydrogen Collider (LHC), the world’s largest and most powerful particle accelerator. Frisch points out that the experiments done with the LHC are highly coarse grained and localised from the perspective of quantum field theory. Frisch notes that, inside the LHC, a ‘proton beam is taken to interact directly with the highly localised electromagnetic fields produced in various components of the accelerator, such as bending magnets and focusing magnets...yet the micro state of a pot of cheese fondue in a nearby mountain resort is not included in the model’ (2014, p.64).

Relatedly, one could argue that Strevens’ objectivism about causation fails because it rests on the problematic supposition that scientists are primarily interested in deep, metaphysical understanding. This supposition is problematic for a number of reasons. First, as I have already elucidated, Strevens’ account is too easily interpreted as a normative account of causation and explanation in science. Strevens fails to provide sufficient evidence for the claim that scientists are primarily concerned with ontology. It seems much more likely that working scientists in a variety of fields will have a range of interests, which may or may not include ontology. Here, one worries that Strevens misconstrues the regulative ideals that guide scientific inquiry. Second, one could simply argue that the supposition that scientists are interested in deep, metaphysical understanding while lacking sufficient evidence is not a supposition at all. Rather, it is a philosophical inclination being

⁵Examples abound in practical sciences like psychology, medicine, immunology, and law. See Pickering, (1995) *The Mangle of Practice: Time, Agency & Science*.

imposed on science.

The second reason that Strevens' defence of objectivism about causation fails is that standalone explanations themselves rely on frameworking. Rather than making causation relative to a model consisting of high-level, coarse-grained variables (like Woodward, 2003), Strevens' account of difference-making makes causal explanations relative to models composed of 'kairitic', fundamental, lower-level processes. Granted, Strevens' account of difference-making would considerably widen the explanatory framework. But Strevens' account relies on frameworking nonetheless. In the end, Strevens' argument amounts to the idea that explanations in science (*should*) rely on fundamental level, physical properties, and not coarse grained higher level variables that are often taken as paramount.

4.4 The *Either-Or* Mentality

If objectivism about causation fails, what then? If there is no frame-independent way for us to think about causation, wouldn't our causal claims become radically relativistic and subjective? Also, if objectivism fails (and we are forced to accept relativism about causation), do we not have a good reason to be sceptical of causal claims in science?

In this section, I will argue that to be an objectivist about causation requires one to think in terms of a specific (*Either-Or*) dichotomy.⁶ The *Either-Or* mentality, which says that *either* causation is frame-independent and objective *or* it is radically relativistic and subjective, is a mistake that often leads to unwarranted and misguided assumptions. Perhaps the most common, and indeed the most severely misguided, is the assumption that frame-independence means objectivity, by contrast, that relativism means subjectivity.

In his book *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis* (1983), Richard Bernstein explains that the *Either-Or* mentality (he uses this term) is a product of Cartesian Anxiety. Descartes' *Meditations*, Bernstein notes, is the '*locus classicus* in modern philosophy,' for the metaphor of the "foundation" and the conviction that 'the philosopher's quest is the search for the Archimedean point upon which we can ground our

⁶The *Either-Or* mentality results from apparently sound reasoning: *either* x is the case *or* y is the case. However, the conclusion, $(x \vee y)$ is not logically inclusive. For example, it could be the case that $(\neg x \wedge \neg y)$. I present this as an alternative to $(x \vee y)$ in 4.5.

knowledge' (p. 16). Significantly, Bernstein sees objectivism as the product of Cartesian Anxiety. He says

It would be a mistake to think that the Cartesian Anxiety is primarily a religious, metaphysical, epistemological, or moral anxiety. These are only several of the many forms it may assume. In Heideggerian language, it is “ontological” rather than “ontic,” for it seems to lie at the very center of our being in the world. Our “god terms” may vary and be very different from those of Descartes. We may even purge ourselves of the quest for certainty and indubitability. But the heart of the objectivist position, and what makes sense of his or her passion, is the belief that there are or must be some fixed, permanent constraints to which we can appeal and to which we are secure and stable (ibid, p. 19).

I suspect that this belief—the belief in some fixed, permanent constraints—is what leads Strevens to defend the counterintuitive idea that we need advanced courses in physics to understand a “real” causal relationship.

It seems likely that Strevens suffers from Cartesian Anxiety and falls prey to the *Either-Or* mentality. Without sensing the anxiety, Strevens’ criticism of Woodward’s interventionist approach can seem misplaced. Indeed, Woodward acknowledges the subjectivity inherent with his approach (more on this below), yet he makes numerous attempts to show how interventionism avoids a radical or untoward subjectivity; Woodward has repeatedly said that any viable theory of causation should be ‘agent-independent’ and ‘objective’ (2003, 2007, 2013).⁷ Yet, once we see Strevens’ anxiety for what it is (i.e., an attempt to ground causation in some fixed, permanent constraints), we can understand why he is so critical of Woodward (and others) who fail to *sufficiently* ground or secure the notion of cause. Rather simply, my contention is that Strevens’ position, along with his criticism of Woodward, rests on the *Either-Or* mentality.

In the light of the Cartesian Anxiety, another objectivist about causation becomes illuminated: Russell (1913). Russell’s (1913) position on causation,

⁷Woodward’s position on objectivity leads to criticism from both sides of the debate. Strevens (2007) argues that Woodward’s approach is *insufficiently* objective, while Price (2017) argues that Woodward’s approach is *too* objective. In 4.6–4.9 I will clarify these issues and reveal a consensus between Woodward and Price (and Collingwood) regarding objectivity.

i.e., eliminativism, can easily seem misguided. Indeed, something about it often seems obviously wrong (for example, see Field, 2003; Hitchcock 2007). Yet, when we situate ourselves within the *Either-Or* mentality, Russell's conclusion seems quite reasonable. What Russell did in 1913 was to show how there are no *absolute* causal relations. Russell was an objectivist about causation because he tried to ground the notion of cause in such a way that it could be described as frame-independent (recall that Russell thought that causal generalisations, which are *localised*, were 'much too vague' and 'otiose' (1913, p. 8–10). And, when Russell concluded that this type of grounding was impossible, he thought that the notion of cause should be eliminated. Assuming the *Either-Or* mentality, Russell's conclusion makes sense; because there were no absolute structures to ground the notion of cause, the notion of cause had to be eliminated—to do otherwise would lead to vagueness and scepticism, both of which Russell thought were inimical to science.

Interestingly, Russell later came to see his objectivism as an error. Indeed, it was a mistake that Russell corrected. In *Human Knowledge* (1948), Russell maintains his position on physical causation. He says, 'the concept 'cause', as it occurs in the works of most philosophers, is one which is apparently not used in any advanced science' (p. 399). Yet, Russell's stance in 1948 represents a more pragmatic approach to causation compared to his objectivist approach in 1913. Even though causation was not grounded in fundamental laws, Russell said in 1948, '[the concept 'cause'] still has importance as the source of approximate generalisations and pre-scientific inductions, and as a concept which is valid when suitably limited' (ibid). These generalisations are 'much weaker' than fundamental laws, but the 'assumption' that A causes B 'may suffice to give a high degree of inductive probability' (p. 400).

It is interesting to note that, as we will see in the next section, Russell's pragmatic stance on causation is compatible with the notion of objectivity in modern science. The objectivist position of Russell (1913) and Strevens (2009) is not. This is because objectivism is a value that is not commonly held in modern science. Objectivity *is* valued in modern science, but objectivism does not necessarily mean objectivity.

4.5 Objectivity in Modern Science

In this section, I want to further explore the two senses of objectivity that I have identified. Particularly, I want to situate both absolute and relative

objectivity in scientific practice by exploring the idea of objectivity in science.

In their book *Objectivity* (2007), science historians Lorraine Daston and Peter Galison note that scientists in the nineteenth century took on the ideal of absolute objectivity (what Daston and Galison call *mechanical objectivity*) because they wanted a more accurate and factual representation of nature. For these scientists, objectivity could be secured by eliminating human agency and frameworking. Agent-independence was thought to confer objectivity by providing impartial, accurate, and compelling evidence, while the subjective, by contrast, was thought to eliminate objectivity for the reason that something about the subject was being included within our judgments (for a useful review, see also Moser, 1999, pp. 19–28; Keller, 1997, pp. 313–332). For example, the idealised images of human skeletons presented in the *Tables of the Skeleton and Muscles of the Human Body* (1747), which were beautifully drawn by the professor of anatomy at Leiden, Bernhard Siegfried Albinus, were regarded with suspicion a century later for being overly subjective. Mechanical devices, such as the single lens camera, came to replace hand-drawn images because they were thought to represent nature more accurately. A *real* skeleton could be captured by a photograph but not by a human-made drawing.

The emerging idea of objectivity at the turn of the 20th century can be seen in a debate between two histologists (histology is the microscopic study of tissues). In 1906, Cajal and Golgi shared the Nobel Prize for medicine. The scientists had both used the new staining method of black-reaction to make visible the nerve cells in the human brain. The debate was centred on the structural accuracy of the images that had been created, independently, by Golgi and Cajal. Golgi had altered the images by adding hand-drawn highlights and colour variations to reveal what he deemed to be significant. Cajal, who refrained from changing the images in any way, charged that Golgi's images utterly failed to accurately represent the correct and objective images that he had elicited. As Cajal claimed that the goal of science was to 'see nature clearly,' by which he meant independent of the human subject, he subsequently charged that Golgi's interventions were *subjective* and ultimately unscientific.

Given the absolute sense of objectivity, a subjective judgment is a judgment or belief supported by evidence that was compelling for some rational beings and not others. For example, you may believe that the monument is on the left, while I think that the monument is on the right. You say that the image in figure 4.1 is a picture of a duck, while I say that it is a picture

of a rabbit.

In contrast to absolute objectivity, relative objectivity relates objectivity to a particular context and means non-arbitrary. It's important to note that if the idea of relative objectivity is successful (in the sense that is supported by modern science), the (*Either-Or*) dichotomy described in the previous section, which represents the objectivists' position, can be shown to be false. Below, I argue that the idea of relative objectivity *is* successful in the sense that it is coherent with scientific practice. The importance of the distinction between absolute and relative objectivity is to see that the objectivist is working under the assumption that objectivity *means* absolute objectivity. A vital question is whether or not scientific practice supports this hypothesis. I argue below that it does not.

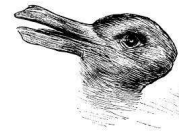


Figure 4.1:

In *Rethinking Objectivity* (1997), the historian Allen Megill calls the relative sense of objectivity *Dialectical Objectivity*.⁸ Megill says that dialectical objectivity, 'involves a positive attitude toward subjectivity,' because subjectivity 'is indispensable to the constituting of objects' (1997, p. 9). I call this type of objectivity *relative* objectivity because something is objective *given*, and relative to, some parameter, perspective, point of view, or framework. *Given* your position on the street, it is an objective fact that the monument is on the left. *Given* the human visual system (and given that it is functioning properly in an appropriate environment), it is an objective fact that certain flowers are coloured.

In *Objectivity*, Daston and Galison primarily explore how the notion of objectivity, in relation to scientific practice, was first developed in the nineteenth century and subsequently transformed over the following centuries. Daston and Galison distinguish three unique and overlapping senses of the word 'objective', while in *Rethinking Objectivity*, Megill distinguishes four senses of objectivity. The takeaway message from both of these studies is that the history of objectivity in science teaches us that the word 'objective' is often confounded as it is continually being adapted and re-adapted to fit varying conceptions in science.

The modern sense of objectivity, as perceived by scientists beginning

⁸For a defence of dialectical objectivity, see Pickering, 'Objectivity and the Mangle of Practice', 1997; Barnouw, 'The Shapes of Objectivity', 1997; Fabian, 'Ethnographic Objectivity Revisited', 1997; and Galison, 'The Journalist, the Scientist, and Objectivity', 2014.

around the 1930's (see Daston and Galison, pp. 265–315), represents a move away from absolute objectivity. In the modern sense, objectivity means *justification* and *reliability* and is defined by the elimination of arbitrary judgments (ibid, p. 309–62). Objectivity in the modern era (what Daston and Galison call objectivity as *trained judgment*) is modest when compared with the absolute idea of objectivity. This is because objectivity is no longer viewed as a binary notion. Something can be *more or less* objective as it is a matter of degree. The reason for this is that objectivity as trained judgment is concerned with justification and reliability rather than accurate representation and truth.

One of the most significant contrasts between absolute objectivity and objectivity as trained judgment concerns the human subject and the notion of perspective. The absolute idea views the subjective and perspectival as *fundamental obstacles* to accurate representation and real science. By contrast, the idea of objectivity as trained judgment, much like dialectical objectivity, sees the subjective and perspectival as *necessary supplements* to useful science. What Daston and Galison make clear is that scientists working in the second half of the twentieth century began to focus explicitly on the identification of the salient. Things like atlases, maps, and scientific textbooks needed to include trained judgment based on familiarity and experience, so that the *user* could understand the salient features of certain scientific practices.

In the sense of trained judgment, a belief or theory is said to be objective if it is justified, and a belief or theory has justification if it is appropriately functional. The evidence that provides justification, and therefore objectivity, is often a working distinction. To illustrate, when V.A. Firsoff published his *Moon Atlas* (1961), he argued that expert opinion was ineliminable. Because the lights and shadows on the moon were so significantly affected by its libration, many decisions were made concerning the magnetograms or pictorial representations of the moon's magnetic field. Of course, these decisions were partly subjective as they involved human judgment. Nevertheless, trained judgment was needed to make the image useful. Firsoff's argument was important because what scientists wanted in 1961 was a *reliable* map of the moon—one that could, for example, distinguish between a mountain and a crater. In certain operational contexts, e.g., given the question, 'where should one land a spacecraft on the moon', the subjective, or the human subject, was no longer thought to be inimical to science. Scientists viewed the *real*, i.e., mind-independent, map of the moon as insignificant because it

could not provide the necessary distinction.

The key aspect that separates absolute objectivity from objectivity as trained judgment is a notable shift in what scientists perceived to be their principal motives and regulative ideals. Scientists working in the nineteenth century tried to discover and represent a world untouched by human hands.⁹ Significantly, scientists in the modern era began to reject the motive of eliminating human interaction in science and thus moved away from absolute objectivity—the elimination of the subjective and perspectival—because the motive for these scientists had shifted towards *use*. By including trained judgment, the human subject became vital to scientists who wanted to convey to apprentices ‘the means to classify and manipulate’ (Daston and Galison, p. 322).

These developments are significant for a number of reasons. First, we learn that objectivity in the modern sense is in accordance with the relative idea of objectivity and not the absolute. Human interaction, perspective, judgments, and values that were once viewed as inimical to science are now viewed as necessary additions to useful and functional science. Second, the historical development of scientific ideals in the twentieth century shows us that science is amicable to the idea of frameworking. Scientists are interested in the salient features of certain systems or objects because their motive is use or functionality. This supports Frisch’s idea in §3. Frisch’s argument was that physicists relied on localised causal models because this enabled them to answer important questions and make useful predictions. Indeed, the claims of Collingwood, Price, and Woodward that causation is relative to a point of view and that relativity is needed because we are focused on *use*, far from being problematic, shares an important similarity with modern science. Third, these developments are significant because they show us that the *Either-Or* dichotomy is false. It is not true, according to modern science, that something is either aperspectival, agent-independent, and therefore objective *or* it is radically subjective. Thus, we discover that a rejection of absolute objectivity does not lead to a radical subjectivity. Indeed, modern science presents us with an alternative view: something can be relative and objective.¹⁰

⁹Daston, Galison, and Megill all note that philosophers in the twentieth century would make this motive even stronger by adding the idea that science should represent the world *as it really is*.

¹⁰I will argue in the following section that this is true for causation. I will also argue it is a position held by Collingwood, Price, and Woodward.

Two points are also worth noting. First, Russell seems to have shifted from absolute to relative objectivity. Russell's 1948 account of causation is structured in terms of how we make inferences. Thus, causation becomes relative to certain epistemic constraints. Yet, Russell thinks the notion of cause is highly useful nonetheless, which is a major shift from his objectivist position in 1913. Second, we can see more clearly why Strevens' ontological approach to causation fails as a purely descriptive account. As I noted above, Strevens is forced to retain some level of frameworking in science because it is a practical necessity. Again, Strevens thinks that standalone explanations are superior because they have ontological precedence. But the history of science indicates that ontology no longer has the precedence in science that it once enjoyed. At the very least we can say that what also has precedence is *functionality*—the very thing that validates relativity and frameworking.

4.6 The Debate Between Woodward and Price

An important question we might ask now is, what is the relationship between the manipulation approach to causation and objectivity? Notably, Woodward and Price hold contrasting views on the status of the manipulation approach and whether it should be given a subjectivist or an objectivist interpretation. On the one hand, Price (1993, 2017) argues that the relationship between causation and manipulation entails a subjectivist position on causality, while on the other, Woodward (2003) argues that his manipulation approach implies an objectivist position on causality. The discrepancy between Woodward and Price on objectivity is problematic for two reasons. First, it suggests that my claim of solidarity between Woodward and Price is somehow mistaken. Second, and more significantly, the discrepancy between Woodward and Price, that is, the general lack of consensus on the status between manipulation and objectivity, gives credence to the critics of the agency approach. For example, Alyssa Ney (2009) argues that the so-called subjectivity of manipulation approaches to causation provides support for physical theories of causation, which she takes to be more objective. Ney says that 'insofar as one has reason to believe that facts about [manipulation] are subjective...it can provide a powerful case for foundationalism' (2009 p. 758; I pick up this argument from Ney in chapter 5).

The debate between Woodward and Price boils down to a matter of emphasis. By focusing on the alleged concerns regarding objectivity in the

absolute sense, i.e., given the objectivist criteria, Price emphasises a subjectivist position on causation. In contrast to Price, by placing emphasis on the positive aspects of objectivity in the *relative* sense, Woodward stresses an objectivist position. Because Price and Woodward are working without the appropriate distinctions, i.e., absolute *and* relative objectivity, they simply misinterpret each other’s position. In the end, my arguments will show the two positions—a *rejection* of absolute objectivity and a *acceptance* of relative objectivity—are compatible. Indeed, the rejection of absolute objectivity and the acceptance of relative objectivity is a stance held by Collingwood, Woodward, and Price.

The distinctions between absolute objectivity, relative objectivity, and absolute subjectivity are critical to understanding the debate between Woodward and Price. By relying on these distinctions, I argue that we can actually resolve the debate and build an objective agency approach to causation. In the debate between Woodward and Price, both end up misinterpreting each other’s position. Price is not a subjectivist, as Woodward claims, and Woodward is not an objectivist, as Price claims. If objectivity means *relatively objective*, the correct way to interpret Price’s “subjectivism” is in defence of the *relative* in “relatively objective”, and the right way to interpret Woodward’s “objectivism” is in defence of the *objective* in “relatively objective”.

4.7 Price’s Perspectivism

Up to this point, I merely suggested that Price accepted relative objectivity, despite the fact that he explicitly defends a form of subjectivity. To show that my claim is correct, I must demonstrate that Price’s defence of subjectivity is a result of his rejection of absolute objectivity. Second, and more challenging, it must be demonstrated that Price is amenable to some form of objectivity.

It is commonly supposed that Price supports a subjectivist position on causation (e.g., Woodward, 2003, 2013). However, this supposition about Price is misleading. Using the distinction between absolute and relative objectivity and pure subjectivism, we can see how a rejection of objectivism or the idea of absolute objectivity would not necessitate a subjectivist position on causation. For example, one could reject absolute objectivity and still accept objectivity in the weaker, relative sense. My claim is that this rejection-acceptance model—that is, reject absolute objectivity and accept

relative objectivity—is the correct interpretation of Price.¹¹

One could immediately object, citing Price’s 1993 essay ‘Causation as a Secondary Quality’, that my claim simply misses the point. Clearly, one could argue, Price wants to teach us that there is something *fundamentally subjective* about our concept of cause. If, as Price argues, causation is a secondary quality, then Price’s point is that *it is something about us as human agents* that gets us to causation. To this I would reply in favour: Indeed, Price’s point is precisely that it is something

about us as human agents that gets us to causation. What I would add is that this merely prohibits us from holding an absolute position on objectivity about causation. If, as in the case of colour, we think of objectivity as relative to the context of certain dispositions, e.g., the human visual system, then there is still a sense in which we are allowed to speak about objectivity. For example, there may be no absolute sense in which we can sensibly talk about the colour of an object; however, it remains an objective fact, *given our disposition to see colour*, that the apple in figure 4.2 is red. Given the (normally functioning) human visual system, to call the apple green or blue in this context would simply be false.

To get a better grasp of what Price means by subjectivity, consider Price’s 2001 essay, ‘Causation in the Special Sciences: The case for Pragmatism,’ where he explicitly defends a type of subjectivity. Price advocates what he calls a *practice-subjectivity* about causation. *Practice-subjectivity* is meant to be contrasted with a stronger form of *ontological-subjectivity*. The distinction is useful because Price maintains that the concerns for causal realism relate to the ontological axis; once again Price is arguing that when we think about causation, we must turn inwards, as it were, and away from the objectivist’s invocations of reality. The inward turn is necessary because any

adequate philosophical account of causation needs to begin with its role in the lives of *agents*, creatures who have the primitive experience of intervening in the world in pursuit of their ends



Figure 4.2:

¹¹Relative objectivity results from a *rejection, rejection, acceptance* model (reject absolute objectivity, reject absolute subjectivity, and hence accept a relative objectivity). However, for stylistic reasons, I will sometimes speak of relative objectivity in relation to a *rejection-acceptance* model. The *rejection* should be read as a rejection of both absolute objectivity and absolute subjectivity.

(Price, 2001, p. 107).

For Price, causation is practice-subjective because an adequate philosophical account of causation must make

reference to the role of the concept in the lives and practices of the creatures who use it (Ibid, 106).

It is important to note again that Price's defence of subjectivity, thus far considered, merely entails a rejection of absolute objectivity. Indeed, Price is explicit that he *does* reject absolute objectivity. To illustrate, note that Price's defence of practice-subjectivity is meant to contrast with the stronger form of subjectivity, *ontological-subjectivity*. According to Price, causation is *ontologically-subjective* if 'the *existence* of causal relations depends on the minds, speakers, observers or the like; if causal states of affairs are in this ontological sense mind- or observer-dependent' (2001, p. 106). What Price rejects is the idea that causation is frame-independent. Price rejects absolute objectivity and thus (given objectivism) claims that causation is subjective because we have to reference the lives and practices of the creatures who employ causal concepts. Yet, it seems plausible, indeed likely, that once we have made the necessary reference, we can speak about objectivity.

However, Price is somewhat unclear on this point because he chooses to talk about *realism* rather than objectivity. For example, Price takes ontological-subjectivity to counterpoise strong realism. 'Strong realism,' as Price (2001, p. 112) calls it, relies on the so-called *inference to the best explanation* for believing in the reality of theoretical entities and states of affairs postulated by science. Following this view, the best explanation of the available observational data is that these entities and postulates exist or obtain. Price rejects strong realism about causation for a number of reasons, including the reason that it would be viciously circular. However, Price points out that his sympathies lie with another kind of realism—what he calls 'minimal' or 'weak' realism (ibid, p. 113). A minimal realist simply takes the existence claims of science at face value, 'and rejects any additional metaphysical or philosophical viewpoint from which it would make sense to ask, 'Do these things (electrons for example) *really* exist' (ibid). Price suggests that Arthur Fine's defence of *the natural ontological attitude* is compatible with minimal realism. So, we could say that, because ontological-subjectivity *counterpoises* a strong realism, Price's acceptance of a minimal realism is *compatible* with a weaker, minimal form of objectivity. I will return to this point in a moment.

Moving ahead with Price's work, we see the same defence of subjectivity against those who he perceives to be causal realists. 'Some writers (e.g., Pearl, 2000; Woodward, 2001),' Price says, 'agree about the centrality of notions of agency and manipulation to an understanding of causation, but take the resulting view in a more realist or objectivist spirit' (Price, 2004, p. 49). One question we might ask is, why is Price so worried about the 'realist spirit?'

Price motivates a kind of subjectivity because he is concerned that the 'realist or objectivist spirit' will lead to a sort of scepticism regarding causation. He remarks:

I think that Woodward's view that his approach is more 'realist' than that of MP is likely to prove more of a hinderance than a help at this stage, in that it makes him more prone to sceptical worries about whether there is *really* causation inside the sun, or whether causation *really* runs past-to-future in neighbouring galaxies (2017, p. 93).

In a recent paper, 'Causation, Intervention and Agency—Woodward on Menzies and Price' (2017), we can again see that Price's defence of subjectivity is an argument against what I call absolute objectivity—viz., frame or human-independence. Price says that 'it is the subjectivist view that does the better job of recognising the contingencies and limitations of the human standpoint, and the objectivist view that confuses us with gods' (2017, p. 90). Though Price himself does not specify what he means by 'objectivism,' I think it is clear that he means absolute objectivity (frame- or human-independence). Moreover, I think it is also clear that Price relates realism, that is, strong-realism, with absolute objectivity. Subsequently, when Price mentions realism, he means absolute objectivity. So, the subjectivity that Price defends can be understood as a reaction against objectivity in the absolute sense. I will now discuss my claim that Price rejects absolute subjectivity.

It seems to me that Price's so-called minimal realism is interchangeable with relative objectivity. For example, they both suggest that we should take certain claims at face value (e.g., if we take a human point of view regarding causation, we can make true, false, or otherwise factual claims about certain causal relationships, and we need not ask, 'are these *really* true' or 'do these facts *really* obtain'); they both reject the traditional picture of objectivity as

agent-independence, and they are both grounded in practical considerations (whether or not *weak realism* deserves to be classified as a species of realism is another question).

The most compelling evidence for Price's *acceptance* of relative objectivity comes from his idea that causation is perspectival. Price (2007) uses the familiar example of 'foreignness' to introduce the notion of perspective to our understanding of causation. As Price explains, we know that the notion 'foreigner' only makes sense from a certain location or perspective. From our location z , it is appropriate to say that y 's are foreigners—from our location, we apply the concept to them. However, the people on the opposite side of the border apply the same notion of foreignness to us. From their location z^* , it is appropriate to say that x 's are foreigners. Significantly, Price points out that this does not deny the reality of foreigners:

[Foreigners are] not figments of our collective imagination, or social construction, or useful fictions. They're not mind-dependent, and they don't disappear when we don't keep an eye on them. Our 'folk theory' about foreigners isn't subject to some global error, and the term 'foreigners' certainly manages to refer. Some of our beliefs about foreigners are mistaken, no doubt, but only by failing to accord, case-by-case, with the objective reality to which they are certainly answerable. There are many facts still to be discovered about foreigners, such as their precise distribution in space and time. Moreover, these are matters for scientific discovery...in a nutshell, foreigners are as real as we are. (Price 2007, p. 250)

Nonetheless, Price thinks we learn something when,

mind's broadened by travel, we realised that foreigners themselves use the very same concept, but apply it to us!...[T]he reality of foreigners notwithstanding, there's a sense in which foreignness is a less objective matter than we used to think. (2007, pp. 250–51)

Here we can see the precise rejection-acceptance model that relative objectivity requires. 'Foreignness' is not a mind-independent natural kind—it is not aperspectival or human-independent as absolute objectivity would require. Nonetheless, it is not purely subjective. We can speak objectively

about foreigners or ‘foreignness’ *given that we specify context* or given that we understand the concept in terms of a given perspective.

Price claims that the same goes for causation. Causation is perspectival because it relies on ‘our viewpoint, or location, in one way or another’ (Price, 2017, p. 91). Specifically, Price thinks that causation is relative to the context of what he calls *deliberation*. How (or what) we deliberate (about) as human agents is determined and limited by many factors, including how we experience time and what we take to be within our control. Price says that our deliberation as human agents is limited by *options*. Options are propositions that the agent takes herself to have the option of ‘deciding to make true’. Price says that *fixtures*, that is, all matters of fact that are not held to be a matter of choice in deliberation, ‘denotes everything else’ (Price, 2017 p. 87). When I deliberate about getting into better shape, for example, I take as options (i) what I eat and (ii) how often I go to the gym. These are things that I can control or ‘make true’ in one way or another. Fixtures are everything else that I take to be outside of my direct control, for example, the weather, the fat content of certain foods, and so on. So, when you and I are both deliberating about fitness, *given that we share the same options*, if you and I disagree about the effects of diet on fitness, one of us must be wrong. Given this context, it will be an objective fact whether or not eating habits will affect fitness.

In the scenario where you and I are both deliberating—we both share the context of deliberation—but we do not share the same context of options, i.e., we take different objects to be within our control, Price says there is an *intramodal* difference between us.¹² There would be an *extramodal* difference between us if we did not share the context of deliberation. Price explains:

We differ *intramodally* compared to creatures who are also agents but have [e.g.,] the opposite temporal perspective to our own. We differ *extramodally* from creatures who are not agents and therefore lack what it takes to employ the concept of causation in the first place (Price, 2017, p. 88).

So, we differ *intramodally* from creatures who are also agents but who take different objects to be within their control, e.g., space aliens with the

¹²This is quite similar to Collingwood’s (1940) notion of handles in the practical sense of cause. Options, like handles, are relative to what certain agents can bring about, manipulate, or otherwise control. I will investigate this similarity in §9.

opposite temporal perspective to our own or gods like Jehovah and Zeus. We differ *extramodally* from creatures who are not agents at all, e.g., intelligent trees. Generally speaking, Price would say that all human agents share the same options; unlike Collingwood (1940), Price would say that human agents do not differ *intramodally*.

The relevant point to note is that Price's perspectivism can be classed as relative-objectivity. Causation is not mind- or frame-independent as absolute objectivity would require—causation is subjective because it is frame-dependent. Nonetheless, causation is not purely subjective. Given that we specify a relevant context, it is an objective fact whether or not I am a foreigner or whether or not diet affects fitness.

4.8 Woodward's Interventionism

For Price's theory, the question of relative objectivity was whether or not there was evidence of Price's acceptance of objectivity. Placing Woodward in the context of relative objectivity is beset with a different concern. Whereas Price is misunderstood as a full-blooded subjectivist, Woodward is misunderstood as a full-blooded objectivist. So, the question for Woodward is whether or not there is evidence that he rejects absolute objectivity.

Woodward and Price, like Collingwood, are manipulationists. Price's version of the manipulation theory argues that it is something about us *as agents* that gets us to causation; causes are things that *we* regard as handles and provide *us* with a sense of power and control. Woodward's manipulation theory, interventionism, would accept this claim but lose the emphasis.

The notion of an intervention is an abstract representation of a human experimental manipulation that is stripped of its anthropocentric elements and characterised in terms that make no reference to human beings and their activities. (Woodward, 2003, p. 374)

So characterised, we can think of an intervention as:

An idealized experimental manipulation carried out on some variable X for the purpose of ascertaining whether changes in X are causally related to changes in some other variable Y . (ibid, 95)

Interventionism is meant to provide a theory of causal explanation. If we intervene appropriately on X and notice changes in Y we can explain Y by citing X . If I go from eating one bowl of ice cream a week to eating two bowls of ice cream a day, assuming I hold all the other relevant variables constant, and I notice changes to my fitness, I can explain the changes to my fitness by citing my diet—we can say that *diet* has a causal relationship to *fitness*.

Woodward takes the causal relation to be an objective fact. Nonetheless, it is a complicated issue determining just how objective Woodward intends to be. Recall the earlier quote from Price:

Some writers (e.g., Pearl, 2000; Woodward, 2001) agree about the centrality of notions of agency and manipulation to an understanding of causation, but take the resulting view in a more realist or objectivist spirit (Price, 2004, p. 46).

Michael Strevens seems to agree with Price and takes part of Woodward's project to be an account of the *nature* of causation. Referencing Woodward's 2003 book, *Making Things Happen*, Strevens says:

[Woodward] argues that according to the manipulation account, causation is mind-independent, or objective, on the grounds that the characterisation of causation offered by the account makes reference only to counterfactual facts that are 'not dependent on human attitudes or beliefs'. This argument clearly supposes that the Woodwardian characterisation of causation tells us about the nature of causation (indeed, it had better tell us *everything* about the nature of causation, or else we might miss some mind-independent truth-maker for causal claims) (Strevens, 2007, p.15).

It is clear that both Price and Strevens take Woodward's project to include a defence of objectivism and absolute objectivity according to my definitions of the terms. Indeed, there are a handful of passages where Woodward speaks about objectivism and suggests that he endorses an objectivist view. Consider the two passages from Woodward (2003) where Strevens and Price base their allegations:

The counterfactuals on which causal claims are based seem to be true in a mind-independent way (Woodward 2003, p.118).

I would claim that this is essentially the objectivist position regarding causality and agency that I have endorsed (ibid, p.125).

Indeed, there are a handful of passages where Woodward speaks about objectivism and suggests that he endorses an objectivist view. This can be misleading. Indeed it has been. Both Price and Strevens take Woodward to mean that he is an objectivist according to my understanding of the term. But this is not what *Woodward* means—it seems quite clear from reading the passages above that there is more to say about Woodward’s “objectivism”. Indeed, Woodward writes that causal claims *seem* to be mind-independent because, as he explains three lines earlier,

I suggested above that there may be a limited sense in which [subjectivism] may be true: which causal claims we accept as true (or at least *readily* accept) are influenced by what we take to be a “serious possibility”. However, once we fix which possibilities are serious, it seems to me that...there is no other sense in which counterfactuals about the outcomes of hypothetical experiments associated with typical causal claims are in some way dependent on human attitudes and beliefs. In other words, the counterfactuals on which causal claims are based seem to be true in a mind-independent way (ibid p.118)

Unless we are willing to contend that Woodward entirely contradicts himself within the same paragraph, i.e., as accepting subjectivism and yet somehow emerging as an objectivist, we should view Woodward’s comments on objectivity as a statement of relative objectivity. For it is only when we *fix* the so-called serious possibilities that we can make objective claims about counterfactuals and causal relationships.

It is worth considering the idea of serious possibilities in more detail. According to Woodward (e.g., 2003), what we consider to be a serious possibility is based on a variety of factors: the various questions we ask, what we are interested in explaining, and the conditions and factors that we hold to be invariant all determine a sense in which we hold some possibilities to be more serious than others. Consider the example of a patient who died from a failure to receive an antibiotic. In the case of the patient’s death, we do not hold strangers from distant cities who fail to administer the shot causally responsible. To understand why this is so, that is, why a counterfactual such

as, *If Y, who is not a doctor and a complete stranger to patient X, had administered z to X, X would have lived* fails to provide a plausible explanation for the patient's death, we have to consider our ordinary interests. In such cases, we are interested in *preventing* similar deaths. So, while it may be true that the patient would have lived had a stranger (Y) administered the drug, we do not consider strangers to have a causal effect on the patient because we do not hold strangers accountable and/or responsible for preventing/failing to prevent similar deaths.¹³

As I argued in chapter 3, Woodward thinks that the idea of serious possibilities leads to a kind of relativity. 'What we want to explain and what we take as serious possibilities is relative to our interests', Woodward says. However, relativity, Woodward points out, should not be considered as a threat to objectivity:

it does not follow that for a fixed explanandum M and for a fixed explanans E , whether E explains M is itself interest-dependent. Obviously, it is not puzzling and no threat to the "objectivity" of explanation that the explanans E may explain M but a different explanans E' may be required to account for M' (2003, p.230).

On Woodward's account, interest-relativity will simply determine which factors or variables we include in a causal model. Thus, interventions are relative to a variable set V , where any particular model can reveal the causal relationships between the relevant variables. Interventions (so construed) allow us to answer what Woodward calls *what-if-things-had-been-different-questions* (*w-questions* for short). In terms of my fitness, the *w-question* that I want to answer is how changing [*diet*] will effect [*fitness*]. If we change variable X [*diet*] and notice changes in Y [*fitness*], we can say that there is a causal relationship between X and Y —in some sense, we can explain Y by citing X . This type of explanation, Woodward admits, is somewhat weak. We may not discover, for example, the precise mechanism that links diet and fitness when we attempt to answer the *w-question* about diet and fitness. Nonetheless, the explanation is not inconsequential or arbitrary. If a change to X produces a change in Y , there is a causal relationship between X and Y .

¹³I will explore the relationship between accountability and serious possibilities in chapter 9.

Seemingly, interventionism is open to a Strevens-type worry (see §3), which claims that it is flawed because it does not provide an adequate account of causation at the fine-grained, fundamental level. However, Woodward does not see this as a threatening charge since ‘the manipulationist theory assigns a more limited significance to correctness at the level of fundamental ontology’ (Ibid, p.231). The limited concern with ontology is eclipsed by the concern for *functionality*—knowing how changes in one variable will affect another is practical and highly useful in the context of ordinary life and the special sciences. Thus, a large part of Woodward’s project (see Woodward, 2007) is focused on motivating what he calls a *macroscopic* view of the causal relation. Analysing causation from the macroscopic perspective of ordinary life and the practical sciences, while highly pragmatic, makes causation *coarse grained* and causal explanations *weak* (as we saw in chapter 2). Nonetheless, Woodward calls himself a *moderate realist* (2007, p. 103) about causation and claims that the causal relation is objective and real.

The claim that causation is coarse grained and relative to a certain macroscopic perspective (see Chapter 2§5) seems to contradict Woodward’s claim that causation is objective and real. However, as I claimed above, unless we are willing to contend that Woodward is open to such an obvious contradiction, I think we can more accurately view Woodward’s defence of objectivity as a defence of relative objectivity. Indeed, in his 2007 essay ‘Causation with a Human Face,’ we find the precise *acceptance-rejection* model that relative objectivity requires. In his article, Woodward argues that causal claims are relative to ‘small worlds’—that is, macroscopic systems where agents are able to intervene—and that *from this context*, modal notions such as probabilities, chances, and causes can be interpreted as objective and real (p. 102). However, Woodward points out that if we take our criterion for ‘objectivity’ to be fundamental physics in a deterministic world, where the notion of small worlds does not apply, then probabilities, chances, and causes are subjective (see pp. 98–103). Speaking about the chances that we ordinarily associate with gambling devices such as a roulette wheel, Woodward says that

If our criteria for the objective reality of such chances is whether they appear in fundamental physical laws then the answer is clear: the chances we associate with such devices are not objectively real, however useful they may be for summarising the behaviour of the devices and guiding betting behaviour. (ibid, 100)

Woodward rejects this criterion for objectivity. Notice that this posi-

tion merely entails that Woodward rejects the absolute sense of objectivity. Woodward suggests that if we take our criteria for objectivity to be mind- or frame-independence, then causation is subjective. However, Woodward is keen to point out that if we take different criteria for objectivity, then we can claim that causation is objective. If we mean that objective probabilities, as Woodward *does mean*, are probabilities that reflect patterns ‘that a macroscopic agent is able to impose or learn about’ (ibid)—probabilities that are stable and invariant from *our* perspective as agents—then there are nontrivial objective probabilities associated with the roulette wheel’ (ibid).

Regarding causation, Woodward says that there are objective facts about causes given that we take as our criteria some macroscopic perspective. Also, because we know that the relevant perspective for the interventionist is causal models, we can interpret Woodward as saying that we can think of the causal relation as objective and real in relation to a causal model and a variable set V . Thus, Woodward *rejects* absolute objectivity but *accepts* the idea of objectivity which says that our criterion for objectivity is frame-dependent.

A large part of Woodward’s project is meant to teach us that causation *is* a concept that strictly applies to small-world standards, and subsequently that causation is objective. We can be moderately realist about causation because a (functional) causal relationship, even though it is relative to a certain context, is objective and real.

For Woodward, causation is objective, but Woodward is not an objectivist. What matters for Woodward is the functional relationship between individual variables in a causal model. Given that causal explanation and causal relationships applies to small worlds, causation is *relative*. However, once we have specified a small-world-context, viz., a causal model and a variable set V , we can make *objective* judgments about the causal relation.

4.9 Collingwood and Objectivity

Labelling Collingwood as a relative objectivist about causation is somewhat complicated. The complication stems from the fact that Collingwood did not speak of either objectivity or subjectivity in relation to causation (or in relation to anything else, as far as I am aware). Despite the fact that Collingwood does not explicitly address the topic of objectivity in relation to causation, my argument is that *the best interpretation* of Collingwood is as a relative objectivist. This claim is based on a distinct argument: inves-

tigating the relationship between the different ideas or senses of causation that Collingwood identifies in *An Essay on Metaphysics* (1940), with the two senses of objectivity and subjectivism that I have identified, entails that Collingwood is a relative objectivist about causation.

For my argument to be successful, I must show that Collingwood rejects both absolute objectivity and absolute subjectivity and hence that he accepts a form of relative objectivity. To begin, I will focus on Collingwood's metaphysics and its relationship to his idea(s) of causation.

Recall that Collingwood's idea of metaphysics was the search for absolute presuppositions (see chapters 2&3). As I explained in chapter 2, Collingwood uses the term "cause" to illustrate how different sciences, each with their own unique presuppositions, hold contrasting ideas about causation. By investigating the idea of causation in historical, practical, and theoretical sciences, Collingwood speaks of three distinct meanings of the term "cause." In the historical sciences (sciences that deal with the mind) 'that which is caused is the free and deliberate act of a conscious and responsible agent, and causing him to do it means affording him a motive for so doing' (1940, p. 285). The sense of causation that governs practical sciences like medicine and engineering assumes 'that which is caused is an event in nature and its cause is an event or state of things by producing or preventing which we can produce or prevent that whose cause it is said to be' (ibid, pp. 296–7). While theoretical sciences assume 'that which is caused is an event or state of things and its cause is another event or state of things such that (a) if the cause happens or exists, the effect must happen or exist even if no further conditions are fulfilled (b) the effect cannot happen or exist unless the cause happens or exists' (ibid, 285–86).

Here, Collingwood shows that the term 'cause' is relative to certain practices and ultimately the presuppositions that guide each form of enquiry. In other words, to be a doctor or an engineer *requires* presupposing that it is possible to change the course of nature and thus that a cause 'is an event by producing or preventing which we can produce or prevent that whose cause it is said to be'. To be a historian requires presupposing that man is rational and thus that 'that which is caused is the free and deliberate act of a conscious and responsible agent.'

So how does all of this relate to objectivity? Like the term 'cause', it is clear that Collingwood would take the concept of objectivity to be relative to certain presuppositions and thus to certain forms of enquiry. When we try to relate the notion of objectivity to Collingwood's metaphysics, specifically his

ideas of cause, we must situate ‘objective’ within the specific sense of ‘cause’ in question. So, in one sense, *objectivity is relative* because it is non-absolute. That is, we must apply the notion of objectivity to specific contexts, viz., certain disciplinary contexts that rely on specific presuppositions. Yet, as we have seen, relativity does not entail absolute subjectivity. For instance, we might say that *causation is objective* once we apply the notion of objectivity to a specific sense of cause. For example, we might say that ‘A causes B’ is an objective fact in the practical sense if a cause is something that allows an agent to produce or prevent. So, if I can manipulate A to bring about or prevent B, then ‘A causes B’ will be an objective fact.

Of course, this assumes that more than one agent will share the same frame of reference regarding manipulation (i.e., producing and preventing). If causation is relative to what an agent can produce or prevent, in many (general) cases, human agents will share the same frame of reference. In more technical cases like the automobile accident, there may be a greater degree of relativity, which I call the relativity of expertise (see chapter 2§3). In the case of the automobile accident, different agents disagreed about the handle or the cause of the event. Yet, even in cases where the context is expertise, there is bound to be more than one expert, and thus when one of them disagrees about the cause of an event, at least one of them must be wrong. So, for Collingwood, there is a compatible frame for relative objectivity—either human agents in general or expertise.

In the scenario where two agents who share the same frame of reference disagree, and hence that one of them must be wrong, we come to see how Collingwood avoids absolute subjectivity. It is incorrect to view the practical sense of cause as absolutely subjective because what causes what is not relative to individual agents.

It might be argued that Collingwood accepts the idea of absolute objectivity once we apply the notion of objectivity to the theoretical sense of cause. Remember that theoretical causes are taken to be agent-independent and absolute. So, according to the theoretical sense of cause, we might say causes can be absolutely objective when a cause is something that exists wholly in nature, independent of any human agents. However, this argument would not hold because Collingwood rejects the theoretical sense of cause for being nonsensical. Russell (1913) had argued the notion of cause was nonsensical because it was outdated and incompatible with advanced sciences like physics. Collingwood believed that the (theoretical) idea of cause was nonsensical for a different reason. Collingwood argues that the idea of

theoretical causation results in a contradiction. The notion of theoretical causes, where a cause is absolute and exists wholly within nature, (i) rests on the idea of *compulsion*—a cause is something that compels the effect. Yet, Collingwood points out that (ii) the idea of compulsion is derived from our experience as agents. He says that the idea of compulsion ‘is derived from our experience of occasions on which we have compelled others to act in certain ways...and conversely, occasion in which we ourselves have been compelled to act’ (ibid, p. 322). Collingwood concludes that causation in sense three is an anthropomorphic idea applied to nature: ‘Causal propositions in sense III are descriptions of relations between natural events in anthropomorphic terms’ (ibid).

For Collingwood, causation is completely agent-relative. So, we would not apply the notion of absolute objectivity to causation (i) because causation is essentially anthropocentric and agent-relative; also, (ii) the notion of objectivity is itself contextual. Yet, causation is not absolutely subjective because once the notion of objectivity has been applied to a certain context, e.g., the practical sense of cause, whether or not A causes B will be an objective fact.

Indeed, Collingwood holds the same position on objectivity and causation as Woodward and Price. The consensus that I have revealed in this chapter regarding causation and objectivity is significant for two reasons. First, the debate between Woodward and Price is resolved—they emphasise different aspects of the objective/subjective debate, but I have shown that their respective positions support the same conclusion, viz., that causation is relatively objective. Second, the agency approach to causation is not subjective, either by design or by accident. This last result is quite significant. That the agency approach can capture a coherent sense of objectivity means that the agency approach does not entail a radical or untoward subjectivity, which is a common assumption in the literature on causation. Indeed, as I argued in chapter 3, any viable theory of causation must capture some sense of objectivity. Notably, we can now say that the agency approach is able to do so.

Chapter 5

Overcoming Foundationalism

5.1 Introduction

In the last chapter, I argued against the form of absolutism called objectivism. I argued that objectivism about causation was impracticable, based on an invalid assumption, and inimical to modern science. In this chapter, by focusing on the work of Alyssa Ney, I will argue against another form of absolutism, the doctrine known as foundationalism. Foundationalism is the idea that whatever exists can be reduced to one basic property or level.

To achieve this, I will answer three questions:

1. What is Ney's theory of foundationalism?
2. Can Ney avoid extreme foundationalism?
3. Is Ney's foundationalism successful?

§2 will introduce the traditional idea of foundationalism about causation and a modified version of foundationalism about causation provided by Alyssa Ney. In §3, I will outline Ney's moderated foundationalism. In §4, I will describe how Ney allegedly avoids traditional foundationalism by focusing on her ideas about difference-making. In §5, I will reveal why Ney's foundationalism fails; §6 will be used to motivate an alternative position.

5.2 Foundationalism About Causation

The debate regarding foundationalism about causation usually concerns whether our common sense causal claims or causal claims in the special sciences have a basis in fundamental physical theory. A traditional causal foundationalist believes the following:

Fundamental physical laws supply a causal foundation for all of the causal claims occurring in the special sciences and...every application of a fundamental physical theory must be interpretable in terms of a notion of 'cause,' possessing all of the features of the notion that figures in common sense and the special sciences. (Woodward 2007, p. 70)

By contrast, anti-foundationalists believe that causal claims of common sense and the special sciences have no basis in physics. For example, Woodward (2007) (and also, Hitchcock [2007] and Elga [2007]) argues that our ordinary causal claims have to reference the notion of an *intervention*. Since Woodward does not believe that the notion of an intervention can be analysed in terms of the concepts of fundamental physics, Woodward is an anti-foundationalist.

One of the most serious problems facing the traditional foundationalist concerns Russell's localisation argument. Fundamental physical laws are highly abstract and schematic. According to fundamental physics, for any high-level event e , the kind of event that factors into our ordinary experience, the range of physical influence(s) acting on e is so vast that it becomes practically immeasurable. Russell thought that this fact showed us that fundamental science did not support the conventional notion of cause. Another possibility, mentioned by Field (2003), is that, according to fundamental physics, we are simply wrong about causation. Field has us imagine two people standing near a fire. Both bystanders want to extinguish the fire. One bystander, Suzy, grabs a hose and begins to spray the fire with water. Another bystander, Sam, decides to pray for the fire go out. According to fundamental physics, our intuition that Suzy's spraying will make a difference to the fire and that Sam's praying will not is mistaken. We are mistaken because, according to fundamental physics, Suzy and Sam both make a difference to the fire. That is, both Sam and Suzy, regardless of their actions, have a physical influence on the fire.

The problem for the traditional foundationalist is that they are forced to reject the distinction we ordinarily insist on making between causes and background conditions or between significant and insignificant difference makers. As we have already seen, the distinction is vital to things like human action and effective strategies.

Traditional foundationalism is extreme and, unless one would be willing to concede that we are wrong about causation, indefensible. However, in a recent paper, ‘Physical Causation and Difference-Making’ (2009), Alyssa Ney argues for a moderated foundationalism that attempts to avoid some of the problems with traditional foundationalism, specifically the issue of localisation. I call Ney’s foundationalism about causation *moderated* because she allows for frame-dependence regarding the causal *relata*. Ney attempts to account for localisation by taking into consideration the ordinary *function* of causation. By taking into account specific interests—whether explaining, predicting, or acting—Ney believes that we can determine the causal *relata* and thus localise a particular causal relationship. Ney’s theory qualifies as foundationalism because she thinks that a (local) causal relationship, the relationship between *c* and *e* (a *cause* and an *effect*), is reducible to and ultimately grounded in physical facts. Because Ney claims that the causal relation depends on physical facts, she says that a causal foundationalist is anyone who *looks to physics* to discover the nature of causation or the nature of the relationship between *c* and *e*.

To support her theory, Ney identifies two kinds of causal facts: difference-making facts and physical facts. Ney’s argument is that difference-making facts *depend* on physical facts and that the nature of causation is physical. Ney says that

I will call this view about the relationship between physical causation and difference-making: ‘foundationalism about causation’. Causation is fundamentally a microphysical phenomenon, which can be discovered by looking directly at our fundamental physical theories. The obtaining of facts about difference-making depends on the obtaining of facts about physical causation (740).

Ney thinks that the most compelling argument for foundationalism is ‘based on the physicalist point’ that

Physics does not just provide us with a comprehensive account of what exists in the universe but an account as well of why

these events occur. That is, physics provides us with not only an ontology and laws for allowable synchronic states but dynamical laws as well. These laws single out those features of systems that are causally relevant to the production of effects. (2009, p. 757)

Dynamical laws tell us, given the initial conditions of some event, how a physical system develops or alters over time. Linear dynamics, for example, can explain the acceleration of a moving object, such as a bullet, by identifying (among *many* other things) the force of the small explosive charge in the primer of the bullet. Ney's idea is that dynamical laws can tell us about the nature of causation.

My arguments will show that Ney's foundationalism does not lead to the conclusion that, by looking *directly* at our fundamental physical theories, we can capture the nature of causation. Rather, what Ney's account reveals is that we can only make sense of a particular causal relationship in terms of physical facts *given* that we have specified some initial system by taking into consideration certain facts about difference making. Accordingly, Ney's conclusion that the nature of causation is physical is somewhat misleading. Ney repeats many times throughout her essay that for the foundationalist, 'causation is fundamentally physical' (p. 760). This gives the reader the impression that causation is *nothing above or beyond* physical facts. Yet, as I will reiterate, Ney thinks that we have to consider facts about difference making and ordinary pragmatics—the things we are interested in doing—to make sense of the causal relation. Assuming that *these* kinds of facts are capable of being described in terms of physical properties, Ney's argument may hold. However, I also intend to show that if Ney's argument does hold, what she says about the causal relation makes the nature of causation interactional (I elaborate on this point in 5.6). In other words, I intend to show that if Ney's argument is successful, the causal relation is *not* absolute or mind/agent-independent.

5.3 Ney's Foundationalism

We can already summarise Ney's basic position. This is my understanding of her primary claims:

Ney's Foundationalism Thesis (NFT):

(*P1*): When we want to capture the nature of causation, i.e., the relationship between *c* and *e*, we should look to physics.

(*P2*): When we look to physics, we can capture the nature of causation.

(*C*): Therefore, the nature of causation is physical.

According to Ney, (*P1*) is a necessary and sufficient condition for foundationalism about causation—one qualifies as a foundationalist by accepting (*P1*). Ney is a ‘positive’ foundationalist because she accepts (*P2*). A positive foundationalist also accepts (*C*), the idea that the nature of causation is physical. If one accepted (*P1*) but rejected (*P2*), they would be a ‘negative’ foundationalist. Ney calls Russell (1913) a negative foundationalist (Ney, 2009, p. 748). Collingwood, Woodward, and Price all reject (*P1*) and qualify as anti-foundationalists.

Throughout her essay, Ney assumes physicalism, the thesis that everything is physical. To understand how *NFT* avoids being a mere tautology, one must consider Ney’s important distinction between causal facts. For Ney, there are two kinds of causal facts: facts about difference-making and facts about physical causes. Naturally, Ney takes the physical facts to be fundamental. *NFT* is thus also a thesis regarding the relationship between difference-making facts and physical facts, viz., the claim that difference-making facts depend on physical facts. In essence, Ney believes that when we say (in ordinary language) that *A* caused *B*, *A* makes a difference to *B* because of the physical relationship between *A* and *B*. So, *NFT* will need to be modified to accommodate Ney’s ideas about the relationship between difference-making facts and physical facts, which I will do below after we have learned more about the alleged the relationship between difference-making facts and physical facts.

Ney considers her antagonist to be anyone who holds the view that ‘the physical causal facts are not fundamental’ (p. 740). Causal facts, from this view, ‘are necessarily facts about difference-making’ (ibid).¹ Furthermore, Ney argues that ‘Russellians’ about causation, essentially negative-foundationalists, are misguided. Ney thinks that Russell was right to look to physics to understand causation, but she argues that he was wrong to conclude that there was nothing in physics to ground causation (p. 747–52). Ney says that,

¹Ney suggests that David Lewis (2004) and Barry Loewer (2001) hold this view.

When we are looking for a physical theory of causation, we are looking for something in the fundamental physics that can tell us how a later event may evolve out of an earlier one (p. 751).

Note Ney's use of the term 'event'. One thing that distinguishes *NFT* from Russell (1913) is that it *assumes* localisation. That is, instead of asking whether or not our ordinary notion of cause is grounded in physics, like Russell, Ney asks, *given* that the causal relata is of the category *event*, can causation be grounded in physics?

There is a significant difference between the two questions. Russell's question was whether or not our ordinary notion of cause, which is localised and asymmetric, had a basis in fundamental physics. Ney's question differs from Russell in that she *assumes* localisation and asymmetry. She assumes localisation and asymmetry by granting our ordinary notion of events. While Ney correctly points out that she and Russell are in agreement regarding *P1*, it is implicit within her account that she and Russell are also in agreement regarding the answer to *Russell's* question. Our ordinary notion of cause (when causal relata are of the category event) is not grounded in fundamental physics. If we look to physics *tout court*, Ney says

[T]here are a lot of causal relations at this world, perhaps a lot more than we ordinarily assume. The fields of our best physical theories are spread out across the entire universe and interact with everything in their reach. They link small events like your leaving the house this morning with those more significant ones transpiring in Iraq a little later and more distant ones farther away in the galaxy. It is not quite true on this picture that 'everything causes everything,' but things come close (p. 741).

Ney says that the above quotation represents 'the first part of the foundationalist's account of causation, *but, not the whole story*' (ibid, my emphasis). To get the rest of the story and to see why Ney's position differs from Russell's, we must consider the alleged relationship between difference-making facts and physical facts.

5.4 Physical Facts and Difference Making

According to Ney's theory, difference-making facts and physical facts 'bear an important relationship to each other' (p. 739). On Ney's account, difference-

making facts are important because they eliminate most of the ‘multitude of causes’ (p. 741) that make up an event. Ney can avoid the extreme version of foundationalism, she thinks, because her theory accounts for facts about difference making and thus solves the problem of localisation.

Consider again what Ney says about events. When she says that, ‘we are looking for something in the fundamental physics that can tell us how a later event may evolve out of an earlier one’ (p. 751), the important question is, how do we determine the relevant event? If we *just* look to physics, as Ney is well aware, our concept of ‘event’ has no meaning (indeed, this is the lesson that we learn from traditional foundationalism). Events are localised, and fundamental physical relationships are not. One possible answer is that we can determine the events by considering facts about difference making *before* we look to physics. For example, according to Ney’s theory, it is correct, given particular facts about difference making, to say that Suzy’s spraying the fire with a hose made a difference to the fire and that Sam’s praying did not. Perhaps once we have considered these difference-making facts, we can look to physics to tell us how two events, e.g., spraying the hose and the fire going out, are fundamentally related.

While this interpretation seems plausible, Ney would argue that it mistakenly inverts her view. She would say that the foundationalist does, in fact, look to physics to tell us about causation *before* she looks to facts about difference makers. We look to physics and discover the vast array of physical causes that exists at the fundamental level, *then* we ask, which out of the wide range of physical causes made a difference to the effect? I think that this is what Ney would have us believe. Indeed, she says that

According to the foundationalist, a theory of causation must start with an account of what fundamental physics tells us about the causal relations that obtain at our world. Physics does not just tell us what there is and what happens. In other words, it does not merely give us an inventory of the fundamental properties or kinds of entities there are, and the types of events that take place. It also tells us why these events occur—what is responsible for what in a causal sense. This is the first part of the foundationalist’s account of causation, but not the whole story (p. 741).

In reply to Ney’s paper, Luke Glynn (2013) argues that Ney’s theory indicates that we must *first* look to difference-making theories, likely because

this will allow us to make sense of the notion *events*. Glynn argues that, given his interpretation of Ney’s theory, what it shows, contrary to Ney’s claim, is that the nature of causation is difference making. However, given that Ney explicitly denies this, I will grant her position—a foundationalist theory of causation must start ‘with an account of what fundamental physics tells us about the causal relations that obtain at our world’ (ibid). However, if we are ‘looking for something in fundamental physics to tell us how a later event may evolve out of earlier one’ (p. 751), the question of how to determine the relevant events remains. Ney does provide an answer. Ney thinks that we have to consider the standard *function* of causation. That is, she thinks the foundationalist needs to take into consideration how we are using the notion of cause, *given* our particular interests. Ney says that ‘in ordinary circumstances,’ we are interested in ‘the explanation and prediction of large macroscopic events, and the planning of our own actions’ (p. 742). So, by considering our interests—be it explaining, predicting, or acting—the foundationalist can establish the relevant events.

A summary of Ney’s position thus far looks like this: when asked, ‘why did the fire go out?’, the foundationalist can look to physics to understand how the event, *Y* there being no fire, evolved out of the earlier event, *X* there being a fire. *Then* a difference-making theory can tell us which of the fundamental facts made a difference. The dynamical laws that tell us about the relevant difference-making facts reveals the nature of causation.

Initially, we said that Ney’s thesis, the idea that the nature of causation is physical, followed from Ney’s idea that the only way to understand causation was to look to physics. Given what we have learned, *NFT* must be modified.

Ney’s Foundationalism Thesis (NFT*)*:

- (P1): When we want to capture the nature of causation, i.e., the relationship between *c* and *e*, we should look to physics **and to facts about difference making**.
- (P2*): The causal relata depend on our interests. By considering our interests, we can determine the causal relata and the relevant events. (hidden premise)
- (P3): Facts about difference making supervene on physical facts.
- (C): Therefore, the nature of causation is physical.

To accurately capture Ney's argument, the original argument *NFT* had to be modified. We added $(P2^*)$ so that we can determine the relevant events to show how, again in Ney's words, 'one event evolved out of an earlier one' (751). We also added the caveat, 'and to facts about difference making', to $(P1)$.

5.5 The Fate of Ney's Foundationalism

The problem for moderated foundationalism should now be apparent. Ney's thesis that *the nature of causation is physical* does not easily follow from $(P1)$ – $(P3)$. On the surface, *NFT** appears to be invalid.

A further problem arises when we consider what Ney initially claimed about foundationalism about causation. If you recall the quote from §2, Ney claimed that 'causation is fundamentally a microphysical phenomenon *that can be discovered by looking directly* at our fundamental physical theories' (p. 740, my emphasis). This is a claim that seems to directly contradict *NFT**. It is not true, given $(P1)$ – $(P3)$, that causation, if it is a microphysical phenomenon, can be captured by looking directly at our fundamental physical theories. If we want to derive a physical theory of causation from $(P1)$ – $(P3)$, it cannot be obtained by simply looking to physics. According to Ney's theory, we also consider our ordinary interests to determine the causal relata as well as certain facts about difference making, which tell us about contextually related background conditions (I argue for this point regarding contextually related background conditions below).

However, despite the fact that we cannot derive a physical theory by looking directly at physics, Ney could still argue that *NFT** holds. *NFT** is plausible, according to Ney, because she thinks that we must assume $(P3)$, viz., that facts about difference making supervene on physical facts. Ney accepts supervenience because she assumes physicalism, the idea that everything is physical (she says as much in a footnote on page 740). If we grant $(P3)$, it is plausible that (C) follows.

The problem for the moderated foundationalist is that they have not provided a compelling reason for us to allow an *exclusively* physical description of causation. In fact, it could be argued that both traditional and moderated foundationalism provide compelling reasons to *reject* a physical interpretation of causation. Indeed, as a physicalist, Ney is intensely aware of the problem. She understands how traditional foundationalism is both extreme

and undesirable. Accordingly, Ney must admit that ‘there is a bit more that needs to be said about causation *beyond whatever can be learned from physics*’ (741, my emphasis).

As we have seen, the *bit more* is just what a standard counterfactual or probabilistic theory of difference making tells us about causation. For example, a counterfactual or probabilistic theory of difference making will say that the cause of the fire going out was likely that Suzy sprayed the hose. However, before we make this judgment, the foundationalist believes there is still more to be said. We must specify particular interests, i.e., how we want to use the notion of cause—whether we want to explain something, predict something, or perform an action. My argument is that once we are forced to rely on standard difference-making ideas and ordinary pragmatics, the claim that *the nature of causation is physical* is misleading and ultimately dubious. To see why, consider what we are typically interested in doing with causation. As agency theorists like to point out, the notion of cause is useful because it serves a unique function. We want to explain things like why the forest burned down or why forests sometimes burn down; we want to predict earthquakes, stock market crashes, and sporting events, and we want to prevent things like cancer, wars, poverty, and famine. Note that the foundationalist will not be interested in doing any of these things *per se*. Foundationalism is purely a metaphysical theory about the nature of causation.

Before we move on, one particular point needs to be made clear. As a physicalist, Ney is limited to a narrow metaphysics. She believes that (or accepts something like the thesis) ‘everything is physical’ or ‘the real world consists simply of the physical world’. For the physicalist, everything supervenes on or reduces to the physical. The mind supervenes on the physical properties of the brain, colour supervenes on physical properties of light and (some combination of) physical properties of observers. Note: I do not take Ney’s thesis, *the nature of causation is physical* to be the claim that, (i) whatever causation is, it is physical. Nor do I take that claim to be overly controversial. Rather, by claiming that the nature of causation is physical, I take Ney to be arguing that (ii) we can capture the essence of the causal relation, i.e., the indefeasible quality which determine its being so, by understanding a (certain) combination of facts. Claim (i) is not something most philosophers, including agency theorists, would deny (causation isn’t ‘spooky’ for example). Furthermore, if Ney were arguing for (i) her argument would be viciously circular. Her position (ii) differs from (i) in that she is

claiming that we can capture the causal relation simply by understanding the physical properties or facts of which it is associated. The important thing to note, something I have been trying to indicate above, is that Ney's thesis (moderated foundationalism) takes the causal relation to be associated with *more* than just whatever our best physical theories tell us about the (mind-independent) world. The *more*, as I did indicate above, concerns facts regarding difference making relationships and ordinary pragmatics. Ney is a positive foundationalist because, *once she has made the move to include difference making and pragmatic facts into the causal relation*, she can claim (*via [P3]*) that the nature of causation is physical. Seeing Ney's position in this way always us to discover a critical point. The point is this: what Ney is really saying is that we can capture the causal relation by looking at physical facts *given* that we include facts about difference making and facts about pragmatics and *given* that these facts supervene on physical properties.²

Allow me to explain all of this with an analogy. Consider the ordinary notion of colour. Many philosophers and scientists (e.g., Descartes, Locke, Galileo) have argued that colour is a secondary quality. That is, colours (like all secondary qualities) are something that can only be understood in terms of how they appear. Primary qualities (Locke uses the example of shape) can be understood independently of how they appear. In terms of an objects' qualities, say, those of a chair, its primary qualities, e.g., its height and its length, are taken to be mind-independent. That is, if all observers were removed, the chair itself would not disappear. This is not so for secondary qualities. The chair may have some "power" to affect our senses—say, that it looks blue. But blue is a secondary quality because if we removed the observer(s), the "blueness" of the chair would likewise be removed. So, and to my point above, if we wanted to capture the nature of colour (e.g., blueness) by looking to physics, we couldn't *just* look to those physical facts which were mind-independent. We would also have to consider some complicated set of physical facts about the observer. In relation to foundationalism about causation, the traditional foundationalist about causation seems to think of the causal relation as a primary quality—something in the natural environment that exists independently of the mind. What Ney seems to do with her modified account of foundationalism is to put the observer, so to speak,

²In some way, Ney's thesis is circular (e.g., she assumes physicalism and argues that the nature of causation is physical). Whether or not it is viscously so will depend on how close one takes her to be defending (i).

back into the equation. It's as if Ney is saying, following the colour analogy, we can capture the nature of colour in terms of physical facts *given* that we consider (along with the physical properties of light) certain (physical) facts about the observer. In fact, this is just what we see in terms of her thesis regarding causation. When we want to capture the nature of causation, Ney admits that we cannot just look to our best physical theories, perhaps as Russell did in 1913. Ney argues that to make sense of the causal relation in terms of physical facts (like we would with colour) we have to put something else into the equation. To reiterate once more, the something else in question is facts about difference making and ordinary pragmatics.

Following the colour analogy, one could argue that Ney's theory amounts to the idea that causation is something like a secondary quality. For example, one could argue that if we removed human agents from Ney's theory, we would lose the ability to sensibly speak about causation (i.e., we would be forced to accept traditional foundationalism). However, I won't pursue this argument here, as it could easily misconstrue Ney's emphasis on objectivity. Although I did indicate in the previous chapter that we could speak objectively about colour (and causation) *given* that we specify context, I think a more accurate representation of Ney's theory is to say that it's *interactional*. What I mean to convey by 'interaction' is that Ney's theory shows us that the causal relation is something that emerges from, or is constituted by, a specific interaction. I will explain this in more detail below.

Now that this is hopefully more clear, I want to investigate the relationship between physical facts, difference making, and the nature of causation. In doing so, I will argue that Ney's thesis leads to the conclusion that the nature of causation is interactional.

5.6 Foundationalism and Interactionism

Until now, the relationship between physical facts, difference making, and the nature of causation has seemed mostly untroubled. I want to challenge this conception and argue that, given the foundationalist's reliance on pragmatics and difference-making theories, the idea that the nature of causation is physical is somewhat fraudulent.

If the nature of causation is physical, as the traditional foundationalist claims, causes can be thought of in terms of natural or physical relationships that exist in the world independent of human beings. Much like Colling-

wood's theoretical sense of cause, the traditional foundationalist pictures causes as *real*, or, as Putnam (1982) remarks, some 'furniture of the world'. My argument is that once we are forced to consider difference-making facts and ordinary pragmatics, the idea of mind-independent, *real* causes existing in nature is inconsistent and false. It may be that a specific causal relationship depends on physical facts, but these facts are not absolute or mind-independent.

Mackie claimed that what we call a cause is 'what makes the difference in relation to some assumed background or causal field' (Mackie 1974, p. 35). To illustrate, consider what Hart and Honoré (1985) called *relativity to the context of inquiry*. The context of inquiry tells us that one and the same situation elicits different judgments about difference making depending on the type of enquiry. Here is an example used by Menzies, in the article, 'Difference-Making in Context' (2004), where he defends Hart and Honoré's idea that causation is relative to context:

The Indian Famine

The cause of a great famine in India may be identified by an Indian peasant as the drought, but the World Food Authority may identify the Indian government's failure to build up reserves as the cause and the drought as a mere condition (Menzies, 2004, p. 144).

When we ask the question, *what caused the famine*, we will have two different answers.

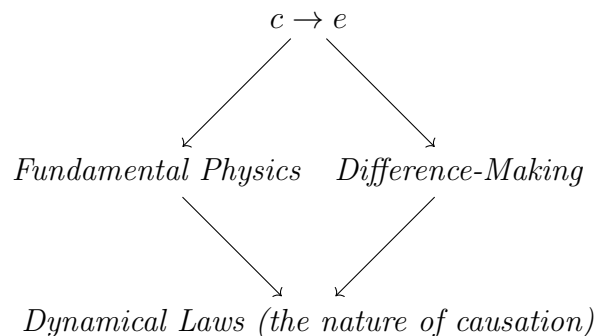
In a similar example, Putnam tells a story about a fire. Putnam has us imagine that Venusians land on Earth and observe a forest fire. One of the Venusians says, '*I know what caused that—the atmosphere of the darned planet is saturated with oxygen.*' Putnam says that this shows us that one man's (or extraterrestrial's) background conditions can easily be another man's cause. What is and what is not a cause or an explanation 'depends on background knowledge and our reason for asking the question' (Putnam, 1982, p. 150). When we ask the question, *what caused the forest fire*, we will have two different answers.

The idea that causation is relative to a type of context is something that agency theorists have embraced. Agency theorists argue that causation is relative to a context and that the context in question has something to do

with human agency. Collingwood, for example, describes causes as *handles* because they allow us to produce/prevent certain effects. Handles (or causes) are relative to what an agent knows how to produce or prevent. In other words, something makes a difference because it acts as a handle. In the example of the car accident, different agents come up with different answers to the question of what caused the accident. For Collingwood, there is no such thing as a *real* cause because the practical idea of causation presupposes the notion of a handle. Also, *real* causes are thought to be absolute and independent of context, which repudiates the idea of a handle.

Ney could argue that the relativity of causation is not problematic for the moderated foundationalist. She could argue that, given $P2^*$, we can determine the relevant context (e.g., the person asking the question). Furthermore, she could argue that, given that we have considered certain facts about difference making in $P1$, we can also determine the relevant background conditions. Also, assuming $P3$, we can grant C (this would amount to a rejection of the claim that the foundationalist holds an absolute position on causation or has an inherent interest in *real* causes).

Consider a representation of Ney's theory, as outlined above:



The representation was constructed as follows: $c \rightarrow e$ represents the idea that we have granted the events that make up the *causal relata* (by taking into consideration our ‘ordinary interests’). After this, we move to **Fundamental Physics** to discover ‘the multitude of causes that exist at the fundamental level’, and to **Difference-Making** facts that provide us with the necessary means to determine the relevant background conditions and which of the physical causes made a difference. Lastly, because Ney believes we can describe the relevant difference-making facts in terms of physical, **Dynamical Laws**, she claims that the nature of causation is physical.

Table 5.1: Causation and Objectivity

<i>Subjectivism</i>	<i>Relative Objectivity</i>	<i>Absolute Objectivity</i>
	Collingwood Price Woodward	Strevens Ney

Notice again the position of Ney’s theory, table 5.1.

Ney is placed under *absolute objectivity* because she thinks that the causal relation is physical and absolute. The important thing to note is that, given my critical review of her theory, Ney’s position in the table is incoherent. That is, much like Strevens, what Ney *actually says* about the causal relation does not lead to the conclusion that the causal relation is frame-independent or absolute. Because Ney has to rely on difference making theories and ordinary pragmatics, we cannot say that her position, foundationalism about causation, represents absolute objectivity about causation.

If my argument holds, Ney’s reliance on difference-making facts would amount to a rejection of the idea that the causal relation is mind-independent and real, and indicate that the causal relation is *interactional*. That is, according to my analysis of Ney, the causal relation appears to be something that emerges from a particular interaction. To better understand what I mean by ‘interaction’, consider what Stephen Palmer says about colour interactionism in his book, *Vision Science: Photons to Phenomenology*

The colors we see are based on physical properties of objects and lights that cause us to see them as colored, to be sure, but these physical properties are different in important ways from the colors we perceive. Color is more accurately understood as the result of complex *interactions* between physical light in the environment and our visual nervous systems (1999, p. 95, my emphasis).

Palmer is arguing that colour is something which emerges from an interaction between the physical world and the human visual system.

In Ney’s case, I’m arguing that the interaction involves human agents—what kinds of questions we ask, what we want to explain, or how we might act—and the physical world. For example, the truth-makers for the claim, ‘the cause of the fire going out was Suzy’s spraying’, are not mere (physical)

facts about the world. If they were, it would also be true (to some extent) that the cause of the fire going out was Sam's praying (this is because both Sam and Suzy, by their mere presence, have an effect on the fire). Rather, the truth-makers for the claim, 'the cause of the fire going out was Suzy's spraying' are based on facts about (the interaction between) human agents and the physical world. Ney's theory is an asymmetric version of interactionism because she emphasises the physical aspect due to her other philosophical inclinations (viz., physicalism). Even still, if we removed the human agents from the interaction, there would be no way to make sense of difference making facts and the ordinary pragmatics of causal enquiry. If we removed these facts, as Ney herself indicates, the 'bit more' that needs to be said in order to avoid traditional foundationalism—which Ney rejects—would be lost. And once this is lost, according to Ney's own theory, we lose the ability to capture the nature of causation.

Because difference-making ideas are contextual and mind/agent-dependent, it follows that the causal relation is frame-dependent and non-universal. I did not place Ney under *relative objectivity* because she does not endorse a relative position, although it may be the case that foundationalism about causation fits better under this category. Again, the important point to note is that her theory, along with Strevens', does not amount to a coherent position for *absolute objectivity* (I have placed a line through both Ney and Strevens to represent this).

This leaves the moderated foundationalist with a dilemma. She can (i) resort back to traditional foundationalism, or (ii) reject any form of foundationalism altogether. Given my arguments, I take the most reasonable choice to be (ii), a rejection of foundationalism.

It is interesting to note that the rejection of foundationalism, the idea that the nature of causation is physical, does not lead to the idea that the nature of causation is difference making (as Ney thinks) or to the idea that the nature of causation is somehow anthropocentric. Rather, what it demonstrates is that causation involves a complicated relationship between different kinds of facts. In other words, it supports the idea that causation requires a certain kind of *interaction*. I will explore and defend a version of *causal* interactionism in chapters 7 & 8.

Though it is clear that Ney herself does not support an agency approach to causation, I take her position (as I have outlined it) to be compatible with an agency approach. If my argument, which says that Ney's position leads to the conclusion that the nature of causation is interactional, is successful,

there are obvious parallels between moderated foundationalism and more familiar agency approaches (e.g., Collingwood, 1940). Given my arguments, what I wish to point out is that, rather than being seen as an argument against the main tenets of the agency approach, Ney's position offers latent support for the idea that causation is agent-dependent.

5.6.1 A Note on Fundamentality

This chapter has focused on Ney's versions of foundationalism. However, a brief note on her use of the term 'fundamental' may be helpful. Note what Ney says about fundamentality. She claims that 'Causation is *fundamentally* a microphysical phenomenon, which can be discovered by looking directly at our fundamental physical theories' (p. 740, my emphasis).

So, what does Ney mean by 'fundamentally'? The notion of fundamentality (in metaphysics) typically aims to capture the idea that there is something basic or primitive in the world. Given Ney's claim, and the use of 'fundamentally' above, we could interpret Ney to mean that causation, at the most basic level, is physical. I take it that Ney would support this claim. However, given the worries for traditional theories of foundationalism about causation presented above, particularly those related to Russell's localisation argument, the claim that 'causation is fundamentally physical' needs to be elaborated.

Ney is, of course, aware of this, as her essay is rightly taken to be a defence of foundationalism, specifically one that can handle localisation requirements. Since one essential task for the notion of fundamentality is to help us articulate the view that there is a hierarchical structure to reality, whatever Ney means by 'fundamental', she *is* indicating that she takes physical facts about the world to (somehow) ground the causal relation. However, as we saw in this chapter, Ney's foundationalism is nuanced by the fact that her theory relies on facts about difference-making. While there are currently many ideas regarding fundamentality (e.g., *Primitivism* [see Glazier, 2016; Schaffer, 2009], *Absolute Independence* [see Bernstein, 2016], and a weaker version of fundamentality sometimes called *Restricted Independence* [see Dixon, 2016; Bennett 2017]), Ney's foundationalism aligns best with a version of fundamentality known as *Well-Foundedness* (see Raven [2016]; Jago [2018]).

As I mentioned above, one of the critical tasks for the notion of fundamentality is to capture the idea that there is a foundation of being and

that everything else depends on the fundamental entities. This idea about fundamentality is often expressed concerning well-foundedness. A standard formulation of this general idea is as follows: a priority/dependence chain is well-founded if and only if it terminates, i.e., has an end constituted by one or more entities that do not depend on any other entity.

Regarding Ney's use of 'fundamentally', I take her claim to be that the 'end' or entities that do not depend on any other entities are physical facts. Likewise, she maintains that facts about difference-making are non-fundamental in the sense that they depend on physical facts.

Granting its plausibility, I do not take Ney's claims regarding fundamentality to threaten the agency approach to causation. As I have indicated, whatever a particular causal relationship turns out to be (say a relationship between two events), it is a further question to inquire about the well-foundedness of these events. Ney's thesis should thus be taken as two distinct arguments. She argues (1) that we need facts about difference-making to establish causal claims (in regards to the question, how can we improve traditional theories of foundationalism?). And she argues (2) that these claims are well-founded in physics (in regards to the question, what are the basic entities that make up the causal relation?). Note that the second question is one that manipulationists are generally not interested in pursuing. But the lack of interest should not indicate a refutation of the claim that causal relationships (once they have been established) are well-founded in physics. The claim that agency theorists refute, in regards to the *first* question, is that the causal relation is *merely* physical—that agency does not do any of the work in establishing causal claims.

Note further that, for Ney, these two questions have to be asked simultaneously. Given Ney's moderated version of foundationalism, we have to ask both questions. For if we only inquire about the basic entities that make up the causal relation, that is, if we inquire about well-foundedness *without* establishing specific causal relationships, we would be led to traditional foundationalism, which Ney rightly rejects.

In regards to Ney's foundationalism, my argument has been that Ney's foundationalism is misleading because her two arguments often seem conflated. Ney's repeated claim that we can discover causation by looking directly at physics is about the question regarding fundamentality. In regards to the first question, how can we improve traditional theories of foundationalism, part of the answer is *not* by looking directly at physics. The other part of the answer to this question, as Ney herself is aware, is that we must rely

on difference-making facts and human agency. Moreover, I don't see any reason why agency theorists should be worried about further enquires regarding fundamentality. Indeed, they should be able to accommodate whatever our best scientific theories have to say about the matter.

Chapter 6

Agency and Anthropocentrism

6.1 Introduction

In chapter 3, I argued that the agency approach faces two critical problems: the problem of subjectivity and the problem of scope. In chapter 4, I suggested that the problem of subjectivity could be overcome by distinguishing between two kinds of objectivity: absolute objectivity and relative objectivity. I defended the idea of relative objectivity by arguing that the notion of absolute objectivity (regarding causation) was misguided, incompatible with modern science, and the result of a false dichotomy. My arguments supported the idea that the agency approach can be seen as objective in the relative sense. In the previous chapter, I showed how foundationalism about causation, the idea that the nature of causation is physical, is flawed, and I argued that Ney's moderated theory is compatible with an agency approach.

In this chapter, I will suggest that the problem of scope rests on a certain *type* of agency, pointing out that the problem does not necessarily arise for an agency approach to causation. Because the problem of scope is similar to and often conflated with the so-called problem of anthropocentrism (Woodward, 2003; Price, 2017), I will explain anthropocentrism in the context of the agency approach to causation and reveal its relation to the problem of scope. I do not intend to solve the problem of scope in this chapter (I do attempt to solve it in Chapter 8). In this chapter, I show how the problem of anthropocentrism *can* lead to the problem of scope, and I suggest that the agency approach can avoid the problem. My arguments will show that the agency approach is anthropocentric, but not in a way that is untoward or in

such a way that it necessarily leads to the problem of scope.

Subsequently, I will clarify the notion of ‘anthropocentric’ and investigate the anthropocentrism associated with the agency approach to causation. To support my argument, I will distinguish between three different views of anthropocentrism: a common view, a projective view, and a metaphysical view. I will also distinguish between two kinds of agency: overt agency and mental agency. These distinctions will allow me to show that Woodward’s interventionist account is anthropocentric according to the common view and that Woodward’s interventionist account relies on the notion of agency. I will achieve this by investigating three questions:

1. How can we understand ‘agency’ and ‘anthropocentrism’?
2. In what way are the accounts of Collingwood, Price, and Woodward anthropocentric?
3. Does interventionism involve agency?

In §2, I will reveal three different ways to think about anthropocentrism: *the common view*, which says that causation is anthropocentric when we give human agents a privileged status (i.e., we take a human point of view) regarding causation; *the projective view*, which says that causation is anthropocentric because we project human qualities onto the natural world; and *the metaphysical view*, which says that causation is anthropocentric because causal relationships are created by human agents (in some way, human agents “make” the truth-makers of causal claims). I will also distinguish between two kinds of actions or agency: overt actions, which include bodily movements like walking or throwing, and mental actions, which include reasoning about the effects of bringing something about, e.g., reasoning about what *would* happen if I were to ϕ . In §3, I will locate the anthropocentrism within the manipulationist accounts of Collingwood, Woodward, and Price. I will show that all of the accounts are anthropocentric according to the common view, only Price’s account is anthropocentric according to the projective view, and how all of the accounts avoid anthropocentrism according to the metaphysical view. §4 will be used to clarify the notion of agency, and to elaborate and defend the two notions of agency that I endorse. I will argue that mental agency is real agency by relating the notion of mental agency to the standard definition of (overt) agency. Finally, in §5, I will clarify and highlight the agency found in Woodward’s interventionist account, arguing

that Woodward's account of interventionism relies on the notion of (mental) agency, and I will conclude by showing that the agency approach does not support an untoward anthropocentrism, contrary to Woodward's claim.

6.2 Agency and Anthropocentrism

Thus far, I have merely suggested, following the criticism of Woodward, that the agency theory is unduly anthropocentric. It will be important now to see whether or not this is true, and to clearly understand the connection between agency, manipulation and anthropocentrism.

To introduce the problem of (untoward) anthropocentricity, consider a well-known objection made by Woodward. Regarding the manipulation approach to causation, Woodward (2003, 2013) thinks that agency theory fails because it is unacceptably anthropocentric. He says:

[Price's agency theory] flies in the face of any plausible version of naturalism: it makes agency out to be a fundamental, irreducible feature of the world...it leads us toward an undesirable anthropocentrism or subjectivity regarding causation (Woodward, 2003, p. 123).

Although Woodward seems to conflate the issue of anthropocentrism with subjectivity, it is important to understand what Woodward means by *undesirable*. Notably, Woodward's criticism of Price's use of agency is in regards to Price's solution to the problem of spurious causes, specifically in the context of a manipulationist approach to causation. If you recall from chapter 3, Price argues that we can overcome the problem of spurious causes by introducing the notion of agency, i.e., a 'free action,' into a probabilistic theory of causation. According to Price, a free action can break the correlation between X (an alleged cause) and Y (an alleged effect of X) by creating an 'independent causal history' (ibid). In the context of spurious causes, Price is arguing that a free act can break the correlation between X and Y by acting on or manipulating X in such a way that the correlation between X and Y disappears. Allegedly, the correlation will disappear because the agent's free act will create a level of independence or separation (an independent causal history) between X and Y . Woodward's criticism is that the notion of *acting on X* to create the necessary independence, and thus break the correlation

with Y , is too anthropocentric and thus fails as a solution to the problem of spurious causes.

What this means is that Woodward thinks that Price's use of agency, *as a solution to the problem of spurious causes*, is unduly anthropocentric. If we break the correlation between X and Y by a free act, it simply does not make sense to talk about causation between events where an agent's free act is incapable of creating the necessary independence, i.e., when an agent would not be capable of acting in such a way as to distinguish causes from correlations.

Significantly, Woodward's argument that Price's approach is unduly anthropocentric leads to the problem of scope. The problem is not that Price's account relies on agency *per se*. Rather, according to Woodward, the problem is that when we base the solution to the problem of spurious causes on the notion of free action, the solution becomes dependent on the manipulation capabilities of human agents. According to Woodward, there is no way for us to distinguish between causes and correlations that do not involve a human action, and thus we unduly limit the scope of what we are allowed to call causes.

While Woodward's criticism of Price is significant, I believe that we have to be much more clear about what we mean by *anthropocentrism* and by *agency* because there is no single understanding of the word 'anthropocentric,' nor is there a unique understanding of the word 'agency.'

There are three ways that something can be anthropocentric.

The Common View: Anthropocentrism means human beings enjoy a central and privileged status in the world. For example, many Christians believe they were created in the *imago dei* and that this gives them a privileged status over animals and the natural world. In terms of causation, this view says that human agents are regarded as having a central and privileged status (e.g., over things like metaphysics—i.e., the nature of causation—and the relationship between causation and the natural world). Anthropocentrism entails a privileged status for the human 'point of view' regarding the causal relation.¹

¹Because these terms can easily be misconstrued, my use of 'metaphysics' will mean *ontology*; also, when I say the 'natural world', I mean *physical facts* about the world. Moreover, I say that human agency is given a privileged status over things like metaphysics and the natural world rather than things like astrology and 14th century literature because there is a common conception in the literature on causation that when thinking about

The Projective View: Anthropocentrism means human characteristics are being *projected* onto something that is non-human. For example, when ancient people used to think volcanic eruptions were caused by gods, or when the pilgrims blamed a witch for their poor harvests, their thinking was anthropocentric because they believed that these events came about by some specialised agency. Regarding causation, the projective view says that we project a kind of power that naturally belongs to human agents onto the natural world.

The Metaphysical View: Anthropocentrism means anti-realism. The anti-realism is taken to mean that the world is only knowable through our cognitive faculties. This is true for Berkeley, Kant, and Nelson Goodman, to name only a few. To illustrate, Goodman's 'irrealism', a development of Kant's transcendental idealism, stated that the empirical world was somehow human, 'made by us' as he put it (Goodman, 1978). Regarding causation, the metaphysical view says that the causal relation is somehow created by human agents.

To help us understand Woodward's charge of unacceptable anthropocentrism and the problem of scope, we can also distinguish between two kinds of agency.

Overt Agency: Overt agency includes intentional actions. Intentional actions are deliberate bodily movements made by human beings. These actions include raising one's arm, moving one's leg, and more complex movements such as walking, throwing, opening, pushing, carrying, pulling, taking, etc. Overt actions allow us to control or achieve certain outcomes or effects. For example, I can raise my heart rate by running up a large hill, or I can reduce the chances of catching a cold by taking vitamin C. We can say that overt agency provides us with practical control—when I do *X*, I will achieve *Y*.

causation, we are at least tacitly concerned with metaphysics and the natural world and not with things like astrology or 14th century literature. As I mentioned in chapter 3, influential theories of causation such as Salmon's mark theory of causation and Dowe's process theory of causation suggest that what's important about causation is physical facts. And as we saw in chapters 4 & 5, people like Strevens and Ney, among others, have argued that what's important about causation are metaphysical facts—ontological theories about what causation is or what grounds causation in the natural world.

Mental Agency: Mental agency is the process of reasoning about the effects of bringing something about. We reason about the effects of bringing something about to enable *understanding*. Significantly, we can reason about the effects of *possible* actions, such as changing your diet, and *impossible* actions, such as altering the distance between the moon and planet Earth (we can also reason about non-actions, like imagining that you are in France). Reasoning about possible or impossible actions enables control over some outcome or effect. For example, you have control over your weight after you reason about the effects of a particular diet, or you have control over the tides on Earth after you reason about altering the distance between the moon and Earth. We can say that mental agency enables *hypothetical control*—*if* I did *X*, I *would* achieve *Y* (I will discuss the concept of mental agency in §4).

Having made these distinctions, we can now assess the anthropocentrism and agency that is associated with the accounts of Collingwood, Woodward, and Price. I will also reveal why Woodward interprets Price's use of agency as problematic by arguing that Woodward takes 'agency' to mean overt agency.

6.3 Anthropocentrism in Collingwood, Woodward, and Price

I will begin my discussion of anthropocentrism by looking at Collingwood's practical idea of a cause. Collingwood is the most explicit about anthropocentrism, and his account shares important similarities with both Woodward and Price. In many ways, as I will argue in §5, Collingwood can serve as a useful model when thinking about things like agency, anthropocentrism, and metaphysics. Following the discussion of Collingwood, I will analyse Woodward's interventionist account and finish with Price, who holds a slightly more complicated view of agency and causation.

6.3.1 Collingwood's Anthropocentrism

Collingwood tells us that 'that which is caused (in the practical sense) is an event in nature, but the word 'cause' still expresses an idea relative to human conduct, because that which is caused is something under human control' (1940, p. 296).

To better understand this sense of cause, consider again Collingwood's example of the automobile example. When a county-surveyor, an automobile manufacturer, and a driver meet at the site of the accident, they all have different answers to the question of what caused the crash. Why? The answer lies in Collingwood's definition of a cause in the practical sense. A cause is something in nature, but it is relative to what an agent can produce or prevent. The driver will think that speeding caused the crash because he can prevent speeding; the surveyor will think that defects in the road caused the crash because he has an in-depth understanding of roads, and the manufacturer will think that the car's design caused the crash, since she is an expert in car design.

Now imagine that a bystander entered the scene and said, 'It is true that you feel a sense of control over the things you understand and can manipulate, but beyond that, beyond the mere control of things, there is a *real* cause of the car crash, and, ultimately, that is what we should be concerned with. So what is the real cause of the crash?'

As we have seen, Collingwood thinks the idea of a *real* cause is nonsensical. Unlike the generality, and relativity, of the word 'cause' in the practical sense, theoretical causes, which are purportedly real causes in nature, are 'tight' or absolute (see Chapter 2§3). Collingwood thinks that there is no sensible way for us to answer the question regarding "real" causes because it is an absolute presupposition of practical sciences like law and engineering that causes are handles. So, in the practical sense, we are concerned with what agents can manipulate and control rather than metaphysical truths or "causes" that exist independently of human agents.

This reveals that Collingwood's account is anthropocentric according to the common view. Regarding causation, human agents enjoy a privileged status. That is, we investigate the causal relation not by looking to nature itself but by looking at nature through the eyes of a practical agent. Collingwood clarifies this point nicely. He says that the practical idea of cause rests on an important idea about the relationship between man and nature. The idea is this:

The anthropocentric idea that man looks at nature from his own point of view; not the point of view of a thinker, anxious to find out the truth about nature as it is in itself, but the point of view of a practical agent, anxious to find out how he can manipulate nature for his own ends.

Ultimately, the word cause means the following:

Something that exists in nature, though [something] that is relative to what an agent can produce and prevent (Collingwood, 1940, p. 321).

Technically, and if we follow Collingwood's philosophy, it is not correct to view Collingwood's account as non-realist or anthropocentric according to the metaphysical view described above. Collingwood's idea of metaphysics was the historical investigation of absolute presuppositions and precluded ontology. To call Collingwood's account of causation 'metaphysically non-realist' would be a mistake. Collingwood was interested in describing the presuppositions that guide our practical investigations, not ontological truth (here, I only intend to show how Collingwood avoids metaphysical anti-realism. In a later chapter (7), I will explain my position on realism).

It should be noted that Collingwood thinks the idea of *theoretical* causes (causation in sense three) is anthropocentric according to the projective view. Theoretical causes are anthropocentric, Collingwood argues, because when the animistic conception of nature, the idea that natural events were caused by supernatural agency, was replaced by the mechanical conception in the seventeenth century, the agency or *power* that naturally belonged to man and was once attributed to supernatural beings was transferred to the natural world. Thus, Collingwood says

When we come to Newton, and read (e.g.) the *Scholium* appended to his Definitions, we find him using as a matter of course a whole vocabulary, taken literally, ascribes to 'causes' in nature a kind of power which properly belongs to one human being inducing another to act as he wishes him to act. Causes are said, in the twelfth paragraph in that *Scholium*, to be 'forces impressed upon bodies for the generation of motion. True motion is neither generated or altered, but by some force pressed upon the body moved'. The cause, for Newton, is not that which impresses the force, but the force itself. (ibid, p. 325)

Collingwood thinks that theoretical causes that are purported to be agent-independent are anthropocentric in the sense that we project a power that naturally belongs to human beings onto nature. Therefore, theoretical causes are nonsensical because the idea of causation involves 'a relic of animism' (ibid, p. 327) and yet is claimed to be agent-independent.

6.3.2 Anthropocentrism in Woodward

Recall what was said about interventionism in §2.5. Interventionism allows an agent or a scientist interested in a particular set of questions to discover effective strategies, or what Woodward sometimes refers to as *difference-making* strategies. The idea behind difference-making strategies is that causal claims are understood as claims about what would happen (or what would be different) ‘under interventions on...one or more variables’ in a causal model (ibid, 73).

To illustrate, the causal model in Figure 2.1 shows that an intervention on X would make a difference to Y and Z , whereas an intervention on Z would make no difference to Y . Thus, according to the model, we can infer a causal relationship between X and Y and a non-causal relationship between Y and Z .

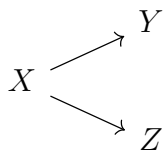


Figure 2.1

In Figure 2.1 and 2.2, a researcher who only observes the correlations X , Y , and Z will be unable to determine which figure is correct. Both figures agree about which correlations exist (for example, between Y and Z), but they disagree about what would happen under certain interventions. Figure 2.2 claims that an intervention on X will change Z , that there is an intervention on Y that would change both X and Z , but that no intervention on X will change Y . By contrast, Figure 2.1 claims that an intervention on X will change both Y and Z but that there is no intervention on Y that will change Z .



Figure 2.2

As we have already seen, Woodward describes interventionism as a counterfactual theory and claims that when we discover a causal pathway from X to Y , we can explain Y by citing X .

According to Woodward (2007), causation is best understood *from a human point of view*—like Putnam’s internal realism, it is said to have a *human face* (see Chapter §2.5) because the explanations that are most useful to us are ones that align with a human view of the world—a macroscopic view that is essentially coarse grained.

Thus, we can say that interventionism is anthropocentric according to the common view. When thinking about causation, human agents are given a privileged status. Because interventionism is explicitly described as coarse grained and weak,² Woodward thinks that it is incorrect to view interventionism as a metaphysical theory of causation (recall the debate with Strevens). Woodward believes that interventionism is a pragmatic theory of causation that is fundamentally concerned with causal discovery and causal explanation (see Woodward, 2003, p. 331–33). Consequently, it is incorrect to view interventionism as anthropocentric according to the metaphysical view.

6.3.3 Price’s Anthropocentrism

Now that we have seen the complexity of anthropocentrism in Collingwood and Woodward, we can juxtapose their accounts with Price.

It is easy to understand how Price’s account is anthropocentric in the common sense (see Chapter 3§1). Price is a pragmatist who believes that philosophers should begin by examining the role of a problematic concept, like ‘cause’, in the practical and cognitive life of the creatures who use it (see Price [2011]). Moreover, Price thinks that concepts are tools, and the types of tools we use depend on the kinds of creature we are. In a sense, the agency approach to causation that Price defends simply links our concept of cause to the fact that we are a certain type of agent. For Price, agency is a *contingent* feature that grounds the concept of cause. Agency is contingent because our so-called *situation* or experience as agents could have been different (see Price, 2007) So, as epistemically limited beings who are *situated* in time, the concept of cause directly reflects *our* experience of agency: acting towards an uncertain future guided by an incomplete knowledge of the past. Moreover, Price thinks that by emphasising the notion of agency in the context of a

²Recall from chapter 2 that interventionism is weak in the sense that causal explanations do not amount to, nor do they specify, nomologically sufficient conditions. Rather, according to Woodward, causal explanations will include the kind of invariant generalisations that we experience as macroscopic agents.

theory of probabilistic causation, we can overcome significant problems (see Chapter 3§2).

So, Price's approach, like Collingwood's and Woodward's, can be categorised as anthropocentric according to the common view. When thinking about causation, human agents are given a privileged status. Price would say that it is *our* concept, in which the "our" stands not so much for us as speakers and concept users but to *our* unique experience or situation as human agents.

Price's approach is also anthropocentric in the projective sense. The "projection" is rather different from the Newtonian sense discussed above (in terms of Collingwood's theoretical sense of cause). In that sense, we stained or gilded reality with an anthropomorphic 'power'. For Price, what we project is not an anthropomorphism but our natural *view of the world*—the way we see the world as agents but also as human beings with a unique physiological makeup. To illustrate this point, consider an analogy with models. When we model something, for example, when we model an event, the variables we include into the model are ones that we deem relevant and necessary to that event. The model is distinct from the actual world—the model represents a part of the world, but it is not the world itself—this being true from an *external* perspective, or a view from outside of the model. However, given a view of the model from the inside, i.e., given a *situated* perspective in the model, we could say that the model represents a unique ontology. Likewise, when we project our point of view or perspective as agents, we project our reality *in the world* to some reality outside of it. Price says, 'we think of projection as the production of a fictional ontology' (2004, p. 57).

If we follow Price's thinking, Price is not anthropocentric in the metaphysical sense. This may seem counterintuitive. Because Price thinks that causation is relative to our experience of agency, one could argue that causation becomes relative to our cognitive makeup, as Kant did, or that we 'make the world,' as Goodman did. Price argues that his position, although similar to the views of Kant and Goodman, avoids similar metaphysical results because he advocates a global expressivism (see Price [2013]). An expressivist is someone who takes the primary function of certain statements to 'express' an attitude rather than a matter of fact. For a *causal* expressivist like Price, the primary function of causal language is the expression of an evaluative *attitude* rather than a truth regarding a mind-independent reality. For example, a moral expressivist would interpret the claim 'abortion

is not murder' as an expression that reflects the cultural and philosophical attitudes of a given speaker. Similarly, Price believes that we should interpret the claim 'X causes Y' as an expression that reflects the situation or 'attitude' of a given agent. For Price, causal vocabulary is not intended to be descriptive or factual.

Price takes the resulting position to entail a certain kind of subjectivity regarding causation, but he argues that it is incorrect to correlate subjectivity with metaphysics. As we saw in Chapter 4, Price (2017) defends his position by arguing that Woodward's 'objectivist' position is misguided precisely because it leads to metaphysical concerns.

It is important to note that Price clearly endorses the common sense of anthropocentrism that is supported by both Collingwood and Woodward. Moreover, because Price believes that we project a kind of fictional ontology, I take it that his "projectivism" is simply a continuing effect of his anthropocentrism in the common sense (e.g., in terms of causation, we give the agent's 'fictional' ontology a privileged status). We are not staining or gilding reality with an anthropomorphic power whereby human agents create causal relationships; rather, we are simply projecting a *human point of view* of the world onto the world itself. This does entail some kind of subjectivity (as does Woodward's account, which similarly states that we should investigate causation from a human point of view). Yet, it does not entail the metaphysical view that causation is created by human agents. Rather simply, I take Price's projectivism to mean that the claim 'A causes B', while possibly true from a human point of view, is false given a mind-independent sense of reality. So, when we say (it is true that) A causes B, we are projecting a human point of view—a fictional ontology—and subsequently take 'A causes B' to be an objective fact about the world (i.e., objective in the relative sense discussed in Chapter 4).

Why Woodward is Worried About Anthropocentrism

Before I reveal the agency that is implicit within interventionism, I want to consider an important question: why, according to Woodward, is Price's notion of agency so problematic?

The answer concerns Price's notion of manipulation, viz., free action. Recall from chapter 2 that Price introduces the notion of 'free actions' as a way to abstract away from the evidential import of an event (i.e., from ideas about what causes what). As an illustration, recall the earlier example—a

structure in which atmospheric pressure (variable Z) is a common cause of the reading X of a barometer and the occurrence of a storm Y , with no causal link between X and Y . According to Price, manipulating X *via* a free act would make the correlation between X and Y disappear.

Woodward argues that the notion of a ‘free act’ in this context is much too vague. For example, Woodward says that ‘if we wish to follow [Price] in defending the claim that if an association between A and B persists when A is given the right sort of “independent causal history” or is “manipulated” in the right way, then A causes B , we need to be much more precise by what we mean by the quoted phrases’ (Woodward, 2013, p. 13).

Woodward believes Price’s concept of agency, *as a solution to the problem of spurious causes*, is unduly subjective or unacceptably anthropocentric. If we break the correlation between X and Y by a free act, it simply does not make sense to talk about causation between events where an agent’s free act is incapable of creating the necessary ‘independence’, i.e., when an agent would not be capable of acting in such a way as to break the correlations and discover whether X causes Y or whether X is merely correlated with Y . In the situation in which an agent would not be able to perform a manipulation or free action, e.g., inside the sun, there is no way to distinguish between causes and correlations.

Significantly, Woodward’s argument that Price’s approach is unduly anthropocentric leads to the problem of scope. The problem is not that human agents or human beings are given privileged status in regards to causation. Nor is the problem that Price’s account relies on human agency (*per se*) or even that human agents project a fictional ontology. Rather, the problem, according to Woodward, is that when the solution to the problem of spurious causes relies on *overt* manipulations, the free act of an agent, there is no way for us to make sense of causal claims that do not or cannot involve an overt human action. The problem of scope arises (given Woodward’s interpretation of the agency approach) due to the fact that we cannot sensibly talk about causal relationships where an agent’s (overt) act is incapable of distinguishing causes from correlations. Thus, the problem of scope, from this interpretation, comes to rest on the idea that causation relies on the notion of overt agency, and we cannot sensibly talk about causes that do not or cannot involve an overt action.

I will pick up the issue of scope in a later chapter (8). For now, I want to defend my claim that Woodward’s concern is not about agency and anthropocentrism *per se* by revealing the agency implicit within interventionism.

To do so, I will defend the claim that agency includes both overt and mental actions.

6.4 How to Understand the Concept of Agency

My aim in this section is to clarify the notion of agency and to explain how agency involves mental actions.

For an agency theorist, one who accepts the idea that causation is agent-dependent, it is often taken by default—due to a lack of positive alternatives—that ‘agency’ means overt action. As I have previously explained, if agency means overt action, then the agency approach has several, possibly fatal, flaws. However, as I will argue below, *there are numerous ways to think about agency*. I will defend the idea that agency includes more than overt action by first considering the standard theory of overt or bodily agency and then relating the standard theory to the notion of mental agency. I conclude that mental agency qualifies as real agency because mental agency meets the standard criteria for agency.

The standard conception of agency (see Anscombe, 1957; Davidson, 1963, 1971; Schlosser, 2015) says that actions include *intentional* or *deliberate* bodily movements. Mele (1997) calls these actions ‘overt,’ because they are ‘actions that essentially involve peripheral bodily movements’. Consider this well-known passage from Davidson:

This morning I was awakened by the sound of someone practicing the violin. I dozed a bit, then got up, washed, shaved, dressed, and went downstairs, turning off a light in the hall as I passed. I poured myself some coffee, stumbled on the edge of the rug, and spilled a bit of coffee fumbling the *New York Times*. (Davidson, 1971, p. 3)

Which of these events include agency? Davidson says that ‘some of these items record things I did; others, things that befell me’ (ibid). Among the things that Davidson did were getting up, washing, and going downstairs. Among the things that befell or merely happened to him were being awakened and stumbling on the edge of the rug. Davidson’s *actions* include what *he did*, not what merely happened to him. That is, Davidson’s actions were bodily events that were intentionally made by him and not the events that were automatic or unintentional. I display agency when I intentionally raise

my arm, but I do not when you move my arm by accident or when I move it while I am asleep.

When thinking about agency, it is crucial to understand what philosophers mean by ‘intentional.’ In *Intention* (1957), Anscombe argues that when we speak of human agency, we are referring to something that the agent does with intention. An agent acts intentionally when their action comes under the description of a ‘Why’ question. A ‘Why’ question can be answered by giving *reasons* for doing X. For Anscombe, reasons stand in place of something like a means-end utility function, which usually can be captured by the words *because* or *so that I could...*, where these words specifically reference some *future-oriented* goal or intention. (Doing X is the *means* for achieving some *end* Y.) So, we can say that Davidson acted by walking downstairs and not when he stumbled on the rug because we can capture *Davidson walked downstairs* in terms of some (future-oriented) intention. In other words, if we ask Davidson why he walked downstairs, he could reply that he walked downstairs *because* he wanted to get some coffee or *so that he could* read the news. Also, it is plausible to think that we could not capture *Davidson stumbled on the edge of the rug* in the same manner.

If we take actions to be ‘basic,’ that is, as mere bodily movements, then there is a way to subsume Davidson’s stumbling as a distinct action. Intentional actions can then be seen as distinguishable from basic movements like stumbling. Anscombe called basic movements (e.g., stumbling on the edge of the rug) *unintentional* actions. The term ‘agency’ is typically used in a narrow sense to denote the performance of intentional actions. So, we can say that, according to the standard theory, something is an action iff it has a *reason* explanation (e.g., in terms of the agent’s desires, beliefs, and intentions).³ One way to make the distinction between basic and intentional action more clear is to say that Anscombe and Davidson, and indeed philosophers today, are interested in the distinction between *intentional* and *unintentional* actions.

³This entails a coarse grained (or minimising) view (Anscombe 1957; Davidson, 1963) on the individuation of actions, in which you perform an action under a description (Anscombe thought the description should provide reasons, where Davidson thought it should provide causes). When I flick the switch, turn on the light, illuminate the room, and alert the bugler, I perform only one action under the description *I flick the switch*. Under a fine-grained view, how many actions you perform depends on how many act-properties are instantiated. If you instantiate four act-properties, then you perform four distinct actions (Goldman, 1970).

So, agency includes overt actions. But is agency limited to them? In the contemporary literature on agency, it is common to accept that some forms of thinking, or so-called mental actions, also involve agency. In her essay, ‘Is There a Sense of Agency for Thought?’ (2009), Joëlle Proust asks the question, ‘can we ascribe a sense of agency to thought?’ Her answer is positive. Proust says that a sense of bodily agency arises when ‘a subject has a feeling of agency for her movements when she is in a position to anticipate and evaluate the consequences...[that] are associated with them’ (ibid, p. 256). Implicitly, Proust recognises the standard criteria for action by saying that we evaluate the *consequences* of certain movements, movements which we can assume to be directed at a future-oriented end or goal. In other words, Proust is saying that a sense of bodily agency arises when we act intentionally and that we lack a sense of agency when we act unintentionally.

According to Proust, we also have a sense of mental agency—that is, a sense of agency that arises when we think. Proust’s argument is that ‘what holds for bodily actions should hold for mental actions as well’ (ibid, p. 266). In other words, Proust thinks that if we can show that mental actions meet the standard criteria for agency, then we can ascribe a sense of agency to thought. To do this, she distinguishes intentional thoughts from unintentional ones. ‘Automatic thinking,’ as Proust calls it, are unintentional thoughts. These thoughts arise as part of a ‘past conditioning’ and are distinguishable from thoughts that are intentional. Intentional thoughts are ‘active’ or directed towards some end. When Davidson stumbles on the edge of the rug and thinks, ‘yikes!’ his thinking is automatic. When he thinks or rather wills himself to walk downstairs, his thinking was active. It is in such a *willing* that we come to feel a sense of agency for thought. To distinguish the directed mental event from the automatic, Proust says that, in the former case, ‘the operation is called,’ while in the later, ‘it is activated.’ As a matter of definition:

A *willing* or *trying* is a mental event through which an operation is (1) called because of its instrumental relationship to a goal, and (2) is thereby made available to executive processes. In bodily action, the goal is [that an external change be brought about in virtue of this trying]. In mental action, the goal is [that an epistemic—or motivational change be brought about in virtue of this trying] (ibid, p. 267).

The ‘executive processes’ that Proust refers to is a kind of reasoning.

Proust says that ‘mental agency occurs in the context of critical reasoning’ (ibid, p. 260). The ‘motivational change’ that she speaks of applies to our *beliefs* or *attitudes*. Proust’s conclusion is that in the context of critical reasoning, which effects our attitudes and/or beliefs, we experience a sense of agency for thought.

Mele (1997; 2009) also argues that mental actions involve agency. In ‘Agency and Mental Action’ (1997), Mele argues that certain forms of reasoning are intentional and therefore qualify as real agency. Examples of mental agency include solving a chess problem in one’s head (1997, p. 231), deliberating about accepting a job offer (ibid), or deciding to *A* (1997, pp. 204–3). Like Proust, Mele argues that mental agency results from the process of *trying*. ‘Trying to *x*,’ Mele says, ‘is making an effort to *x*’ (2009, p. 18). Types of *trying* like *solving*, *deliberating*, and *deciding* qualify as agency because they are intentional and directed towards a possible end or goal. Relying on the standard conception of agency, Mele argues that some mental actions qualify as agency because mental actions can be intentional and goal oriented.

Like Mele and Proust, Pamela Hieronymi claims that certain forms of thinking qualify as agency; however, unlike Mele and Proust, Hieronymi argues that mental agency amounts to a unique *kind* of agency. In the article, ‘Two Kinds of Agency’ (2009), Hieronymi argues that mental agency qualifies as real, yet distinct, agency because it gives us a unique sense of control, which Hieronymi calls *evaluative* control. For example, whereas I have actual or practical control over the state of the window when I throw my computer at it, there is another sense in which I control the window by *deliberating* about what will happen when I throw my computer at it or *deciding* to do so. Hieronymi argues that mental actions like deliberating and deciding warrant evaluative control because they provide us with control over our attitudes. Attitudes embody a person’s answer to a question or a range of questions. To control our attitudes, Hieronymi says that, ‘you must simply bring it about that [you] settle positively the question of whether *p* or whether to ϕ ...we bring it about that we are committed to *p* as true or to ϕ -ing’ (ibid, 141). If I believe that *p*, *throwing the computer at the window will cause the window to break*, and, in the case that I want to break the window (and my computer), I will have settled the question of whether to ϕ .

Indeed, the idea that mental actions involve agency has been widely endorsed (see, for example, *Mental Actions*, 2009). Most of these arguments take the general form of an argument made by Peacock:

Mental action is a genuine subspecies of action in general. The differences between mental action and bodily action are fundamentally only the differences between the mental and the bodily. (2009, p. 1).

Peacock argues that mental actions can qualify as real agency, given that we can form a link between the standard concept of (bodily) agency and mental action. Rather simply, and like Mele and Proust, Peacock argues that if we can show that certain forms of thinking are intentional and goal oriented, then they will qualify as real agency.

6.4.1 Hypothetical Agency

Given that agency includes both overt and mental actions, I want to make a distinction that can help improve the agency approach to causation. I want to argue that certain kinds of thinking amount to *hypothetical* agency. Hypothetical agency is a type of mental agency that involves deliberate reasoning about the effects of *bringing something about*.⁴ Our mental lives are filled with hypothetical action. You reason about what would happen if you miss a meeting, whether or not to attend the wedding, to wear the blue dress or the red one, to give to charity, to walk next door, etc. Hypothetical action, in general, is the process of reasoning about the effects of *bringing about*.

Reasoning about the effects of bringing about is a type of active deliberation. Active deliberation affects our beliefs (what we believe to be true) and possibly our attitudes (how we intend to act). My argument is that hypothetical agency is real agency because it is intentional (active) and oriented towards an end or goal, viz., understanding the effects of bringing something about. Significantly, understanding the effects of bringing something about enables what Pearl called ‘deep understanding’.

If hypothetical agency is real agency, we should be able to form a link between it and the standard concept of agency. According to the standard concept of agency, an agent acts intentionally when their action comes under the description of a ‘Why’ question. As I said earlier, a ‘Why’ question

⁴In many cases, hypothetical agency—that is, deliberate reasoning about the effects of bringing something about—will involve reasoning about the effects of some overt action. However, hypothetical agency is not limited to reasoning about overt actions. I can reason about the effects of *imagining* that I was back in France. For example, I reason that If I were to bring it about that I imagine being in France, I would become cheerful.

can be answered by giving *reasons* for doing X. Reasons stand in place of a means-end utility function, which usually can be captured by the words *because* or *so that I could...* One could say, ‘I did X (as the means) so that I could...’ (achieve some end). The important question is, in terms of hypothetical agency, can we answer the ‘Why’ question? Indeed, we can. So, how can we answer *why did you X?* where X means *reason about the effects of bringing something about?* The answer to this question, following Anscombe, is that I did X—that is, I reasoned about the effects of bringing something about—so that I could understand what would happen if I *were* to bring this particular something about (Pearl and Woodward would note that we do X—that is, reason about the effects of bringing something about or what they call hypothetical manipulations—so that we can make predictions and enable deep understanding).

I conclude that hypothetical agency is real agency because it meets the standard criteria for agency. Moreover, that satisfying the criteria for the standard concept of agency amounts to real agency is commonly supported in the literature on agency (e.g., Mele, 1997, 2009; Peacock, 2009; Proust, 2009).

The standard concept of (overt) agency is linked with the notion of control, i.e., overt actions allow me to practically control individual objects and/or events. We might ask: Can we form a connection between hypothetical agency and control? The answer is also positive. Hypothetical agency entails hypothetical control. Hypothetical control stems from reasoning about what would happen *if* I were to bring something about, where an approximate understanding of the effects of bringing about Z, for example, provides me with hypothetical control over Z’s effects. Note this quote from Pearl:

When we have such understanding, we feel “in control” even if we have no practical way of controlling things. For example, we have no practical way to control celestial motion, and still, the theory of gravitation gives us a feeling of understanding and control, because it provides a blueprint for hypothetical control (Pearl, 2000, p. 345).

Say that I want to know what will happen if I throw a cannonball at my wall. I do not have to throw the cannonball to know some of the effects of doing so. I only need some approximate knowledge about the effects of large, heavy objects coming into contact with permanent ones. Alternatively, I

may have some approximate knowledge regarding Newton's laws of motion and the materials used to construct the wall. Seemingly, this is all that is required for me to know that throwing the cannonball at my wall will damage the wall. Thus, by having approximate knowledge of the effects of bringing something about, I gain (hypothetical) control over individual objects or events. In the example of the car crash, the driver, the motor-manufacturer, and the surveyor gain control over the car crash by reasoning about what they understand. The control they have over the car crash is hypothetical, but the control is real nonetheless.⁵

It is worth explicitly mentioning that hypothetical actions do not necessarily lead to overt actions. When you reason about what would happen if you miss a meeting, your hypothetical actions can result in overt action, for example, you go to the meeting, or you go home. However, when you deliberate about what would happen if you threw your computer at the window, you could refrain from doing so. You might avoid throwing the computer simply because the effects of doing so are undesirable.

As it happens, hypothetical actions do not necessarily involve even *possible* actions. Possible actions are the overt actions that are possible for a human agent. Possible actions include simple movements of the body, e.g., moving my arms or legs. Possible actions also include more complex actions, like helping in the construction of a jumbo jet or flying one from London to New York. Impossible actions are overt actions that cannot be performed by human agents. Impossible actions include running at 35 mph or flying a jumbo jet to Mars. Hypothetical actions include both possible and impossible human actions. For example, I may reason about the consequences of throwing a 2-kg cannonball at the wall at 25 mph (something that is humanly possible). Or, I might reason about what would happen if I threw a 75-kg cannonball at the wall at 200 mph (something that is humanly impossible). I can reason that if I were to throw the 2-kg cannonball at the wall at 25mph, I would create a large hole in my wall. I can also reason that if I throw the 75-kg cannonball at the wall at 200 mph, then I would produce a large hole in my wall along with a loud noise and severely damage my neighbour's car parked on the opposite side of the wall.

But, what happens if we do not have an approximate knowledge or understanding of the effects of bringing something about? Interestingly, inter-

⁵I disagree with Pearl that hypothetical control amounts to a sense or mere 'feeling' of control. If hypothetical agency is real agency, hypothetical control is real control.

ventionists like Woodward and Pearl are interested in discovering knowledge of this kind. They want to know whether X is the cause of Y or whether X is merely correlated with Y . They want to settle the following question: Is X causally related to Y ? As we have seen, causal modelling allows them to do this. To illustrate, if we learn that X is causally related to Y , then we can reason about the effects of some action related to X and gain (hypothetical) control over Y . And recall, discovering causal relationships, according to Woodward, allows us to learn answers to ‘what if things had been different questions’ (Woodward, 2003). I will explain in the next section how causal modelling involves agency and interaction with the world.

Consider the relationship between hypothetical control and ‘deep understanding’. To illustrate, consider two things that I cannot physically manipulate (or practically control): the gravity of massive bodies in space (the gravitational field due to/experienced by bodies) and the tides on Earth. While it is true that I cannot overtly manipulate the moon in such a way as to effect the tides, given that I have some approximate knowledge of Newton’s theory of gravity, I *can* hypothetically control the tides on Earth. For example, given Newton’s theory, I can understand what would happen if I doubled the moon’s mass or what would happen if I moved the sun a million miles closer to Earth. This understanding enables hypothetical control.

So, agency includes (i) overt actions, which involves intentional bodily movements, and (ii) hypothetical actions (a type of mental agency), which involves reasoning about the effects of bringing something about. Now that we understand that agency includes hypothetical action, an interesting question follows: How do different manipulation approaches utilise the notion of agency?

On one hand, standard agency theories, e.g., Price (1993, 2007) and Collingwood (1940) are commonly thought to utilise the concept of overt agency. For example, as I explained in chapter 3, by defining a cause as something that agents can ‘bring about,’ Price is often interpreted as using agency to mean overt agency. Similarly, Collingwood, it could be argued, utilises overt agency because he describes a cause as a ‘handle,’ where the handle in question provides the agent with *practical* control.

Although it is common to suppose that Collingwood’s theory exclusively employs overt agency, it should be noted that his theory has the tools to make use of hypothetical agency as well. Collingwood is easily understood, or perhaps easily misunderstood, as exclusively utilising overt agency. Causes are described as *practical* and as handles that allow us to produce and prevent

‘in nature.’ All of this gives the impression that when Collingwood speaks about agency, he is talking about overt agency. However, this impression is misleading. Again, in the example of the car crash, the driver, the motor-manufacturer, and the surveyor arrive on the scene of the crash and try to determine the cause. What they end up doing is *reasoning* about the effects of some action. The surveyor, for example, does not physically manipulate the curvature of the road and then determine the cause of the crash. Rather, he reasons that *if* he had manipulated the road, he *would* have prevented the accident.

However, it is important to note that Collingwood denies reasoning about impossible human actions. In one example, Collingwood speaks about the failure of a car to drive up a large hill. He says that, *had* we been able to stomp down hills, we could say that the cause of the car’s failure to drive up the hill was the hill itself. As Collingwood points out, because we *cannot* stomp down hills, the hill cannot be considered a cause. So, Collingwood can utilise hypothetical actions, but only hypothetical actions that are humanly possible.

As I will explain in more detail in the following chapter, Price’s theory clearly relies on the concept of overt action. In regards to the problem of scope, Price, like Gasking and von Wright, argues that we can overcome the problem by analogical reasoning. As we will soon see, analogical reasoning entails that we reason—by analogy—between events that can be overtly manipulated and events that cannot. I will pick up this argument in Part 3, viz., chapter 8.

It is often difficult to say how interventionists like Pearl and Woodward do or do not rely on the concept of agency. My argument is that interventionism differs from standard agency approaches because they rely on, and therefore utilise, the notion of hypothetical agency. This is because interventionists reason about the effects of manipulations within a causal model. The correct way to structure Woodward’s relationship to the notion of agency, given his theory of interventionism, is to say that Woodward does not rely on the concept of *overt* agency like Collingwood and Price but that he does rely on the concept of *mental* agency. This will be fully revealed in §5.

Before I explain the connection between interventionism and agency, it will be helpful to briefly clarify who the concept ‘agent’ refers to and why.

6.4.2 Who's an Agent?

In the literature on agency, it is usually taken by default that ‘agency’ refers to *human* agency. When other kinds of agency are discussed—mental, shared, or collective—it is still in the context of human agency. I have already discussed the arguments for mental agency. *Shared* agency occurs when two or more individuals do something together (such as sing a song). *Collective* agency occurs when two or more agents act as a group (such as a social movement). The one notable exception in the literature is the discussion of artificial agency. The central question in relation to artificial agency is whether or not robots have the appropriate internal states to ground mental representation and intentional agency. Whether or not a robot can be called an agent (in terms of the standard concept) depends on whether or not robots can form intentions (reasons for action).

According to less demanding views, the standard concept of agency provides an interesting and central *kind* of agency. Barandiaran et al. (2009) claim that the standard concept is compatible with the claim that there are more kinds of agency, including kinds of agency that do not require representational mental states. Another kind of agency is what Barandiaran et al. calls ‘minimal agency.’ A (minimal) agent is an entity that is distinguishable from its environment and that is doing something by itself in accord with a particular goal. For example, a beaver, which builds a damn, or a lion, which kills a gazelle, could be called a minimal agent. This view departs from the standard conception in its characterisation of action—‘doing something’—in terms of the agent’s environment and its adaptation features. Barandiaran et al. suggest that animals and even very simple organisms can be said to have the fundamental goal *to be*: to continue their existence.

Similarly, Harry Frankfurt (1978) has pointed out that the behaviour of animals constitutes a low-level doing or ‘acting’. When a spider walks across the table, the spider *directly controls* the movements of his legs, and they are directed to take him from one location to another. Likewise, flowers that bloom at certain times of the day display ‘action’ in some weak sense. Frankfurt’s influence on this point has been the elucidation of ‘full-blooded’ human agency.

Intuitively, this seems plausible. If we called a lion or a seal an ‘agent,’ we would be referring to its minimum capacity to hunt, eat, or mate. Based on the behaviour of other animals, such as chimpanzees and orcas, the distinction between minimal and full-blooded agency becomes blurred.

In general, however, the term ‘agent’ applies to any being with the appropriate mental states—the ability to form intentions and act for known reasons. We can only speculate that “other” agents exist—perhaps intelligent aliens or intelligent fish. In the meantime, ‘agency’ has been reserved for distinctly human actions.

Regarding the literature on causation and agency, philosophers have also been primarily concerned with *human* agency (though indeed, what *human agency* entails has remained frustratingly vague). Despite the occasional foray into hypotheticals, only human action has been discussed. One interesting hypothetical, which I introduced in chapter 5, was put forward by Putnam (1982). Putnam has us consider a group of aliens who travel to planet Earth and happen upon a forest fire. Assumed to be full-blooded agents like us humans, the aliens determine that the cause of the fire is the abundance of oxygen in the atmosphere (we’re told the aliens are from Venus, where there is little oxygen). The alien’s conclusion about the fire is meant to strike us as odd. *We* humans would not come to the same conclusion. Speaking of Price’s agency theory, Jon Williamson (2007) argues that this potentially represents a problem for the agency approach. Are we to conclude that there are two *kinds* of causation—causation for Earthlings and causation for Venusians? Williamson’s concern, I presume, is that the agency approach seems to lead to the worrying conclusion that causation *varies* with different agents, perhaps with manipulation abilities or with what an agent considers to be a serious possibility.

Given that different full-blooded agents are likely to have differing manipulation abilities and contrasting “serious possibilities,” I would argue that what varies is the causal relata. As I will argue in the following two chapters, different agents will interact with various worlds in varying and contrasting ways. How we interact with any possible world depends on *facts about the agent*. Collingwood thought that causes would vary with *control*, that is, what an agent can produce or prevent. Price believes causes will vary with the agent’s *situation*, e.g., the way an agent is situated in time. Woodward also thinks causes will vary with the kinds of questions we ask and what we model (which depends on what we take to be serious possibilities). This variation does not entail that there are different kinds of causation, just different ways of interacting. I will defend the idea that the causal relata (and thus what we take to be causes) will vary alongside variations in interaction in chapter 8.

How interactions vary may depend on something about the *agent them-*

selves, for example, the agent's situation or their manipulation abilities. In an example used by Price, what human agents think of as causes will vary with aliens from a faraway galaxy, where the direction of time is reversed. Human agents and time-reversed agents differ in terms of their situation in time. What human agents take to be a cause, the time-reversed aliens take to be an effect. In Putnam's example, causes vary with the *variables* being taken into consideration rather than with variations pertaining to the agent's themselves. In an ad-hoc model of the fire, the Venusians, who are surprised by the level of oxygen in the area, consider this to be a significant factor in the fire and determine the cause of the fire to be oxygen. Human agents do not consider this variable important (they do not include it in their ad-hoc model of the fire); instead they come up with a different conclusion about the cause of the fire.

As I will explain in more detail in chapters 7 & 8, causal variation is simply a predictable feature of the agency approach. Contrary to any initial concerns, I will show that this is a benefit of the agency approach that I defend (Causal Interactionism) and the agency approach in general.

6.5 Do Interventions Involve Agency?

I have noted important similarities between Collingwood's and Price's theories with Woodward's theory of interventionism. All of the theories are correctly regarded as manipulation approaches; both the agency theory and the interventionist theory consider the discovery of genuine causal relationships as a primary motivation, and all three of the accounts are anthropocentric according to the common view. What I want to do now is bring to light the agency that is implicit within interventionism.

6.5.1 Interventionism and Agency

Making the agency within interventionism more explicit will be useful. To do this, I will focus on one of Collingwood's important ideas—the notion of a *handle* and the associated notion of control. After I have fully articulated the use of these terms following Collingwood, the association with interventionism should become apparent.

A cause in Collingwood's practical sense was said to concern events in nature. However,

The word ‘cause’ [practical sense] still expresses an idea relative to human conduct, because that which causes is something under human control, and this control serves as a means whereby human beings can control that which is caused. In this sense, the cause of an event in nature is the handle, so to speak, by which human beings can manipulate it (Collingwood, 1940, p. 298).

Significantly, Woodward agrees with Collingwood that causes are handles. Woodward says that ‘for manipulation theories of causation...causes are to be regarded as handles or devices for manipulating effects’ (2013, p. 1).

Indeed, Collingwood’s idea of *handles* fits nicely with Woodward’s notion of an intervention. Recall that the motivation behind Woodward’s (and Pearl’s) idea of an intervention was the discovery of genuine causal pathways—distinguishing between causes and correlations. The way that we go about doing this, Woodward claims, is by making a change in X *via* an intervention within a causal model and producing a change in Y . Significantly, once we have discovered a causal relationship between X and Y , we can *in theory* control Y . Thus, X becomes a ‘handle’ in Woodward’s sense of an intervention.

The practical sense of cause takes the Baconian mantra that *knowledge is power*. Woodward says that we want to determine genuine causal relationships so that they ‘might be exploited for purposes of manipulation and control’ (2007, p. 72). Pearl’s answer to the so-called second riddle of causation (chapter 2§5), ‘what difference would it make if I told you that a certain connection was causal,’ is that knowing causal relationships makes a difference to *how we act*. ‘The central theme [of causation],’ Pearl writes, ‘is to view causation as a computational scheme devised to facilitate prediction of the effects of actions’ (2000, p. 347)

Formulated as *hypothetical manipulations*, interventions allow us to discover what *would* happen if we did X or what *would* happen if we did Z (this is one way that interventionism leads to what Pearl calls *deep understanding*. See Chapter §2.5). Whether or not we overtly do X or Z does not actually matter: if we know that $\neg c \rightarrow \neg e$ (or $c \rightarrow e$) we can effectively explain e by citing c , and c becomes, albeit *hypothetically*, a handle for e . Pearl sums up these ideas by saying,

Deep understanding means knowing not merely how things behave yesterday but also how things will behave under new hypothetical circumstances, control being one such circumstance.

Interestingly, when we have such understanding we feel “in control” even if we have no practical way of controlling things (2000, p. 348).

Handles and the associated notion of control can be captured by what Pearl has called ‘small world’ systems. To illustrate, consider a simple model of the tides on Earth. When we model the tides, we include variables such as the mass of the moon, the mass of the sun, the oceans, and Newton’s law of universal gravitation. We do not include things like the weight of the Eiffel Tower or the number of whales in the Pacific Ocean. Interventionists model small-world systems because ‘handles’ would disappear if we considered the universe as a whole. Pearl says:

If we wish to include the whole universe in the model, causality disappears because interventions disappear—the manipulator and the manipulated lose their distinction (2000, p. 350).

If we go back to Collingwood, we can again note that it is only through a certain form of relativity that we can make sense of ‘handles’. For Collingwood, the relativity of causation was explicitly represented by the practical nature of causation. Upon reflection, he says, ‘we know that the conditions of any given event are quite possibly infinite in number, so that no one could thus martial them for selection even if he tried’ (Collingwood, 1940, p. 303). Of course, we know that this is not a problem for Collingwood. He says:

If I find that I can get a result by a certain means I may be sure that I should not be getting it unless a great many conditions were fulfilled; but so long as I get it, I do not mind what these conditions are (ibid).

What Collingwood is expressing here is that we consider a very limited range of factors when we set out to determine—and potentially utilise—causes. In the case of the automobile accident, all three of the witnesses ended up considering slightly different conditions. Collingwood suggests that some of the conditions that they come to consider are those in which they have some sense of understanding or expertise—those that give them a sense of control. The motor-manufacturer understands car design, and this gives her a viable sense of control over car crashes. She would represent the handle as *safer design* because in some sense that is what she understands—and

that is what she can control. A crude representation for the manufacturer might look like the following:

Condition 1: *Car crash*

Condition 2: *Defective design in the car*

Handle: *Safer design*

This is likely all the motor-manufacturer would need to consider to achieve her result. If the result in question were preventing automobile accidents, the manufacturer would have an acceptable means (represented by her conditions above).⁶ This leads Collingwood to say that causality in the practical sense is relative to a certain *point of view*.

It is important to remember that Collingwood's idea of causation (practical sense) can accommodate the notion of mental agency and the associated notion of hypothetical control. For example, it is plausible that when the surveyor considers the car crash, given his expertise and the conditions he represents for the crash, he determines the cause by *reasoning*—he reasons that changing the road to such and such a geometry *would have* prevented the car crash. He does not overtly manipulate the road and *then* determine the cause of the crash. Rather, he simply reasons that *if* he did alter the condition of the road, then he *could* have prevented the crash. The same is true for the driver and the manufacturer. Given the conditions that they consider for the crash, they reason that if they did manipulate the cause in the right way, they could have prevented the crash.

I have been arguing that this is also true for interventionists. The interventionist simply replaces what Collingwood calls 'conditions' with a variable set *V*. On an interventionist model, the driver, the surveyor, and the manufacturer all get different causal explanations for the crash because they included different variables in their model of the crash. Pearl and Woodward would argue that we should model *all* of the conditions (i.e., variables) then intervene on selectively isolated pairs and determine a more objective explanation. Indeed, interventionism, although similar to Collingwood's analysis of causation, is a significant development of Collingwood's general idea that causation

⁶Alternatively, the county surveyor understands the geometry of roads, and he comes to represent Condition 1, replaces Condition 2 with *Dangerous road*, and replaces the **handle** with *Safer roads*. The driver would represent Condition 1, add condition 2, *Speeding*, and replace the **handle** with *Safer driving*. All three witnesses would hold things like the surface temperature of the sun and gravity to be invariant.

is relative to individual judgment. Woodward agrees that causal judgment is itself relative. However, Woodward thinks that causation is relative to our judgment regarding difference makers and ‘serious possibilities’—these being key factors to our judgment of what variables to include into a causal model. To illustrate, recall the earlier example: if a patient dies at a nearby hospital because someone neglected to give them their daily shot, we do not hold strangers in a distant city who failed to administer the shot causally responsible because we do not consider their doing so a ‘serious possibility’ (see Woodward, 2003 pp. 39–57).

Woodward’s definition of a cause shares a kind of relativity with Collingwood’s practical sense of cause, where our causal judgments represent a certain *point of view*. The difference is that Collingwood takes causation to be relative to human agents with unique skill sets and understanding, whereas Woodward takes causation to be relative to a model or a modeller with a unique set of questions and assumptions about serious possibilities.

Both kinds of relativity represent a form of utility. Steven Sloman, in his 2009 book *Causal Models*, captures this insight perfectly when he writes that, ‘animals are selective in what they attend to. They tend to the properties that serve their goals as best they can, while ignoring properties that are irrelevant’ (2009 p. 11).

The fact that causal judgments rely on a form of relativism leads Woodward to characterise interventionism as weak and incomplete (see Chapter 2§5). If *weakness* and *incompleteness* become worrisome, recall the suggestion from Collingwood in the quote above.

So how does interventionism involve agency? Unlike Price’s explicit use of agent probabilities, Woodward’s use of agency is implicit but no less significant. Woodward’s account relies on agency in two ways. Firstly, and more generally, Woodward approaches the concept of cause from the perspective of human practice—things we *do* as human beings (viz., explain and predict). Secondly, and more significantly, interventionism relies on a notion of hypothetical agency (and thus hypothetical control). Causes are regarded as handles because we think that *if* we could manipulate the cause in the right way, there would be an associated change in the effect. Woodward writes:

Interventionist accounts take as their point of departure the idea that causes are *potentially* means for manipulating their effects: if it is possible to manipulate a cause in the right way, there would be an associated change in its effect (Woodward, 2007, p. 20).

Basically, Woodward suggests that causes can be thought of as (i) practical handles, where overtly doing A effects B, and (ii) *potential* (or hypothetical) handles, because *if it were possible* to manipulate A in the right way, we could effect B. Potential handles represent causes that cannot be overtly manipulated. Yet, they are still considered handles because we know that, via interventionism, *if we were* to manipulate A in the right way, *we could* control B.

Potential handles provide us with hypothetical control—*if I could do A, I could control B* because they rely on mental agency. When we learn that A causes B relative to *V*, we reason that if we did A, we would control B. For example, given the model of the tides that was specified above, an interventionist would come to regard the moon as a handle to the tides on Earth. It is true that we cannot overtly manipulate the moon and that we do not have practical control over the tides on Earth. However, an interventionist will reason that *if* we did bring about A, alter the position of the moon relative to Earth, then we could control B, the tides on Earth. Interventionism thus agrees with Collingwood's practical notion of causes as handles and with Collingwood's idea of control. The significant difference is that interventionism explicitly *extends* the idea of control to potential handles by utilising mental agency.

Although I will introduce my specific argument regarding the problem of scope in a later chapter (Chapter 8), we can note here that once we introduce and apply the notion of mental agency into an agency approach, we can potentially locate a solution to the problem of scope. Indeed, by highlighting the similarities with Collingwood's practical idea of causation and arguing that Woodward's account relies on agency—and thus that interventionism should be considered an agency approach to causation—I am suggesting that the agency approach can successfully overcome the problem of scope by extending the causal relation to include events that do not or cannot rely on overt actions but which do rely on mental actions.

This chapter has shown that the agency approach is not untoward (as a manipulation approach) simply because it is anthropocentric or because it makes an essential use of agency. Indeed, all of the manipulation accounts that we have considered are anthropocentric, and they all rely on some form of agency.

Part III

**Building a Coherent Agency
Approach**

Chapter 7

Causal Interactionism

7.1 Introduction

In this chapter, I want to bring together the work that has been done thus far and outline a viable agency approach to causation. The position that I defend is called *Causal Interactionism*. My argument is that any viable approach to causation must make a reference to agency because causation involves an *interaction* between human agents and the world.¹

We have learned that manipulationists like Woodward and Price agree with Collingwood that causes are handles and that causation is best understood from a particular point of view. The point of view in question is often taken to be that of a purposeful agent—an agent with some particular end or goal. Still, the connection between causation and agency is ambiguous. Does agency *create* causation? Does agency *track* causation? Or, is causation somehow *comprised* of agency?

This chapter is an attempt to clarify a position regarding the relationship between causation and agency. My position is that causation is constituted by agency; I say that causation should be understood as an *interaction* between human agents and the physical world. Causal Interactionism is the idea that the causal relation is something that emerges from the interaction

¹By ‘world,’ I mean something like the physical environment, or what Collingwood called ‘nature.’ I take it that Collingwood was correct when he says ‘that which is caused is an event in nature, but the word cause still expresses an idea relative to human conduct’ (1940, p. 296). I interpret Collingwood as saying that causation involves an interaction between the world—something in nature—and human agency. For stylistic purposes, I will sometimes refer to the world as the physical world or the physical environment.

of human agents with the physical world.

I seek to address the following questions: given that there is a relationship between causation and agency, how we should understand the relationship? What does the relationship between causation and agency amount to? What are its consequences?

My hope is that by bringing clarity to these issues, I can show that the agency approach can avoid an extreme position. Recall that one of the biggest problems with the agency approach (see chapter 3) is that it appears to result in an extreme and untoward metaphysics; if causation relies on agency, then we cannot make sense of causal claims (in this world or any other) where agency does not exist.

The common supposition is that the agency approach is extreme or untoward because it flies in the face of any respectable science (one wants to say, ‘of course there’s causation without agency!’). My thesis is that Causal Interactionism can avoid an extreme metaphysics. I will show that, by taking science, viz., physics, seriously, we come to see that there is no such thing as *agent-independent causation*. The resulting position is a form of non-realism, where ‘realism’ can be taken as the claim that the objects and theories of scientific enquiry exist in such a way that they are mind- and agent-*independent*² (Part 2 of my project was used to defend the idea that causation is not mind- and agent-independent). I will argue that the resulting position avoids untoward consequences by motivating the idea that causation is comprised—*not created*—by human agency. It may be the case that those with strong intuitions for causal realism disagree with or reject such a position. While I hope that those with such intuitions will be open to my arguments and possibly troubled by what I have to say, they will not be able to reject Causal Interactionism because it amounts to an extreme or

²Non-realism can be broadly construed as the denial of specific *existence* claims, for example, the denial that *witches exist* or as the denial of specific *independence* claims, for example, the denial that colours exist independently of observers. The non-realism that I advocate primarily concerns *independence*. To explain, consider again the analogy with colours. Most philosophers, I imagine, would not deny the existence of colours. Non-realism about colours usually takes the form of rejecting the *independence* of colour properties. That is to say, most of us would deny that colours exist in the world in such a way that they are independent of something like a visual system. The non-realism I support about causation is meant to work in the same way. I do not deny that the objects or physical mechanisms which make up certain causal relationships, say the moon, gravity, and the tides on earth, exist. Rather, what I deny is that causal relationships exist in such a way that they can be described as agent-independent.

untoward metaphysics.

This chapter will be structured around three questions:

1. How do agents interact with the world *via* causal models?
2. What is my theory of Causal Interactionism?
3. What are the metaphysical implications of Causal Interactionism?

This chapter will be laid out as follows. In §2, I will defend the idea that causal modelling involves agency and that human agents can interact with the world *via* causal models.

In §3, I will introduce my theory of Causal Interactionism. I will note important similarities between Causal Interactionism and traditional agency theories like Price (1993, 2007) and Collingwood (1940) as well as other manipulation approaches such as Woodward (2003, 2007). I will also note where Causal Interactionism differs from both traditional agency theories and Woodward's manipulationist approach. By defending the idea that causation involves an interaction between human agents and the physical world, and by extending the notion of agency to include mental actions, my argument is that Causal Interactionism is the most plausible interpretation of the agency approach to causation.

I will end in §4 by offering some motivational support for the agency approach by considering the metaphysical implications of Causal Interactionism. By contrasting Causal Interactionism with Price, who explicitly denies any benefits of a discussion or analysis regarding the metaphysical implications of the agency approach (and thus does not provide any such discussion), I will argue for a metaphysical picture that is not only consistent with the claims made by Causal Interactionism, but it is one that can help mitigate worries regarding the agency approach to causation. In the following chapter, I will further investigate the prospects for Causal Interactionism.

7.2 How do Human Agents Interact with the World *via* Causal Models?

In chapter 6§4, I suggested a link between causal modelling and mental agency. In this section, I want to explore that relationship in more detail

and explain how causal modelling involves mental agency. I will also show how causal modelling allows us to interact with the world.

In a sense, my arguments rely on the idea that there are multiple ways for agents to interact with the world. We can interact with the world by acting *overtly*, by doing ϕ in ω^* where ϕ is an intentional bodily movement and ω^* is the actual world. We can explain the subsequent interaction as a man named Jon doing $\phi(\omega^*)$ and resulting in $\psi(\omega^*)$ —Jon’s throwing the ball caused the window to break. However, we can also interact with the world *via* causal models ω where ω represents something like what Pearl called the world in the model. We manipulate causal models by doing $\phi(\omega)$, carrying out an experimental manipulation within the model. The manipulation of the model allows us to reason about the effects of $\phi(\omega^*)$. We can explain the subsequent interaction as a women named Mary doing $\phi(\omega)$ and resulting in $\psi(\omega)$. Or, (note the hypothetical reasoning) *if* Mary could do $\phi(\omega^*)$, it would result in $\psi(\omega^*)$. For example, Mary’s manipulation of the distance between the earth and the moon (within the model) led to a difference to the tides on earth; *if* Mary could alter the distance between the earth and the moon, she could affect the tides.

Recall from chapter 2 that Pearl argues that we can gain knowledge about causation by causal modelling (I listed three basic ideas to Pearl’s approach in chapter 2§5). But how can we apply the notion of agency and interaction to causal models? The answer partially depends on what kind of model is being used. There are many kinds of models. For simplicity, we can distinguish between two kinds of models: physical models and non-physical models. Physical models are called *material* models; non-physical models are called *abstract* models. Unlike physical models, which are constructed as physical objects, abstract models are built using mathematical equations.³

It is easy to understand how scientists can interact with the world *via* material models. Material models such as Watson and Crick’s structural

³Abstract models are sometimes thought of as stories (Hartmann, 1999), fictions (Frigg 2010; Fine, 1993; Godfrey-Smith, 2009, 2006), or abstract objects (Giere, 1988) that contain a deliberate *simplification* of some complicated phenomena, with the goal of making it more tractable. Abstract models allow us to strip away, in our imagination, the properties of concrete objects that we find irrelevant. Abstract models may also contain deliberate *distortions*. Isolated systems, frictionless planes, and perfectly rational agents are some well-known examples. While the debate over how to classify abstract models is ongoing, a proponent of Causal Interactionism need not be wedded to the idea that abstract models are mathematical objects or stories or even fictions. All that matters for Causal Interactionism is that there is a sensible link between abstract models, agency, and interaction.

model of DNA or a scale model of an aeroplane can be overtly manipulated because they are physical objects. However, it is less clear how human agents can interact with the world *via* non-physical objects such as abstract models. As an illustration of how human agents *can* interact with the world through abstract models, consider Volterra's predator and prey model.

To understand why certain types of fish populations had been reduced in the Adriatic Sea during the light fishing of World War I, Volterra created a mathematical model representing a simple environment of two populations of fish, one predator population and one prey population. The predator and prey model was composed of two differential equations. These equations included variables for the size of the prey population V , the size of the predator population P , and the intrinsic growth rate r and death rate m for both predator and prey populations (Weisberg, 2007).

By experimenting with the model, Volterra was able to (hypothetically) manipulate the fish in the Adriatic Sea. He was able to do this by changing the values of r and m . The changes inside the model were meant to represent variations in the *target* environment, i.e., the actual fish populations in the Adriatic Sea. Volterra was able to represent real world conditions, e.g., heavy fishing, by making changes to r and m inside the model; when Volterra increased the value of r , he was (hypothetically) manipulating the growth rate of the fish population by changing the level of fishing in the Adriatic Sea. Significantly, by manipulating the level of fishing in the model environment, Volterra was able to discover the effects of these changes. It just so happened that when Volterra increased the amount of fishing, he discovered corresponding changes to both V and P . Through the manipulations that were made inside the model, Volterra was subsequently able to gain insight into how light fishing may have affected the fish population in the Adriatic Sea.

Although Volterra's manipulations were hypothetical, his strategy captures what Woodward considers to be the intuitive core of the manipulation approach. Woodward says that 'A causes B if and only if B would change if an appropriate manipulation *were* to be carried out.' (Woodward, 2013 p. 13, original emphasis).

Thus, causal modelling allows us to reason about the world (e.g., the moon and the earth's tides or fish populations in the Adriatic Sea) *via* hypothetical manipulations. Performing interventions inside causal models allow us to say, 'if X were to be changed by an intervention to such and such a value, the value of Y would change'. From this, we reason that, *had* we overtly

changed X , the change would effect Y . Woodward’s argument is that this type of counterfactual reasoning is both necessary and sufficient for capturing causal relationships. The implicit suggestion is that human agents do not have to (overtly) perform manipulations to gain a knowledge of causation—it is sufficient for us to reason about hypothetical manipulations.

So, we can understand how scientists utilise the concept of agency and how they interact with the world by manipulating abstract models (I state why this amounts to an interaction with the world below). By thinking about mental agency, we can say that human agents interact with the world *via* causal modelling by manipulating something within the model and reasoning about the effects of the manipulation in the actual world.

7.3 My Theory of Causal Interactionism

Causal Interactionism develops the claim that the causal relation can be analysed in terms of a particular interaction between human agents and the world.

In essence, Causal Interactionism says that causation is something that exists iff agents interact with the physical world (admittedly, this will initially sound vague, but I will clarify this particular point as we move ahead). In any case where we say that A causes B , A will involve manipulation and human agency and B will involve a reaction of some kind—something that happens in the physical environment (i.e., something the world communicates to us in response to A). We can think of the interaction like this: If X is a human agent and Y is the world, when X and Y interact, X is doing something to influence Y , and Y is doing something that influences X (when Y influences X , it is typically in terms of information and understanding).

Allow me to demonstrate this concept using two examples. When we say that *throwing the rock caused the window to break*, A , *throwing the rock*, involves human agency. B , the window breaking, is something that happens in the world as a result of A . In regards to the interaction, we could say that the agent did A and the world does B .

Now consider the example of the extinction of the dinosaurs. We say that the dinosaurs went extinct (B) because a large asteroid slammed into the earth near the Mexican Peninsula (A). We think that had we prevented the meteorite from hitting the earth, we could have prevented the extinction. When we reason about preventing A (bringing about $\neg A$, say by going back

in time and blowing up the asteroid when it was a million miles away from Earth), the world reacts by doing ($\neg B$), that is, the world reacts by having dinosaurs *extant* or still in existence. In regards to the interaction, we could say that, had I brought about $\neg A$, the world would bring about $\neg B$ ⁴.

To better understand Causal Interactionism, consider below how the theory relates to, and ultimately differs from, other manipulation approaches.

7.3.1 How does Causal Interactionism Relate to Other Manipulation Approaches?

Causal Interactionism is similar to what Collingwood says about the practical idea of causation. In the practical sense, a cause is something that exists in nature but is relative to human agency. I agree with Collingwood that a cause is something that exists in nature and that a cause is relative to human agency. What I would add, following Causal Interactionism, is that because the cause A is something in nature, and because it is relative to agency, the first part of the “interaction” between agents and the world involves our manipulation of the world. The second part of the “interaction” involves the world’s “reaction” to our manipulation, which is B . B is something that happens in the world as a result of A . Sometimes A is an activity, such as ‘throwing a rock’ caused B ; sometimes it is a thing or situation—*had* we manipulated A , we could have produced or prevented B . Significantly, the interaction between the world and human agents can help to explain the relationship between causation and agency. What Causal Interactionism shows, in line with Collingwood’s philosophy, is that causation is *comprised* of agency. Causation *involves* agency, but it also *involves* the world.

Causal Interactionism is analogous to Menzies’ and Price’s (1993) idea that causation is a secondary quality. One could say that causation, like colour, emerges from a particular interaction. Like colour, which emerges from the interaction between the human visual system and light, causation is something that emerges from the interaction between human agents and the physical world.

Causal Interactionism is also analogous to Price’s idea that causation is perspectival. As I argued in chapter 4, causation (the causal relation) is relative to a frame of reference. The frame in question for Price is the ‘agent’s perspective,’ where the agent’s perspective takes into account the

⁴Of course, it could be the case that the dinosaurs went extinct for another reason.

directionality of causation by taking into account something about our epistemic makeup, viz., how we deliberate (I will discuss the relationship between Causal Interactionism and frame-working below).

Causal Interactionism, too, can make sense of the directionality of causation. Bearing in mind the standard concept of agency, we can capture the asymmetry of causation by focusing on the ‘Why’ question. Intentional actions, and thus agency, assumes directionality. The concept of agency entails that agents act for *reasons*, and that reasons stand in place of a means-end utility function. The utility function can be captured by the words *because* or *so that I could...*, where these words specifically reference *future* oriented goals. What Causal Interactionism would add, is that the goal in question can be practical (as in, tangible, e.g., bringing something about) or theoretical (as in, understanding what would happen *if* I were to ϕ).

Causal Interactionism corresponds to interventionism in that it says the causal relation can be captured by reasoning about the effects of manipulations within a causal model. Like Woodward (2003), Causal Interactionism says that we can capture causation by focusing on causal models. When we carry out an experimental manipulation within a causal model, we enable hypothetical reasoning. This type of reasoning allows us to understand ‘what would be different,’ as Woodward says, or, as Causal Interactionism says, what would happen if we were to bring something about.

Causal Interactionism is similar to all of the manipulation approaches that we have considered, i.e., those of Collingwood, Price, and Woodward, in that causation can be labelled as relatively objective. Indeed, one of the benefits of Causal Interactionism is that it can make the relative objectivity of causation quite clear. Consider again the analogy with colour. When we say that colour is objective, for example, *given* that we take into consideration the (normal functioning) human visual system, it is abundantly clear how colour is relatively objective. Colour is relative in the sense that the concept applies to a particular frame of reference, viz., the human visual system. It is objective in the sense that we can make factual (true or false) claims about colour given the appropriate frame. In chapter 4, the apple on page 72 was, in fact, red. To say that it was green or blue would be false.

Causal Interactionism can similarly demonstrate how relative objectivity applies to causation. Because the causal relation results from an *interaction* between human agents and the physical world, what causes what will be an objective fact about the world given that we take into consideration certain facts about human agency (viz., our interaction with the world). It was

unclear in the last chapter what to make of *facts about human agency* and thus the appropriate frame of reference. For example, the frame of reference for Collingwood was what agents could produce or prevent. This sometimes specified individual agents, based on their expertise, and other times human agents in general. The frame of reference for Price was human agents in general. Also, the frame of reference for Woodward was causal models.

Given the picture of relative objectivity presented in chapter 4, so-called facts about causation could vary depending on which theory one looked at and ultimately with the frame of reference that was being used. For Collingwood, the curvature of the road, the design of the car, and the speed of the vehicle all come out as possible causes of the car accident. However, the objective facts about the cause of the accident will change depending on who you ask. For (qualified) surveyors, the cause is the curvature of the road; for the driver, the cause is speeding, etc. One potential advantage to Woodward's theory, as we saw, is that interventionism can take into account a wider frame of reference. For example, by modelling the car accident, an investigator could manipulate multiple variables, e.g., the road, the car, and the driver, and determine objective facts about causation without relying on individual point of views. Given a model of the crash, we might discover that, given the curvature of the road, any car moving faster than 70 mph will not make the turn and will crash into the hill. Thus, if we discover that the car was travelling at 75 mph when it attempted to make the curve, we could say that *the car crashed due to the curvature of the road and the speed of the vehicle*. One individual may think differently, as Collingwood suggests, but the italicised statement is the objective fact of the matter.

Because Causal Interactionism extends the concept of agency to include hypothetical action, Causal Interactionism *can* make use of the widest possible frame of reference (or 'point of view, as it were'), which is sometimes causal models (that include as many variables as we find interesting or relevant)⁵. However, according to Causal Interactionism, we can say that causation is relative because what causes what is relative to *interaction* (i.e., how human agents interact with the world). According to Causal Interactionism, how human agents interact with the world varies (as we will see in more detail in the

⁵I say that Causal Interactionism *can* make use of the widest possible frame of reference, but I do not mean to imply that Causal Interactionism states that frames will not vary or that frames will always be as wide as possible. The frame of reference *will* vary for Causal Interactionism, the point worth noting is just that Causal Interactionism *can* make use of causal models, which potentially represents a wide frame of reference.

next chapter), and thus the frame of reference grounding (relative) objectivity will also vary. In cases where we *can* utilise causal modelling, the causal model will represent the objective facts about the causal relation. However, in cases where we are not utilising causal models (like Collingwood's idea of individuals reasoning about the cause of the car crash), we can still make objective claims about the causal relation, although the objectivity of the causal claims may be somewhat limited.⁶

The benefit of Causal Interactionism is that it can explain *how* frames of reference vary by noting variations in interactions. While Causal Interactionism can accommodate hypothetical agency and causal modelling, this accommodation does not indicate, nor should it, that Causal Interactionism represents a normative agency approach to causation, where the causal relation is said to be objective or discoverable *merely* from the perspective of causal models. Rather, as we will see in the next chapter, the benefit of Causal Interactionism is that it can explain *what is happening* or what the differences are when we look at causation from competing or contrasting reference frames. Causal Interactionism can accommodate causal modelling, which potentially represents a uniquely broad frame of reference; but, in other cases where causal claims do not rely on causal modelling in a technical sense, Causal Interactionism can also accommodate and explain what is happening. In the example of the car crash, it is not incorrect to think that different agents will obtain contrasting ideas about about the causal relation based on what they understand and ultimately how they model the crash (in other words, how they interact with the world). Even so, Causal Interactionism has the tools to model the crash using a broader frame of reference, as an investigator might do according to Woodward (2003).

7.3.2 How Does Causal Interactionism Differ From Other Manipulation Approaches?

Perhaps the most significant difference between Causal Interactionism and other manipulation approaches to causation is that Causal Interactionism

⁶If you recall the idea of objectivity in the modern sense according to Daston and Galison (objectivity as 'trained judgment'), something could be *more or less* objective because it was a matter of degree. This idea fits nicely with Causal Interactionism. The driver may make an objective claim about a causal relationship given his interaction with the world, but it may be less objective than the claim made by an investigator utilising a causal model.

clearly shows how agency is related to causation. However, there are more subtle variations between the theories, which I will now discuss.

Causal Interactionism is different from Collingwood's practical sense of causation in two ways. The first, which is a matter of clarification, is that Causal Interactionism can explain cases like the automobile accident. Collingwood fails to explain the distinction between cases like the accident, which relies on hypothetical agency and hypothetical control, and how this differs from cases where we rely on the notion of overt agency and practical control. Collingwood's theory allows for hypothetical agency in that the surveyor, the driver, and the manufacturer are all seen as reasoning about the effects of certain manipulations. Causal Interactionism differs from Collingwood in that Causal Interactionism makes clear what is going on in these cases, e.g., we learn that they involve hypothetical agency and the notion of hypothetical control.

The second way that Causal Interactionism is different from Collingwood's approach is that Causal Interactionism can extend hypothetical reasoning to impossible human actions. For example, if we included the incline of the hill into a causal model, we could determine that the hill was a contributing cause to the car's failure to advance, despite the possibility that we are unable to overtly manipulate the incline of the hill. Thus, Causal Interactionism can extend Collingwood's practical idea of causation to situations where objects or events are outside of the practical control of human agents. We say that the gravitational effect of the moon on the earth's oceans cause the tides because we know, given certain information regarding Newton's law of gravitation, that if we manipulated the moon in the right way, we could effect the tides. It is true, of course, that we cannot overtly manipulate the moon in such a way that would effect the tides. But, we know via causal modelling (including certain variables like the mass of the moon and the movements of tides on Earth into a model) and hypothetical agency (reasoning about the effects of manipulations within the model) that if we could manipulate the moon in the right way, the tides would change. Other cases might include 'the lightning caused the field to catch fire' and 'the movements of the continental plates caused the 1989 San Fransisco earthquake'.

Causal Interactionism differs from Woodward's theory of interventionism in that Causal Interactionism explicitly involves human agency (in the overt sense and that extended sense which includes hypothetical agency). As we have seen, Woodward has continually denied that interventionism involves agency. In an important way, Causal Interactionism can clarify what

Woodward means by this. Essentially, what Woodward denies is that interventionism involves (or relies on) *overt* agency. However, as I have made clear (in chapter 6 and here), interventionism does depend on the concept of agency, viz., hypothetical agency. So, Causal Interactionism differs from interventionism in that Causal Interactionism recognises interventionism as an agency approach to causation.

Regarding Price's agency theory, Causal Interactionism differs primarily in terms of clarification. Unlike Price, Causal Interactionism removes ambiguity regarding the notion of agency. According to Causal Interactionism, 'agency' means both overt and hypothetical agency. Causal Interactionism also clarifies the relationship between causation and agency. Causal Interactionism says that causation results from an interaction; Causal Interactionism shows that causation is comprised of agency; and Causal Interactionism illustrates how to conceptualise cases of causation that do not involve overt, or even possibly overt, human actions. As I will show in the next chapter, Causal Interactionism also has a clear solution to the problem of scope and a clear explanation for the so-called *experience of agency*, which Price and other agency theorists (e.g., von Wright and Gasking) rely on to defend the idea that causation is linked with human agency. As this represents a significant prospect for Causal Interactionism, it will be discussed in more detail in the following chapter.

However, Causal Interactionism is quite different from what Menzies and Price say about the causal relation. The causal relation for Menzies and Price is explicitly defined as a *means-end* relationship. For instance, Menzies and Price claim that *A* causes *B* means that human agents can *bring about B* by *bringing about A*. This definition has led opponents of the agency approach to worry that a causal relationship is somehow *created* by overtly doing *A*—a causal relationship is *created* by an occurrence of overt human agency. For example, Woodward seems to interpret Menzies and Prices as claiming that agency creates the causal relation when he says that their definition of cause makes agency a magical quality—'according to which our ability to manipulate *X* somehow makes it the case that a means-end connection comes into existence between *X* and *Y*' (Woodward, 2003, p. 120). In other words, Woodward takes Menzies and Price to mean that a causal relationship "exists" only in conjunction with overt action (Woodward could argue that a similar problem arises for Causal Interactionism. Indeed, I say that causation only "exists" if human agents interact with the world. But there is nothing untoward or magical about this. Indeed, Causal Interactionism is very clear

that while causation is constituted by agency—human agency makes up part of the interaction that gets us to causation—but agency is not the *whole* story. Causation also involves the world).

As I discussed in chapter 4, Price (2017) has pointed out that Woodward’s theory of interventionism seems to imply that agency tracks a kind of mind-independent causation. The idea is that causation (whatever it is) *really* exists, it exists in such a way that it is mind-, agent-independent, and that, as agents, we learn to track real causal relationships (perhaps through some form of behaviourism). The problem with suggesting that causation is mind-independent is that it leaves manipulationists, who support a form of relativism and a kind of subjectivity regarding causation, open to the worry that they make causation *unduly* relativistic and subjective (recall the worries raised by Strevens).

So, contrary to Menzies and Price, we shouldn’t imply that we have to *do A* in the overt sense and that our *action* somehow brings about *B*. Nor should we suggest, perhaps in some sense like Woodward, that human manipulation tracks mind-independent causal relationships. Rather, according to Causal Interactionism, we should say that we manipulate *A*, either overtly or hypothetically, and that *B* is something that happens in the physical world. In other words, we should say, much like Collingwood, that causation is comprised (i.e., constitutive) of agency. Given these various interpretations and suggestions, the landscape looks like Table 7.1:

Table 7.1: Relationship Between Agency and Causation

Relationship	Price	Woodward	Collingwood	Causal Interactionism
<i>Tracks</i>		×		
<i>Creates</i>	×			
<i>Comprised</i>			×	×

7.4 The Metaphysics of Causal Interactionism

One further difference between Price’s agency theory and Causal Interactionism has to do with metaphysics. One of the biggest challenges with understanding Price’s theory relates to his claim that causation is anthropocentric

alongside his claim that the agency theory should be seen as metaphysically quiet. In other words, following Price's approach, it is difficult to understand *why* agency is so important to causation. Price, of course, argues that causation is pragmatically anthropocentric. But, to my point, the idea that causation is pragmatically anthropocentric, i.e., that causation is practically useful to us as agents, perhaps in terms of effective strategies, is not something most philosophers would deny. Nor, more importantly, has it been something so compelling that it leads large crowds to support the agency approach.

What the agency approach to causation needs, I think, is a stronger claim, that is, a claim based on compelling evidence to motivate and support the idea that causation is comprised of agency. Consider the following possibility.

I say that causation is the result of an interaction between human agents and the physical world. Causal Interactionism implies that causation is contingent on human interaction. I said earlier that the first part of the interaction involves human agency. So, Causal Interactionism states that causation *exists* iff there is agency. Thus, Causal Interactionism offers compelling motivation for the claim that agency is necessary for understanding causation. Price, who recognises the contingent nature of causation (Price, 1993, 2007, 2017), argues that the metaphysical meat—represented by such words as *exists*—should be kept off the bone.

Keeping with my own predilections...I think [we should] explicitly disavow that the project of the agency theory should be seen as metaphysics in the first place. Rather it should be seen as what I have sometimes called philosophical anthropology: the task of explaining why creatures in our situation come to speak and think in certain ways—in this case, in ways that involve causal concepts. I think this is one of a range of philosophically interesting cases in which the useful questions turn out to be questions about human thought and language, not questions about other aspects of the world (such as the nature of causation). (Price 2017, p. 76)

I agree with Price that the agency approach to causation is well placed to answer all sorts of anthropological questions (e.g., why human agents come to think and speak in causal terms, why human agents need the notion of cause). But, I disagree with Price that the agency approach has nothing interesting to say about metaphysics or the nature of causation. In fact, the

agency approach that I prefer, Causal Interactionism, *benefits* from saying something about the nature of causation—namely, following Russell (1913) and arguments made against Ney in chapter 5, that the nature of causation is not physical. The nature of causation is not physical because causation is agent-dependent.

The metaphysical claim that causation is not physical *puts us in the appropriate context* to investigate the essential features of causation that are anthropocentric. In other words, once we understand that fundamental physical theories cannot capture the nature of causation on their own (as I have argued), it becomes easier to appreciate the fact that causation has something to do with us as agents—a fact that provides the agency approach with much-needed motivation.

As an example, imagine two possible worlds: (i) an autonomous world in which agents are unable to intervene and (ii) a world exactly like our own but with no human agents.

What could we say about causation in world (i)? It might be argued that the causal relation is agent-independent because we cannot apply the notion of manipulation or intervention. What happens in this world is simply a matter of mechanics, or a way the system functions independently of any human agent. There are two things to note about this world. One, when we apply *our* notion of causation to this world—a notion that involves manipulation and intervention—then we can say that there is causation in this world. For example, we could say that, if we *could* intervene in world (i), then we could alter the system. This would, of course, amount to *changing* the metaphysics of the system, i.e., it would no longer be an autonomous world. So, in one sense, we can say that there is causation in world (i) but that it is no longer autonomous; when we project an agent or agent-based reasoning onto world (i), we enable causal reasoning but alter the metaphysics of the system. Second, keeping in line with the definition of Causal Interactionism, we could say that *if we stick to the metaphysics of the original system*, that is, if we keep world (i) autonomous, then there is a sense in which there is no causation in this world. In this world, we could say that ‘everything influences everything’ (e.g., what happens in one place effects everything else to some degree) or that the system is self sufficient and just keeps on running. But, we could not say that there is causation in this world because, technically speaking, there are no human agents, and thus there is no frameworking and no way for us to speak about manipulation, difference making, or “the” cause of a particular event. In other words, we could not speak about causation as

we usually do.

The same applies to world (ii), that is, the world with no agents. One may contest and say, ‘so what happens in world (ii) when a tree falls in the woods and kills a squirrel? Are we really going to claim that the tree was not somehow causally responsible for the squirrel’s death? In reply, I would say, ‘yes, of course we *could* say that the tree caused the squirrel’s death.’ But, when we reason about the cause of the squirrel’s death, we are projecting agent-based reasoning into world (ii), and again we are left claiming that there is causation in this world by changing its metaphysics, i.e., we apply *our* concept of cause onto the world with no agents, thereby making it a world with at least one agent, e.g., the agent doing the reasoning. However, if we stick to the original metaphysics, we could not claim that the tree caused the squirrel’s death or that anything causes anything else. Again, this is because, without agent-based reasoning, we do not allow for frameworking, manipulation, intervention, or difference making, and it does not make sense to apply our notion of cause into this particular world. In the world with no agents, things may and probably do happen, just as they do in world (i). But, to the point of Russell’s localisation argument (1913), in a world with bare physical facts—a world where the only ingredient so to speak is physical facts—there is no way to apply the standard notion of cause. To do so, we would need a second ingredient, human agency.

So, what Causal Interactionism reveals in these two examples is that *if we are being technically consistent*, that is, if we are sticking to world (i) and world (ii)’s metaphysics, there is a sense in which we can say there is no causation in worlds without agents.

My argument is that the agency approach needs a stronger metaphysical claim. But, what if my claim that there is no causation without agency is *too* strong? It is indeed a strong claim, but all that matters for the agency approach is whether or not it is consistent. I claim that it is. Consider an analogy with colour. Imagine a world exactly like our own, except that every creature in this world is blind or there is no creature with a visual system. In this world, there is *one ingredient* for colour, viz., physical light. But, without a visual system, there is no sensible way that we can speak about colour. As with causation, we could apply *our* standard concept of colour to this world and say, ‘of course, if the world is exactly like our own, there is green grass and red leaves’. But, as I have already explained, this simply amounts to changing the metaphysics of the original world. In the original world (according to its metaphysics), *there is no colour*. There is the

potential for colour, i.e., there is physical light, but there cannot be colour itself without a second ingredient, a visual system to process physical light into distinct colours.

The argument regarding colour is meant to be equivalent to my argument regarding causation. Technically speaking, in a world with no agents, there is no causation because it lacks the requisite ingredients—physical facts *and* agency. Worlds without agents have the *potential* for causation, there are physical facts, but there cannot be causation itself without a second ingredient, agency or agent-based reasoning.

Intuitively, we do want to say that worlds just like our own but with no agents or with no sighted creatures have causes and colours. I agree with the intuition. Yet, if we look closely enough at what's really going on in these cases, for example, we investigate some of the assumptions we are relying on when we say that worlds like our own yet with no sighted creatures have colours, we do come to see that we are simply projecting *our* own experience onto these worlds and then assuming, via the projection, that there are colours.

Perhaps it is more difficult to accept the argument regarding causation. Allow me to make one final point. Perhaps you disagree with my claim that it would be nonsensical to say that the tree was not the cause of the squirrel's death. Fair enough. But ask yourself why you think the tree has to be the cause. Likely, you're assuming that the tree is the thing we could have manipulated to prevent the animal from dying—it was the tree that made a difference. Yet, we know that in the world with no agents the concept of manipulation does not apply. So we cannot say that the tree is the cause because it is the thing that would allow us to prevent the squirrel's death. Perhaps the cause is some general event which ultimately led to the tree fall. It may be that there was a storm and lightning hit the tree and that's why it fell over, killing the squirrel. For the same reason, this doesn't work either, but also because the storm-event cannot be the cause because our notion of 'event' similarly doesn't apply to this world. You may see where this is going. The last thing we might say is that everything within light-distance to the fallen tree is the cause. Perhaps this does come out as the proximate "cause" of the squirrel's death (and likewise some kind of "cause" to movements in world [i]). My argument has been that if it is, it is nothing like what we would ordinarily call 'a cause'. And thus, there is a real sense in which there is no causation in worlds without agents.

Chapter 8

Prospects for Causal Interactionism

8.1 Introduction

In this final chapter, I want to highlight the leading prospects for Causal Interactionism. These include the following:

1. Overcoming the problem of scope.
2. Clarifying “the experience of agency.”
3. Explaining causal variation and ascription.

Agency theorists are well aware of the problem of scope. Indeed, there have been numerous attempts to overcome it. In §2, I will highlight the various efforts to overcome the problem of scope and some of the challenges associated with them. I will focus on what I consider to be the main issue for the agency theorist. The issue is how to extend causation beyond our capacity for manipulation while remaining *consistent*. In essence, the agency theorist must do two things. First, they must show how to extend causation to events or objects that lack a clear connection with human agency. Second, they must explain how the extended cases involve agency. In other words, a satisfactory solution to the problem of scope must extend causation to cases that seemingly *lack* human agency while also showing that these cases *do* involve agency. I will argue that agency theorists have been unsuccessful. Following the work of the previous chapter, I will offer a unique solution

to the problem of scope, which I claim is sufficient for solving both issues. Following Causal Interactionism, I will show how to extend causation to events or objects that lack a clear connection with human agency, and I will show how these cases involve agency.

In §3, I will pick up a related problem regarding the issue of scope as it relates to our *experience* as agents—or what Gasking and Price call ‘the experience of agency’. There are two issues at stake regarding the matter of scope as it relates to the experience of agency. First, perhaps the most critical, is that all human agents lack the experience of manipulating the properties of certain objects or events that we know to be causal, e.g., the mass and movement of large bodies, the weather, and the continental plates. Second, different agents will lack comparable experience. The lack of comparable experience entails that different agents, depending on *their* unique experience, will have conflicting ideas about the causal relation. The easiest way to understand this problem is by relating it to a problem we have already seen. I have already noted that one of the biggest problems with Collingwood’s practical account of causation is that causation depends on *what an agent can produce or prevent*. Agents with different capacities for producing and preventing will disagree about the cause of an event (recall the automobile accident). We can say the same thing about the experience of agency. Agents with different experiences will disagree about the cause of an event. Following Causal Interactionism, I will offer an interpretation of the ‘experience of agency’, I will argue that this interpretation is the most plausible, and I will explain how it avoids the issue of scope.

Lastly, in §4, I will pick up the issue regarding difference-making theories and what we call a cause. As I mentioned in the previous chapter (and above), agency approaches lead to the idea that causes will vary alongside certain parameters, e.g., what an agent can produce and prevent, an agent’s history of manipulation, the agent’s situation, or variable selection within a causal model. We can say that the cause of an event varies when, given the same event (E), two agents disagree: one agent determines the cause of (E) to be *P*, and another agent determines the cause to be *Q*. That this is so was made clear in chapter 4, where I discussed the relativity of causation and argued that causal relation is relative to a certain frame of reference.

However, in light of what was said about Causal Interactionism in the previous chapter, this needs further elaboration. One may point out, regarding Causal Interactionism, that a cause is taken to be the thing that, if *we* could have produced or prevented it, we *could* have produced or prevented

the effect. This leads to a particular concern. The concern is that there are too many causes, for there are indeed many things I could produce or prevent to control some effect and that what we ascribe as a cause is simply arbitrary. I will situate this concern in terms of difference-making theories because a similar concern arises for these theories, and there has already been significant work on how to explain and resolve the issue. The issue is that we need a way to distinguish between non-causes, non-significant difference makers, and significant difference makers. I will argue that the matter has not been sufficiently explained. For example, Menzies claims that we can resolve the issue by thinking about context. I believe that Menzies is correct. It helps to think about causation in terms of context, but thinking about context does not explain why we make such distinctions (e.g., between non-causes, insignificant causes, and causes) in the first place. Following Causal Interactionism, I will argue that we can capture the distinction by thinking about interaction and human agency. In the end, I will show how to overcome the concern and explain why this is a particular benefit of Causal Interactionism.

8.2 Overcoming the (Analytic) Problem of Scope

There are two ways that agency theories can lead to the problem of scope. First, following agency theories such as Collingwood (1940), Gasking (1955), von Wright (1971), and Price (1993), the problem of scope arises due to the particular *analysis* of causation. A cause is something an agent can produce or prevent (Collingwood), a recipe that involves bodily movement (Gasking), something an agent manipulates (von Wright), or something an agent can overtly bring about (Price). An event that one cannot produce, manipulate, or bring about cannot be considered causal. Second, following Gasking and Price, the problem of scope arises due to the particular claim that causation is connected to *the experience of agency*. What “the experience of agency” entails is unclear. However, Woodward (2003, 2013) seems to think that the experience in question relates to an individual agent’s *history of manipulation*. An agent’s history of manipulation includes the events or objects that an agent *has* the experience of overtly bringing about or manipulating. When an agent has the historical experience of bringing about or manipulating an object or event, then these experiences make up the events or objects

that what we are allowed to call causes. In other words, the causal relation will depend on our history of manipulation. Any event or object that is not including in the agent's historical experience of manipulation cannot be considered causal.

Here are the two ways that agency theories can lead to the problem of scope:

Analytic problem of scope: the problem of scope arises because some theory states that a cause is something that an agent can manipulate M . However, there are many cases where we say that H is a cause, but M is not applicable to H .

Experiential problem of scope: the problem of scope arises because some theory states that causation is connected to our experience as agents. But there will be many cases where the relevant experience does not extend particularly wide—in relation to the problem of scope, wide *enough* to capture the multitude of known causal relationships.

In this section, I will focus on the problem of scope according to the *analytic problem*. In the following section, I will focus on the problem according to the *experiential problem*.

Price tries to overcome the problem of scope, which he calls the issue of 'unmanipulable causes,' by 'weakening' the agency theory to allow for a particular kind of reasoning. Price argues that we can reason by analogy from the 'intrinsic features' of two events and overcome the problem of scope. The idea seems to be that we can extend the agency theory—A causes B iff I can bring about B by bringing about A—by analogy. We can reason, by analogy, between events that we can manipulate and events that are unmanipulable.

The analogical reasoning approach is initially found in Gasking (1955) and von Wright (1971). Gasking argues that A causes B if manipulating A allows us to produce or prevent B. Though causation is a 'recipe' for producing and preventing, Gasking acknowledges that we sometimes want to make a 'theoretical claim' about causes that we cannot manipulate. He says that

One can sometimes properly say of some particular happening, A, that it caused some other event, B, even when no-one could have

produced A, by manipulation, as a means of producing B. For example, one may say that the rise in mean sea-level at a certain geological epoch was due to the melting of the Polar ice-cap. But when one can properly say this sort of thing it is always the case that people can produce events of the first sort as a means to producing events of the second sort. For example, one can melt ice in order to raise the water level in a certain area (Gasking, 1955, p. 485).

Following Gasking, we could say the melting of the polar ice caps caused the mean sea level to rise because we have the similar experience of raising the water level in a bucket by pouring more water into it.

Likewise, von Wright claims that we can understand the eruption of Vesuvius and the destruction of Pompeii in terms of claims about causes that human beings *can* manipulate. He uses the example of impacts of falling stones on human heads. Von Wright notes

The eruption of Vesuvius was the cause of the destruction of Pompeii. Man can through his action destroy cities, but he cannot, we think, make volcanoes erupt. Does this not prove that the cause-factor is not distinguished from the effect-factor by being in a certain sense capable of manipulation? The answer is negative. The eruption of a volcano and the destruction of a city are two very complex events. Within each of them a number of events or phases and causal connections between them may be distinguished. For example, that when a stone from high above hits a man on his head, it kills him. Or that the roof of a house will collapse under a given load. Or that a man cannot stand heat above a certain temperature. All these are causal connections with which we are familiar from experience and which are such that the cause-factor typically satisfies the requirement of manipulability (1971, p. 70).

Note the similarity between Gasking and von Wright's approach with Price. Price says that

[I] would argue that when an agent can bring about one event as a means to bringing about another, this is true in virtue of certain

basic intrinsic features of the situation involved, these features being essentially non-causal though not necessarily physical in character. Accordingly, when we are presented with another situation involving a pair of events which resembles the given situation with respect to its intrinsic features, we infer that the pair of events are causally related even though they may not be manipulable. In its weakened form, the agency account states that a pair of events are causally related just in case the situation involving them possesses intrinsic features that either support a means-end relation between the events as is, or are identical with (or closely similar to) those of another situation involving an analogous pair of means-end related events (1993, 204).¹

Price's quote needs some clarification. Price says, following both Gasking and von Wright, that we can reason about unmanipulable causes by analogy with causes that can be *overtly* manipulated. In essence, the problem of scope and the suggested solution looks like this:

¹Hereafter, I will focus on Price's solution to the problem of scope. I do this for two reasons. First, Gasking, von Wright and Price all attempt to overcome the problem of scope using the same strategy, viz., analogical reasoning between two events. Second, Price's account of analogical reasoning is the most developed, and the most widely discussed in the literature on causation.

The Problem

1. The agency theory says that A causes B *iff* a human agent can overtly manipulate A.
2. There are many cases where we say that A causes B, and we cannot overtly manipulate A.

The solution

3. In cases where we say that A causes B and we cannot overtly manipulate A, we can reason about manipulating A *via* analogy with Z, where Z (the cause) can be overtly manipulated and Z (which shares intrinsic features with A) brings about Y, which resembles B.

Furthermore, Gasking, von Wright, and Price must reason by analogy from events that *can* be overtly manipulated to events that cannot so they can retain the essential feature of the agency theory—that causation involves human agency (which they take to mean ‘overt agency’). To keep the agency theory consistent, they must show how events or objects that *cannot be overtly manipulated*—event A—involves agency. To do this, they must find a similar event that *involves human agency*—event Z. When they reason by analogy from Z to A, they can consistently maintain that *causation requires human agency*.

Note the similarity between (3) and what Price says, following the quote above

Clearly, the agency account, so weakened, allows us to make causal claims about unmanipulable events such as the claim that the 1989 San Francisco earthquake was caused by friction between continental plates. We can make such causal claims because we believe that there is another situation that models the circumstances surrounding the earthquake in the essential respects and does support a means-end relation between an appropriate pair of events. The paradigm example of such a situation would be that created by seismologists in their artificial simulations of the movement of continental plates (p. 205).

Following 3, the *friction between continental plates* represents A, an unmanipulable cause. The *artificial simulations of the movement of continental plates* represents Z.

So, Price thinks we can understand cases like *the movement of the continental plates caused the 1989 San Francisco earthquake* because we have the analogous experience of overtly manipulating a model that shares ‘intrinsic features’ with the actual event, viz., the 1989 San Francisco earthquake.

Woodward argues that Price’s solution to the problem of scope fails. Woodward thinks the idea of ‘intrinsic features’ is inherently vague. He believes that if we are forced to rely on artificial simulations or small-scale models to understand causes that cannot be overtly manipulated by human agents, we may ‘fail to capture their causally relevant features because, for example, the models fail to ‘scale up’’ (Woodward, 2003, p. 125).

I agree with Woodward that the notion of ‘intrinsic features’ is vague. If by ‘intrinsic’ Price means ‘essential’, then we are left trying to find a strict analogy with the ‘fundamental elements’ that two events, that is, a *M* event (an event we can overtly manipulate) and a non-*M* event (an event we cannot overtly manipulate), both share. In many cases, this seems overly demanding. Take the simple case of a grass field catching fire after a lightning strike. Surely, we have the comparable experience of making things catch fire, even grass. Farmers can burn down fields by throwing a lit match into it, which is roughly 600-800° (Celsius). Similarly, forest fires occur nearly every year in California because someone throws a cigarette butt (the tip can reach 900°) into dry grass. But, the essential features of lightning are rather different from those of a match or a cigarette. Lightning is the movement of electrical charges and technically does not have a temperature. Lightning causes fires because it heats the air through which it passes. A fire can occur after a lightning strike because the air through which it travels can reach temperatures of 28,000°. Against Price, it is not clear that we can say ‘lightning caused the field to catch fire’ because we have the comparable experience of burning grass using a small flame (from a match or a cigarette). This is because the *M* event *I caused the field to catch fire by throwing a lit match into it* seemingly does not share the fundamental elements of the non-*M* event *the field caught fire because of lightning*. For example, a match or a cigarette has a direct heat source, i.e., a temperature, whereas lightning technically does not. Indeed, the significant question arises: how can we tell

what a good analogy is?²

Price could argue that we have a better analogy—scientists who can recreate the effects of a lightning strike in a laboratory. If scientists artificially created a lightning bolt that hit a small patch of dry grass (or any combustible material), and the grass caught fire, then we can reason by analogy between this event and the real-world lightning event. Let us grant that this is feasible. Even if the relevant *M*-event exists, a problem remains. The strict analogy defence entails that we can only speak about lightning causing fires *after* scientists were able to create the analogous event in their laboratory. Before the analogous event, Price's theory entails that it would have been irrational to say that lightning caused the field to catch fire. This seems untoward. It is likely that we knew, *via* some other form of reasoning, that lightning caused fires before the strict analogy that Price is after. Indeed, the story of Prometheus, c.8th-century BCE, suggests that this is true. In the Prometheus myth, Prometheus steals fire from Zeus by sneaking into Zeus' domain and taking a spark from Zeus' lightning bolt. It is likely that after witnessing many fields, forests, homes, and other objects catch fire after a lightning strike, we reasoned by induction that *lightning strikes lead to fires*. In the particular case that the field caught fire after the lightning strike, we could use abduction (IBE) to explain the fire. Moreover, anyone today who is *unaware* of what scientists do in their laboratories would be deemed irrational in saying that 'lightning caused the fire'. *They* cannot reason by analogy from the appropriate *M*-event, because the event is unknown. Neither would anyone who is unaware of what seismologists do in their laboratory be able to reason that movements in the continental plates caused the 1989 San Francisco earthquake.

There is a related problem for Price's solution to the problem of scope. As I mentioned in the previous chapter, there are cases where we say that A caused B based on the manipulation of abstract (non-physical) models. Because the agency theorist must reason between a *M*-event and a non-*M* event, there will be many cases where Price's analogical approach breaks down. For example, we would not be able to speak about the causes of the decline of fish in the Adriatic Sea, the Earth's tides, or stars exploding as supernovae because scientists do not model these events in such a way that

²There are many other cases where similar problems for Price's solution to the problem of scope can be made evident. For example, how would we apply Price's strict analogy approach to: 'shark bites can be lethal'?

they can be overtly manipulated. Rather, as I showed, these cases rely on the hypothetical manipulations of abstract models. If by ‘artificial simulations’ Price means an abstract model, then his analogy between what scientists do in their laboratory and the 1989 San Francisco earthquake would similarly break down (because M is not applicable to abstract models).

In such situations, there are three available options for the agency theorist. The first choice (1) is to bite the bullet and admit that we cannot speak sensibly about the causes of the earth’s tides or exploding stars. The second option is to show how we can (2) draw the correct analogy between some M event and the real world non- M event. The last option is to (3) show how abstract modelling involves real agency.

I assume that Price would reject (1). Price explicitly defends option (2). However, an obvious problem with option (2) is that it appears to amount to the rather strange claim that we know about the causes of certain events not based on the sophisticated abstract modelling used in science but rather on some analogy based on the manipulation of a clumsy physical model. Price does indicate that his theory can rely on option (3). Evidence for this claim could be Price’s use of ‘simulations’ in the example he uses of the 1989 San Francisco earthquake. Price says that the best example of a means-end relationship concerning the earthquake is a situation ‘created by seismologists in their artificial simulations of the movement of continental plates.’ I suggested above, following Woodward, that by ‘simulations’, Price was referring to a physical model. However, it may be the case that Price intends ‘simulations’ to mean a computer simulation or an abstract model. If this is the case, Price would need an argument to explain how the experimental manipulations of abstract models involve real agency.

Price’s only argument for this position is that Woodward’s theory of interventionism—which utilises the concept of experimental manipulations of abstract models—requires agency. Because interventionism avoids the problem of scope, if Price’s argument stands, he thinks that the agency theory can make use of the same tools. Reasoning by analogy with abstract models would help Price because it would allow him to use abstract models, which are common in science, as the relevant non- M event. Rather strangely, Price’s position, as we saw in chapter 4, is that Woodward’s theory makes causation mind-dependent, and therefore this leads to agent-subjectivity. On one issue, Price is correct: Woodward’s theory *is* mind-dependent. But, Price has not shown how mind-dependance (what I called frameworking) leads to real agency. Sure, Woodward’s theory entails that *we* model events based

on limited factors or variables that *we* find interesting and relevant (I am following Price here in emphasising the ‘we’). But, frameworking and relativity, which apply to Woodward’s theory, do not *entail* real agency. As we saw in the debate between Woodward and Price on objectivity, Woodward *accepts* that interventionism is relative (to a certain extent), but he *denies* that interventionism involves (overt) agency. So, Price himself is unable to utilise those features of interventionism that might help his agency theory overcome the problem of scope simply by arguing that interventionism is subjective. To utilise the same tools as Woodward’s theory, Price would need a specific argument to show precisely *how* interventionism, specifically abstract modelling, involved the notion of agency. Price offers no such argument. Price’s theory is *overly* committed, i.e., exclusively committed, to the concept of overt agency, and, in turn, he is unable to capture the features of interventionism that would otherwise benefit his theory. As I will argue in a moment, Causal Interactionism *does* provide an argument to show how abstract modelling involves real agency. So, a unique benefit to Causal Interactionism is that it can rely on option (3), whereas Price’s version of the agency theory cannot.

Option (3) will not work for Price. The agency theorist is left with a dilemma: Either he accepts that we have to reason from a *M* event to a non-*M* event in order to remain consistent with his analysis of cause (‘A causes B iff bringing about...’), or he must alter his definition of a cause. Option one is insufficient as a solution to the problem of scope. It leaves too many cases unanswered. Option two means the agency theorist admits a small defeat and heads back to the drawing board.

8.2.1 A Solution to the Problem of Scope

In the previous chapter, I said that Causal Interactionism has the tools to overcome the problem of scope. This was somewhat misleading. Actually, Causal Interactionism *avoids* the issue of scope altogether. Causal Interactionism does not need to find a solution to the problem of scope because, as constructed, the problem does not occur. The problem does not occur because I have shown that abstract modelling involves real agency (*viz.*, hypothetical agency). This allows Causal Interactionism to extend the causal relation to casual claims captured by abstract models.

Concerning the issue of scope, the problem with traditional agency theories is that they are forced to reason between a *M*-event and a non-*M* event.

This is problematic for two reasons. First, we can never be sure, as Woodward points out, that the analogy is strong enough. There may be cases, many cases in fact, where the analogy breaks down because we failed to consider one of the essential features in question. Moreover, as shown in the lightning example, there may be cases where we have a good analogy but whether or not the analogy is successful in the sense that it captures the essential features of the two events is unclear. Second, an appropriate analogy could be unavailable, and we would be forced into irrational considerations (such as ‘I don’t know why the field caught fire’ when we have substantial evidence, albeit not a strict analogy, for the claim *the lightning was the cause*).

As an example, consider what Causal Interactionism would say about certain so-called unmanipulable causes, e.g., the lightning and the fire. As I mentioned above, as long as we have an approximate understanding of the effects of lightning, we could reason that *had we prevented the lightning, the grass would not have caught fire*. We can also reason that, *had we produced lightning, we could have caused a fire*. There are two things to point out. First, contrary to the strict analogy approach, all that is necessary to overcome the problem of scope is an approximate understanding of a cause’s (e.g., the lightning’s) effects. Second, given this understanding, we can reason that ‘had I prevented the lightning, I could have prevented the fire’. What this shows is that we can eliminate the need for a strict analogy and, quite significantly, how the alleged non-*M* event—the lightning strike—involves real agency, viz., hypothetical agency.

The same is true for the case of the 1989 San Francisco earthquake. Given an approximate understanding of the effects of plate tectonics, I can use hypothetical agency to reason about the effects of certain (hypothetical) manipulations. For example, I could deliberate about what would happen if I were to manipulate the convergent boundaries (the place where the crust is destroyed as one plate dives under another) in such a way that the force or energy between the two plates accumulated but was not released. If the accumulated energy was equivalent to 99,000,000 tonnes of TNT, when the energy was released, I would produce a magnitude 9.0 earthquake. I can also reason about the earthquake as a cause. Given an approximate understanding of the effects of earthquakes, I can use hypothetical agency to reason about the destruction of cities like Pompeii or San Francisco.

In essence, Causal Interactionism avoids the problem of scope because Causal Interactionism shows us how abstract modelling involves real agency. By reasoning about hypothetical manipulations, Causal Interactionism can

extend causation to events or objects that cannot be manipulated overtly but that can be hypothetically manipulated within a causal model. This is a significant advantage over conventional agency theories, which have traditionally relied on the problematic notion of analogical reasoning. It goes without saying that there are, of course, good physical models and that it is okay to rely on them—the point is that exclusively relying on them does not allow us to overcome the problem of scope. Coming back to the lightning example, Causal Interactionism can use hypothetical reasoning to say that the cause of the fire was the lightning and avoid the need for a strict analogy as well as the subsequent concern about whether or not the analogy is good one.

There are many other cases where perhaps we *could* manipulate certain objects or events overtly, but our understanding of specific causal relationships are based on hypothetical agency and abstract modelling. I used the example in an earlier chapter of the decline of certain fish populations in the Adriatic Sea following the light fishing period of World War I. We were able to understand the relationship between fishing levels and the populations of fish in the Adriatic Sea by hypothetically experimenting with abstract models (recall Volterra's mathematical equations). Causal Interactionism can take the science at face value. Because Causal Interactionism shows us how hypothetical experiments with abstract models involves real agency, there is simply no need to think any further about how the hypothetical experiments relate to real-world, overt manipulations. For example, a traditional agency theorist like Price might say that we *could* carry out overt manipulations—for example, by running costly and time consuming fishing experiments in the Adriatic Sea—but there is simply no reason to view the real-world experiments as preferable to the hypothetical ones. Moreover, it is a common practice throughout science to draw conclusions *via* abstract modelling, and it would be absurd to say that the agency theorist can make sense of (some of) the claims in science (which rely on hypothetical manipulations) because we *could* perform some closely related overt action.

8.3 The Experiential Problem of Scope

In the above section, I explained the problem of scope in relation to a analytic analysis. The problem of scope, according to the analytic analysis, is that there are events or objects that we know to be causal, but these events or

objects fail to qualify as causes based on the specific analysis that a cause is something that an agent can manipulate or bring about. I also explained how Causal Interactionism avoids this problem.

However, following the ideas of Gasking and Price, there is a second analysis of the agency theory, one which is often conflated with the analytic analysis, that seems to lead to the problem of scope. Agency theorists suggest that causation is connected to the *experience of agency*. I will explain the experience of agency in more detail below. However, *prima facie*, relating causation to our *experience* seems to limit what we are allowed to call causes. As I stated in the introduction, if we think of the ‘experience’ as something like our historical experience of manipulation, then anything that falls outside of the agent’s historical experience, that is, anything an agent *does not* have the experience of manipulating, cannot be considered to be causal.

Given the interpretation of ‘experience’ as *an agent’s historical experience of manipulation*, the idea that causation is connected to our experience certainly does unduly limit what we are allowed to call causes. As I will explain below, I think this interpretation is indicative of Gasking’s approach. Whether or not Price endorses this interpretation is not clear. Some of Price’s remarks on the experience of agency seem to lead to the idea that he is thinking in terms of the agent’s historical experience of manipulation, like Gasking. Other remarks by Price appear to indicate a rather different interpretation of ‘experience’. I will parse many comments from Price, note a distinction between two alternative interpretations of ‘experience’, and argue that a single definition of ‘experience’ is not explicitly supported by Price. I will then conclude this section by offering my understanding of the experience of agency, and I will argue that this interpretation avoids any issues of scope.

8.3.1 Gasking and Price

Gasking (1955) argues that the cause-effect relationship should be explained in terms of a ‘producing-by-means-of relation’ (1955, p. 485). Gasking says, we say “A causes B” ‘in cases where one could produce an event or state of the A sort as a means to producing one of the B sort’ (ibid). For example, we say that iron glows by making it hot. For Gasking, the notion of causation is connected with our manipulative techniques for producing results. He famously said that ‘a statement about the cause of something is very closely connected with a *recipe* for producing it or for preventing it’ (ibid,

my emphasis).

Following Gasking, there is a second and separate analysis of the agency approach that leads to the problem of scope. The second analysis is based on what Gasking says about an agent's *experience*. Note the following quote:

We learn *by experience* that whenever in certain conditions we manipulate objects in a certain way a certain change, A, occurs. Performing *this* manipulation is then called: "producing A" (ibid, 486. My emphasis).

Gasking seems to imply that we only say that A causes B when we have the overt experience of producing B by producing A. So, there appears to be a second concern regarding the problem of scope. If causation is directly related to an agent's 'experience', then we can only speak about causes when we have the requisite experience (of production, for example). It is true that Gasking says that 'one says A causes B in cases where one *could* produce an event or state-of the A sort' (my emphasis), but what Gasking says about experience seems to contradict this. We speak of A causing B not when one *could* produce A but in cases where one has the historical experience of producing A. Gasking confirms this position when he speaks of recipes or *techniques* for production. He says the notion of causation 'is essentially connected with our manipulative techniques for producing results' (ibid, p. 483). Of course, we learn such techniques through experience. Originally, we did not find recipes just lying around; we learned recipes by *performing* manipulations.

So, following Gasking and his remarks on experience, the problem of scope arises because there are many cases where we say that A causes B and we lack the requisite experience of producing B by doing A. Examples abound, and it seems that we are right back where we started. Human agents lack the experience of producing earthquakes, Big Bangs, expanding universes (and indeed, in some cases, much more), and yet we know that these events either have causes or act as causes themselves.

Because Gasking says that we speak of producing events of the A *sort*, he might reply that while we lack the direct experience of producing earthquakes and Big Bangs, we do have the experience of producing events of this sort. I will not reply to this because I have already spoken about the problems associated with analogical reasoning—viz., reasoning between a non-*M* and a *M* event. Even if this kind of logic gets us to cases where we lack the

connection to direct experience, a new problem emerges. The problem is that human agents do not share the same experiences, and it is not clear that Gasking would allow us to generalise the direct experience of one agent to another (it is not clear that he would allow us to “share recipes”).

Like Gasking, Price also relates causation to our experience as agents. He says, quite significantly,

[T]he central point is that the concept of causation is to be explained by relation to our experience as agents in the same way that the concept of colour as a secondary quality is to be explained by relation to our experience as observers (1993, p. 193).

Price notes that our experience as agents is fallible. Overt manipulations (of the ‘bringing about’ sort) do not lead to an infallible guide to the existence of causal relationships. Rather, Price’s point seems to be that our experience as agents provides a level of *credence* for certain causal relationships (what he calls ‘agent probabilities’). Nonetheless, Price is committed to the idea that the ‘central point’ of the agency theory is that causation is directly tied to our experience as agents.

That being the case, what does Price mean by ‘experience’? Consider the following quote. Price says that, as agents, we have a very special class of succession:

[T]hose 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. We might say that the notion of causation thus arises not, as Hume has it, from our experience of mere *succession*; but rather from our experience of *success*: success in the ordinary business of achieving our ends by acting in one way rather than another (1993, p. 194)

In one way, Price seems to be saying something rather different from Gasking. When he says that we have the experience of doing one thing and achieving another, Price appears to indicate that ‘experience’ means our experience of *being* agents, that is, our experience of being agents in general. Call ‘being an agent’ the *general experience of agency*. Given the general

experience of agency, the so-called experience of agency would mean the experience of having done one thing and achieving another. However, Price seems to contradict this interpretation when he talks about our experience of *success*. The experience of success seems to indicate that what Price means by experience is something more similar to Gasking. I explained that, for Gasking, we say that A causes B, when we have learned, through *experience*, a recipe for production ('producing by means of...'). If we focus on what Price says about success, it may be that Price is essentially reiterating Gasking's point. Utilising Price's language, we could say that we have good evidence of causal relationships when we have the experience of successfully bringing about B by overtly doing A. In other words, agent probabilities, which indicate causal relationships, strictly apply to the agent's experience of successfully bringing about (some effect). The experience of successfully bringing about A and achieving B indicates that 'experience' means what it did for Gasking—the historical experience of manipulation or (in Price's language) bringing about.

Indeed, if 'experience' meant the *general experience of being an agent*, it is hard to see how Price would cash out 'agent probabilities' in terms of agents generally having done one thing and achieved another. If you recall from chapter 3, according to Price, a free action can 'break' the spurious correlation between two events by creating an independent causal history *via* the free act. The probability that A causes B would be higher when the agent's free action was working through a causal pathway, rather than a correlation. So, at least in terms of agent probabilities, 'the experience of agency' seems to mean *the historical experience of bringing about*. For example, we could not distinguish between spurious and (actual) causal relationships by having the general experience of agency, that is, by merely doing one thing and achieving another. We speak about causal relationships when the agent has *already had* the experience of doing one thing and bringing about another. For example, I know that the rooster is a spurious cause of the rising sun because I actually manipulated the rooster, not because I have the general experience of doing one thing and achieving another, e.g., having the experience of rubbing two sticks together and producing fire. I *could* break the evidential bearing of roosters on rising suns by creating an independent causal history and thus break the usual evidential bearing of roosters on rising suns, but this would require a new and distinct act, *a new experience*, viz., one that deprives roosters of their usual evidential bearing to rising suns.

By relating experience to our success, the suggestion seems to be that ‘experience’ means our historical experience of bringing about a certain event (B) by doing something overt (A). If this interpretation of experience is correct, then the worry arises that agent probabilities (and thus causal relationships) apply only to those events or situations where an individual agent has the experience of performing a manipulation. While I think it is still unclear how we should interpret Price’s use of ‘experience’, it is important to note that some of his critics have made up their minds. Woodward, for example, interprets Price’s use of experience in terms of the *historical experience of manipulation*. Woodward says:

If the only way we understand causation is by means of our prior grasp of the experience (or notion) of agency, then we face the obvious problem about the extension of causal notions to circumstances in which the relevant experience of agency is unavailable (2003, p. 123).

When Woodward speaks of ‘the relevant experience,’ he takes experience to mean the agent’s historical experience of manipulation or bringing about. So, given this interpretation, agent probabilities would apply only to those events or situations where a particular agent *has* the overt experience of bringing about B by doing A. Any event or situation where an agent lacked the relevant experience of bringing about would not be considered causal.

However, Price might be able to avoid some of these complications if by experience he means the *general experience of agency*. If by ‘experience’ Price means the general experience of being an agent, then we could extrapolate our experience as agents, that is, doing one thing and achieving another, to any case where we *could* bring about B by doing A. Indeed, this would avoid the problem of relating causal relationships to an individual’s history of bringing about. However, it would not altogether prevent the problem of scope. If I lacked the experience of manipulating an event that is known to be causal, e.g., the movement of tectonic plates, the analogical reasoning approach that Price defends would still require that I have the experience of manipulating a similar event. There are two problems with this. First, I may lack the experience of manipulating a similar event, and I would not be able to capture a claim such as ‘the 1989 San Francisco was caused by the movement of tectonic plates’. Second, even if I had the experience of manipulating a similar event, I would be left trying to make sense of the

unmanipulable cause by analogical reasoning, and I have already shown that the analogical reasoning approach is insufficient for solving the problem of scope.

If we grant that by ‘experience’ Price means the general experience of being an agent, it may make the problem of scope less *demanding*, that is, there may be fewer cases of unmanipulable causes. However, the significant point to bear in mind is that, even if we give Price the ‘general experience’ analysis of experience, we are still left trying to make sense of all of the causal claims where manipulation is humanly impossible. On the general experience analysis, I may lack the experience of manipulating a particle accelerator. Nonetheless, because I have the experience of being an agent, and because it is (humanly) *possible* to manipulate a particle accelerator, e.g., by working at CERN, I can talk sensibly about the creation of the Higgs-Boson. Yet, even on the general experience analysis, I cannot make sense of causal claims that involve *impossible* human actions, and, again, Price’s solution of analogical reasoning is insufficient to overcome it.

So, whether or not we interpret Price’s views in terms of the *experience of being an agent* or in terms of the *historical experience of bringing about*, he cannot avoid the problem of scope (and he does not overcome the problem by relying on analogical reasoning).

8.3.2 What Does Causal Interactionism Say about the Experience of Agency?

I have argued that the agency approach can overcome the problem of unmanipulable causes by making a distinction between overt and hypothetical agency. Hypothetical agency, which includes mental actions like *reasoning about the effects of bringing something about*, allows us to extend the concept of agency to events where overt agency is impossible. This, in turn, makes it possible to capture causal claims that do not or cannot involve overt agency.

Moreover, I have shown that hypothetical agency allows us to relinquish the limiting requirement of analogical reasoning between two events (a non-*M* and a *M* event). But, what would Causal Interactionism say about the experience of agency? The question is important because the answer reveals that the ‘experience of agency’ is an important component of hypothetical agency, that is, reasoning about the effects of bringing something about. I think the most plausible interpretation of the experience of agency is *the*

general experience of agency. What Causal Interactionism should, and can, make clear is that the ‘experience of agency’ should not be limited to our first-hand experience as agents (that is, our historical experience of manipulation). In the following example, we can see how the experience of agency works in terms of hypothetical cases (i.e., hypothetical agency).

Recall Pearl’s example of the moon. The moon is an object that is impossible for me to overtly manipulate. I cannot overtly manipulate the moon, yet because I have a general understanding of the moon, say that its mass is X and it affects the tides, *and* because I have the experience of being an agent, that is, the experience of having done one thing and achieved another, I can hypothetically manipulate it. The same goes for many of the scientific claims about causes in space: the expansion of the universe, Kepler’s supernova in 1604, or the orbit of Uranus around the sun. Surely, the experience of agency arises due to the (overt) manipulation of ordinary objects. However, we can extrapolate this experience to hypothetical cases. Consequently, in the hypothetical cases, when we reason that *if I could* do A, *I would* produce or prevent B, the experience of agency plays a critical role.

I have used this section on the experience of agency to do three things. First, I made a distinction between two possible ways that the agency theory can lead to the problem of scope. The issue of scope is usually taken to mean the problem in the *analytic* sense, the sense that was considered in §3. However, ‘the experience of agency’ also leads to the problem of scope, and it is often unclear which sense of the problem is being used when one raises the problem (e.g., the quote from Woodward in this section rests on the *experiential* sense). Second, the so-called experience of agency is a central part of Price’s agency theory and, because I have been so heavily focused on Price’s work, brushing aside the ‘experience of agency’ would amount to dismissing a significant component his theory. Third, and most importantly, Causal Interactionism can accommodate the ‘experience of agency’ and show how it avoids the problem of scope.

8.4 Causal Variation and Ascription

In this last section, I want to highlight a problem for manipulation theories by focusing on the idea of difference making. Difference-making theories claim that A is a cause of B because A (counterfactually or probabilistically) makes a difference to B. Woodward (2003) explicitly endorses a counterfac-

tual difference-making theory. What I want to highlight is a problem for difference-making theories. The problem is that there are many variables (using Woodward's terminology), which, by manipulating them, would make a difference to some effect. But, many of the variables that would make a difference to an effect are considered to be insignificant and ultimately non-causal. Thus, there must be a distinction between insignificant difference makers and significant ones. I will highlight this problem and argue that the agency approach that I prefer, viz., Causal Interactionism, can successfully accommodate the distinction and explain why it is important.

Regarding the distinction between difference makers, that is, the need to distinguish between insignificant and significant difference makers, we can situate the problem by asking a question. In cases where there are various difference makers, why are we willing to accept one claim about causation as true and not another when both variations fit the difference-making criteria?

To highlight the problem, consider the following example.

Imagine that you have just returned from a week-long philosophy conference entitled 'Causation and Difference Making'. As you enter your flat, you notice that your favourite plant has died. Before you left for the conference, you had asked Sally, your neighbour, to water the plant while you were away. On seeing the dead plant, you infer that Sally forgot to water it or that she got the dates for the conference wrong. Therefore, the cause of the plant's death is Sally's omission—*Sally's not watering it*. In other words, what made a difference to the plant (either probabilistically or counterfactually) was the fact that Sally did not water it. However, upon reflection, you suddenly realise a problem: had *anyone* watered the plant, it would not have died. So, for example, had the Queen watered the plant, it would not have died. Should you conclude that the cause of the plant's death was the *Queen's not watering it*? The Queen's omission indeed made a difference to the plant, and it is technically correct to conclude, via the difference-making analysis, that the Queen's not watering the plant caused it to die.

Consider another example. Imagine that a hurricane has just levelled Jim's house and every house on his block. Jim's neighbour, Lorenz, who has been out of town, arrives and asks Jim what happened to his home. Jim tells him that a hurricane destroyed every house on the block—his house was destroyed because of the hurricane. While this explanation seems initially plausible, Lorenz, who is a physicist, tells Jim that he is somewhat mistaken. Large events like hurricanes, Lorenz explains, can result from small causes like the movement of a butterfly. The precise strength and direction of the

storm was determined by sensitive initial conditions, which are influenced by minor perturbations such as the flapping of the wings of a distant butterfly. So, given a standard difference-making analysis, we can say that the cause of the damage was the flapping of a butterfly's wings.

What these examples are meant to reveal is that we need a way to distinguish between difference makers. Some difference makers, like hurricanes and Sally's omission, seem more significant than a butterfly and the Queen. The important question is, how do we distinguish between significant and insignificant difference makers?

One thought is that we can assign certain states, e.g., calm (non-hurricane) weather, as defaults. We hold Sally responsible for the dead plant because we think that, by default, people in distant places or strangers like the Queen bear no responsibility to you or to your possessions. Being clear about defaults is useful when thinking about the distinction between difference makers, particularly in the context of causal modelling. Interventionists (Pearl, 2000; Woodward, 2003) model 'small worlds' where causal relationships pertain to a limited number of variables. It may be the case that we do not include things like the Queen and the butterfly into a causal model because of some prior expectation—by default we think that these factors are irrelevant. However, note that this does not *explain* the problem—it simply takes for granted the fact that we hold some things to be more significant than others. We still need to answer the following question: If we think that the Queen's omission made a difference to the plant, why not include it into a causal model?

Before I answer the question, I want to consider a well-known solution related to this problem. Peter Menzies, in his article, 'Difference Making in Context' (2004), points out that it is often taken to be a matter of pragmatics which 'draws a distinction between causes and background conditions' (p. 144). 'Conditions' are best thought of as all of the conditions of the physical environment prior to a particular event. Given a strong realist assumption about causation (e.g., the kind that Russell (1913) rejected), the event in question is taken as an effect, and the prior conditions are all the causal influences that make up or brought about the effect. So, the distinction between causes and conditions is threefold: it includes *irrelevant* factors (non-causes), non-significant difference-making factors (factors that technically have a causal influence on the effect that we reject as causal), and difference-making factors (what we call causes). What Menzies and myself want to show is that non-significant difference-making factors are also irrelevant—they are

simply non-causes. Consider an example used by Menzies:

A person develops lung cancer as a result of years of smoking. It is true that if he had not smoked he would not have developed cancer. It is also true that he would not have developed lung cancer if he had not possessed lungs, or even if he had not been born. But it is absurd to think his possession of lungs or even his birth caused his lung cancer (Menzies, 2004, p. 143).

We might say that we do not think of the Queen's omission or being born as causes because the human mind selectively and pragmatically ranks *the* cause to be some other factor. 'The principle of invidious selection', as David Lewis calls it, ranks most of the causes of an event as mere background conditions, disqualifying them from being difference makers. For Lewis and others, who hold the metaphysical assumption that causation is a real and absolute property in the world, the Queen being born and a profusion of other events *are* causes. According to Lewis's theory, 'any event but for which the effect would not have occurred is one of its causes' (Menzies, *ibid*). Menzies wants to show that this is false. He intends to show that the Queen, being born, and a profusion of other events are not causes.

How does Menzies do this? Menzies calls the distinction between causes and conditions (conditions will include non-significant difference makers) the problem of *profligate* causes (2004, p, 143). Profligate causes are *useless* causes or what Menzies means by non-causes. Menzies argues that we can make the distinction by focusing on context. For Menzies, C makes a difference to E relative to F^m where F is a system or model and *m* represents the normal laws and conditions within a system.

Menzies notes two parameters that provide context for causal claims. First, the *context of enquiry*, and second, the *context of initial conditions* or what he calls the context of 'normal conditions'. The first parameter states that causation is relative to enquiry, that is, the kinds of questions we ask, and the second condition says that what we include into a causal model or a system is based on the normal laws and conditions that make up that model or system.

To illustrate these parameters, consider again the problem of the Indian famine that I discussed in chapter 5:

The cause of a great famine in India may be identified by an Indian peasant as the drought, but the World Food Authority

may identify the Indian government's failure to build up reserves as the cause and the drought as a mere condition (Menzies takes this from Hart and Honoré 1985: pp. 35–6).

This example reveals that the same situation elicits different causal judgments. The idea is similar to what Collingwood brings to light in the practical sense of cause. I will highlight the connection with Collingwood below. For now, note that the example demonstrates that different agents determine that the same events have different causes, depending on the type of enquiry being undertaken (e.g., the peasants are asking, 'why did our crops die?' and the World Food Authority is asking, 'why did the peasants starve?').

Here is another example from Menzies:

If a building is destroyed by fire, it may be true that the fire would not have taken hold but for the oxygen in the air, the presence of combustible material, and the dryness of the building. But these are not significant difference makers. On the other hand, if a fire breaks out in a laboratory or in a factory, where special precautions are taken to exclude oxygen during the experiment or manufacturing process, it would not be absurd to cite the presence of oxygen as a significant difference maker and a cause of the fire. In both situations it may be true that the fire would not have occurred if oxygen had not been present (Menzies modifies this example from Hart and Honoré, 1985: pp. 35–6).

This case reveals that what we include into a causal system is relative to the normal laws and conditions of a particular system. Menzies calls this the 'relativity to the context of occurrence'. The 'context of occurrence', usually implicitly, gives us the 'normal conditions' that arise within a particular system—a system being an indeterminate group of objects bound by *ceteris paribus* laws:

A particular system may consist of a great many objects or very few, of very large objects or very small ones. Astronomers and cosmologists investigate vast systems—solar systems, galaxies, or the cosmos itself. The systems investigated by biologists and economists—economies, markets, species, populations, and so on—are smaller, but still large by human standards. On the other

hand, the systems investigated by particle physicists are small by any standard. It is not always easy to determine which objects belong to a particular system. This is not just because of our epistemic limitations, but because the spatiotemporal boundaries of the system are indeterminate (Menziez, 2004, p.146).

I find much of what Menziez has to say about contextual parameters illuminating. However, there is much more to be said about how, and indeed why, we come to distinguish between difference makers. Why is the context of inquiry helpful for making the distinction? Why is the context of initial (ordinary) conditions likewise useful for distinguishing between causes and non-causes? Are there more contexts that we use to distinguish between difference makers? Below, I argue that we can answer these questions by thinking about human agency and causal reasoning.

Before I offer an explanation for the above difficulties, it will be useful to consider the notion of causal reasoning. Causal reasoning is steeped in abstraction. What Menziez calls a system (what Pearl and Woodward call ‘small worlds’) is highly abstract. We begin to abstract away much of the world because we have particular interests and goals. You may want to avoid being injured or getting ill, and your interest leads you to ask a specific set of questions; this, in turn, leads to the fact that much of the world becomes irrelevant. If you want to avoid getting lung cancer or the flu, for example, you will not be interested in how many universities are in California, the number of bolts that are needed to build a 1957 Chevy, or whether I prefer coffee over tea.

Causal reasoning also involves value judgments. Collingwood references this idea when he talks about the hypothetical case of scientists claiming to have discovered the cause of cancer, but the alleged “cause” does not enable us to prevent cancer. Collingwood rejects the idea of a cause where the notion of a handle is not applicable (where the “cause” cannot be used to produce or prevent). What this reveals is that when we generalise a causal relationship, ‘C causes E’, we are interested in producing a *fruitful* generalisation. That is to say, we are interested in generalisations that serve particular ends. When causal generalisations are fruitful, having control over C (the cause) will allow us (practical or hypothetical) control over E (the effect). We think that if we can produce or prevent C, we can produce or prevent E.

Causal reasoning also relates to what Pearl (2000) calls the essence of causation—prediction and control. To illustrate this, consider the examples

shown above. In all of the examples, I claim that the distinction between difference makers can be captured by considering human agency and the notion of control.

Absurdities Explained

1. The Queen: The Queen's omission is not a cause because there is no obvious sense in which we have control over the Queen or where she has control over the plant (she does not know who you are or where you live or even that you own a plant, for example). Sally's omission is considered causal because there is an obvious sense in which we have some control over what she does, for example, we interacted with her and asked her to water the plant (in Menzies' analysis, we would say that the 'normal condition' is that Sally waters the plant).
2. Lung Cancer: Having lungs and being born are not causes in the sense that there is no obvious or ethical way for us to control lung cancer by removing people's lungs or killing babies. On the other hand, smoking is considered a cause of lung cancer because we can control whether or not we choose to smoke.
3. The Presence of Oxygen: Oxygen is not a cause of the fire in the first instance because there is no obvious way for us to control the level of oxygen. However, the reason why oxygen is a cause of the fire in the laboratory is that in this context, unlike the earlier context, we do have control over the level of oxygen.
4. The Indian Famine: The peasants consider the cause of the great Indian famine to be the drought because the peasants think that if there were more rain, then the famine would not have occurred. Alternatively, the World Food Authority thinks that a lack of food storage is the cause of the famine because that is what they can control. For example, if they had intervened by storing more food, then the famine would not have occurred.

I said that all of these cases could be captured by thinking about the notion of agency and control. However, you would be correct to point out that in the event of the Indian Famine, the peasants cannot (overtly) control the level of rain and thus the occurrence of a drought. The Indian Famine example reveals a problem for the idea that the distinction between difference makers can be captured by agency, viz., that it is impossible to overtly

manipulate and control things like the weather. However, the problem only arises for standard agency theories of causation, e.g., Price's (1993) account. The peasants cannot *overtly* control the level of rain. The problem does not arise for Causal Interactionism. The peasants can *reason*³—and it is likely that this is what they are doing—that if they *could* control the level of rain, they could have prevented the famine (Hindus believe that the deity Indra controls the level of rainfall. In Vedic mantras, Indra is called upon to release rain in times of draught and to stop downpours in times of excessive rain. In the Vedic text, Manusmriti 3.76, we see the belief that worship could affect rainfall, 'An oblation duly thrown into the fire, reaches the sun; from the sun comes rain, from rain food'). This is significant. What this shows is that Causal Interactionism can resolve *all* of the absurdities, whereas traditional agency theories cannot. Indeed, a benefit of Causal Interactionism is that it can achieve the distinction between significant and non-significant difference makers in cases where the reasoning involves hypothetical agency—reasoning about the effects of ϕ where ϕ represents an impossible human action.

So, given that we are thinking in terms of Causal Interactionism, we can explain the distinction between difference makers and overcome the alleged absurdities noted above. The fact that we can show that the distinction between difference makers can be captured by thinking about control and human agency is a benefit of the agency approach. The fact that we can capture the distinction in cases where overt action does not apply (e.g., the Indian Famine and the peasants) is a further benefit of Causal Interactionism.

The last benefit of Causal Interactionism is quite significant, and it's worthwhile to briefly elaborate on what has been said. How we interact with the world as agents is based on context—contexts like enquiry and normal conditions. In the example of the building fire and the Indian famine, different agents determined different causal relationships depending on how they *interact* with the world. That is, the kinds of questions we ask and the 'normal conditions' that we take for granted are all included in the unique way that we interact with the world as practical agents. When we interact with the world as agents, we always have some particular end in mind, and

³The reason the peasants apply this type of reasoning to [rainfall] is because rain is naturally variable. In previous years, we can assume, the level of rain was much higher. The famine occurred, according to the peasants, because the amount of rain varied to such a degree that their crops died. Below, I will further explain why the peasants view the lack of rain as the cause, and why other situations, e.g., the plant and the Queen, do not necessarily involve further elaborations of hypothetical reasoning.

this end will help determine the things we abstract away, based in part on what the agent takes to be the normal conditions. The end we are interested in achieving will also help to determine what kinds of questions we ask. In simple terms, context is merely a natural part of the interactions that we have with the world as agents. So, for Causal Interactionism, the context that Menzies' speaks of is directly built in, and the cause of an event can vary because different agents interact with the world in various ways. Allow me to further explain.

It is true that if the peasants had reasoned *if the World Food Authority had stored more food, then the famine would not have occurred*, then the cause of the famine for the peasants would be the lack of food stores. However, the lack of food stores is not a cause for the peasants because their "normal conditions" do not include a certain level of food stores. This is revealed, in part, by what kind of question the peasants ask about the famine, e.g., why did our crops die?, which is simply a measure of how they interact with the world. But, this also shows a further type of context: the *context of prior expectations*. If the peasants had some prior expectation about the amount of food they would receive from the World Food Authority, they may have reasoned about the food stores. In this case, the lack of food would be a cause for the peasants as well as for the World Food Authority. If the peasants had a prior expectation regarding the level of food stores, they might have asked a different question, e.g., why wasn't there more food stores? Thus, their interaction with the world would be different. But, the peasants had no prior expectation of food stores. They did have prior expectations regarding the amount of rainfall that would occur in a particular season. Thus, and rather simply, a cause is determined or ascribed to a particular event by the kind of interaction we have with the world as agents. The interactions we have with the world as agents include or assume the kinds of questions we ask, what we take to be "normal", and some level of prior expectation.

Imagine that the cause of a house fire was taken to be the running stove—the fire was caused because the stove was left running. In this case, the homeowner may reason that, if she had turned off the stove, she could have prevented the fire. However, it would also be true if the homeowner had reasoned *if the home was never built, then the fire would not have occurred*. This last counterfactual is true. But, the house being built is not taken as a cause because, among other things, the homeowner is asking, why did *this* house burn down? Moreover, when reasoning about the cause of the house fire, she is considering some prior expectations about her behaviour (that she

turns off the stove) and the normal conditions of the home when she is away (that the stove is turned off).

So, the particular cause we ascribe to an event is based on how we interact with the world. In many cases, like above, the kinds of questions we ask, what we consider to be normal, and our prior exceptions will determine which events, objects, or variables we deem to be significant and ultimately causal. If a young boy dies because a stray bullet hit him, there are likely hundreds if not thousands of counterfactuals that could be used to explain his death. Had he not been born, had he not woken up, had he not been at that particular street corner, had the shooter not..., etc. But, what we ascribe as *causal* will be determined by the interactional engagements we have with the world as agents. If the boy's mother sent her child to the corner to get milk because the mother had forgotten to buy it the day before, she might hold herself responsible. Alternatively, when I make a causal model of the event, I do not include the mother's request for milk because I have no prior expectation that a mother would avoid doing such a thing. I would not include the boy being born or the fact that he got out of bed that day for the same reason.

Significantly, the interactions we have with the world as agents can also clarify how we ascribe the notion of control to particular objects or events. Similar to what Collingwood points out, the mother applies the idea of control to her decision to send her son to get milk. For the mother, her prior expectations and her normal conditions lead her to feel a sense of control of her son's movements around the city. The fact that she applies the idea of control to her decision regarding her son leads her to believe that she is causally responsible for her son's death. When we make a causal model of the event, we may lack the same prior expectations and assume different normal conditions. For example, we do not hold the prior expectation that a mother refrains from asking her son to get milk. We may hold the expectation that people will not shoot guns in the city, and this may lead us to think that the shooter is responsible for the boy's death. In other words, we apply the notion of control to the shooter.⁴ Thus, we reason that had I manipulated the shooter, or if we could manipulate further shootings, I believe that I could

⁴This is somewhat different from Collingwood. When I apply the idea of control to the shooter, I may not have a practical way of controlling shooters—maybe if I were a police officer or an FBI agent, I would—but this is not needed. I apply the notion of control to shootings based on my prior expectations and what I take to be normal conditions, not, as Collingwood says, because I can control shootings. This reveals an instance of applying the notion of control in the *hypothetical* sense.

have prevented the boy's death or that I could prevent future deaths from occurring.

8.4.1 Moral Judgment

I said that how we interact with the world as agents includes things like the questions we ask, the ends we are interested in achieving, normal conditions, and prior expectations. All of these contexts or interactions go into causal ascription. These are perhaps some of the most significant factors or contexts that make up our interactions, but there are others. One that is also worth mentioning, and perhaps one that is tied up to some extent with the others, is moral reasoning or moral judgment. Moral reasoning is practical reasoning about what, morally, one ought to do (or in the past tense, what one ought to have done). Moral judgment is the act of ascribing moral value to a particular action—to judge an action as moral or immoral is to say that the action is *right* or *wrong* (worthy of *praise* or *blame*). What I want to do here is draw attention to the fact that in some cases, causal ascription will relate to moral intuitions; I want to highlight the context of *moral judgment*.

To be clear, I argue that we can determine significant difference makers based on judgments regarding moral responsibility. That is, in cases where the significant difference makers have *not yet* been determined, we may determine them by considering moral judgments. There may be cases where our intuitions regarding significant difference makers lead us to determine moral responsibility. In such cases, what is happening is likely that we have used one of the other contexts, e.g., the context or normal conditions, to determine significant difference makers. If a burglar enters my home and breaks my television, using the context or normal conditions, I may reason that the burglar is causally responsible for my damaged television because, under normal conditions, people do not break into my home and try to steal my possessions. In this case, I might *then* judge the burglars actions as wrong because I have already determined him to be causally responsible for damaging my television. I am interested in reversed cases—that is, cases where moral judgments allow us to determine significant difference makers.

A possible example of the context of moral judgment is the case presented above. It may be the case that the mother feels morally responsible for her son's death and then ascribes the notion of control to her decision regarding her son. In this case, her moral reasoning leads to her conclusion that she is causally responsible for her child's death—i.e., she takes her decision re-

garding her son to be a significant difference maker. This being the case, the mother is saying, ‘what I did is wrong; therefore, I’m causally responsible’. In other words, if the mother did not feel a sense of morally responsibility for her son’s death, she would not consider her action to be a significant difference maker.

However, if I were to model the event of the child’s death, I may reason differently. I could argue with the mother, ‘you did nothing wrong; therefore, you are not causally responsible.’ My argument would reveal that I do not share the same context of moral judgment as the mother. I may judge that the shooter acted wrongly and therefore that his actions are worthy of blame. In this case, I am reasoning that the shooter was wrong and that the shooter is causally responsible for the child’s death—the shooter’s actions represent a significant difference maker. I do not deem the mother’s decision regarding her son to be wrong, and thus, I do not consider the mother’s action to be a significant difference maker.

The context of moral judgment may also be relevant to the case of the plant and the Queen.⁵ When I ask Sally to water my plant for me while I am away, assuming she agrees and that we have the dates right, I consider Sally to be *morally responsible* for my plant’s well-being while I am away. I do not consider strangers like the Queen to bear any morally responsibility for my plant’s well being. This sense of moral responsibility leads me to judge Sally’s omission as wrong, and this, in turn, leads to my judgment that Sally is causally responsible for my plant’s death. Sally is a significant difference maker, not the Queen, because Sally acted wrongly.

Something resembling the context of moral responsibility is also present in Collingwood’s historical sense of cause. Collingwood, recall, tells the story of the Fall—of Adam and Eve in the garden. When God asks Adam why he ate the forbidden fruit, Adam replies that he did so because Eve told him to; he sinned because Eve persuaded him to do so. What we see from this, implicitly, is that Adam does not feel morally responsible for his sin. Adam puts the blame on Eve because he thinks that her request for him to eat the fruit makes her *morally responsible*. According to Adam, it was Eve’s actions, not his own, which led to the Fall. Because Eve acted wrongly, and not Adam, Eve’s actions represent a significant difference maker. However, God reasons differently. God holds Adam and Eve morally responsible for their

⁵Assuming we have not already determined relevant difference makers using another context, e.g., the context or normal conditions which was used above.

own actions, and he subsequently banishes both of them from the garden. For God, Adam and Eve both acted wrongly, and both of their actions represent significant difference makers.

The brief foray into moral reasoning and moral judgment is meant to highlight the fact that our moral judgments will often help to determine significant difference makers. In other words, actions that we judge wrong and blame worthy will often go into our considerations of causal ascription, as I have tried to indicate with these examples. A simple way of expressing this is to say that how we interact with the world as agents will sometimes include a particular kind of moral reasoning or moral judgment, and these judgments will go into causal ascription.

Chapter 9

Conclusion

The primary work of my thesis can be broken into two parts.

First, I have tried to show that there is, in fact, a significant and essential relationship between causation and agency. I take it that Collingwood is correct, along with Woodward and Price who agree with him, that causation is agent-relative. Causation is agent-relative because causes are handles—that is, they are things in nature that give us control (over other things in nature). In essence, the causal relation is agent-dependent because causes (as handles) are essentially connected to (overt or hypothetical) production or prevention.

There are numerous implications for the causal relation. As I argued in chapter 4, against Strevens, the causal relation is not mind-independent and absolute. As Strevens himself notes, frameworking, which involves localising a causal relationship, is fundamental to scientific practice—indeed, even to physics. It follows that the causal relation is relative. Though it is unclear just based on relativism that there is a constitutive relationship between agency and causation, I argued that causation is relative to practices like frameworking, and thus human agency, because, as Collingwood and Pearl note, handles (and thus causes) would disappear if we take the causal relation as absolute and frame-independent.

Following the work of chapter 5, we can also claim that the causal relation is not, *fundamentally*, physical. In other words, the causal relation doesn't supervene directly on the physical—agents get in the way, even in physics. Of course, in some cases, a causal relationship will clearly reference physical facts. Indeed, a causal relationship may be reducible to physical facts. Agency theorists can happily accept the claim that sometimes facts

about a causal relationship will be physical. However, in other cases, there may be no straightforward physical connection at all. We can assume that the stock market crash in 1929 had something to do with physical facts, but it is impossible to say what all of these facts are or even why they are significant. More importantly, we saw that practices like frameworking represent an equally significant role when thinking about the nature of causation—so much so that we can say that the causal relation *essentially* relies on practices like frameworking. For example, we can only say that C causes E when C and E have both been localised (or made relative) to some kind of frameworking. Now, we can see much more clearly what this means: frameworking results from a particular *interaction* between agents and the world, and thus human agency is also an essential element of the causal relation. By critically analysing Ney’s theory of foundationalism, we are able to make a significant and substantial metaphysical claim: the nature of causation is not (entirely) physical.

In essence, chapters 4 and 5 were used to motivate the idea that causation is agent-dependent or agent-relative by supporting the idea that the causal relation is not mind-independent, absolute, or fundamentally physical.

Secondly, I defended an agency approach to causation against two critical concerns: the problem of scope and the problem of subjectivity. As I explained in chapter 6, because the problem of scope can be taken as the result of an untoward anthropocentrism, I also had to explain how the agency approach avoided an anthropocentrism of this kind. To do so, I argued that the agency approach, which includes Woodward’s theory of interventionism, is anthropocentric in what I called the common sense—it is anthropocentric because human agents are given a privileged status regarding causation. Significantly, I was able to show that the agency approach avoided an untoward anthropocentrism, that is, metaphysical anthropocentrism, by rejecting the claim that an agency approach stipulates that the causal relation is somehow *created* by human agents.

By highlighting the debate between Woodward and Price, I further demonstrated that the agency approach is capable of capturing a sense of objectivity. I argued that the agency approach is objective in a relative sense while explaining how Woodward, Price, and Collingwood all reject the notion that causation is objective in the absolute sense. By revealing the ways in which all three philosophers rejected the idea that causation is absolutely subjective, I concluded that the agency approach was, indeed, objective. Given that the causal relation is relative to certain interactional parameters, we

can say that the claim ‘A causes B’ is an objective fact. It is not a mind- or agent-independent fact, as Price argues, but it is a fact about the world nonetheless, as Woodward rightly suggests. This is significant for two reasons: I was able to resolve the debate between Woodward and Price while overcoming the problem of subjectivity.

By demonstrating that a cause is something an agent can manipulate—either overtly or hypothetically—the agency approach that I prefer, Causal Interactionism, is able to capture well-known scientific claims about causation, which, as I said in the introduction to this thesis, appear to have nothing to do with human agency. It *appears* that claims such as ‘a large asteroid caused the extinction of the dinosaurs’ has nothing to do with agency because such claims do not rely on *overt* agency. Nonetheless, we do think that if we had prevented the asteroid from hitting the earth near the Mexican peninsula, we could have prevented the extinction. Significantly, by demonstrating how such hypothetical reasoning involves real agency, I developed a method to extend causation to events that do not or cannot involve overt human agency and thus solved the problem of scope.

Lastly, I pointed out that Causal Interactionism allows us to make a strong (and clear) metaphysical statement about the causal relation. Because the causal relation is neither physical nor anthropocentric, the nature of the causal relation becomes *interactional* (Chapter 7). Indeed, the interactional nature of causation fits well with Price’s ‘minimal’ and Woodward’s ‘moderate’ realism. Two statements follow: the causal relation is (1) something that is a part of the natural world (as Woodward likes to point out), but it is (2) not something that exists independently of human agents (as Price continually maintains). In many ways, Causal Interactionism represents a coalescence and unites the ideas of both Woodward and Price. But, the two key ideas have been dormant in Collingwood all along. Indeed, the bulk of this work boils down to two foundational claims that Collingwood stated over 75 years ago—that a cause is something which exists in nature but is agent-relative. In many ways, and as we have seen, Price’s work has been focused on the later of Collingwood’s two claims, while Woodward has tended to focus on the first (though they both accept each one). Once statement (2) has been sufficiently motivated, all that was needed was (i) a way to show that both (1) and (2) could be consistently maintained in light of apparent counterexamples, like the dinosaurs’ extinction (see §2 & §3 of this Chapter), and (ii) that an agent-relative approach to causation could somehow maintain a sense of objectivity.

In light of the work of Causal Interactionism, Collingwood's words become appreciably well founded. Causation, Collingwood says,
rests on...the relation between man and nature (1940, p. 310).

Bibliography

- [Anscombe, 1957] Anscombe, G.E.M. (1957). *Intention*, Oxford: Basil Blackwell.
- [Ayer, 1971] Ayer, A.J. (1971). *Language, Truth, and Logic*. Penguin Books.
- [Baghramian, 2010] Baghramian, M. (2010). ‘A Brief History of Relativism’, in M. Krauz, *Relativism: A Contemporary Anthology* (ed). Columbia University Press. 31–52.
- [Barandiaran, *et al*, 2009] Barandiaran, X.E., E. Di Paolo, and M. Rohde, (2009). ‘Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action’, *Adaptive Behavior*, 17(5): 367–386.
- [Barnouw, 1997] Barnouw, D. (1997). ‘The Shapes of Objectivity’, in A. Megill (ed), *Rethinking Objectivity*. Duke University Press, 127–50.
- [Beebe, 2007] Beebe, H. (2007). ‘Hume on Causation: The Projectivist View’, in Corry and Price (eds.), *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, Oxford: Oxford University Press, 224–249.
- [Beebe, 2006] Beebe, H. (2006) *Hume on Causation*, Oxford: Oxford University Press.
- [Bennett, 2017] Bennett, K. (2017). *Making Things Up*, New York: Oxford University Press.
- [Bernstein, 1983] Bernstein, R.J. (1983). *Beyond Objectivism and Relativism*. University of Philadelphia Press.
- [Bernstein, 2016] Bernstein, S. (2016). ‘Grounding is not Causation’, *Philosophical Perspectives*, 30(1): 21–38

- [Bernstein, 2017] Bernstein, S. (2017) ‘Causal Idealism’, in T. Goldschmidt and K. Pearce (eds.), *Idealism: New Essays in Metaphysics*. Oxford University Press.
- [Carr, 1961] Carr, E.H. (1961). *What is History?* (Penguin Books).
- [Cartwright, 1979] Cartwright, N. (1979). ‘Causal Laws and Effective Strategies’, *Noûs* 13 (4): 419–37.
- [Collingwood, 1940] Collingwood, R. G. (1940). *An Essay in Metaphysics*. Oxford: Oxford University Press.
- [Collingwood, 1946] Collingwood, R.G. (1946). *The Idea of History*. Oxford: Oxford University Press.
- [Collingwood, 1960] Collingwood, R. G. (1960). *The Idea of Nature*. New York: Oxford University Press.
- [Daston and Galison, 2007] Daston, L. & Galison, P. (2007). *Objectivity*. MIT Press: Cambridge.
- [Davidson, 1963] Davidson, D. (1963), ‘Actions, Reasons, and Causes’, reprinted in Davidson (1980) *Essays on Actions and Events*, Oxford: Clarendon Press. 3–20.
- [Davidson, 1971] Davidson, D. (1971). ‘Agency’ in Binkley, R.W., Bronaugh, R.N., Marras, A., (eds) *Agent, Action and Reason*: University of Toronto Press. 3–26.
- [Dixon, 2016] Dixon, S. (2016) ‘What Is the Well-Foundedness of Grounding?’ *Mind*, 125(498): 439–468.
- [Dowe, 2000] Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.
- [Elga, 2007] Elga, A. (2007). ‘Isolation and Folk Physics’, in Corry & Price (eds.), *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, Oxford: Oxford University Press, 106–119.
- [Fabian, 1997] Fabian, J. (1997). ‘Ethnographic Objectivity Revisited’, in A. Megill (ed), *Rethinking Objectivity*. Duke University Press, 81–108.

- [Field, 2003] Field, H. (2003). ‘Causation in a Physical World’, in M.J. Loux & D. Zimmerman (eds.), *Oxford Handbook of Metaphysics*: Oxford University Press.
- [Fine, 1986] Fine, A. (1986). *The Shaky Game: Einstein, Realism, and the Quantum Theory*, University of Chicago Press.
- [Fine, 1993] Fine, A. (1993). ‘Fictionalism’, *Midwest Studies in Philosophy* XVIII.
- [Frankfurt, 1978] Frankfurt, H. (1978). ‘The Problem of Action’, *American Philosophical Quarterly*, 15: 2 157–162.
- [Frigg, 2010] Frigg, R. (2010). ‘Models and Fiction’, *Synthese* 172: 251–268.
- [Frisch, 2014] Frisch, M. (2014). *Causal Reasoning in Physics*. Oxford: Oxford University Press.
- [Galison, 2014] Galison, P. (2014). ‘The Journalist, the Scientist, and Objectivity’ in Padovani, F., Richardson, R., & Tsou, J (eds) *Objectivity in Science: New Perspectives from Science and Technology Studies*. Springer: New York.
- [Gasking, 1955] Gasking (1955). ‘Causation and Recipes’ *Mind* Vol. 64, No. 256 pp. 479–487.
- [Gaukroger, 2012] Gaukroger, S. (2012). *Objectivity*, Oxford: Oxford University Press.
- [Geertz, 1984] Geertz, C. (1984). ‘Anti Anti-Relativism’ *American Anthropologist*, New Series, Vol. 86, No. 2. 263–278.
- [Gillies, 2005] Gillies, D. (2005) ‘An Action-Related Theory of Causality’, *British Journal of Philosophy of Science* 56: 823–842.
- [Giere, 1988] Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.
- [Glazier, 2016] Glazier, Martin, 2016, ‘Laws and the Completeness of the Fundamental’, in *Reality Making*, Oxford: Oxford University Press: 11–37

- [Glynn, 2013] Glynn, L. (2013). ‘Causal Foundationalism, Physical Causation, and Difference Making’, *Synthese* 6: 1017–1037.
- [Godfrey-Smith, 2006] Godfrey-Smith, P. (2006). ‘The Strategy of Model-based Science’, *Biology and Philosophy*, 21: 725–740.
- [Godfrey-Smith, 2009] Godfrey-Smith, P. (2009) ‘Models and Fictions in Science’, *Philosophical Studies* 1: 101–116.
- [Goldman, 1970] Goldman, A. (1970). *A Theory of Human Action*, Englewood Cliffs, NJ: Prentice-Hall.
- [Goodman, 1978] Goodman, N. (1978). *Ways of Worldmaking*. Hackett Publishing Company.
- [Gopnik and Schulz, 2007] A. Gopnik and L. Schulz (2007). (Eds), *Causal Learning: Psychology, Philosophy, and Computation*. New York: Oxford University Press.
- [Hart and Honroé, 1985] Hart, L.A. and Honroé, T. (1985). *Causation in the Law*. Oxford: Oxford University Press.
- [Hartmann, 2009] Hartmann, S. (2009). ‘Models and Stories in Hadron Physics’, in Mary M. & Margaret, M. (eds) *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge University Press, 326–346.
- [Hausman, 1988] Hausman, D. (1998). *Causal Asymmetries*. Cambridge: Cambridge University Press.
- [Hitchcock, 2007] Hitchcock, C. (2007). ‘What Russell Got Right’, in Corry & Price (eds.), *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, Oxford: Oxford University Press, 45–65.
- [Hieronymi, 2009] Hieronymi, P. (2009). ‘Two Kinds of Agency’ in *Mental Actions*, L. O. Brien and M. Soteriou, (eds.), Oxford: Oxford University Press, 138–162.
- [Honoré, 2010] Honoré, A. (2010). ‘Causation in the Law’, *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition).

- [Jago, 2016] Jago, M. (2016) (ed.), *Reality Making*, Oxford: Oxford University Press.
- [Jaynes, 2003] Jaynes, E.T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.
- [Kant, 1998] Kant, I. (1998). *The Critique of Pure Reason*, G. Paul & W. Allen (eds). Cambridge: Cambridge University Press.
- [Keller, 1997] Keller, E.F. (1997). ‘The Paradox of Scientific Subjectivity,’ in A. Megill (ed), *Rethinking Objectivity*. Duke University Press, 313–332.
- [Krauz, 2010] Krauz, M. (2010). *Relativism: A Contemporary Anthology*. Columbia University Press.
- [Lewis, 1973] Lewis, D. (1973). *Counterfactuals*. Oxford: Basil Blackwell.
- [Lewis, 2004] Lewis, D. (2004). ‘Void and Object’, in Collins *et al.* (2004), 277–90.
- [Lower, 2001] Loewer, B. (2001). ‘Review of *Mind in a Physical World*’, *Journal of Philosophy*, 98, 315–24.
- [Mackie, 1974] Mackie, J.L. (1974). *The Cement of the Universe*. Oxford: Clarendon Press.
- [Megill, 1997] Megill, A. (ed) (1997). *Rethinking Objectivity*, Duke University Press, second edition.
- [Mele, 1997] Mele, A.R. (1997). ‘Agency and Mental Action’, *Philosophical Perspectives*, 11: 231–249.
- [Mele, 2009] Mele, A.R. (2009). ‘Mental Action: A Case Study’, in *Mental Actions*, L. O. Brien and M. Soteriou, (eds.), Oxford: Oxford University Press, 17–37.
- [Menzies and Price, 1993] Menzies, P. and Price, H. (1993). ‘Causation as a Secondary Quality’, *British Journal for the Philosophy of Science*, 44: 187–203.
- [Menzies, 2004] Menzies, P. (2004). ‘Difference-Making in Context’, in J. Collins, N. Hall, & L. Paul (eds) *Causation and Counterfactuals*. 139–80.

- [Menzies, 2007] Menzies, P. (2007). ‘Causation in Context’, in *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, Price & Corry (eds.), pp. 191–223.
- [Mill, 1875] Mill, J.S. (1875). *A System of Logic*. London: Parker and Son.
- [Moser, 1999] Moser, P.K. (1999). *Philosophy After Objectivity*. New York: Oxford University Press.
- [Ney, 2008] Ney, A. (2008). ‘Physicalism as an Attitude’. *Philosophical Studies*. 138: 1–15.
- [Ney, 2009] Ney, A. (2009). ‘Physical Causation and Difference-Making’, *British Journal of Philosophy of Science* 60: 737–764.
- [Noonan, 1999] Noonan, H.W (1999). *Hume on Knowledge*, London and New York: Routledge
- [Norton, 2006] Norton, J. (2006). ‘Do the Causal Principles of Modern Physics Contradict Causal Anti-Fundamentalism?’
- [Norton, 2007] Norton, J.D. (2007). ‘Causality as a Folk Science’, in Price, H & Corry, R. (Eds.) *Causation, Physics, and the Constitution of Reality*: Oxford University Press. 11–44.
- [Palmer, 1999] Palmer, S.E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT press.
- [Peacocke, 2009] Peacocke, C. (2009). Mental Action and Self Awareness (II): Epistemology, in *Mental Actions*, L. O. Brien and M. Soteriou, (eds.), Oxford: Oxford University Press, 192–214.
- [Pearl, 2000] Pearl, J. (2000). *Causality*. New York: Cambridge University Press.
- [Pickering, 1995] Pickering, A. (1995). *The Mangle of Practice: Time, Agency & Science*. Chicago: Chicago University Press.
- [Pickering, 1997] Pickering, A. (1997). ‘Objectivity and the Mangle of Practice,’ in A. Megill (ed), *Rethinking Objectivity*. Duke University Press, 109–126.

- [Potochnik, 2011] Potochnik, A. (2011). ‘Explanation and Understanding: An Alternative to Strevens’ *Depth*. *European Journal of Philosophy of Science*, 1: 29–38.
- [Proust, 2009] Proust, J. (2009). ‘Is There a Sense of Agency for Thought?’, in *Mental Actions*, L. O. Brien and M. Soteriou, (eds.), Oxford: Oxford University Press, 253–280.
- [Price, 1991] Price, H. (1991). ‘Agency and Probabilistic Causality’, *British Journal of Philosophy of Science*, 42: 15–76.
- [Price, 1992] Price, H. (1992). ‘Agency and Causal Asymmetry’ *Mind*, 101: 501–20.
- [Price, 1992b] Price, H. (1992). ‘The Direction of Causation: Ramsey’s Ultimate Contingency’. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, *Volume 2: Symposia and Invited Papers*, pp. 253–267.
- [Price, 1996] Price, H. (1996). *Time’s Arrow and Archimedes’ Point*. New York: Oxford University Press.
- [Price, 2001] Price, H. (2001). ‘Causation in the Practical Sciences: The Case for Pragmatism’ in Suppes, Costantini, Galavotti (eds), *Stochastic Causality* Stanford University Press. 103–121.
- [Price, 2004] Price, H. (2004). ‘Models and Modals’, in Douglas Gillies (ed.), *Laws and Models in Science*, Kings College Publications, 49–69.
- [Price, 2007] Price, H. (2007). ‘Causal Perspectivalism’, in Corry and Price (eds), *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, Oxford: Oxford University Press. 250–91.
- [Price, 2011] Price, H. (2011). *Naturalism Without Mirrors*. Oxford University Press.
- [Price, 2013] Price, H. (2013). *Expressivism, Pragmatism, and Representationalism*, Cambridge University Press.
- [Price, 2017] Price, H. (2017). ‘Causation, intervention and agency: Woodward on Menzies and Price’, In H. Beebe, C. Hitchcock, and H. Price (eds), *Making a Difference*. Oxford University Press, 73–98.

- [Price, 2017b] Misak, C & Price, H. (eds) (2017). *The Practical Turn: Pragmatism in Britain in the Long Twentieth Century*. Oxford University Press.
- [Putnam, 1982] Putnam, H. (1982). ‘Why There isn’t a Ready Made World’, *Synthese* 51:2, 141–167.
- [Putnam, 1990] Putnam, H. (1990). *Realism with a Human Face*. Harvard University Press.
- [Ramsey, 1929] Ramsey, F (1929). ‘General Propositions and Causality’, in D.H. Mellor (ed.) (1978). *Foundations: Essays in Philosophy, Logic, Mathematics, and Economics*. London: Routledge and Kegan Paul, 133–51.
- [Raven, 2016] Raven, M. (2016). ‘Fundamentality Without Foundations’, *Philosophy and Phenomenological Research*, 93(3): 607–626
- [Rorty, 1979] Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton University Press.
- [Russell, 1913] Russell, B. (1913). ‘On the Notion of Cause’, *Proceedings of the Aristotelian Society*, 13: 1–26.
- [Russell, 1948] Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. Routledge: London & New York.
- [Russo, 2011] Russo, F. (2011). Review of: *Depth. An account of scientific explanation*, Michael Strevens. *Theoria*, 71: 261–263.
- [Salmon, 1984] Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.
- [Salmon, 1998] Salmon, W. (1998). *Causality and Explanation*. Oxford: Oxford University Press.
- [Schaffer, 2009] Schaffer, Jonathan. (2009). “On What Grounds What”, in Chalmers, *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Clarendon Press. 347–383
- [Schlosser, 2015] Schlosser, M, (2015). ‘Agency’, *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.).

- [Sloman, 2009] Sloman, S. (2009). *Causal Models: How to Think About the World and Its Alternatives*, Oxford University Press.
- [Strevens, 2007] Strevens, M. (2007). ‘Review of Woodward, *Making Things Happen*’, *Philosophy and Phenomenological Research*, 74: 233–49.
- [Strevens, 2008] Strevens, M. (2008). ‘Comments on Woodward, *Making Things Happen*’, *Philosophy and Phenomenological Research*, 77: 171–92.
- [Weisberg, 2007] Weisberg, M. (2007). ‘Who is a Modeler?’, *British Journal for Philosophy of Science* 58: 207–233.
- [Woodward, 2003] Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.
- [Woodward, 2007] Woodward, J. (2007). ‘Causation With a Human Face’ in R. Corry and H. Price (eds), *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, Oxford: Oxford University Press. 66–105.
- [Woodward, 2007b] Woodward, J. (2007). ‘Interventionist Theories of Causation in psychological Perspective’, in A. Gopnik and L. Schulz (eds), *Causal Learning: Psychology, Philosophy, and Computation*. New York: Oxford University Press.
- [Woodward, 2008] Woodward, J. (2008). ‘Response to Strevens’. *Philosophy and Phenomenological Research*, 77: 193–212.
- [Woodward, 2009] Woodward, J. (2009). ‘Agency and Interventionist Theories’, in *The Oxford Handbook of Causation* Beebe, Hitchcock, and Menzies, eds. Oxford: Oxford University Press.
- [Woodward, 2012] Woodward, J. (2012). ‘A Functional Account of Causation’, *Philosophy of Science*, 81 (5): 691–713.
- [Woodward, 2013] Woodward, J. (2013). ‘Causation and Manipulability’, *The Stanford Encyclopaedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.).
- [Williamson, 2010] Williamson, J. (2010). *In Defence of Objective Bayesianism*. New York: Oxford University Press.

- [Williamson, 2007] Williamson, J. (2007). 'Causality' in D.M. Gabbay and F. Guentner (eds), *Handbook of Philosophical Logic*.
- [von Wright, 1975] von Wright, G. (1975) 'Causality and Determinism'. New York: Columbia University Press.
- [von Wright, 1971] von Wright, G.H. (1971). *Explanation and Understanding*: Routledge Library Editions.