Causes as Deviations from Actual Standards

A Perspectival Account of Causation

Georgina Statham

This thesis is presented for the degree of Master of Philosophy of

The University of Western Australia

School of Humanities

Discipline of Philosophy

2013

Abstract

This thesis defends an account of a kind of token causal judgement that is common in everyday discourse, according to which causes are deviations from the normal course of evolution of a system. The account is based on a theory of causation developed by Peter Menzies, which I reinforce with a notion of *normal* defined by Sarah McGrath. McGrath claims that on the relevant sense of 'normal', events (or states) are normal if they adhere to certain *actual standards*, which can be descriptive or normative. Combining Menzies' account of causation with McGrath's notion of *normal* results in an account (which I call the 'Menzies–McGrath model') according to which token causes are deviations from the normal course of evolution of systems that are governed by actual standards.

I argue that the Menzies–McGrath model is not consistent with a position that I call the 'natural network model'—a metaphysical picture that underpins most of the accounts of causation that have recently been defended. According to the natural network model, the causal history of the universe consists of a single network of events and two-place causal relations, to which true causal judgements directly refer. I defend an alternative metaphysical picture, according to which the truth values of token causal judgements are relative to the kinds of systems described above. That is, to systems that are open to intervention, and governed by actual standards. Which kind of system is relevant to a particular causal judgement is determined by the purpose of the individual (or group) making the judgement—that is, by what I call a 'purpose-dependent perspective'. The truth values of token causal judgements are thus not completely mind-independent, but perspectival.

Contents

Acknowledgements

I would like to thank my supervisors—Nic Damnjanovic, Miri Albahari and Nin Kirkham—for all their help over the course of my Masters research. Nic, thanks for your questions, which always seemed to get to the heart of an issue, and for continuing to read drafts after leaving UWA; Miri, thanks for your diligence, thoroughness, and enthusiasm; and Nin, thanks for your encouragement, and astute philosophical advice.

Thanks also to Stewart Candlish, for carefully copy editing the entire thesis; and to Peter Menzies, for reading an early draft of much of this thesis and offering valuable advice and discussion. Diane Stringer and Neil McDonnell also provided helpful comments on drafts of one or more chapters.

I would not have been able to undertake this Masters degree without the financial support of an Australian Postgraduate Award and a Theresa Symons Postgraduate Scholarship, both of which I am very grateful to have received.

Finally, thanks to Jacqui Boaks, for putting up with me in the office, and ensuring that I (mostly) stayed sane.

Introduction

In a recent series of papers, Peter Menzies has defended an account of the concept of causation based on the idea that we conceptualise parts of the world as instances of systems, each with a normal course of evolution, but subject to external interventions.[1] On Menzies' account, causes are interventions that make a difference to the normal course of evolution of these systems. In this thesis, I will defend a modified version of Menzies' theory as an account of one (but not the only) form of token causal judgements. The modification I make to Menzies' account involves reinforcing it with a notion of *normal* developed by Sarah McGrath, according to which normal events are those that adhere to *actual standards*. These actual standards can be either descriptive or normative—so, controversially, the resulting account of causation (which I call the 'Menzies–McGrath model') allows normative factors to play a causal role.

In Chapter 1, I defend the Menzies–McGrath model as a psychologically accurate account of (at least a large proportion of) the token causal judgements made in everyday life. In other words, the account of token causal judgements provided in Chapter 1 is an account of the semantics of our everyday concept of causation. I show that the Menzies–McGrath model is better able to account for three phenomena of causal discourse—context sensitivity, judgements of causation by omission, and judgements in which normative factors appear to play a causal role—than the position that currently dominates the metaphysics of causation, which I refer to as the 'natural network model'.

According to the natural network model, the causal history of the universe is a single structure (or network), consisting of events linked by mind-independent, two-place

[1] Peter Menzies, "Difference-Making in Context" *Causation and Counterfactuals*, eds. J. Collins et al. (Cambridge, MA: MIT Press, 2004); Peter Menzies, "Causation in Context" *Causation, Physics, and the Constitution of Reality*, eds. H. Price and R. Corry (Oxford: Clarendon Press, 2007); Peter Menzies, "Platitudes and Counterexamples" *The Oxford Handbook of Causation*, eds. H. Beebee et al. (Oxford: Oxford University Press, 2009).

causal relations. Our causal discourse *represents* this structure, in that truth values of causal sentences are directly determined by the network. That is, the sentence '*c* caused *e*' is true if and only if the events *c* and *e* are both contained in the network, and are linked by a series of causal connections leading from *c* to *e*. Finally, in this metaphysical picture, causation is (often implicitly) held to be a *natural* relation. Exactly what it takes to be a natural relation in the sense intended by those who accept the natural network model is hard to define, but Menzies has pointed out that natural relations are typically thought to possess the following collection of features: first, natural relations are *contingent* and *a posteriori*; second, they form *scientific natural kinds*—that is, they are objective features of the world, and are therefore suitable targets of scientific investigation; and finally, natural relations are *non-normative*, in that whether a particular natural relation obtains is not influenced by human norms or values.[2]

Developing an account of the metaphysics of a particular phenomenon is always a trade-off between competing factors. In the case of causation, close observation of causal discourse reveals a tension between the semantics of everyday causal judgements and a number of assumptions that underpin the natural network model, including the claims that the causal history of the universe is a single relational structure, and that causation is a natural relation. In other words, there is a gap between the semantics of token causal judgements, and the metaphysics of the natural network model.

In Chapter 2, I present a series of considerations that all individually support the conclusion that this gap is unbridgeable. That is, the majority of Chapter 2 consists in a

[22] Menzies argues that most metaphysicians of causation implicitly accept that causation is a natural relation (where to be a natural relation is to have the features described above) in Menzies, "Platitudes and Counterexamples". For example, Beebee, who explicitly endorses the network model of causation, asserts that causation is non-normative, and that the (metaphysical) concept of causation 'carves nature at its joints'; she thus implicitly endorses the claim that causation is a natural relation. Helen Beebee, "Causing and Nothingness" *Causation and Counterfactuals*, eds. J. Collins et al. (Cambridge, MA: MIT Press, 2004).

convergent argument against the natural network model; an argument that, I claim, provides sufficient reason to reject the natural network model in favour of an alternative metaphysical picture evoked by the Menzies–McGrath model. On this alternative, everyday token causal judgements refer to open, partial systems, rather than to a single relational structure.

In the remainder of the thesis, I develop the components of this alternative to the natural network model. Because the Menzies–McGrath model holds that causes are *interventions* in the normal course of evolution of a system, this account can be considered as an agency, or manipulability, theory of causation. As a way of surveying the metaphysical landscape, I discuss the metaphysics of three manipulability accounts in Chapter 3: Menzies and Huw Price's agency theory, James Woodward's interventionism, and Price's later perspectivalism. The latter two positions can be considered as developments of Menzies and Price's account—Woodward's in the direction of increasing realism or objectivity, and Price's in the direction of increasing antirealism or subjectivity.

I reject Woodward's objectivist interventionism, and endorse Price's claim that causation is perspectival. That is, the truth values of causal judgements are dependent on features of us, as well as on features of the mind-independent world. However, Price claims only that the *direction* of causation is perspectival, and thus (implicitly) that the perspectivalism of the concept of causation is *intersubjective*—it arises from the epistemic position we occupy as agents embedded in time.

In Chapter 4, I discuss the perspectives involved in causal reasoning, and argue, contrary to Price, that the perspective from which we make and evaluate causal judgements is not intersubjective, but *purpose-dependent*. We represent the world differently, depending on our interests and purposes at any particular time, and these

different representations result in different token causal judgements. In the second half of the chapter, I consider the structure of the purpose-dependent perspective, and conclude that this perspective consists of a model of the normal course of evolution of a kind of system, as well as a set of *possible interventions* in these systems.

Finally, in Chapter 5, I defend an account of the metaphysics of (one kind of) causation, a position that I refer to as 'perspectival realism'. According to the Menzies–McGrath model, in order for a particular token causal claim to be true, three conditions must be met. Roughly, these are as follows: first, the cause and effect must be the actual values of the relevant variables; second, it must be true that if the cause had not occurred, the effect would not have occurred, relative to the normal course of evolution of the relevant kind of system; and finally, the purpose-dependent perspective must be an appropriate model of a kind of system instantiated in the actual situation.

I argue that these conditions can only be met if the world contains open systems of a certain type, and if humans are cognitively constituted such that we are able to represent these systems. However, the same part of the world is generally a component of multiple systems (for example, there are different systems at different levels of description) and different systems are relevant to different causal inquiries. The truth or falsity of a causal judgement thus depends on which system has been picked out—that is, on the perspective from which the judgement is made—which is always a partially mind-dependent matter. For this reason, causation is perspectival—the truth values of causal judgements depend on features of us, as well as on features of the mind-independent world. However, because the systems involved in token causal judgements are constitutionally mind-independent, the perspectival realism suggested by the Menzies–McGrath model is a form of causal realism, rather than anti-realism.

Chapter 1: Causes as Deviations from Actual Standards

According to Menzies, causes are deviations from the normal course of evolution of a system, a claim that he develops into an account of token causal judgements.[3] The aim of this chapter is to suggest one modification to Menzies' account, and to defend the resulting theory as an account of a kind of token causal judgement that is common in everyday causal discourse. The focus of this chapter is, therefore, not the metaphysics of causation, but the intuitive, largely implicit theory of causation that we apply when making causal judgements in everyday life.

The chapter is structured as follows: in §§1.1 and 1.2, I discuss the motivation for Menzies' account of causation, and its position within the philosophy of causation more generally. §1.3 consists in a detailed outline of Menzies' account of causation. Then, in §1.4, I introduce a notion of *normal* that McGrath argues is central to our judgements of causation by omission. According to this notion of *normal*, an event is normal if it adheres to some actual standard, which may be descriptive or normative. Finally, in §1.5, I combine Menzies' account of causation with McGrath's notion of *normal*, to give a modified account of causation that I call the 'Menzies–McGrath model'.

1.1 Normal and deviant causation

In developing his account of causation, Menzies takes the following quote from H. L. A. Hart and Tony Honoré as his starting point:

> The notion, that a cause is essentially something which interferes with or intervenes in the course of events which would normally take place, is central to the common-sense concept of cause ... Analogies with the interference by human beings with the natural course of events in part control, even in cases

[3] Menzies, "Difference-Making in Context"; Menzies, "Causation in Context"; Menzies, "Platitudes and Counterexamples".

> where there is literally no human intervention, what is to be identified as the cause of some occurrence; the cause, though not a literal intervention, is a *difference* to the normal course which accounts for the difference in the outcome.[4]

Menzies sees Hart and Honoré's model of the everyday concept of causation as having three components. First, applying the concept of causation to any particular event requires that we represent that event as part of a *system*; second, we assume that this system, if it is not subject to any outside intervention, will follow a *normal*, or *natural* course; and third, we identify a cause as something that *makes a difference*, in that it corresponds to some kind of *intervention* in the normal course.[5]

This final condition, which stipulates that a cause *makes a difference*, places Menzies' account of causation in the counterfactual tradition. According to counterfactual accounts of causation, causes and effects are related by *counterfactual dependence*. The most basic counterfactual account states that an event, *c*, is a cause of a second event, *e*, if and only if it is true that if *c* had not occurred, *e* would not have occurred.[6] Combining this basic counterfactual account with the central thesis of Menzies' theory—that causes are interventions that represent a deviation from the normal course of evolution of a system—leads to a very rough statement of Menzies' account: a cause, *c*, is an abnormal intervention in a system, such that if *c* had not occurred, and the system had been left to follow its normal course, the effect, *e*, would not have occurred, either.

---

[4] H. L. A. Hart and Tony Honoré, *Causation in the Law* (Oxford: Clarendon Press, 1959): 27. (Italics in the original.)

[5] Menzies, "Causation in Context": 201-02.

[6] The counterfactual tradition is seen as descending from Lewis's seminal paper: David Lewis, "Causation" *Philosophical Papers*, vol. 2 (Oxford: Oxford University Press, 1986). For an overview of the current literature on counterfactual accounts of causation, see the papers in John Collins et al., eds., *Causation and Counterfactuals* (Cambridge, MA: MIT Press, 2004). Many of the papers in this book are attempts to address a number of significant, and well known, objections to counterfactual accounts of causation, most notably *pre-emption*, however, the problem that preemption poses for counterfactual accounts will be put aside in this thesis.

Intuitively, this rough definition does capture at least part of everyday causal reasoning. Consider this commonplace example:

> I get into my car, turn the key, and the engine coughs, splutters, and fails to start. Frustrated, I immediately start to try to work out what is wrong with my car.

In this situation, the failure of my car to start is a deviation from the normal course of events, in which turning the key starts the engine. In order to determine the cause of this problem, it is necessary to understand the car (or the car's ignition system) as a mechanical system, consisting of a large number of components, which all have a particular role to play. We assume that the cause of the car's failure to start is an intervention in one of the components of the system, which changes the state of this component, such that it no longer operates within normal parameters, does not play its role within the system, and therefore makes a difference to whether or not the car starts.

The kind of causal reasoning described in the above example is common in everyday life, and Menzies and Hart and Honoré are right that it is a central component of the concept of causation. Notice, however, that as soon as we conceptualise part of the world as a system with a certain normal course of evolution, it is possible to enquire about instances of causation *within* the normally functioning system, as well as causes that are deviations from this system. For example, I can ask what caused my car to start yesterday, when it *was* working properly, which is just to ask about the causal structure of the system consisting in the normal running (or at least starting) of my car. That is, not *all* particular instances of causation are deviations from the normal. This suggests that our everyday causal reasoning needs to be divided into (at least) two different kinds—the first in which causes are *part* of the normal course of evolution of a system, and the second in which causes are *deviations* from the normal course of evolution.

1.1.1 Type versus actual causal judgements

It is necessary to say something about how these two different types of causal reasoning fit into the literature on causation more generally. In the philosophy of causation, a distinction is generally made between *type* and *token* (or *actual*) causal claims. Type causal claims are held to describe generalisations that hold between kinds of events (e.g. 'Hard blows to the head cause concussion'), whereas token causal claims describe particular situations, and state that one event is causally responsible for another event (e.g. 'A brick falling on John's head caused his concussion'). However, as Woodward has recently pointed out, aligning the distinction between type and actual causal judgements with the type/token distinction is misleading.[7]

The reason this alignment is misleading, according to Woodward, is that there can be a *type* causal relation between two kinds of events, *C* and *E*; *tokens* of both of these kinds of events, *c* and *e*, can be present in a particular situation; and yet it can still be true that *c* did not cause *e*. For example, shots to the head typically cause death. However, it is possible that Samantha is shot in the head and dies an hour later, but that the shot to the head was *not* the cause of her death—perhaps she would have survived the shot, but was given a lethal injection of tetrodatoxin just afterward. In that case, the cause of Samantha's death was the poisoning, not the shooting. Woodward claims that the fact that type and token causal judgements can come apart in this way suggests that the two kinds of causal judgements play different roles in human reasoning.

According to Woodward, *type causal judgements* are typically forward-looking—when making type causal judgements, we typically start by considering a particular event, and then reason *forward* in time to its effects. Thus, type causal judgements are used to

[7] James Woodward, "Psychological Studies of Causal and Counterfactual Reasoning" *Understanding Counterfactuals, Understanding Causation*, eds. C. Hoerl et al. (Oxford: Oxford University Press, 2011): 38-39.

make predictions, decide on manipulations, and explain repeatable events.[8] For example, we know that there are type causal relationships that hold between certain factors and heart disease. A doctor may use this information to predict that a particular middle aged patient has a high risk of heart disease because he is overweight, does not exercise much, and has a family history of heart problems.

Judgements of the actual cause, on the other hand, are (according to Woodward) typically backward-looking, and closely connected to the notion of responsibility. That is, when making actual causal judgements, we often focus on an event, and work *backwards* in time to determine that event's cause. The causes picked out are the events that are judged to be (causally) responsible for the effect. Woodward also points out that attributions of actual causation typically refer to causes that are a deviation from some norm. For example, when examining a patient who is suffering from dizziness, a doctor would try to work out what is responsible for the dizziness in this particular case, and might conclude that the cause was an ear infection. Here, both the dizziness and the ear infection (the cause and effect) are deviations from the norms of proper functioning of a healthy person.

Woodward argues that type and actual causal judgements are studied by different research programs in the *psychology* of causal reasoning.[9] The first research paradigm, which investigates type causal reasoning, aims to work out how we learn about the causal structure of the world. In a typical experiment, subjects (often children) are exposed to unfamiliar causal relationships, and given the task of figuring out how they work (which they demonstrate either by verbal report, or by actively intervening to

[8] Woodward, "Psychological Studies of Causal and Counterfactual Reasoning": 39.
[9] Woodward, "Psychological Studies of Causal and Counterfactual Reasoning": 40. (Apologies to the reader for the repeated citations here and throughout this thesis. Unfortunately, the latest version of EndNote does not permit the use of 'Ibid.')

manipulate the system in question).[10] The second research paradigm, which is focused on judgements of the actual cause, aims to determine which events we pick out as 'the cause' in complex scenarios. In a typical experiment, subjects are presented with a vignette in which it is clear that a number of causal factors contribute to a particular outcome, and are asked to identify the cause.[11] In these experiments, subjects are more likely to pick out abnormal events as causes.

It is likely that Woodward is right that type and actual causal judgements involve different forms of reasoning, and that these different forms of reasoning are associated with different research paradigms in psychology. However, the difference between type and actual causal judgements comes into clearer focus if these categories are formulated against a conceptualisation of the world in terms of systems that are susceptible to intervention. Viewed in relation to this conceptualisation, the form of reasoning that underlies *type causal judgements* can be seen to consist in learning about, and representing, the causal structure of the kinds of systems that we typically encounter (including the types of factors that can intervene in the normal course of evolution of these systems). *Judgements of the actual cause*, on the other hand, involve noticing, and reasoning about, interventions in these normal systems in particular cases.

Once we distinguish between type causal judgements and actual (or token) causal judgements based on whether or not a cause is part of the normal course of evolution of a system, it becomes clear (as noted above) that there are some token causes that are *not* deviations from the normal course of evolution of a system. To return to the earlier example, we can talk about the cause of my car's starting yesterday, when it functioned

[10] For a good introduction to this research, see Alison Gopnik and Laura Schulz, *Causal Learning: Psychology, Philosophy and Computation* (Oxford: Oxford University Press, 2007).
[11] For example, see David R. Mandel, "Mental Stimulation and the Nexus of Causal and Counterfactual Explanation" *Understanding Counterfactuals, Understanding Causation*, eds. C. Hoerl et al. (Oxford: Oxford University Press, 2011); Daniel Kahneman and Dale T. Miller, "Norm Theory: Comparing Reality to Its Alternatives" *Psychological Review* 93 (1986).

normally. This shows that there are some causal judgements that do not fit into either of Woodward's categories.

Taking into account both Woodward's distinction between type and actual causal judgements, and the claim that (many) token causes are deviations from the normal course of evolution of a system, leads to the suggestion that token causal judgements fall into two types. The first category of token causal judgements refers to causes that are deviations from the normal course of evolution of a system. I will call judgements of this type 'deviant token causal judgements'. The second category refers to causes that are part of the normal course of evolution, and will be called 'normal token causal judgements'. Because the distinction between deviant and normal causal judgements is orthogonal to the distinction between type and token causes, there is (at least) a fourfold taxonomy of causal judgements: normal type, deviant type, normal token and deviant token. This taxonomy is summarised in Table 1.

Normal type e.g. 'The tides are caused by the moon's gravitational field.'	Deviant type e.g. 'Flat batteries are a cause of cars failing to start.'
Normal token e.g. 'The high tide this morning was caused by the moon's gravitational field.'	Deviant token e.g. 'The battery's being flat caused the failure of my car to start this morning'

Table 1: A taxonomy of causal claims

The category 'deviant token causal judgements' corresponds to the category Woodward calls 'actual cause judgements', which is also referred to in the literature as 'token causation' (or just 'causation'). Deviant token causal judgements, like judgements of the actual cause, are often backwards-looking, and are also attributions of responsibility, in the sense that to make a deviant token causal judgement is to claim that one event is

causally responsible for another event. In other words, 'deviant token causation' roughly corresponds to the concept that many philosophers are attempting to analyse when they provide accounts of actual (or token) causation. For this reason, although other people working in the philosophy of causation do not often distinguish between deviant and normal token causes,[12] the account of deviant token causal judgements defended in this chapter can be directly compared to the existing accounts of token causation in the literature (with the proviso that there are some token causes (i.e. normal token causes) that the account defended here is not intended to capture).[13]

The aim of the remainder of this chapter is to give an account of deviant token causal judgements. Before providing the details of this account, I will show that dividing token causal judgements into two categories, deviant and normal, can help us respond to some of the problems that beset traditional theories of token (or actual) causation. In particular, I will argue that an account based on the thesis that token causal judgements are relative to kinds of systems with a normal course of evolution is more consistent with a number of features of causal discourse than accounts of actual (or token) causation that are based on the natural network model of causation.

1.2 The natural network model of causation

As discussed in the introduction to this thesis, the dominant theories of causation in the philosophical literature are based on a common 'metaphysical picture' of causation, which represents the causal history of the universe as a giant, mind-independent 'relational structure' of events. This is sometimes referred to as the 'network model of

[12] There are several other philosophers who build the distinction between deviant and normal (or default) states of affairs into their accounts of actual causation, in various ways. For example, see: Hart and Honoré, *Causation in the Law*; J. L. Mackie, *The Cement of the Universe* (Oxford: Oxford University Press, 1980); Christopher Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason" *Philosophical Review* 116 (2007); Ned Hall, "Structural Equations and Causation" *Philosophical Studies* 132 (2007).

[13] The accounts of actual (or token) causation in the literature include counterfactual theories, as well as other kinds of accounts, most notably probability and process theories. For an example of each, see: Lewis, "Causation"; Ellery Eells, *Probabilistic Causality* (Cambridge: Cambridge University Press, 1991); Wesley C. Salmon, "Causation without Counterfactuals" *Philosophy of Science* 61 (1994).

causation'.[14] The network model is an intuitively compelling account of causation, probably because it is consistent with a number of platitudes about causation: namely that causation is a relation between events, that this relation is two-place, and that it is mind-independent, or natural.[15] To reflect the fact that the network is typically assumed to consist of *natural* relations, in this thesis, the picture of causation just described will be referred to as the 'natural network model'.

The claim that the network model lies at the heart of our concept of causation is not consistent with the idea that causal judgements are based on a conceptualisation of parts of the world as systems, which are subject to external intervention. This is because, as I will go on to argue (in §2.3.1, and then in more detail in Chapter 4), any single spatiotemporal region contains a number of different causal systems, so the system referred to in a particular deviant token causal judgement depends on the perspective of the person making the judgement which, in turn, depends on the purpose of the causal inquiry. The truth values of causal statements are, therefore, partially dependent on which systematization the speaker has in mind, rather than the entire causal history of the universe (a claim that will be defended in more detail in Chapter 5). In this section, I will argue that adopting the systems model, rather than the network model, as the basis of our causal reasoning, can help account for three phenomena that raise problems for the network model.

1.2.1 Example 1: the fire in the lab

The first feature of causal discourse that generates a problem for the network model is *context sensitivity.* As an illustration of this feature, consider the following scenario:

[14] The term 'network model' is due to Steward: Helen Steward, *The Ontology of Mind* (Oxford: Clarendon Press, 1997): Chapter 7. However, the basic idea can be traced back to Lewis: Lewis, "Causal Explanation". This model of causation lies behind theories of causation as diverse as counterfactual accounts (where it motivates the use of 'neuron diagrams') and process theories.
[15] The sense of 'natural' intended here is defined in the Introduction.

1. A fire breaks out in a lab.

In most cases, it would be strange to mention the presence of oxygen as a cause of the fire. However, if we make the example more elaborate, and imagine that the fire breaks out in a special fume cupboard from which oxygen is supposed to be excluded, it now *would* be appropriate to cite the presence of oxygen as a cause.[16] This example highlights a distinction made in everyday causal discourse between *causes* and *conditions*, where conditions are events (or states) that are necessary for an effect, but are not typically cited as causes. More importantly for current purposes, this example also shows that the distinction between causes and conditions is *context sensitive*—the presence of oxygen changes from being a condition to a cause when we alter the context in which the fire takes place.[17]

Neither the fact that we distinguish between causes and conditions, nor that this distinction is context sensitive, is disputed. What is disputed, however, is how we should account for this context sensitivity. Theorists who (either explicitly or implicitly) endorse the natural network model tend to take one of two approaches at this point: the pragmatic response, or the ambiguity response. The pragmatic response claims that the distinction between causes and conditions (and any other context sensitivity in causal discourse)[18] can be completely accounted for by conversational pragmatics. The idea is that it is true to say that the presence of oxygen caused the fire in both situations described above, it is just that in the first situation (as in most cases) the presence of oxygen is assumed, and is simply not conversationally *relevant*. The ambiguity response, on the other hand, (unsurprisingly) accounts for the context sensitivity of

[16] This example is due to Hart & Honoré. Hart and Honoré, *Causation in the Law*: 33.
[17] Perhaps, on reflection, you think that oxygen *is* a cause of the fire in both situations, at least in some sense. This intuition can be explained by saying that the presence of oxygen is a normal token cause (as opposed to a deviant token cause) of the fire in most situations.
[18] For examples of other types of context sensitivity observed in causal discourse, see Jonathan Schaffer, "Causal Contextualism" *Contrastivism in Philosophy*, ed. M. Blaauw (Hoboken: Taylor and Francis, 2013); Menzies, "Causation in Context".

causal discourse by positing an ambiguity in the term 'cause'. According to this response, our causal discourse is not context sensitive at all—what *is* context sensitive, however, is which causal concept is being applied.

In the next chapter, I will argue that both the pragmatic and ambiguity responses are inadequate. Briefly, the pragmatic response is unsuccessful because it fails to account for the *purpose* of the causal concept, which is to enable us to manipulate the world. The ambiguity response, on the other hand, cannot account for *all* of the context sensitivity that is evident in causal discourse (and in particular, for the distinction between causes and conditions).

Putting the pragmatic and ambiguity responses aside, the idea that causal judgements are based on a conceptualisation of the world in terms of systems (as opposed to a single network) offers a promising means of accounting for context sensitivity, because the relevant systems are themselves context sensitive. The same causal factor may therefore be deviant with respect to one system, but not with respect to another. For example, the presence of oxygen represents a deviation from the normal course of evolution of an oxygen-free fume cupboard, but not from the normal course of evolution of most situations in which a fire might break out. Thus, assuming that in both situations we are applying deviant token causal reasoning to determine whether oxygen is a cause of the fire, the systems model can account for the context sensitivity of the above example.

It is worth mentioning that a number of other alternatives to the natural network model, designed to account for the context sensitivity of causal discourse, have recently been developed.[19] The most prominent of these alternatives is contrastivism, according to which the grammar of causal sentences is (three- or) four-place, rather than two-place,

[19] For a summary of positions according to which causal judgements are semantically sensitive to context, see Schaffer, "Causal Contextualism".

so that fully articulated causal claims have the form: *c* rather than *c** caused *e* rather than *e**, where *c** and *e** are contrast events (i.e. non-actual alternatives to the cause and effect). Contrastivism (and its relationship to the account of causation defended in this chapter), will be discussed in more detail in Chapter 2.

1.2.2 Example 2: the plant-watering

The following example highlights two further problems with the network model of causation:

> 2. I go away for a few weeks, and my housemate, Tim, promises to water my geranium while I'm gone. He forgets, and, as a result, the geranium dies.

The first problem here is that the posited cause of the geranium's death (Tim's failure to water it) is an omission, rather than a distinct positive event. More generally, in everyday causal discourse, we often cite omissions as causes. This is problematic, because according to the network model, causation is a relation between *events*, but omissions are not events, and therefore seem unable to enter into causal relations.[20]

The second problem raised by Example 2 is that it seems that the reason we consider Tim's failure to water the plant as a cause of its death, rather than, say, Barack Obama's failure to do the same, is that Tim *promised* to water it, whereas Obama did not. The idea that is the promise that makes the difference between Tim's and Obama's omissions suggests that our judgements about causation can be influenced by our acceptance of moral norms, which contradicts the assumption that causation is a *natural*

[20] For an overview of the problem posed by judgements of causation by omission, including potential solutions, see Jonathan Schaffer, "Contrastive Causation" *Philosophical Review* 114 (2005): 299-300 and David Lewis, "Postscripts to 'Causation'" *Philosophical Papers* vol. 2 (Oxford: Oxford University Press, 1986): 189-93. It is worth pointing out that it is possible to hold that causation is not a relation between events, but facts, in which case causation by omission may not be a problem. However, to pursue this strategy is to accept the existence of negative facts, which comes at a cost in terms of ontological parsimony.

relation—that is, a completely objective relation, independent of human concerns and values.

Although everyday judgements concerning both causation by omission and normative causation raise problems for the network model of causation, neither of these phenomena is inconsistent with the claim that causal reasoning is based on a conceptualisation of parts of the world as systems with a normal course of evolution. Clearly, we can conceptualise normative systems—for example, systems describing the laws or conventions of a particular community—and deviations from the normal course of a system can just as easily be omissions as positive causes. Thus, the idea that token causal judgements are based on a conceptualisation of the world in terms of systems, each with a normal course of evolution, can provide us with a semantic solution to the problems raised by both judgements of causation by omission and judgements in which normative factors appear to play a causal role, as well as context sensitivity. As these three phenomena are significant features of our causal discourse, any account of the causal concept should be able to explain them. The fact that Menzies' approach to causation (understood as an account of the type of causal judgement I have called 'deviant token causal claims') promises to do so is therefore a significant advantage of his account over accounts based on the natural network model.

Because the aim of this chapter is to provide an account of deviant token causal claims, as these claims are used in everyday life, rather than an account of the metaphysics of causation, the account of the causal concept defended here is perfectly consistent with the natural network model, if the latter is construed as an exclusively *metaphysical* (rather than semantic) theory. (However, in the remainder of this thesis I will argue that there are good reasons for giving up the natural network model, even on its metaphysical interpretation.)

1.3 Menzies' account: causes as deviations from the normal

Having outlined the motivations for Menzies' account of the concept of causation, I will now discuss his account in more detail. As noted above, this will involve interpreting Menzies' theory as an account of deviant token causal claims. Recall that his account has three central components: first, events are represented as components of *systems*; second, these systems have a *normal* course of evolution; and third, causes *make a difference* to this normal course. Incorporating these ideas into a counterfactual framework, and considering causes and effects as values of variables X and Y respectively, Menzies arrives at the following definition:

> *Menzies' Account*: $X = x$ *causes* $Y = y$ relative to the default values $X = x'$ and $Y = y'$ if and only if (i) the actual values of X and Y are x and y respectively; and (ii) if an intervention were to change the value of the X variable from x' to x, the value of the Y variable would change from y' to y.[21]

The idea is that the default values (x' and y') are the values X and Y *normally* have. Menzies claims that the default values of variables correspond to a set of *default worlds*. These default worlds are those possible worlds in which the relevant kind of system follows its normal course of evolution.[22]

Menzies uses the idea of the 'default world' to generate a semantics of causal conditionals along the lines of David Lewis's possible world semantics. There is not

---

[21] Adapted from Menzies, "Platitudes and Counterexamples": 360. Menzies' most complete formulation of his account is presented using the structural equations framework, which is a system of causal modelling that is becoming increasingly common in the philosophy of causation. I think that the basics of his account can be presented more clearly without using the structural equations framework. However, in the interests of completeness, I will include Menzies' formulation of the above truth conditions in the language of the structural equations framework:

> A value of a variable *X makes a difference* to the value of another variable *Y* in a default causal model if and only if plugging in the default values of the variables in the structural equations yields $X = x$ and $Y = y$ and there exist actual values $x' \neq x$ and $y' \neq y$ such that the result of replacing the equation for X with $X = x'$ yields $Y = y'$. Menzies, "Causation in Context": 208. (Italics in the original.)

[22] Menzies, "Causation in Context": 216.

space to go into the details of these semantics here.[23] The important point, however, is that we evaluate whether a certain event makes a difference to an effect (and therefore whether it is a cause) *relative to these default worlds*. That is, the closest possible worlds are those that contain the fewest interventions into the default worlds. The idea is that causal claims can be understood as a kind of conditional: *if* the given situation instantiates a kind of system, with a certain normal course of evolution (i.e. if the default worlds accurately represent the normal course of evolution of the relevant situation) *then* $X = x$ causes $Y = y$ if and only if conditions (i) and (ii) (specified above) hold.[24]

Before illustrating Menzies' account with an example, there are some terms that need clarifying. The first of these is 'variable'. Intuitively, the values of variables represent incompatible states of an object or system. That is, they represent events, or states of affairs. Importantly, variables can take a number of forms—they can be discrete (e.g. 'number of participants') or continuous (e.g. 'length'), and can also be binary (e.g. 'has lungs' vs. 'does not have lungs') or take a range of values (e.g. 'colour').

The term 'intervention' also needs clarifying. Much work has gone into giving technical definitions of this term in causal systems, and there is not space to go into the technical details here. Very roughly, then, an intervention in a variable can be understood as a manipulation of that variable from outside the system. That is, an intervention in

[23] For the details of Lewis's possible world semantics, see: David Lewis, *Counterfactuals* (Malden, MA: Basil Blackwell, 1973). Briefly, incorporating the default worlds into the semantics of causal conditionals results in two differences from Lewis's possible world semantics: first, the similarity relation is generated by the default causal model (the closest possible worlds are those that contain the fewest interventions into the sphere of default worlds, as noted above); second, the system of possible worlds is not centred on the actual world, but on the set of default worlds. Menzies, "Causation in Context": 215-18.

[24] Adapted from Menzies, "Difference-Making in Context": 159.

variable X directly fixes the value of X to x, and does not affect the values of any other variables in the system except through its influence on X.[25]

Menzies' account of causation can be illustrated using Example 1, the fire breaking out in the lab. Consider two potential causes of this fire—the presence of potassium (a metal that ignites spontaneously in oxygen), and the presence of oxygen. Oxygen and potassium are both present when the fire breaks out. However, in the normal operating conditions of the lab, oxygen is present, but potassium is not. And, of course, there is not normally a fire. The normal operating conditions of the lab supply the default values of the variables (x' and y') and the details of what actually happens supply the actual values of the variables (x and y). Relative to the default values of the variables, potassium makes a difference, whereas oxygen does not. Thus, Menzies' account can explain why the presence of oxygen is not considered to be a cause of the fire in this situation.

Recall that the point of Example 1 is that whether or not oxygen is considered a cause of the fire is a context sensitive matter. When I introduced this example, I also considered a slightly different scenario, in which the fire starts in a fume cupboard from which oxygen is supposed to be excluded. In this situation, all the actual and default values of X and Y are the same, except for the default value of the variable representing the presence of oxygen (x'), which now reflects the fact that oxygen is not normally present. Re-evaluating the relevant counterfactuals according to Menzies' account, we find that in this modified situation, oxygen and potassium are both causes of the fire. That is, oxygen has changed from being a condition to being a cause. Thus, one of the

[25] Obviously, the above clarification is a long way short of being complete, as I have not defined the notion of a *system*, what it means to directly fix the value of a variable, or what it is for one variable to influence another. For technical explications of these ideas, see James Woodward, *Making Things Happen: A Theory of Causal Explanation* (Oxford: Oxford University Press, 2003); Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge: Cambridge University Press, 2000); Menzies, "Causation in Context".

virtues of Menzies' theory is that it accounts for the context sensitivity of the distinction between causes and conditions, one of the phenomena of causal discourse that raise problems for the network model (introduced in §1.2.1).

The idea that causes are interventions suggests an objection to Menzies' account of causation (which applies equally to the modified account defended below): namely, that, since the term 'intervention' is itself a causal notion, the account is hopelessly circular.[26] It is true that any account of causation that makes use of the idea that causes are interventions will be non-reductive (where a reductive analysis of causation provides a definition of 'cause' in entirely *non-causal* terms). However, Menzies' truth conditions of deviant token causal claims are not viciously circular, because evaluating any particular causal claim does not require prior knowledge of the causal *relation* in question, but of the causal structure of the relevant *system*.

Whether the fact that an account of causation is non-reductive is a problem depends on what you are aiming to achieve in giving an account of causation. If his aim was to show that causal relations reduce to (or supervene on) fundamental facts about the world, the circularity of Menzies' account would, indeed, be a serious cause for concern. However, as he is just trying to account for how the causal judgements we make in everyday life enable us to make predictions and give causal explanations, this circularity is not problematic. As Woodward has compellingly argued, we can make interesting and non-trivial claims about our concept of causation without trying to analyse it reductively.[27]

[26] Hall claims that this objection applies to all accounts of causation based on the structural equations framework, pointing out that: 'causal models merely provide a useful means for selectively representing aspects of an *antecedently understood* counterfactual structure.' Hall, "Structural Equations and Causation": 110. (Italics in the original.)

[27] Woodward, *Making Things Happen: A Theory of Causal Explanation*: Chapter 2.

It is worth pointing out that a number of other philosophers have built the distinction between default and deviant values of variables into their accounts of actual (or token) causation. For example, in 'Prevention, Preemption and the Principle of Sufficient Reason', Christopher Hitchcock outlines a rule that he calls the 'Principle of Sufficient Reason', which states that 'when a set of variables all take their default value, they cannot by themselves cause another variable to take a deviant value.'[28] Hitchcock claims that when a causal model obeys the Principle of Sufficient Reason,[29] counterfactual dependence is a necessary and sufficient condition of actual causation. Here, Hitchcock is giving a technical explication of the idea that we expect systems to follow a normal course of evolution, unless acted on by an external intervention. Hitchcock's defence of the Principle of Sufficient Reason therefore supports the claim that deviant token causal judgements are a paradigmatic example of actual causal judgements.[30]

In summary, I have shown that Menzies' account of causation has a number of advantages over accounts based on the natural network model. Unfortunately, however, Menzies' account has a flaw: it relies on an analysis of *normal*, which he does not provide. I will now elaborate on, and then attempt to resolve, this problem.

1.4 McGrath's analysis of **normal**

Menzies points out that the notion of *normal* embedded in the idea of a 'normal course of evolution' is both diverse and context dependent. For this reason, when defining the relevant notion of *normal* (and thus the content of the default worlds), he confines

[28] Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason": 508.
[29] More precisely, Hitchcock claims that it is *causal networks*, not causal models, that can obey (or disobey) the Principle of Sufficient Reason, where a causal network is a directed path from one variable to another in a causal model. Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason": 509-10.
[30] Hall also defends a counterfactual account of actual causation according to which the counterfactual dependence between causes and their effects is relative to a kind of default world, or normalised situation. See Hall, "Structural Equations and Causation".

himself to listing a few laws and norms that can influence the normal course of evolution of a system. These include physical laws, statistical regularities, routines developed by humans to avoid harm, norms of proper functioning,[31] and legal, social and moral norms.[32] He also suggests that these factors may relate to the way the human mind conceptualises systems. In other words, the relevant notion of *normal* may be partially determined by us, rather than by the world alone.[33]

To strengthen Menzies' account of causation, then, we need an analysis of *normal* that somehow unifies these seemingly disparate factors.[34] In 'Causation by Omission: A Dilemma',[35] McGrath defends an analysis of a notion of *normal* that looks very like the notion Menzies needs, which she employs in giving a counterfactual account of causation by omission.

To examine the details of McGrath's notion of *normal*, let us return to Example 2, the plant-watering. As noted in §1.2.2, it is intuitive that the reason we distinguish between Tim's failure to water my plant, which we judge to be a cause, and Obama's failure to water the plant, which we do not judge to be a cause, is that Tim *promised* to do so, whereas Obama did not.

Of course, it is possible that this intuition is wrong, and that we actually use some factor other than the promise (and therefore the difference in Tim's and Obama's moral statuses) to distinguish between Tim's and Obama's omissions. Perhaps we base the judgement that Tim caused the plant's death, whereas Obama did not, on some non-

[31] Menzies, "Causation in Context": 220-21.
[32] Menzies, "Platitudes and Counterexamples": 358-59.
[33] I take it that Menzies is implying that what is normal depends partly on how we represent the world in the following quote: 'There may, indeed, be laws about the way in which the human mind forms conceptions of the inertial behaviour of systems and the default worlds they exemplify.' Menzies, "Causation in Context": 221.
[34] Other philosophers who build the default/deviant distinction into their accounts of causation have also only given a very rough specification of what counts as a default. See Hall, "Structural Equations and Causation"; Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason".
[35] Sarah McGrath, "Causation by Omission: A Dilemma" *Philosophical Studies* 123 (2005).

normative distinction between the two omissions. The advantage of finding a non-normative basis for judgements about which omissions are causes would be that it might then be possible to account for judgements of causation by omission while maintaining that causation is a natural relation, unaffected by human norms and values. One such suggestion is that in the closest possible world in which my geranium gets watered (and survives) it is Tim who waters it, whereas the possible world in which Obama waters it is really quite remote.[36]

Unfortunately, however, the above suggestion fails, because even if we are provided with the information that Tim is extremely unreliable, and never keeps his promises, we still judge that he caused the plant's death. Other naturalistic explanations, for example that Tim's omission is a cause because he is physically proximate to the plant, whereas Obama is not, also fail. (To see this, note that if Obama *had* promised to water the plant, we *would* cite his omission as a cause, despite his lack of physical proximity.) Thus, it really does seem to be the difference in moral statuses that underpins the judgement that Tim's omission, and not Obama's, caused the plant's death. This intuition is supported by evidence from recent work in psychology, which also concludes that token causal judgements are influenced by normative considerations.[37]

What we want, then, if we are going to take our common-sense judgements about causation by omission seriously, is some analysis of causation by omission according to which a promise, and, by extension, the normative more generally, can be causally relevant. This is what McGrath sets out to provide.

[36] Beebee and McGrath both consider this possibility, but rule it out for the reason given below. Beebee, "Causing and Nothingness": 298-300; McGrath, "Causation by Omission: A Dilemma": 134-36.
[37] See Joshua Knobe and Ben Fraser, "Causal Judgement and Moral Judgement: Two Experiments" *Moral Psychology: The Cognitive Science of Morality*, ed. W. Sinnott-Armstrong, vol. 2 (Cambridge, MA: MIT Press, 2008); Joshua Knobe, "Person as Scientist, Person as Moralist" *Behavioral and Brain Sciences* 33 (2010).

The idea that Tim's promise is the crucial factor that distinguishes his failure to water my geranium from Obama's failure to water the plant suggests that the omissions that we count as causes are all deviations from some norm. Keeping a promise is obviously a moral norm, but perhaps other omissions that are judged to be causes represent deviations from other kinds of norms. Here are some examples:

> Tom's alarm clock failing to ring was the cause of his not waking up at 7:00 this morning.
>
> Zoe failed to stop at the traffic lights, and in doing so, caused a crash.
>
> Greg's dog died because its kidney stopped functioning.
>
> Owing to lack of rain, a gum tree in the Blue Mountains died.

The question is: is there something that links all these examples? McGrath argues that they are linked by the following notion of *normal*:

> The notion of the normal I have in mind is highly abstract and applies very generally: to actions, the behaviour of artifacts, and the behaviour of both biological and non-living systems ... It is normal for x to φ iff x is *supposed* to φ. People are supposed to keep their promises (it is normal for them to keep their promises); alarm clocks are supposed to ring at the set time (it is normal for them to ring at the set time).[38]

We are also supposed to follow the traffic laws; kidneys are supposed to filter our blood; and the climate is supposed to be fairly uniform from year to year.

According to McGrath, these different versions of 'supposed to' are linked because they are all 'imposed by certain *standards*',[39] which vary from case to case. She mentions 'artifactual' standards, which apply to alarm clocks, 'biological' standards, which apply

[38] McGrath, "Causation by Omission: A Dilemma": 138. (Italics in the original.)
[39] McGrath, "Causation by Omission: A Dilemma": 138. (Italics in the original.)

to the functioning of kidneys, and 'physical' standards (i.e. standards set by the laws of nature) in the case of the climate.[40] Presumably, we could add 'legal' standards, which apply to road rules.

Importantly, McGrath's notion of *normal* is different from 'statistically usual'. She points out that: 'What is normal can be *un*usual: for example, it is normal for tadpoles to grow up to be frogs, even though they usually get eaten first, and it is normal even for unreliable alarm clocks to ring on time.'[41] This is what makes her notion *normative*, rather than *descriptive*—it is based on what is *supposed* to happen, not necessarily what usually *does* happen. Also, the situation is often complicated because conflicting standards apply. For example, consider a bonsai fig tree. According to biological standards, it is abnormal for fully grown trees to be that small—that is not how they are supposed to be. But according to the standards of bonsai cultivation, the fig tree is normal—it is supposed to be that size. That is, which standards are relevant is a context sensitive matter.

Having defined the relevant notion of *normal*, McGrath puts it to work to distinguish between those omissions that are causes, and those that are not. She argues that there is causation by omission when there is counterfactual dependence between an effect, and the absence of an event that would have been normal, according to some actual standard.[42] So, for example, Tim's failure to water my geranium counts as a cause of the plant's death because Tim's watering the geranium would have been normal, according to the standard of promise keeping.

[40] McGrath, "Causation by Omission: A Dilemma": 138-39.
[41] McGrath, "Causation by Omission: A Dilemma": 139. (Italics in the original.)
[42] McGrath, "Causation by Omission: A Dilemma": 142.

1.4.1 Objections to McGrath's notion of **normal**

The most obvious objection to McGrath's notion of *normal* (and thus her account of causation by omission) is that, because she does not distinguish between prescriptive norms (e.g. keeping promises), and descriptive norms (e.g. consistent weather patterns), her analysis of *normal* is ambiguous. The first point to make in response to this objection is that we often do combine descriptive and prescriptive norms in causal sentences: for example we say things like 'The cake is light because the egg whites were beaten properly.' We also take prescriptive norms to constrain the possible outcomes of systems, as, for example: 'The bishop can move diagonally over any unoccupied square', and 'If you make a promise, you should keep it'.

The idea that causal judgements can be governed by both descriptive and prescriptive norms also has indirect experimental support. For example, Joshua Knobe and Ben Fraser tested people's intuitions about the following vignette:

> The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own.
>
> The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly e-mailed them reminders that only administrators are allowed to take the pens.
>
> On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message … but she has a problem. There are no pens left on her desk.[43]

[43] Knobe and Fraser, "Causal Judgement and Moral Judgement: Two Experiments": 443.

In this vignette, the lack of pens on the receptionist's desk at the end of the day counterfactually depends on both Professor Smith's taking a pen, and the administrative assistant's taking a pen. However, Knobe and Fraser found that subjects were significantly more likely to agree with the statement 'Professor Smith caused the problem' than 'The administrative assistant caused the problem'.[44]

Hitchcock and Knobe consider the results of this experiment in their recent paper 'Cause and Norm'.[45] They argue that the reason we are more willing to say that Professor Smith caused the problem is because Smith violated a rule. His action was therefore a deviation from a norm, and was thus abnormal, in a certain sense. The sense of 'normal' that Hitchcock and Knobe have in mind is very similar to McGrath's notion of *normal*, in that it does not distinguish between descriptive and prescriptive norms. Hitchcock and Knobe argue that although this notion can seem ambiguous, descriptive and prescriptive norms are connected in terms of the way we understand, and classify, the world.[46] For example, if we label someone as 'abnormal'

> [W]e do not typically have in mind either a purely statistical judgement or a purely prescriptive one. Instead, we seem to be making a single overall judgement that takes both statistical and prescriptive considerations into account.[47]

Hitchcock and Knobe's explanation for the connection between descriptive and prescriptive norms is that we use these different types of norms as heuristics for each

[44] Knobe and Fraser, "Causal Judgement and Moral Judgement: Two Experiments": 443.
[45] Christopher Hitchcock and Joshua Knobe, "Cause and Norm" *Journal of Philosophy* 106 (2009).
[46] Hitchcock and Knobe use the term 'statistical norm', instead of 'descriptive norm'.
[47] Hitchcock and Knobe, "Cause and Norm": 598.

other, and may even combine them in a single representation, for example in 'scripts' of stereotypical situations.[48]

Notice that Hitchcock and Knobe's 'representations' sound remarkably similar to Menzies' 'default worlds'—both are models of the normal course of evolution of a kind of system, and both include prescriptive norms, as well as descriptive norms. Hitchcock and Knobe's argument therefore lends support to the idea that our causal judgements are relative to something like Menzies' default worlds. That is, although it sounds strange to talk about physical events in normative language (e.g. 'the climate is *supposed* to stay uniform'), it is less strange to claim that we base our everyday causal judgements on a model of the world that is partly normative—that is, on a default world, in which all norms (both descriptive and prescriptive) are upheld.

A second objection to McGrath's notion of *normal* is that it is vague. This, of course, is true. However, it is worth pointing out that her analysis is vague in a number of different ways, some of which require further clarification, but others of which may actually be helpful.

One sense in which McGrath's notion of *normal* is vague is that it is not clear how we determine which standard applies in a given context. (In Menzies' terminology, it is not clear how we determine which standards are included in a particular set of default worlds.) This does need further clarification. We can go some way towards addressing this problem by specifying which actual standards will apply in at least some contexts. For example, in legal contexts, the relevant laws and legal precedents will determine the actual standards, and in physics, the relevant standards will be the laws of physics. In everyday life, the standards are harder to specify, because our purposes in speaking are less clearly defined. However, this problem is not restricted to McGrath's notion of

[48] Scripts are 'mental representations of familiar sequences of activity'. Douglas A. Bernstein et al., eds., *Psychology*, VIII ed. (Houghton Mifflin, 2008): SIG-25.

normal. Other normative notions, for example *correctness*, are also applied in a wide range of cases, relative to standards that vary from case to case.[49] The fact that (most of the time) we manage to unthinkingly determine which standards of correctness apply in a given situation suggests that it is reasonable to suppose that we apply a similar procedure for determining which standards of normality apply in a given situation, and use this procedure in our causal reasoning. This does not answer the question of *how* we determine which standards apply, but this seems (at least in part) to be an empirical matter, which could most profitably be investigated by psychologists, rather than philosophers.[50]

A second sense in which McGrath's analysis is vague is that it does not pick out a single event and say, 'this is what is normal in this context'; instead, it demarcates a range of normal events. This is because most norms themselves are not totally determinate, in that they pick out a boundary between what is normal and what is not, rather than specifying one particular event as normal.[51] This is the kind of vagueness that may actually be helpful, because it mirrors our actual reasoning about causation. To put this point in Menzies' terminology, the set of default worlds picked out by a particular context may admit of significant variation. For example, the set of default worlds representing the system 'a week in Kate's life', will include worlds in which Kate is doing one of a multitude of possible activities at any given time, none of which strays

[49] McGrath points out that 'chess moves, dance steps, quiz answers, beliefs, baseball pitches, ways of beating eggs and stitching hemlines can all be correct or incorrect. Further, something is correct or not relative to certain *standards*, with the standards varying from case to case. There are *chess* standards governing the movement of the rooks (moving a rook diagonally is moving it *incorrectly*), and there are *culinary* standards governing the beating of eggs (beating eggs with a spoon is beating them *incorrectly*).' McGrath, "Causation by Omission: A Dilemma": 139. (Italics in the original.)

[50] In fact, the role that the normal plays in causal reasoning has already been investigated by psychologists, although the results so far do not seem to be conclusive. For example, see Kahneman and Miller, "Norm Theory: Comparing Reality to Its Alternatives"; Mandel, "Mental Stimulation and the Nexus of Causal and Counterfactual Explanation".

[51] Most social norms, at least, are not totally determinate, although this may not be true of norms in the physical sciences.

out of normal bounds. It seems to me that *this* is the kind of causal system from which we pick out deviant causes in everyday life.

It is worth pointing out that incorporating McGrath's notion of *normal* into Menzies' account of causation involves significantly extending the application of this notion, as McGrath herself only uses it in an account of causation by omission. However, it seems likely that if a particular notion of *normal* is central to our judgements about causation by omission, it is also used in causal reasoning more generally.

Finally, although McGrath does not make this point herself, notice that in all the cited examples of causation by omission, the *effect* is also a deviation from a standard of normality. It is normal to get to work on time (or at least, there is a norm that says I should get to work on time); it is normal not to have a crash; it is normal for a kidney to function properly; it is normal for a tree that was alive yesterday to still be alive today.[52] This suggests a connection with Menzies' account of causation, in which both causes and effects are deviations from the normal. In fact, the notion of *normal* defined by McGrath in her account of causation by omission corresponds exactly to the notion of *normal* needed to complete Menzies' account of token causation. Thus, in the next and final section of this chapter, I will formulate a modified version of Menzies' account, by combining his account with McGrath's notion of *normal*.

1.5 The Menzies–McGrath model of deviant token causal claims

Incorporating McGrath's notion of *normal* into Menzies' account of causation requires a slight modification to McGrath's analysis, since she formulates it specifically in terms of causation by omission. Also, in order to conform to Menzies' terminology, the

[52] Note that the relevant standards of normality are dependent on a time frame—it is normal for everything to die *eventually*. And thus, for example, for a tree that was alive 500 years ago to be dead today.

analysis needs to be translated into the language of variables, rather than that of events. This is easily achieved, however, as follows:

> *Default value*: x' is a default value of a variable X if and only if $X = x'$ according to some relevant *actual standard*, *S*.

Again, the idea is that finding the cause of an abnormal effect requires conceptualising the relevant situation in terms of a set of default worlds, which represent a kind of system with a normal course of evolution—that is, a kind of system that evolves according to actual standards. The default worlds are therefore possible worlds in which the relevant descriptive and/or normative standards are unbroken. Importantly, whether the standards are descriptive laws, descriptive norms, or prescriptive norms, they must *actually* be instantiated in the situation in question. We can combine these ideas to arrive at a definition of default worlds, as follows:

Conceptualising part of the world as an instance of a kind of system, *K*, governed by actual standards, *S*, gives rise to a set of default worlds, *W*, such that:

i) each member of *W* contains a counterpart of *K*; and

ii) each member of *W* evolves according to the relevant actual standards, *S*, without external intervention.[53]

Finally, a cause is a factor that makes a difference, relative to this set of default worlds.

Because incorporating McGrath's notion of *normal* just involves changing the specifications of the default worlds, the truth conditions remain the same as those given in Menzies' original account. That is:

> $X = x$ causes $Y = y$ relative to the default values $X = x'$ and $Y = y'$ if and only if (i) the actual values of X and Y are x and y respectively; and (ii) if an

[53] This is a modification of the definition of 'default world' given in Menzies, "Causation in Context": 216.

intervention were to change the value of the *X* variable from *x′* to *x*, the value of the *Y* variable would change from *y′* to *y*.

Again, the idea is that causal claims can be understood as a kind of conditional: *if* the given situation instantiates a kind of system, *K*, governed by actual standards, *S*, *then X* = *x* causes *Y* = *y* if and only if (i) and (ii) hold.

The explicit inclusion of *normative* standards as part of the default worlds entails that the structure of the kinds of systems that comprise the default worlds will often be slightly more complicated than it would have been if the default worlds had contained only descriptive standards. This is because some norms are restrictive, rather than stipulative—these norms act to constrain the range of permissible (and therefore normal) actions, rather than to impose a specific action as required. This point is related to the sense in which the vagueness of McGrath's notion of *normal* can be helpful (discussed in the previous section). For example, compare the moral norms 'We should not commit murder' and 'We should keep promises'. The former of these norms merely rules out a certain class of actions, whilst the latter entails that certain actions are required (depending on the content of the promises, of course). I will not attempt to give an account of the way in which these different types of standard are combined into a single default world, except to point out one consequence, which is that there will sometimes be more than one default value of a given variable.

I will call the account of causation just outlined, which combines Menzies' systemisation of the idea that a cause is a deviation from the normal with McGrath's analysis of *normal*, the 'Menzies–McGrath model'. The characterisation of both 'default worlds', and 'kinds of systems' provided in the above account is still rough, and largely intuitive. These terms are more precisely defined in §§4.7 and 5.2,

respectively. Now, I will illustrate the Menzies–McGrath by returning to Example 2, the sadly neglected geranium.

Recall that in this example, Tim breaks his promise, does not water my geranium whilst I am on holiday, and the plant dies. In the default worlds, however, Tim keeps his promise, waters the plant, and it stays alive. Tim's not watering the geranium is therefore judged a cause of the plant's death, because his omission is a deviation from the set of default worlds, which makes a difference to whether or not the plant survives, relative to these default worlds.

Now consider the other omission initially considered—Barack Obama's failure to water my plant. There is no actual standard according to which it would be normal for Obama to water my plant, so he does not water my plant in the default worlds. Thus, Obama's omission does not represent a deviation from the default worlds, and is therefore not judged to be a cause of the geranium's death. Thus, again, the intuitive verdict is obtained. That is, the Menzies–McGrath model can deal with judgements of causation by omission, including the distinctions we make between those omissions we are willing to cite as causes, and those we are not.

It is worth briefly pointing out that the Menzies–McGrath model deals with Example 1 in very much the same way that Menzies' unmodified account does (described in §1.3). In this case, the standards used to determine the default values of the variables are descriptive norms specifying the amount of oxygen, potassium and fire present under the normal operating conditions of the lab.

In conclusion, in this chapter I have presented the Menzies–McGrath model as an account of deviant token causal claims. In doing so, I have shown that the Menzies–McGrath model can account for three features of our causal discourse that traditional theories of causation (especially those based on the natural network model) find

problematic. This provides support for the claim that token causal reasoning, and therefore the concept of causation, is based on a conceptualisation of the world in terms of systems with a normal course of evolution. In the next chapter, I will provide further support for this claim, by explicitly arguing against the main alternative—the natural network model of causation.

Chapter 2: Objections to the Natural Network Model of Causation

According to the Menzies–McGrath model, the truth values of deviant token causal claims are relative to context sensitive default worlds. That is, the account entails that these causal claims are semantically sensitive to context—the same causal sentence can have different truth values in different contexts. Accounts of causation that endorse this claim are collectively known as *contextualist* accounts. My defence of the Menzies–McGrath model in the previous chapter, and, in particular, of the claim that the Menzies–McGrath model can account for the context sensitivity of causal discourse, amounts to a positive argument for contextualism. However, I am yet to discuss the arguments put forward by *invariantists*, who claim that the truth of token causal claims is insensitive to context.[54] Invariantists are aware that causal discourse exhibits context sensitivity, and tend to account for this context sensitivity by adopting one of two strategies: the ambiguity response or the pragmatic response. Thus, in §§2.1 and 2.2, I will strengthen my defence of contextualism about causation by arguing that each of these responses is inadequate.

The natural network model of causation (introduced in Chapter 1) is a form of invariantism. According to the natural network model, the causal history of the universe consists of a single, mind-independent relational structure—a network of two-place causal relations. True token causal claims succeed in representing this structure, in the sense that the network determines the truth value of every token causal sentence. To accept that the causal concept is semantically sensitive to context is therefore to reject the natural network model: it is not possible to claim *both* that causal judgements directly represent a mind-independent causal structure (i.e. that the truth values of token

[54] The terms 'contextualism' and 'invariantism' have their origins in the literature on epistemic contextualism, and have been imported into the philosophy of causation by Schaffer. See Schaffer, "Causal Contextualism".

causal claims are determined by a mind-independent structure), *and* that causal discourse is semantically sensitive to context (i.e. that the same causal claim can have different truth values in different contexts).

The natural network model is intuitively compelling, and is the default position in the metaphysics of causation. To defend a contextualist account of causation, it is therefore necessary to show that the natural network model is mistaken, and how it is mistaken. For this reason, in §2.3, I provide an argument against the natural network model, and gesture towards an alternative metaphysical picture. Finally, in §2.4, I consider *contrastivism*, a form of contextualism that has recently been defended by a number of philosophers, including Jonathan Schaffer, Hitchcock, and Cei Maslen.[55] I show, first, that contrastivism can be thought of as a combination of semantic contextualism and metaphysical realism; a combination that is achieved by holding that causation is a three- or four-place relation, rather than a two-place relation. Second, contrastivism is a midway position between invariantism and the Menzies–McGrath model—a position that, in Chapter 3, I will argue is unstable.

2.1 Invariantism option 1: the ambiguity response

According to the ambiguity response, the context sensitivity of causal discourse can be accounted for by the fact that the word 'cause' is ambiguous. In this section, I will discuss two versions of the ambiguity response. The first, provided by Helen Beebee, is not actually designed to account for the context sensitivity of causal discourse, but for judgements about causation by omission. The second version, which is defended by Hitchcock, *is* explicitly intended to account for context sensitivity.

[55] Jonathan Schaffer, "Contrastive Causation" *Philosophical Review* 114 (2005); Christopher Hitchcock, "Farewell to Binary Causation" *Canadian Journal of Philosophy* 26 (1996); Cei Maslen, "Causes, Contrasts, and the Nontransitivity of Causation" *Causation and Counterfactuals*, eds. J. Collins et al. (Cambridge, MA: MIT Press, 2004).

2.1.1 Beebee's version of the ambiguity response

Beebee defends a version of the ambiguity response according to which our use of the word 'cause' in everyday discourse is ambiguous between two concepts: *causation* and *causal explanation*. That is, we typically fail to distinguish between the constructions '*c* causes *e*', and '*e* because *c*'. Consequently, not all singular causal sentences are reports of causal relations.[56]

Beebee holds that the concept of *causation* always refers to a relation between events—it picks out a segment of the giant relational structure that is the complete causal history of the universe. However, she argues that the related concept, *causal explanation*, enables us to appeal to more general facts about an event's causal history, and does not directly refer to individual segments in the mind-independent causal structure. In other words, Beebee accepts the natural network model of causation, but claims that the second concept, causal explanation, does not refer to particular causal relations.[57]

Beebee's defence of the ambiguity response is offered as a solution to the problem of causation (or at least apparent causation) by omission. As mentioned in §1.2.2, judgements of causation by omission pose a problem for the natural network model of causation. This is because (assuming there are no negative events), omissions are not events, and are therefore incapable of acting as nodes in a mind-independent relational structure. In 'Causing and Nothingness', Beebee argues that by applying the distinction between the concepts 'causation' and 'causal explanation', the natural network model can be shown to be consistent with judgements of causation by omission. The idea is

[56] Beebee, "Causing and Nothingness". In distinguishing between the concepts 'causation' and 'causal explanation', Beebee is following Davidson, who argues that 'causation' is a two-place relation between events, whereas 'causal explanation' is a sentential connective. See Donald Davidson, "Causal Relations" *Causation*, eds. E. Sosa and M. Tooley (Oxford: Oxford University Press, 1993): 86.

[57] Beebee uses the term 'network model', not 'natural network model'. However, she makes it clear that she considers causation to have the features of a natural relation, so the position she defends is an example of the natural network model.

that when we cite omissions as causes, we are actually applying the concept of causal explanation, rather than the concept of causation itself.

Before considering the *metaphysics* of causation by omission, Beebee discusses the role that omissions play in everyday causal discourse. Her aim in doing so is to show that judgements of causation by omission pose just as much of a problem for accounts of causation according to which causation is *not* a relation between events,[58] as for the natural network model.

After considering which omissions we are willing to cite as causes in everyday life, Beebee arrives at the following definition of causation by omission:

> The absence of an *A*-type event caused *b* if and only if
>
> (i) *b* counterfactually depends on the absence: Had an *A*-type event occurred, *b* would not have occurred; and
>
> (ii) the absence of an *A*-type event is *either* abnormal *or* violates some moral, legal, epistemic, or other norm.[59]

Notice the similarity between Beebee's and McGrath's accounts of our everyday judgements about causation by omission: both claim that we cite omissions as causes when there is counterfactual dependence between an event and an absence that is abnormal in some sense. The only significant difference between the two accounts is that McGrath combines the violation of statistical norms (absences that Beebee calls 'abnormal') and prescriptive norms into a single notion of *normal*.

Beebee rejects the idea that the above definition of causation by absence could have anything to do with the *metaphysics* of causation. The problem is that

---

[58] The alternative to the natural network model that Beebee has in mind is Mellor's account of causation, according to which causes and effects are *facts*. For example, see D. H. Mellor, "For Facts as Causes and Effects" *Causation and Counterfactuals*, eds. J. Collins et al. (Cambridge, MA: MIT Press, 2004).

[59] Beebee, "Causing and Nothingness": 296. (Italics in the original.)

> There just *isn't* any objective feature that some absences have and others lack in virtue of which some absences are causes and others are not. So *any* definition of causation by absence that seeks to provide a principled distinction between absences that are and are not causes is bound to fail: No such definition will succeed in carving nature at its joints.[60]

Beebee's point here is clearly correct—because we tend to cite only *abnormal* omissions as causes, any account of causation by omission that reflects our actual causal discourse must make a distinction between normal and abnormal events (or states); and, as both Beebee and McGrath point out, the relevant notion of *normal* includes normative factors, and is thus not perfectly objective.

According to Beebee, because there is no objective distinction between those omissions we cite as causes and those we do not, it must either be the case that *all* omissions are causes, or that *no* omissions are causes. Thus, causation is no more of a problem for the natural network model (according to which omissions cannot be causes) than for accounts of causation that claim that causation is not a relation between events. These latter accounts must hold *either* that no omissions are causes, *or* (implausibly, according to Beebee) that all omissions are causes.

However, showing that there is no *objective* distinction between those omissions that we cite as causes and those omissions we do not consider to be causes does not, by itself, entail that our judgements about causation by omission are not instances of the central, metaphysically respectable concept of causation. This further conclusion follows only on addition of the premise that there is an objective distinction between events that are causes and events that are not causes. Instead of rejecting the claim that judgements of causation by omission are instances of the metaphysical concept of causation, as Beebee

[60] Beebee, "Causing and Nothingness": 300. (Italics in the original.)

does, it is possible to reject the claim that causation is objective (or non-normative)—that is, to conclude that there is not an objective feature that all causes have and that non-causes lack. If there is no objective feature that all causes have in common, then the fact that there is no objective distinction between omissions that are causes and omissions that are not causes does not entail the falsity of the claim that some omissions are causes and others are not. This is where McGrath and Beebee diverge—Beebee insists that causation must be an objective, natural relation, whereas McGrath tentatively accepts that normative factors play a role in determining what is a cause.[61]

Having concluded that omissions cannot be causes, Beebee needs to provide an alternative explanation of judgements of causation by omission. To this end, she claims that all sentences involving talk of causation by omission are actually applications of the concept *causal explanation*, and that these sentences refer to the fact that the causal history of a particular event does not contain any events of a certain type.[62] For example, Beebee would say that the statement 'Tim's failure to water my geranium *caused* it to die' should be rephrased as 'My geranium died *because* Tim failed to water it', and that this sentence indicates that there are no events of the kind 'Tim's watering the geranium' in the causal history (over the relevant period of time). Furthermore, if an appropriate number of watering events *had* been included in the causal history, the plant would not have died. But, crucially, Tim's omission is not *literally* a cause of the plant's death.[63]

Notice that Beebee's formulation of the ambiguity response does not provide us with any independent means of determining when we are applying the concept of causation,

[61] McGrath, "Causation by Omission: A Dilemma": 145-46.
[62] More precisely, judgements of causation by omission entail that the causal history of an effect did not include enough events of the relevant type, at the relevant time. In providing this account of judgements of causation by omission, Beebee is making use of a theory of causal explanation articulated by Lewis in Lewis, "Causal Explanation".
[63] Beebee, "Causing and Nothingness": 304-06.

as opposed to the concept of causal explanation, other than to say that we must be using causal explanation whenever the putative cause (or effect) is metaphysically problematic. For example, Beebee's claim that the sentence 'Tim's failure to water my plant caused its death' does not actually describe an instance of causation is plausible only if causation by omission has been ruled out in advance. Beebee is not begging the question, because her reasons for thinking that omissions cannot be causes (discussed above) are independent of her formulation of the ambiguity response. However, Beebee's assertion that omissions cannot be causes is based on a prior acceptance of the metaphysical claim that causation is a natural relation. In §2.3, I will argue that there are good reasons to doubt the assumption that causation is a natural relation, in which case Beebee's ambiguity response to the problem of (apparent) causation by omission is neither necessary nor successful.

Recall that the ambiguity response is supposed to offer a solution to the problem of context sensitivity, not just judgements of causation by omission. In order to explain the context sensitivity of everyday causal discourse, the ambiguity response claims that in different contexts, different concepts of causation are at work, and that the context acts to disambiguate between these different concepts. The second version of the ambiguity response, formulated by Hitchcock, takes this approach.

2.1.2 Hitchcock's version of the ambiguity response

In 'Of Humean Bondage' Hitchcock argues (contrary to the assumption of most philosophers working on the metaphysics of causation) that there is not just one concept of causation, which refers to a unique causal relation.[64] Rather, there are *many* causal concepts, which succeed in picking out *different* causal relations. Hitchcock uses a number of examples to support this claim, including the following:

[64] Christopher Hitchcock, "Of Humean Bondage" *The British Journal for the Philosophy of Science* 54 (2003).

> Two assassins, Captain and Assistant, are on a mission to kill Victim. Upon spotting Victim, Captain yells "fire!", and Assistant fires. Overhearing the order, Victim ducks and survives unscathed.[65]

As Hitchcock notes, the causal *structure* of this scenario is clear (i.e. the causal system described by this example is well defined). The example entails a number of true counterfactuals, including 'If Captain had not yelled, Assassin would not have fired' and 'If Captain had not yelled, Victim would not have ducked,' etc. The causal processes realised in the example are also well defined: for example, we know that there are sound waves that travel from Captain to Victim as a result of Captain's yell, and that on receiving these sound waves, Victim ducks. However, Hitchcock reports that philosophers are divided in their opinion as to whether the captain's yell caused the victim to survive.

According to Hitchcock, the reason this example generates conflicting intuitions is that more than one causal concept can be applied to this scenario, and which concept we apply influences whether we think the captain's yell caused the victim to survive. For example, two different causal relations that might apply to the example are: first, *C is part of a causal chain of events leading to E*; and second, *C increases the likelihood of E, relative to some normal alternative*. The yell is a cause of the victim's survival according to the former concept, but not the latter.

Hitchcock's claim that there is more than one concept of causation, and that this fact explains people's different responses to the above example, is plausible. It is also plausible that this kind of explanation could account for some of the context sensitivity of causal discourse. However, this version of the ambiguity response does not explain *all* of context sensitivity of our causal discourse. In particular, the claim that there is

[65] Hitchcock, "Of Humean Bondage": 10.

more than one kind of causal relation does not explain the context sensitivity of the distinction between causes and conditions (nor is it intended to).[66]

To see this, let us return to the case of the fire breaking out in the lab, discussed in the previous chapter (§1.2.1). Recall that the sentence 'The presence of oxygen caused the fire' is judged to be false in most contexts. However, if the fire starts in an area of the lab from which oxygen is normally excluded, we judge that this sentence is true. In order for the ambiguity response to succeed in accounting for this context sensitivity, the reason that we judge the presence of oxygen to be a cause in one situation, and not the other, would have to be that we apply different causal concepts in these different contexts. That is, because of an ambiguity in 'cause', the sentence 'The presence of oxygen caused the fire' is false in some contexts, and true in others. However, it is by no means obvious that the posited ambiguity exists. At face value, this sentence seems to involve the same notion of causation in both contexts—it is extremely unlikely that a native English speaker would detect any ambiguity in the term 'cause' here.[67] Thus, positing an ambiguity cannot explain the fact that we judge the sentence 'The presence of oxygen caused the fire' to be false in most contexts, but true when oxygen is not normally present.

In summary, Beebee's version of the ambiguity response can plausibly account for judgements of causation by omission *if* it can be independently shown that causation is a natural relation, a claim that will be questioned in §2.3. However, as an account of the context sensitivity of causal discourse, the ambiguity response is not successful, because

[66] Hitchcock explicitly states that his claim that there are multiple causal concepts is different from the claim that we pick out different causes as *the* cause of an effect in different contexts. Hitchcock, "Of Humean Bondage": 9.

[67] Note that positing an ambiguity *can* plausibly explain the fact that people have varying intuitions regarding the truth of the sentence 'The presence of oxygen caused the fire' in contexts in which oxygen *is* normally present. According to the concept of deviant token causation described in this thesis, this sentence is false. However, according to a different possible concept, on which *C* is a cause of *E* if and only if *C* is part of the causal history of *E*, the sentence is true.

there is at least one form of causal context sensitivity (i.e. the distinction between causes and conditions) to which the ambiguity response does not apply. Therefore, this response is not, by itself, an adequate invariantist solution to the context sensitivity of causal discourse.

2.2 Invariantism option 2: the pragmatic response

The second strategy employed by invariantists to account for the context sensitivity of causal discourse is to claim that this context sensitivity can be entirely accounted for by conversational pragmatics. For example, advocates of the pragmatic response claim that the distinction between causes and conditions is not part of the semantics of 'cause'; rather, conditions are simply causes that are not *relevant* in a particular context. Lewis advocates this response, stating: 'I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My approach is intended to capture a broad and nondiscriminatory concept of causation.'[68] According to the pragmatic response, our tendency to pick out one (or a few) of these causes and label it 'the cause' can be completely explained by conversational pragmatics.

There are two problems with the pragmatic response. First, if we consider the *purpose* of our causal discourse, it turns out that the distinction between causes and conditions is central to our token causal judgements, rather than a discriminatory afterthought. For this reason, it is doubtful that the 'broad and non-discriminatory' concept that Lewis is referring to actually *exists*. Second, it is questionable that pragmatic mechanisms could adequately explain the context sensitivity of our causal reasoning, even if Lewis's unselective concept of causation did exist.

[68] Lewis, "Causation": 162.

2.2.1 The purpose of the concept of causation

To begin to illustrate the first objection to the pragmatic response, notice that if we stop to consider the purpose of the causal concept, it is evident that we use causal discourse to explain past events and predict future happenings because giving accurate causal explanations and making reliable predictions enables us to *manipulate* the world.[69] To be able to manipulate the world, it is crucial to distinguish one (or perhaps a few) causes from a much broader causal history.

Consider the following example:

> Tony has just had a car accident: he was driving on the freeway just below the speed limit, when the brakes failed and he lost control, crashing into a barrier on the side of the road.

Tony wants to know what caused the crash, so he can prevent another accident from happening in the future.

In terms of counterfactual dependence, it is true that if the frictional force between Tony's car and the road had been greater, he would not have crashed. It is also true that if Tony had been driving at 5 km/h, he would not have crashed; that if a hurricane had swept Tony's car off the road five minutes before the time of the accident, he would not have crashed; and that if the brakes had not failed, he would not have crashed. Because all four counterfactuals are *true*, the frictional force on the car, the speed of the car, the absence of a hurricane, and the failure of the brakes, were all equally causes of the crash, according to Lewis's nonselective concept of causation. However, only the final counterfactual is *useful* to Tony's inquiry. This is because, of the four variables considered, whether or not the brakes are working is the only variable that Tony can

[69] James Woodward makes a similar argument in Woodward, *Making Things Happen: A Theory of Causal Explanation*: Chapter 1.

realistically *control*. This is not to say that causes are always events that we can control, but that if the purpose of the causal concept is to enable us to intervene in the world, the distinction between causes and conditions is crucial.

Hitchcock and Knobe also claim that the purpose of actual, or token, causal judgements is to pick out one (or a few) events from an extensive causal structure:

> If the concept of actual causation were entirely egalitarian, we find it hard to see how it could be helpful for people even to have the concept at all ... The value of the concept of actual causation, we wish to suggest, comes precisely from the fact that it is *inegalitarian*. Our concept of actual causation enables us to pick out those factors that are particularly suitable as targets of intervention.[70]

In other words, the distinction between causes and conditions is a central feature of our token causal judgements. Part of the *meaning* of the claim '*c* caused *e*' is that *c* is an especially *notable* part of *e*'s causal history.[71] If this is right, then the distinction between causes and conditions (and thus at least some of the context sensitivity of our causal reasoning) must be semantic, rather than a pragmatic addendum.

Schaffer makes the same point in slightly different way. Responding to Lewis's stated intention to capture a 'broad and non-discriminatory concept of causation', Schaffer notes that:

> [I]t is not obvious that we *have* any such concept as Lewis seeks. Or at least, it is not obvious that our intuitions about causation can provide any *evidence*

[70] Hitchcock and Knobe, "Cause and Norm": 593-94. (Italics in the original.)
[71] If Hitchcock is right that there are multiple causal concepts, just being an event somewhere in the causal history of an effect might be a necessary and sufficient condition of causation according to *one* of these concepts. The point is that this is not the causal concept we are applying when we make claims such as 'The failure of the brakes caused the crash'.

> concerning this 'broad and nondiscriminatory concept', if our intuitions are shot through with selection effects.[72]

Schaffer's point is that, whatever claims we might want to make about the non-discriminatory nature of the *metaphysics* of causation, our causal discourse is not non-discriminatory at all. In order for the pragmatic response to be plausible, it must be possible to separate the central, unselective concept of causation (which corresponds to both the semantics and metaphysics of causation) from our secondary, pragmatic judgements concerning which causes are relevant. Close consideration of causal discourse does not provide any reason to think that this separation is possible—there is no evidence Lewis's egalitarian concept of causation actually exists.

2.2.2 The inadequacy of known pragmatic mechanisms

The second objection to the pragmatic response is more technical, and poses linguistic problems with treating the context sensitivity of our causal discourse as a matter of pragmatics. The problem is that pragmatic mechanisms, as they are currently understood, do not seem capable of doing the necessary work.[73]

To illustrate this objection, let us return, again, to the example of the fire in the lab. The datum that needs explaining is that the sentence 'The presence of oxygen caused the fire' is judged to be true in a context in which oxygen is *not* normally present, but false in most situations, when oxygen *is* normally available. That is, the presence of oxygen is a condition of the fire in most contexts, and a cause of the fire in some specific contexts: namely those contexts in which the presence of oxygen is abnormal. According to Lewis (and other advocates of the pragmatic response), the sentence 'The presence of oxygen caused the fire' is equally true in all contexts. The reason we are

[72] Jonathan Schaffer, "The Metaphysics of Causation" *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition): §2.3. (Italics in the original.) Hart and Honoré's discussion of the same idea can be found in Hart and Honoré, *Causation in the Law*: 10-11.

[73] My discussion in §2.2.2 largely follows Schaffer, "Causal Contextualism": 41-44.

unlikely to *utter* this sentence in most contexts is that the sentence is irrelevant.[74] The pragmatic response therefore attempts to exploit the different between *truth* and *assertability*. Sentences are true if they reflect the way the world actually is, whereas sentences are assertable if it is appropriate to say them in a particular context.

More precisely, uttering the sentence 'The presence of oxygen caused the fire' would be a violation of Grice's maxim of Relevance (or Relation), which says that to be assertable, any contribution to a conversation must be relevant. As Schaffer puts it, the idea is that

> [When] inquiring into the causes of the ... fire, [the inquirer] is *presupposing* that oxygen is present, and wondering about *what ignited the oxygen*. So citing the presence of oxygen is failing to speak to the question under discussion, and hence flouts Relevance.[75]

The problem with this pragmatic explanation of the distinction between causes and conditions is that it cannot explain our intuition that the sentence 'The presence of oxygen caused the fire' is *false* in some contexts.

According to Gricean pragmatics, to violate the maxim of relevance is to utter a sentence that is intuitively true, but irrelevant. A typical example is as follows:

> Paul: Do you think we should take High St or Elphin Rd to get to the movies?
>
> Sue: I always thought that green looked good on you.

[74] Lewis suggests that Gricean maxims (and thus the difference between truth and assertability) can explain our reluctance to utter many causal sentences. See David Lewis, "Causation as Influence" *Causation and Counterfactuals*, eds. J. Collins et al. (Cambridge, MA: MIT Press, 2004): 101.

[75] Schaffer, "Causal Contextualism": 42. (Italics in the original.) Note that Schaffer is discussing a slightly different example, in which a ranger is determining the cause of a forest fire. This does not affect the point being made. Also note that Schaffer should have said 'and hence violates Relevance', rather than 'and hence flouts Relevance'. To flout a maxim is to *deliberately* break the maxim—that is, to exploit the maxim to implicate something that has not been explicitly stated.

The problem with Sue's response is that it has nothing to do with the question that was asked. Her response is true, but irrelevant, and therefore not assertable.

The sentence 'The presence of oxygen caused the fire' (uttered in a context in which oxygen is normally present) does not violate the maxim of relevance in the same way that Sue's response does. For one thing, we are likely to think that the sentence 'The presence of oxygen caused the fire' is actually *false* when the sentence is uttered in a context in which oxygen is normally present. We might even go further and say 'No, the presence of oxygen didn't cause the fire; the cause was the highly reactive potassium.' That is, the *negation* of the original sentence is assertable. The problem here is that Gricean pragmatics can only explain why some *true* sentences are *not* assertable (and not why some sentences that we think are false are actually true). Thus, the maxim of relevance cannot account for the intuition that the sentence 'The presence of oxygen caused the fire' is false, or for the fact that its negation is assertable.

The foregoing discussion shows that appealing to the Gricean maxim of relevance does not, by itself, do enough to establish that the context sensitivity of causal discourse can be accounted for by pragmatics. Advocates of the pragmatic response need to take their own position more seriously, and demonstrate exactly *how* pragmatics can be used to explain the observed context sensitivity. This project is not entirely without potential—there are some applications and extensions of Gricean pragmatics that *do* claim to account for the fact that sentences that seem false are actually true, just not assertable. For example, faced with an analogous situation in the semantics of *conditionals*, Frank Jackson develops a nuanced version of the pragmatic response, which he claims *can* explain the fact that some unassertable conditionals appear to be false, but are nevertheless true. In the next subsection, I will present Jackson's account, and argue that, although the pragmatic response offered by invariantists in the philosophy of

causation could be similarly improved, the improved version would still not be successful.

2.2.3 Jackson's semantics of conditionals

According to basic propositional logic, a conditional is false if its antecedent is true and its consequent is false. In all other situations, conditionals are true. Thus, introductory propositional logic describes a truth-functional semantics of conditionals. However, as is well known, this truth-functional account is not consistent with all the linguistic evidence concerning conditionals. One problem is that the truth table just described entails that from a true sentence, *p*, you can infer a conditional of the form 'If *s* then *p*', where *s* is *any* proposition. For example, from 'Canberra is the capital of Australia', it can be validly inferred that 'If Perth is the capital of Australia, then Canberra is the capital of Australia', a conclusion that seems obviously false. This is one of the *paradoxes of material implication*, the existence of which constitutes evidence that the semantics of conditionals do not obey the simple truth table described above.

Anyone who claims that the truth-functional account *is* the correct account of the semantics of conditionals is thus in a similar situation to invariantists about causation—both parties defend an account of the semantics of a particular concept that is inconsistent with some of the linguistic evidence. In both cases, it is tempting to appeal to pragmatics to explain this inconsistency.

For example, Grice argues that the truth-functional approach to conditionals is correct, and that the difference between this account of the semantics of conditionals and the use of conditionals in everyday discourse can be explained by conversational pragmatics.[76] According to Grice, sentences like 'If Perth is the capital of Australia, then Canberra is the capital of Australia' are literally true, but not assertable.

[76] Paul Grice, *Studies in the Way of Words* (Cambridge, MA: Harvard University Press, 1989).

As it stands, Grice's response is not very convincing, because it does not account for an important difference between our reactions to sentences generated by the paradoxes of material implication, and other sentences that are logically entailed by the truth of a proposition. For example, Gricean pragmatics can explain the fact that the sentence 'Either Canberra is the capital of Australia, or Perth is the capital of Australia' is true, but not assertable. Anyone who understands the meaning of the word 'or' (in its inclusive sense) will accept that this sentence is true—it is just not relevant, once we know that Canberra is the capital. The situation regarding the sentence 'If Perth is the capital of Australia, then Canberra is the capital of Australia', however, is different. We not only refuse to assert this sentence; we also believe it is false. Thus, to give a complete defence of the pragmatic response to the paradoxes of material implication, it is necessary to explain why we *believe* that many sentences generated by these paradoxes are false, although they are literally true.

Again, a defender of the pragmatic response to the paradoxes of material implication is in an analogous situation to those who advocate the pragmatic response to the context sensitivity of causation—he needs to account for cases in which sentences that *seem* false turn out to be true. This is exactly what Jackson sets out to do in his account of conditionals.

Jackson develops an extended version of Grice's pragmatic solution, according to which the semantics of conditionals is given by the truth table described above, but the use of conditionals in everyday discourse also involves a special rule of assertability. This rule states that whether 'If p then q' is assertable depends on the probability of *q*, given *p* (*Pr*(q | p)). When *Pr*(q | p) is low, conditionals are not assertable, even when true.[77] This explains why 'If Perth is the capital of Australia, then Canberra is the capital of Australia' is not assertable: in this case, *Pr*(q | p) is zero.

[77] Frank Jackson, *Conditionals* (Oxford: Basil Blackwell, 1987).

Jackson's pragmatic response to this example of one of the paradoxes of material implications is clearly an improvement on Grice's original proposal, and it is possible that an analogous assertability rule could be used to strengthen the pragmatic response to the context sensitivity of token causal claims. However, there are good reasons to suspect that even this, more sophisticated, response would be unsuccessful. To see this, note that there are two major lines of objection to Jackson's account. First, the account cannot be extended to embedded conditionals (i.e. to conditionals that are not actually asserted).[78] Second, Jackson does not explain why we believe that sentences like 'If Perth is the capital of Australia, then Canberra is the capital of Australia' are false, rather than simply not assertable. In fact, he resorts to an error theory of conditionals, claiming that our beliefs about which conditionals are true (falsely) correspond to the probabilistic assertability condition, rather than the actual truth conditions of conditionals.[79] As Dorothy Edgington points out, this claim is unpalatable to anyone who thinks the correct account of conditionals should have some psychological basis.[80]

Due to the two objections just discussed, Jackson's pragmatic response is not the only, nor the most widely accepted, solution to this paradox of material implication. The other basic strategy is to argue that conditionals are non-truth-functional—that is, that the truth value of a conditional is not entirely determined by the truth values of its antecedent and consequent.[81] The details of non-truth-functional accounts of conditionals are not important here, except to note that pursuing this strategy is analogous to rejecting invariantism and adopting a contextualist account of causation.

As already mentioned, it would probably be possible to similarly improve the pragmatic response to the context sensitivity of causation by adding extra assertability conditions.

[78] An example of an embedded conditional is 'Jane said that if it is sunny, she'll go to the beach'.
[79] Jackson, *Conditionals*: 38-40.
[80] Dorothy Edgington, "Conditionals" *Stanford Encyclopedia of Philosophy* (Winter 2008 Edition): §4.2.
[81] For example, accounts of the semantics of conditionals that make use of modal and relevance logics are both non-truth-functional.

However, it is not obvious that such an account would succeed—after all, Jackson's pragmatic account of conditionals is not widely accepted. More worryingly, the more complicated the pragmatic assertability conditions of a particular concept, the wider the gap between the *semantics* of that concept, and the actual *use* of the concept. For example, in order to explain the intuition that the sentence 'The presence of oxygen caused the fire' is *false* when uttered in a context in which oxygen is normally present, while holding that the sentence is literally true, it would be necessary to endorse an error theory about people's actual causal beliefs (as Jackson does for conditionals). Adopting a more nuanced version of the pragmatic response thus leads back to the first objection to this response: even if this more sophisticated version succeeds in accounting for all token causal judgements, it results in a disconnect between the concept of causation and the causal judgements that are actually made in everyday life, and thus between the truth conditions of token causal claims and the purpose of having a causal concept in the first place. Thus, even an improved version of the pragmatic response would not succeed as a psychologically (or linguistically) realistic account of causal discourse.

The failure of both the ambiguity and the pragmatic responses to account for the context sensitivity of token causal claims provides good reason to reject an invariantist account of causation, in favour of a contextualist account. Because the natural network model is a form of invariantism, this lack of success also provides further reason to question this model of causation. In the next section, I will go further, and argue that the natural network model is mistaken.

2.3 The natural network model reconsidered

According to the natural network model, the causal history of the universe is a network consisting of events and causal relations. On this picture, all causes and effects are part of a *single* system, and, importantly, the truth value of every causal sentence is directly

determined by this system. Although the natural network model lies behind most of the accounts of token causation that have recently been defended in the literature, the discussion so far in this thesis has generated a number of reasons to be sceptical of this model.

First, Chapter 1 defends an account of the majority of token causal claims that is *not* based on the natural network model. According to the Menzies–McGrath model, deviant token causal judgements refer to interventions in the normal course of evolution of idealised, localised systems, rather than events within a relational structure that encompasses the whole universe. The Menzies–McGrath model can better account for two features of causal discourse than the natural network model (i.e. judgements about causation by omission, and context sensitivity), and thus, I argued, offers a better account of (one kind of) everyday token causal judgements than the natural network model.

Second, in §2.2.1, I argued that the purpose of having a concept of causation is to enable us to manipulate the world, and that in order to fulfil this purpose, the concept of causation must be selective—it must allow us to pick out a few events from the causal history of the universe as the causes of any particular effect. According to the natural network model, however, the concept of causation is *un*selective. Those events that are considered to be *causes* of an effect, and those events that are considered to be *conditions* of the same effect, are equally part of the causal history of the universe envisaged by those who accept the natural network model.

Finally, both McGrath's and Beebee's accounts of judgements about causation by omission (§§1.4 and 2.1.1 respectively), as well as experimental studies of causal judgements (§1.4.1), suggest that there are reasons for thinking that non-natural factors,

including human norms, enter into causal judgements—that is, that causation may not be a natural relation, after all.

The three factors just discussed all individually raise questions about the natural network model. Collectively, they show that there are significant costs in adopting this model as an account of the semantics of the causal concept—the natural network model does not provide a particularly good fit to our actual causal discourse. Further, the arguments provided in §§2.1 and 2.2 show that neither the ambiguity nor the pragmatic response—the solutions offered by those philosophers who accept the natural network model to account for the context sensitivity of causal discourse—are successful. This lack of success casts doubt on the claim that the natural network model can account for the context sensitivity of causation, and thus also on the claim that it is possible to reconcile the metaphysics of the natural network model with the semantics of token causal judgements. The aim of this section is to expand these reasons for questioning the natural network model into an argument for the conclusion that the model is mistaken.

2.3.1 An argument against Beebee's formulation of the natural network model

Beebee summarises the network model in the following passage:

> The complete causal history can be represented by a sort of vast and mind-bogglingly complex "neuron diagram" of the kind commonly found in discussions of David Lewis, where the nodes represent events and the arrows between them represent causal relations.[82]

When confronted with this description of the natural network model, a question that is immediately apparent, which is (strangely) only rarely addressed in the literature, is: which events does the causal history consist of? In other words, at what level of

[82] Beebee, "Causing and Nothingness": 291.

explanation can we represent the causal structure to which Beebee refers? There are two possible answers to this question: first, that the causal structure consists of events at the level of fundamental physics, and is therefore best described in the language of fundamental physics (whatever that turns out to be); and second, that the causal structure consists of events at some other level of explanation.

The problem with the first suggestion is that it is unlikely that the events that figure in everyday causal discourse (e.g. 'my plant's death') are salient at the level of fundamental physics.[83] Presumably, the events at the level of fundamental physics are things like fundamental particles' colliding and changing spin. Thus, if the causal history of the universe consists of events at the level of fundamental physics, it is unlikely that even paradigmatic examples of token causal claims—for example 'The white ball's colliding with the eight-ball caused the eight-ball to move'—succeed in describing causal relations—segments of the network that is the causal structure of the universe.

The second possibility is that the causal structure consists of events, not at the level of fundamental physics, but at some other level of description. The problem with this suggestion is that it becomes impossible to non-arbitrarily assign the level of description that we should use to pick out events. Is the collision between two atoms an event? Between two billiard balls? Two planets? These events are all potential causes. However, because atoms are connected to billiard balls and to planets by mereological relations, and not (normally) causal relations, events at these three levels of description are not all plausibly part of the single relational structure described in the passage from Beebee. These kinds of events cannot all 'be represented by a sort of vast and mind-

[83] The above argument is an analogue of the argument made by non-reductive physicalists when they assert that higher level *properties* (or types) do not reduce to lower level properties. For example, see Jerry Fodor, "Special Sciences, or the Disunity of Science as a Working Hypothesis" *Synthese* 28 (1974). It is surprising that this argument is not made more frequently with respect to events.

bogglingly complex "neuron diagram"… where the nodes represent events and the arrows between them represent causal relations', because the relational structure Beebee describes is composed entirely of events and causal relations, and does not contain any mereological relations.

Thus, it is not plausible that the causal history of the universe consists of a single structure composed entirely of events and causal relations. At the very least, the causal history of the universe will include mereological relations, as well. That is, Beebee's description of the network model does not paint a realistic metaphysical picture.

The above argument is not a knockdown refutation of the natural network model, however, because the central thesis of the natural network model is not that the causal history of the universe is a relational structure that consists entirely of events and causal relations, but that, as Menzies puts it, 'causal relations depend completely on a substructure of mind-independent relations.'[84] That is, what is crucial to the natural network is not the claim that the causal history of the universe consists entirely of *causal* relations, but that causal relations are all part of a *single* mind-independent relational structure. However, I will now show that there is good reason to doubt even this, weaker, claim.

2.3.2 Do token causal claims refer to a single relational structure?

The natural network model entails that there is only one correct way of representing the causal history of the universe. This, in turn, entails that the causal history of any particular spatiotemporal region consists of a single relational structure—of a single kind of causal system. This is a bold claim, especially when combined with the claim that the truthmakers of all token causal claims are to be found within this relational structure.

[84] Menzies, "Causation in Context": 192.

Contrary to the claim that each spatiotemporal region instantiates a single causal system, Menzies argues that:

> There are natural kinds, in my view, but it is the job of metaphysics and science rather than conceptual analysis to investigate what they are. However these investigations turn out, a plausible metaphysics is likely to allow that any particular spatiotemporal region instantiates several kinds of systems. Perhaps an extremely austere physicalism committed to the existence of a unified field theory would assert that every situation is to be modelled in terms of a unique physical kind of system subject to the unified field equations. However, any less austere metaphysics is likely to conclude that several, perhaps imperfectly natural kinds of systems may be instantiated in a given spatiotemporal region. In this case, a conceptual analysis should be able to make sense of the alternative causal judgements about these different kinds of systems.[85]

Assuming (as is reasonable) that the correct metaphysics is not the austere physicalism Menzies refers to, a single spatiotemporal region will contain multiple kinds of systems, exhibiting different degrees of naturalness. For example, the spatiotemporal region that contains the solar system includes kinds of systems ranging from the planetary level to the subatomic level. It also includes systems that are described by different fields of science—from physics to biology and psychology, but also computer science and meteorology. Finally, this region contains systems that are governed by normative laws, including the Australian legal system, and the rules of cricket. Menzies' point is that everyday causal judgements can refer to all these different kinds of systems.

Tim Maudlin puts this point in a slightly different way, claiming that:

[85] Menzies, "Difference-Making in Context": 159.

> Talk about "carving nature at the joints" is just shorthand for "finding a macrotaxonomy such that there are reasonably reliable and informative and extensive lawlike generalisations that can be stated in terms of the taxonomy," and the more reliable and informative and extensive, the closer we have come to the "joints".[86]

According to Maudlin, there is no reason to think that there is only one such macrotaxonomy, and thus no reason to think that there is only one way of carving nature at its joints. Further, since different taxonomies correspond to different causal systems, there is also no reason to think that all token causes are events in the same system—the same relational structure. There is not even reason to think that the laws (or standards) that govern all of these systems can be *reduced* to the laws of physics.

For example, Maudlin points out that the system governing the functioning of word processing programs includes a law (or standard) stipulating that when the computer is running, and the word processing program is open, pressing a key on the keyboard results in the corresponding symbol appearing on the screen. As Maudlin notes, this standard can be used to generate counterfactuals: for example, 'If I had hit the letter "z" on the keyboard instead of "s" just before the last colon, the word that would have appeared would have been "counterfactualz".'[87] This standard can also support causal judgements: for example, 'My mistyping caused the sentence to include the word "casual", instead of "causal".' These are perfectly valid counterfactual and causal claims. However, there is no reason to think that concepts like 'keyboard' and 'letter "z"' can be reduced to the vocabulary of physics. Similarly, normative standards such as the rule of cricket that stipulates that if a fielder catches the ball after the batsman hits it,

[86] Tim Maudlin, "Causation, Counterfactuals, and the Third Factor" *Causation and Counterfactuals*, eds. J. Collins et al. (Cambridge, MA: 2004): 433.
[87] Maudlin, "Causation, Counterfactuals, and the Third Factor": 433.

the batter is (*ceteris paribus*) out, cannot be described in the language of fundamental physics, either.

The systems (or taxonomies) described by Menzies and by Maudlin will be defined more precisely in Chapter 5. For now, it is enough to note that the existence of these different kinds of systems, and the observation at the heart of the Menzies–McGrath model—that many token causal judgements pick out causes that are deviations from these different kinds of systems—provides further reason to doubt that there is a simple connection between the semantics of token causal judgements and the metaphysics of the natural network model.

2.3.3 An epistemological problem arising from the natural network model

Finally, notice that the causal network envisaged by those who endorse the natural network model encompasses the *whole* universe. A representation of this structure would therefore describe the universe as a single system, as it appears from a position *sub specie aeternitatis*—the view from nowhere—from *outside* the universe. Given that we do not have access to the view from nowhere—we never see the universe as a whole—it is hard to see how we could be capable of making correct causal judgements, if the truth of these judgements can be determined only from the view from nowhere. Thus, the natural network faces an epistemological problem. This problem can be avoided by claiming that the truth of causal claims is relative to localised systems, which compose only *part* of the universe, as entailed by the Menzies–McGrath model.

In summary, §§2.3.2 and 2.3.3 consist of two convergent arguments in support of the conclusion that the natural network model is not the correct account of the metaphysics of causation. The first argument, endorsed by Menzies and by Maudlin, targets the claim, entailed by the natural network model, that that each spatiotemporal region consists of only one causal system. The second argument questions whether it would be

possible for us to correctly apply the concept of causation, if the natural network model were true. Combined with the arguments against each of the invariantist responses presented in §§2.1 and 2.2, and with the advantages of the Menzies–McGrath model over the natural network model in accounting for judgements of causation by omission, apparent normative causal factors, and context sensitivity that were discussed earlier in this thesis, these two arguments justify the conclusion that much of our causal discourse refers to systems of the kind described by Menzies and by Maudlin, rather than to the single relational structure posited by the natural network model. These arguments, therefore, provide good reason to reject the natural network model, at least if this model is intended to be the metaphysical foundation for an account of everyday causal discourse.

2.3.4 A rejoinder open to advocates of the natural network model

Of course, it is open to advocates of the natural network model to reject the claim that the metaphysics of causation must be connected in any simple way to everyday causal discourse. However, there are two problems with this response. First, to make such a claim is to give up on the idea that there are meaningful connections between causation and the related concepts of prediction, explanation and manipulation, and thus to give up the claim that there is some connection between causation itself and the purpose of causal discourse.

Second, to sever the connection between causal discourse and the metaphysics of causation is also to sever the connection between our causal *intuitions* and the metaphysics of causation. That is, if there is not a close connection between token causal judgement and the metaphysics of causation, we cannot expect to learn anything about the metaphysics of causation by appealing to examples and counterexamples. In other words, conceptual analysis, the staple of metaphysicians in the philosophy of

causation, and the source of the arguments in *favour* of the natural network model, would be rendered useless.

For these reasons, those who endorse the natural network model should (and do) accept the claim that there is a close connection between the metaphysics of causation and everyday causal discourse. That is, to give an account of the metaphysics of causation is to provide an account of the phenomenon (or phenomena) referred to in causal discourse, and, in particular, by those causal sentences that are considered to be uncontroversially true. I have argued that from this assumption it follows that the natural network model is not the correct account of either the semantics or the metaphysics of causation.

The target of the above argument is not causal realism: the claim that causation is a feature of the mind-independent world. More specifically, the claim that the natural network model is not the correct account of the semantics or metaphysics of token causal claims is consistent with the claim that there *is* a single mind-independent causal structure. Rather, the argument shows that such a structure does not connect to our causal discourse in any simple way—it is not the case that every true causal sentence directly represents a segment of a single relational structure. The concept of causation does not carve nature along a single set of joints. Further, causation is not a natural, non-normative relation. The systems that token causal judgements *do* refer to can consist of prescriptive, as well as descriptive, norms.

2.3.5 Back to Beebee's version of the ambiguity response

In §2.1.1, I considered Beebee's ambiguity response to the problem posed by judgements of causation by omission, and argued that this response is only successful if causation is a natural relation. The conclusion just established—that the natural network model is mistaken, and, in particular, that causation is *not* a completely natural

relation—therefore entails that Beebee's version of the ambiguity response is not successful.

To see why this is the case, consider the sentences: 'Tim's failure to water my plant caused its death' and 'The possum's devouring the leaves of my plant caused the plant's death.' Beebee would claim that the former sentence is actually an application of the concept *causal explanation*, does not literally refer to a causal relation, and should be paraphrased as 'My plant died because Tim failed to water it'. The latter sentence, on the other hand, *is* an application of the concept of causation, and *does* succeed in referring to a causal relation.

The problem with Beebee's response is not with her assertion that the former sentence does not describe an instance of causation, but with her (implicit) assertion that the latter sentence does. The event 'the possum's devouring the leaves of my plant' is not part of a single causal structure, any more than 'Tim's failure to water my plant' is. More generally, the problem is that, if the natural network model provides the correct metaphysics of causation, almost *all* token causal claims will turn out to be applications of the concept of causal explanation, rather than causation. This is because almost all token causal claims refer to partial systems (or macrotaxonomies) at arbitrary levels of explanation, rather than a single relational structure.

2.3.6 An alternative metaphysical picture

The Menzies–McGrath model, and the passages from Menzies and Maudlin included in §2.3.2, support an alternative to the natural network model. According to this alternative metaphysical picture, the world is divided into many different causal systems, which can be represented by different levels of description. For example, a game of billiards (including the table, balls, and players) is one causal system; the physiology of a scarab beetle is another causal system; as is the psychology of an individual person. These

systems are all *open* (they are all subject to interventions), they are all *partial*, and they can all be viewed from a perspective that is located *within* the universe.

The remainder of this thesis will be devoted to elaborating on, and defending, this metaphysical picture, beginning with a discussion of contrastivism, a form of contextualism that combines semantic contextualism with metaphysical realism. The aim of the next section is to compare contrastivism to the Menzies–McGrath model.

2.4 Semantic contextualism + causal realism

According to semantic contextualism, the context sensitivity of a particular area of discourse is explained by the claim that the same *sentence* picks out different *propositions* in different contexts (where a proposition is the linguistic entity that is expressed by a declarative sentence—it is what we believe, if we believe a particular declarative sentence is true). It is clear why this position is referred to as *semantic* contextualism—context sensitivity is accounted for by the relationship between different semantic entities (i.e. sentences and propositions).

Semantic contextualism aims to elucidate the link between the sentences we take to be true in a certain area of discourse, and the propositions these sentences express in different contexts. It is possible to hold contextualism as a purely semantic position—pure semantic contextualism remains neutral as to whether the propositions expressed have mind-independent truthmakers, and thus avoids all discussion of the metaphysics of the concept being accounted for. This position is popular as an account of knowledge claims.[88] However, pure semantic contextualism is a less attractive position with regard to causation, because the philosophy of causation has (for obvious reasons) historically been thought of as a metaphysical pursuit. Anyone defending a contextualist account of

[88] For example, see Keith DeRose, "Contextualism and Knowledge Attributions" *Philosophy and Phenomenological Research* 52 (1992); Jonathan Cohen, "Contextualism, Skepticism, and the Structure of Reasons" *Philosophical Perspectives 13: Epistemology*, ed. J. E. Tomberlin (Malden: Blackwell Publishers, 1999).

the causal concept is therefore expected to consider the metaphysical implications of his account. Traditionally, there are two options here: realism or antirealism. Unsurprisingly, the most popular version of causal contextualism—contrastivism—is a combination of semantic contextualism and metaphysical realism. The motivation for defending such an account of causation is to bridge the gap between context sensitive causal discourse and an objectivist (or realist) metaphysics of causation.

2.4.1 Contrastivism

The guiding idea behind contrastivist accounts of causation is that causal judgements are always relative to contrasts. Contrastivists claim that when uttering a token causal sentence, we are actually expressing a proposition of the form: *c* rather than *c** caused *e* rather than *e**. That is, although causal *sentences* are typically two-place, the *propositions* these sentences express are four-place (although both the number and category of the causal relata are disputed).[89] The idea is that the contrasts (in this case *c** and *e**) are generally not explicitly mentioned in our causal sentences, but are implicitly supplied by the context.[90]

Adding extra places to the grammar of causal propositions creates room for context sensitivity to enter into the semantics of the causal concept. For example, contrastivist accounts of causation deal with the contextual sensitivity of the distinction between causes and conditions by claiming that in different conversational contexts, different relevant alternatives are in play. I will illustrate this idea by returning to the example of the fire breaking out in a lab.

[89] I will focus mainly on Schaffer's version of causal contrastivism. However, other theorists have put forward slightly different contrastivist accounts. According to Maslen, for example, causal propositions are three-place, i.e. *c* rather than *c** caused *e*. Maslen, "Causes, Contrasts, and the Nontransitivity of Causation". For another contrastive account, see Hitchcock, "Farewell to Binary Causation".

[90] We do sometimes make the contrast event explicit—for example, the sentence 'Using margarine rather than butter caused Dave's cholesterol to be lower than it would otherwise have been' is a four-place causal sentence.

According to contrastivists, when determining the cause of a fire that takes place in normal atmospheric conditions, we presuppose that oxygen is present, and do not consider the replacement of oxygen with some other gas as a possible contrast event. So, in the case of the fire breaking out in the lab in which oxygen is normally present, the full causal proposition, including contrast events, is something like:

> The presence of potassium rather than the presence of sodium caused there to be a fire rather than a controlled reaction.[91]

However, if the fire were to break out in a fume cupboard in which oxygen is normally replaced by argon, the presence of argon *would* be an obvious candidate for the contrast event. In this case, the four-place causal proposition would be:

> The presence of oxygen rather than the presence of argon caused there to be a fire rather than a controlled reaction.

The idea is that the truth conditions for these four-place causal propositions can be cashed out in terms of counterfactual dependence. That is: *c* rather than *c** caused *e* rather than *e** if and only if, if *c** had occurred, then *e** would have occurred.

As noted above, one of the motivations for defending a contrastivist account of causation is to bridge the gap between context sensitive causal discourse and an objectivist (or realist) metaphysics of causation. For this reason, contrastivists generally take pains to make it clear that in introducing context sensitive contrast events into their account of causation, they are not implying that causation is subjective.[92] For example, Maslen, who defends a three-place version of contrastivism, states that: 'The causal structure of the world is an objective, mind-independent three-place relation … between

[91] Sodium is less reactive in oxygen than potassium.

[92] In the above paragraph (and throughout this thesis) the term 'subjective' is being used as a contrast term. The intended contrast is 'objective', where 'objective' means *mind-independent*, in the sense that the extension of the relevant term (or the truth value of the relevant proposition) is independent of human beliefs and concerns.

causes, contrasts, and effects.'[93] According to Maslen, once all three places are filled, the truth conditions of causal statements are perfectly objective. Schaffer similarly intends his four-place version of contrastivism to entail that causal sentences refer to an objective four-place relation.[94] He claims that when we utter a true, two-place, token causal sentence, this *sentence* does not directly map onto mind-independent reality. However, once the contrast events have been specified by the context, we arrive at a three- or four-place *proposition*, which (if true) *does* directly map onto mind-independent reality.[95]

Thus, like those invariantists who endorse the network model, contrastivists claim that our causal discourse mirrors, or represents, mind-independent reality—it is just that this representation is carried out by three- or four-place propositions, rather than two-place propositions. Contrastivism is therefore largely consistent with the natural network model of causation—in metaphysical terms, contrastivists are just claiming that the network is more complicated than initially thought. More precisely, contrastivists modify both the semantics and the metaphysics of causation, relative to the natural network model—they claim that just as the logic of causal sentences is three- or four-place, rather than two-place; metaphysically, causation is a three- or four-place relation.

2.4.2 Schaffer's account of causation in the law

According to contrastivists, the contrast events are (typically) determined by the context—they are the relevant alternatives to the events that we pick out as causes and effects, in the sense that the contrast events could have replaced these causes and effects. For completeness, however, a contrastivist account of causation needs to

[93] Maslen, "Causes, Contrasts, and the Nontransitivity of Causation": 341.
[94] Schaffer, "Contrastive Causation": 291-92; Jonathan Schaffer, "Contrastive Causation in the Law" *Legal Theory* 16 (2010).
[95] According to contrastivism, the link between the semantics and metaphysics of causation is analogous to that of velocity, which also has an extra place that is not usually made explicit, specifying the frame of reference. For example, in everyday discourse, we might say 'The Ferrari is travelling at 150 km/h'. However, to be fully explicit, we should say 'The Ferrari is travelling at 150km/h, relative to the surface of the Earth.'

include some explanation of *how* the context determines the contrasts. This is a difficult question, to which different contrastivists have provided answers of varying degrees of comprehensiveness.[96] I do not have space to evaluate all these answers here, but will just focus on one account of the way the contrasts are determined, which applies to *legal* contexts, and which is provided by Schaffer.[97] Schaffer does not intend this explanation of contrast selection to cover all token causal judgements, and in fact provides an alternative, and more general, account of contrast selection elsewhere.[98] However, Schaffer's account of causation in the law is particularly relevant to this thesis, because it has obvious similarities to the Menzies–McGrath model.

In 'Contrastive Causation in the Law', Schaffer argues that the concept of causation used in the law (as in other areas of discourse) is contrastive. However, in legal contexts (unlike many other contexts), we determine the causal contrast event by specifying an alternative *lawful* action, and the effectual contrast event by asking what the outcome would have been if the defendant had acted lawfully. For example, in liability cases, in which it may be necessary to demonstrate that the defendant's negligence caused the plaintiff to have been harmed, what must be shown is that 'if the defendant had acted lawfully, the plaintiff would have met a better fate.'[99] In this case, c is the actual breach of duty, c^* is the corresponding lawful conduct, e is the harm that actually occurred, and e^* is the outcome that would have occurred if the defendant had acted lawfully. Schaffer calls the above construction the 'schema for responsibility',[100] and argues that we naturally use this schema in legal contexts.[101]

[96] For example, see Maslen, "Causes, Contrasts, and the Nontransitivity of Causation": 346-47.
[97] Schaffer, "Contrastive Causation in the Law".
[98] Schaffer, "Contrastive Causation": 318-20.
[99] Schaffer, "Contrastive Causation in the Law": 259.
[100] Schaffer, "Contrastive Causation in the Law": 266.
[101] It is worth pointing out that reference to a single concept of causation in the law may well be misleading—it is possible that lawyers use different concepts of causation in different areas of law (e.g.

Notice the similarity between Schaffer's contrastive account of causation in the law and the Menzies–McGrath model. In both cases, the causes and effects are deviations from the normatively defined normal course of evolution (the lawful course of action), and the contrasts are the corresponding lawful actions. To make this similarity even more explicit, note that Schaffer says: 'it seems that causal judgements in the law are based on a comparison between the actual course of events and an alternative scenario in which the defendant acts lawfully.'[102] This hypothetical world in which the defendant acts lawfully is none other than a default world—a world in which the relevant system follows its normal course of evolution. That is, Schaffer's addition of the schema for responsibility to his contrastive account of causation results in an account that looks very much like the Menzies–McGrath model.

Although the notion of *moral* responsibility is restricted to human actions (and Schaffer's discussion in 'Contrastive Causation in the Law' is restricted to legal contexts), it is clear that the notion of *causal* responsibility applies to a far wider range of events, including events that are not human actions. For example, the sentences 'The lightning caused the fire' and 'The lightning was responsible for the fire starting' have the same meaning. This is why Woodward argues that our actual (or token) causal judgements are closely connected to the concept of responsibility (see §1.1.1).

In fact, 'lawful conduct' can be interpreted in the light of McGrath's notion of *normal*, according to which a lawful event is an event that adheres to a relevant standard. In this light, it is apparent that the schema for responsibility applies to all those cases in which c and e are deviations from actual standards, and c^* and e^* are events (or circumstances) that adhere to the actual standards—that is, c^* and e^* are the values the relevant variables take in the default worlds.

---

crime, torts, equity). The concept of causation used by lawyers may also vary according to jurisdiction (e.g. US, UK, Australia).

[102] Schaffer, "Contrastive Causation in the Law": 272.

These observations suggest that Schaffer's 'schema for responsibility' applies to exactly those causal judgements I have called 'deviant token causal claims'. That is, the schema applies to token causal judgements that pick out causes that are deviations from the normal course of evolution of a system. If this suggestion is right, the semantics of the Menzies–McGrath model is not so much an *alternative* to a contrastive semantics, as an *elaboration* of contrastivism, applied to deviant token causal claims. I think this is the best way to understand the semantics of the Menzies–McGrath model. However, the *metaphysics* of the Menzies–McGrath will turn out to be different from the metaphysics of contrastivism.

Although Schaffer is aware that the difference between lawful and unlawful actions depends on human norms, and is thus not entirely mind-independent, he argues that the addition of a normative component into the selection of the contrast events does not entail that causation is not a natural relation. He draws a sharp distinction between the selection of the contrast events, and the evaluation of causal propositions, given the contrasts, and claims that:

> [N]ormativity enters only in the values we tend to be interested in for c^* and e^*. But for any given setting of these values, contrastive causation is a completely objective matter. Thus [Schaffer's contrastive account] can reconcile the normative elements of causal judgements with an objective metaphysical image.[103]

Schaffer's application of the contrastive account of causation to legal contexts is therefore still largely consistent with the metaphysics of the natural network model of causation—his idea is that the application of the (contrastive) concept of causation to a

[103] Schaffer, "Contrastive Causation in the Law": Footnote 25.

legal context involves certain 'conceptual filters',[104] which specify that the contrast events are lawful actions. However, these conceptual filters are not part of the semantics of token causal judgements. According to the Menzies–McGrath model, on the other hand, the default values of the cause and effect variables *are* part of the semantics of deviant token causal claims.

It is useful to think of invariantism, contrastivism and the Menzies–McGrath model as representing different positions on a continuum, in which the boundary between semantics and pragmatics is being shifted such that an increasing amount of contextual information is included in the semantics. According to invariantism, all the context sensitivity of our causal discourse is a matter of pragmatics, rather than semantics. Contrastivists concede that context sensitivity enters into the semantics of 'cause', by including contextually specified contrast events in the grammar of causal sentences. However, contrastivists do not place any semantic restrictions on these contrast events—specification of the contrast events is left to pragmatics. The Menzies–McGrath model goes further, in that determination of the default values of the variables (the equivalent of the contrast events) is held to be part of the semantics of the causal concept. That is, the role of pragmatics, relative to semantics, is further reduced.

In the next chapter (§3.3.2), I will argue that the separation Schaffer relies on, between a pragmatic 'conceptual filter' and an objective semantics of 'cause', does not hold up to scrutiny. For this reason, the midway position between invariantism and the Menzies–McGrath model that Schaffer (and other contrastivists) defend is not coherent—Schaffer's defence of the claim that causation is a completely objective, natural relation does not succeed.

[104] Schaffer, "Contrastive Causation in the Law": 292.

In this chapter, I have provided an argument against the natural network model, and suggested an alternative metaphysical picture, based on the claim at the heart of the Menzies–McGrath model—that token causal judgements are relative to partial, open systems. In the next chapter, I will temporarily step back from the Menzies–McGrath model, and consider the advantages and disadvantages of three different positions in the metaphysics of causation, as a way of more precisely locating the metaphysics of the Menzies–McGrath model within the existing literature on causation.

Chapter 3: Manipulability Accounts of Causation

This chapter will focus on the metaphysics of three accounts of causation: Menzies and Price's agency theory, Woodward's interventionism, and Price's later perspectivalism.[105] These three accounts are all *agency*, or *manipulability*, accounts of causation. That is, the authors all claim that our causal reasoning is grounded, in some sense, in our status as agents—in our ability to *manipulate*, or *intervene in*, the world. However, they employ this idea in different ways, arriving at quite different conclusions about the metaphysics of causation.

According to Menzies and Price's agency theory, causation is a secondary quality, in the sense that an account of the causal concept must include essential reference to human capacities. Whether Menzies and Price regard causation itself as a metaphysically mind-dependent or mind-independent feature of the world (and thus whether their account is metaphysically realist or antirealist), however, is unclear.

In his interventionist theory, Woodward agrees with Menzies and Price that it is our perspective as agents that enables us to engage in causal reasoning. However, he claims that in successfully carrying out this reasoning, we latch onto an objective relation—a feature of the mind-independent world. His interventionist theory is therefore explicitly realist, or objectivist.

In formulating his later theory—causal perspectivalism—Price extends the agency theory he developed with Menzies in the direction of increasing antirealism or subjectivism, arguing that the direction of causation, at least, is a feature of the human perspective, rather than a feature of the mind-independent world itself.

---

[105] Peter Menzies and Huw Price, "Causation as a Secondary Quality" *The British Journal for the Philosophy of Science* 44 (1993); Woodward, *Making Things Happen: A Theory of Causal Explanation*; Huw Price, "Causal Perspectivalism" *Causation, Physics and the Constitution of Reality*, eds. H. Price and R. Corry (Oxford: Clarendon Press, 2007).

The three accounts introduced above can be considered as members of a single evolutionary tree, in that Woodward's interventionism and Price's perspectivalism are both refined versions of Menzies and Price's agency theory—the former in the direction of increasing objectivity, and the latter in the direction of increasing subjectivity. In the two chapters following this one, I will formulate the metaphysics of the Menzies–McGrath model, and argue that this is a further-refined position belonging to the same evolutionary tree, a position which makes use of the best features of both Woodward's and Price's theories.

3.1 Agency theories of causation

According to agency and manipulability theories of causation, the ability of humans to act, or intervene in the world, is central to the concept of causation. These theories are based on the intuitive idea that one event (or the value of one variable), *c*, causes a second event (or the value of a second variable), *e*, if and only if intervening to change *c* would result in a change to *e*. Less formally, the idea is that if you wiggle *c*, *e* will wiggle too. For example, exposure to asbestos is a cause of lung cancer, because by intervening to prevent people being exposed to asbestos, we can prevent occurrences of lung cancer.

Early versions of the agency theory were defended by Collingwood, Gasking, and von Wright.[106] According to Menzies and Price, the central claim of all these agency accounts is that:

> [A]n 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*.[107]

[106] R. G. Collingwood, *An Essay on Metaphysics* (Oxford: Clarendon Press, 1940); Douglas Gasking, "Causation and Recipes" *Mind* 64 (1955); Georg Henrik Von Wright, *Causality and Determinism* (New York and London: Columbia University Press, 1974).

So, according to agency theories, *c* is a cause of *e*, if and only if intervening to bring about *c* is a reliable method of ensuring that *e* occurs.

These early versions of the agency theory are subject to a number of serious objections, the three most significant of which are: first, that the theory fails to account for causes that are *not* possible human actions; second (and relatedly), that agency accounts render causation excessively anthropocentric; and third, that the agency theory is viciously circular.[108] The final objection was addressed in §1.3, where I argued that this circularity is not problematic if we are not attempting to give a *reductive* account of causation. The first and second objections will now be outlined, and discussed in detail below.

First, the problem of unmanipulable causes questions whether the agency theory can be extended to causes that are not manipulable by humans. For example, the earth's rotation is a cause of the daily alternation between day and night. However, humans are not capable of intervening in the earth's rotation to either prevent, or change, this alternation—that is, the earth's rotation is an unmanipulable cause. Since an agent is not able to manipulate the earth's rotation in order to change the diurnal cycle of day and night, it is not clear that the earth's rotation is a cause of this cycle, according to the agency theory.

The second objection to agency theories—that these theories render causation excessively anthropocentric—is related to the problem of unmanipulable causes. According to this second objection, agency theories imply that there would be no causation if there were no human agents (or that the concept of causation would pick

---

[107] Menzies and Price, "Causation as a Secondary Quality": 187.

[108] For example, these objections are raised in J. L. Mackie, "Review of *Causality and Determinism*" *Journal of Philosophy* 73 (1976); Daniel M. Hausman, "Causation and Experimentation" *American Philosophical Quarterly* 23 (1986).

out different causal relations, if it was employed by different agents); a claim, the objection goes, that is obviously false.

Both the problem of unmanipulable causes and the problem of unacceptable anthropocentricity are serious objections, to which a successful version of the agency theory must respond. In 'Causation as a Secondary Quality', Menzies and Price argue that these objections can be overcome if we accept that causation is a secondary quality, as, for example, colour has been traditionally been regarded.

3.2 Menzies and Price's agency theory

The agency theory defended by Menzies and Price in 'Causation as a Secondary Quality' is a probabilistic version of the theory (as opposed to earlier, deterministic accounts). To formulate their probabilistic theory, Menzies and Price extend the basic claim of agency theories (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*')[109] by elaborating on what it is for one event to be an *effective means* of bringing about another event. They claim that bringing about an event, *A*, is an effective means of bringing about another event, *B*, if rational decision theory would prescribe that an agent brings about *A*, rather than ~*A*, in the face of an overriding desire that *B* takes place.[110] The detail of Menzies and Price's position is not important, because the conclusions of 'Causation as a Secondary Quality', and the claims that are relevant to the Menzies–McGrath model, are independent of any specific version of the agency theory. These claims are: first, causation is a secondary quality; and second, if causation is understood to be a secondary quality, the major objections to the agency theory can be overcome. To argue for these claims, Menzies and Price rely

[109] Menzies and Price, "Causation as a Secondary Quality": 187.
[110] Menzies and Price, "Causation as a Secondary Quality": 190.

on an analogy between causation and colour. Thus, in order to explain their analogy, I will briefly introduce the philosophy of colour.

3.2.1 A brief introduction to the philosophy of colour

One of the problems at the heart of the philosophy of colour is that there is a discrepancy between our common sense conception of colour and our scientific conception of colour. According to common sense, colours are one-place properties possessed by objects in the world. We say, and think, 'Fire engines are red', and we mean to assert that a particular object (the fire engine) possesses the property of being red. However, the science of colour shows that there is no single, intrinsic, mind-independent property that demarcates those objects we call 'red'. To explain why a set of objects appear to share the property 'redness', we need to refer to features of the human visual system, as well as to features of the objects themselves. According to a scientific understanding of colour, the reason fire engines look red to us is partly due to the reflectance properties of the fire engine itself (i.e. of the paint it is coated in), and partly due to facts about the relation between the human visual system and certain wavelengths of light.

This discrepancy between the common sense and scientific conceptions of colour gives rise to a debate in the metaphysics of colour concerning where we should locate colour properties. Are they objective features of the mind-independent world? Subjective features of human experience? Dispositional properties? Relational properties? A convenient fiction? All of these answers have their advocates, and I will not enter into the debate here.[111]

It is important to separate this issue in the *metaphysics* of colour from a second issue in the *semantics* of colour terms. The metaphysical debate about colour aims to determine

[111] For an overview of the metaphysics of colour, see the essays in Alex Byrne and David R. Hilbert, eds., *Readings on Color*, vol. 1: The Philosophy of Color (Cambridge, MA: MIT Press, 1997).

what (if any) kinds of things colour properties are—how colour properties are constituted. The semantic debate, on the other hand, asks how we determine *which* (if any) objects are picked out by the concept 'red'—that is, how we fix the *extension* of the term 'red' (and other colour terms). Colour science teaches us that there is no mind-independent, objective property common to every object that we classify as red—there is no neat correlation between the extension of the term 'red' and a single mind-independent, physical property. For this reason, any account of the colour concepts that is consistent with everyday ascriptions of colour terms will have to include reference to the human visual system, as well as the external world.

Menzies and Price claim that to show that a particular quality is secondary, rather than primary, it is enough to show that the relevant quality is *conceptually* dependent on human capacities and responses. For example, they say that 'any account of colour which makes the notion of colour at least *conceptually* dependent on human capacities and responses would have sufficed for our purposes' (where their purpose is to provide an account of a secondary quality against which to compare the agency theory of causation).[112] In other words, for Menzies and Price, the fact that the extension of colour terms is at least partially mind-dependent is enough to ensure that colour is a secondary quality. That is, they assume that whether or not a particular quality is primary or secondary turns on the answer to the *semantic* question, rather than the metaphysical question. Menzies and Price's interpretation of the terms 'primary quality' and 'secondary quality' is different from the traditional interpretation of these terms, according to which a concept refers to a secondary quality if it is *constitutionally* mind-dependent—that is, if the answer to the *metaphysical* question includes mind-dependence.

[112] Menzies and Price, "Causation as a Secondary Quality": Footnote 12. (My italics.)

Menzies and Price's unusual account of secondary qualities can lead to confusion. For example, in 'An Objectivist's Guide to Subjectivism about Colour', Jackson and Pargetter defend a physicalist account of colour, according to which colour properties are physical properties of objects in the external world.[113] (More specifically, Jackson and Pargetter say that redness is the property that causes an object to look red to a human subject in a particular circumstance). By traditional lights (and by Jackson and Pargetter's lights) Jackson and Pargetter's account entails that colour is a primary quality, because they hold that colour properties are constitutionally mind-independent and objective. However, Jackson and Pargetter also claim that the extension of the term 'red' is determined by human responses, so their account entails that colour is a secondary quality, according to Menzies and Price's definition of the term.

3.2.2 Causation as a secondary quality

Taking their unorthodox interpretation of the primary/secondary quality distinction into account, it is important to note that in claiming that causation is a secondary quality, Menzies and Price are not necessarily denying that causation is 'out there in the world'—that causation is constitutionally mind-independent. Rather, they claim that just as 'an adequate account of colour will need to make some reference to human perceptual states or capacities,'[114] agency accounts of causation entail that causation is 'to be explained by relation to our experience as agents.'[115] Thus, as I read them, it is enough to satisfy Menzies and Price's claim that causation is a secondary quality to show that the extension of the term 'cause' is partly determined by human capacities and responses.

---

[113] Frank Jackson and Robert Pargetter, "An Objectivist's Guide to Subjectivism About Colour" *Readings on Color*, eds. A. Byrne and D. R. Hilbert, vol. 1: The Philosophy of Color (Cambridge, MA: MIT Press, 1997).
[114] Menzies and Price, "Causation as a Secondary Quality": 188.
[115] Menzies and Price, "Causation as a Secondary Quality": 193.

One question to keep in mind when evaluating Menzies and Price's agency theory is therefore whether they are actually addressing the *metaphysical* question at all. By asserting that causation is a secondary quality, Menzies and Price are claiming that any satisfactory account of causation will involve reference to our status as agents. However, it is not clear that Menzies and Price have made any commitment regarding the *constitution* of the causal relation.[116] After all, according to them, the claim that *colour* is a secondary quality leaves open all positions ranging from full-blown objectivism (or realism) to subjectivism (or antirealism). Thus, as I understand them, Menzies and Price's intention is not specifically to advocate either realism or antirealism about causation, but to assert that, just as the claim that the extension of colour terms is dependent on the constitution of the human visual system is metaphysically acceptable, so are theories of causation that claim that the extension of 'cause' is determined by our status as agents.

Having said this, it is worth noting that Menzies and Price compare the agency theory of causation to a *dispositional* account of colour—that is, to an account of colour according to which colour is a secondary quality on both their own definition, and the traditional definition (and, therefore, according to Jackson and Pargetter). Everyone agrees that dispositional accounts of colour are metaphysically subjectivist, because dispositional accounts hold that colour properties are constitutionally mind-dependent. Thus, to the extent that Menzies and Price's conclusion that causation is a secondary quality specifically rests on a comparison with the dispositional account of colour, rather than an account in which it is only the extension of colour terms that has a subjective contribution (e.g. Jackson and Pargetter's physicalist account), Menzies and

[116] Price has, at least once, suggested that the agency theory he and Menzies defend in 'Causation as a Secondary Quality' is better seen as an attempt to do 'philosophical anthropology', rather than metaphysics, and describes philosophical anthropology as aiming at 'explaining why creatures in our situation came to speak and think in certain ways'. Huw Price, "Causation, Intervention and Agency—Woodward on Menzies and Price": §2.1.

Price's conclusion may imply that causation is metaphysically subjectivist after all.[117] In other words, because they do not carefully distinguish between the metaphysical and semantic debates, Menzies and Price's conclusions about the metaphysics of causation are unclear.

Finally, notice that if Menzies and Price are correct in their assertion that causation is a secondary quality, the sense in which the truth of causal statements depends on our status as agents is something that applies because of general features of the human condition. This is presumably determined by some combination of our anatomy, perceptual system and psychology. Another way to put this point is to say that their account of causation is based on an *intersubjective* perspective, an idea that will become important below.

3.2.3 The problem of unacceptable anthropocentricity

Menzies and Price use an analogy between their agency account of causation and a dispositional account of colour to respond to the problem of unacceptable anthropocentricity. That is, to show that the agency theory can be extended to times and places where there are no human agents:

> [T]he dispositional theory of colour is to be understood as stating that an object is red just in case it is true that *if* a normal observer *were* present and were to observe the object under standard conditions, it *would* look red to her; and an agency theory of causation is to be understood as stating that a causal relation exists between two events just in case it is true that *if* a free agent *were* present

[117] According to Menzies and Price, their defence of the agency theory does not depend on a dispositional account of colour being correct—they say that they chose to compare the agency account of causation to a basic dispositional account of colour because the two accounts have a similar structure, and therefore face similar objections, to which the dispositional account of colour has recognised responses. Menzies and Price, "Causation as a Secondary Quality": 188-89.

> and able, she *could* bring about the first event as a means to bringing about the second.[118]

Their point is that, just as dispositional accounts of colour make use of counterfactuals to extend the concept of colour to objects that are not actually observed, agency accounts of causation can use counterfactuals to extend the concept of causation to times and places in which there are no humans to intervene.

However, this response does not address a more sophisticated version of the objection, which concerns possible worlds in which agency itself is different. In these worlds, the objection goes, agents have different capabilities. For example, there are possible worlds in which agents have very limited powers of agency—they could be like people with locked-in syndrome, only able to blink their eyes. There are also possible worlds in which agents have more extensive powers of agency. For example, agents could be telekinetic—able to move physical objects just by thinking. According to the more sophisticated version of the problem of unacceptable anthropocentricity, just as the extension of colour terms would be different in worlds in which the human visual system was differently constituted, the agency account of causation implies that in worlds in which agency is different, the extension of the concept of causation would be different, too. The objectors claim that in the case of causation, this implication is unacceptable.

Menzies and Price accept that agency theories entail that there are worlds in which the extension of the term 'cause' is different, even though the word 'cause' has the same meaning as in the actual world (i.e. the same process is followed to determine the extension of 'cause'). However, they deny that this implication of agency theories is unacceptable—their response to the problem of excessive anthropocentricity is thus to

[118] Menzies and Price, "Causation as a Secondary Quality": 198. (Italics in the original.)

accept that causation is an anthropocentric concept. In defence of this claim, Menzies and Price argue that there is a *disanalogy* between the dispositional account of colour and the agency theory of causation when it comes to the amount of variation to be expected in the extension of colour concepts if the human visual system was different, versus the amount of variation to be expected in the extension of the causal concept if cognitive beings had different powers of agency.

Menzies and Price point out that in a world in which humans had a different type of visual system (tetrachromat versus trichromat, for example, or a visual system that was sensitive to light at wavelengths above 900nm), our true colour claims might be quite different. For example, it might no longer be true to say that tomatoes and fire engines are the same colour. However, in a world in which our ability to act was different (we had either much more limited, or more extensive, powers of agency) our concept of causation would be likely to pick out the same relations. This is because Menzies and Price claim that the *degree* of agency possessed by a linguistic community is unlikely to affect the relations that are picked out by the concept of causation—*any* ability to intervene in the world would result in the concept of causation having the same extension as in the actual world.[119] For this reason, the only worlds in which the extension of the term 'cause' is different are worlds in which there are cognitive beings with no powers of agency at all. Menzies and Price therefore conclude that, although causation is a secondary quality, causation is significantly *more* objective than the traditional secondary qualities, including colour.[120]

[119] The reason Menzies and Price claim that the degree of agency possessed by a linguistic community is unlikely to affect the extension of the term 'cause' is related to their solution to the problem of unmanipulable causes (to be discussed in §3.2.4). Briefly, Menzies and Price claim that *c* is a cause of *e* if *either* bringing about *c* is a reliable means of bringing about *e*, *or* the situation involving *c* and *e* is sufficiently similar to a situation involving such a means–end relation.

[120] Menzies and Price, "Causation as a Secondary Quality": 198-202.

A different way of putting this point is to note that, if Menzies and Price are correct, the extension of colour terms is determined by an intersubjective perspective, which includes all *actual* humans, whereas the extension of the causal concept is determined by a perspective which includes all *possible* agents. Because the perspective from which the extension of the term 'cause' is determined has a wider scope than the perspective from which the extension of colour terms (e.g. 'red') is determined, the concept of causation is more objective than the colour concepts.

Notice that the claim that causation is more objective than colour—that all possible agents would possess a concept of causation with the same extension—is incompatible with the Menzies–McGrath model. According to the Menzies–McGrath model, causation is semantically sensitive to context, because the events we pick out as causes are relative to the normal course of evolution of a particular system. Thus, if the Menzies–McGrath model is right, the extension of the causal concept differs among *actual* human agents. This provides reason to think that the concept of causation is actually *less* objective than colour concepts, in certain respects. This point will be important (and discussed in much more detail) in the next chapter. For now, it is enough to note that Menzies and Price's claim that the term 'cause' has the same extension for all possible agents is not compatible with the Menzies–McGrath model. According to the Menzies–McGrath model, causation is more anthropocentric than Menzies and Price are willing to countenance.

3.2.4 The problem of unmanipulable causes

The second major objection to agency theories is the problem of unmanipulable causes. This objection claims that agency theories fail to account for causes that are not *in principle* manipulable by humans. The example Menzies and Price use to illustrate this objection is 'the claim that the 1989 San Francisco earthquake was caused by friction

between continental plates.'[121] To account for this kind of situation, Menzies and Price say:

> [W]e 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 ... The agency account can be weakened to allow for the application of this principle ... 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.[122]

Thus, Menzies and Price's solution to the problem of unmanipulable causes is to revert to a slightly weaker version of the agency theory. They are not saying that a cause, *A*, is *always* an event that an agent could have used to bring about an end, *B*; rather, causes are *either* events that an agent could have brought about as a means to a particular end, *or* events that have something in common with pairs of events that support a means–end relation.

Unfortunately, Menzies and Price do not provide much elaboration on the 'basic intrinsic features' that situations involving manipulable and unmanipulable causes have in common, so it is not clear what these features are intended to be, in metaphysical

[121] Menzies and Price, "Causation as a Secondary Quality": 195.
[122] Menzies and Price, "Causation as a Secondary Quality": 197. (Italics in the original.)

terms. Perhaps, in fact, Menzies and Price intend to be neutral on this point. However, if their account is intended (or interpreted) as an account of the metaphysics of causation, it is important to consider the possible candidates for these features.

In later papers, Woodward and Price have both discussed these 'basic intrinsic features', and both agree that there are two possibilities.[123] The first possibility (and the one that Menzies and Price themselves intend) is that the basic features are non-causal. That is, we extend the concept of causation from manipulable to unmanipulable causes by noticing some non-causal resemblance between the two situations. This possibility is best discussed with reference to another passage in Menzies and Price's original paper:

> 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.[124]

The idea is that the model created by seismologists resembles the actual occurrence of the earthquake in virtue of certain non-causal features that the two situations share, and it is because of this resemblance that the concept of causation can be extended from the model to the earthquake itself.

However, Woodward points out a problem with this suggestion:

---

[123] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 124-26. Price, "Causation, Intervention and Agency—Woodward on Menzies and Price": §4.
[124] Menzies and Price, "Causation as a Secondary Quality": 197-98.

> It is well-known that small-scale models and simulations of naturally occurring phenomena that superficially resemble or mimic those phenomena may nonetheless fail to capture their causally relevant features because, for example, the models fail to "scale up"—because causal processes that are not represented in the model become quite important at the length scales that characterise the naturally occurring phenomena. Thus, when we ask what it is for a model of simulation that contains manipulable causes to "resemble" phenomena involving unmanipulable causes, the relevant notion of resemblance seems to require that the same *causal* processes are operative in both.[125]

Woodward's point is that for a model to successfully represent a situation involving causal processes, the resemblance between the model and the actual situation must be *causal*—the model must have the same (or a very similar) *causal structure* as the situation in question. If Woodward is right, then it appears that the basic features that manipulable and unmanipulable causes have in common are causal, rather than non-causal.

Consider two different models of the movement of continental plates at the time of an earthquake: one made out of plastic, and one made out of jelly. The former (manipulable) model may well resemble the causal structure of the actual (unmanipulable) situation at the time of the earthquake, while that latter will almost certainly not. However, one reason for the difference in performance of the two models is just that the rigidity of the continental plates is a causally relevant factor. There are many other features in virtue of which models of the continental plates could differ (e.g. the colour of the model) which are simply not causally relevant. And the only way of distinguishing between relevant and irrelevant factors is to determine whether manipulating different models has the *effect* observed in the naturally occurring

[125] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 125. (Italics in the original.)

situation that the model is designed to resemble. That is, the relevant factors are those that resemble part of the *causal* structure of the situation being modelled. Thus, I think Woodward is right to conclude that the resemblance between the model and the actual situation requires that the relevant properties are causal. This rules out Menzies and Price's suggestion that the seismologist's model resembles the earthquake because of non-causal factors, and thus that the extension from manipulable to unmanipulable causes involves non-causal features that the two situations have in common.

This leaves us with the second possibility—that the 'basic feature' that manipulable and unmanipulable causes have in common is itself causal. Woodward argues that if all token causes share a common causal feature, then 'there is a certain kind of relationship with intrinsic features that we exploit or make use of when we bring about *B* by bringing about *A*.'[126] That is, to accept this possibility is to revert to a traditional realist (or objectivist) account of causation.

According to Menzies and Price's interpretation of the term 'secondary quality', the difference between primary and secondary qualities is that reasonable accounts of the latter involve reference to human capacities, whereas accounts of the former do not. In other words, causation is a secondary quality if and only if the extension of the term 'cause' is determined by factors that are at least partially mind-dependent. However, if all instances of causation had an objective feature in common, this objective feature would determine the extension of the term 'cause'. That is, to claim that all causal relations have an objective, intrinsic feature in common is just to give a realist account of a *primary*, rather than a secondary, quality. Thus, Woodward concludes that Menzies and Price's agency theory collapses into an objectivist account of causation—causation is, after all, independent of human experience and capacities.

[126] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 125.

But Woodward moves too fast in drawing the above conclusion. He is assuming that, if unmanipulable causes resemble manipulable causes in virtue of some causal feature that situations involving the two types of causes share, this must be because there is a particular relation in the world—a mind-independent relation—that we pick out in both cases. This mind-independent relation would of course just be the causal relation, in which case the metaphysics of causation is objectivist (and causation is a primary quality). However, there is another possibility: it could be that what all instances of causation have in common is not that they are instantiations of a particular mind-independent relation, but that they fill the same role within a certain kind of system. For example, consider a seismologist's model of the San Francisco earthquake, a mechanic's model of the engine of a Honda Civic, and a psychologist's model of drug addiction. According to the Menzies–McGrath account, what these three models have in common is that they represent part of the world as a *kind of system* with a normal course of evolution. Deviant token causes are then *interventions* in these systems. That is, the feature that all deviant token causes have in common is a *relational* property, between the cause itself and a particular kind of system.

If the models used to represent these kinds of systems are *accurate*, these models can be used to predict the effects of interventions, even in the case that a particular system is not manipulable by humans. To ensure this accuracy, the kinds of systems that can be used in making deviant token causal judgements must be constrained by structure of the mind-independent world. However, the relevant kinds of systems must also be able to be represented by humans. Thus, the structure of these systems is also constrained by human cognitive capacities.

Thus, the possibility that Woodward overlooks is the metaphysical picture suggested by the Menzies–McGrath model. That is, that what manipulable and unmanipulable causes

have in common is that they play the same role relative to kinds of systems that are both instantiated in the world, and represented by human cognitive structures. The Menzies–McGrath model is therefore an alternative to the objectivist metaphysics of causation that Woodward envisages. Defending this alternative will require providing far more details regarding both how the relevant systems are picked out, and the structure of these systems—tasks that will be undertaken in Chapters 4 and 5. Now, I will leave Menzies and Price's agency theory, and evaluate a second agency, or manipulability, account of causation: Woodward's interventionism.

3.3 Woodward's interventionism

Woodward's interventionism has been influential in the last few years, amongst psychologists as well as philosophers.[127] His approach has its roots in agency accounts of causation—like Menzies and Price, he argues that our ability to reason causally is grounded in our ability to intervene in the world as agents. However, by discarding the idea that causes are defined in relation to human actions and claiming, instead, that causes are values of variables that are *manipulable* (or open to intervention) in the right way, Woodward develops the agency theory in the direction of increasing objectivity. According to Woodward, free human actions are just one example of an intervention. One way to understand Woodward's account of causation is thus to see his interventionism as a development of Menzies and Price's agency theory, in which one significant improvement is Woodward's much more detailed account of the notion of an *intervention*.

The precise details of Woodward's account of causation are not important here. Roughly, however, he claims that:

[127] To get an idea of the influence Woodward's account of causation has had in psychology, see the papers in Gopnik and Schulz, *Causal Learning: Psychology, Philosophy and Computation*.

> $X = x$ *causes* … $Y = y$ if and only if (i) $X = x$ and $Y = y$ are the actual values of these variables; and (ii) there are values of the variables X and Y—say, x' and y'—such that if an intervention were to change the value of the X variable from x' to x, [the] value of the Y variable would change from y' to y.[128]

These truth conditions can be explicated by way of an example. Consider the plant-watering example, first introduced in Chapter 1:[129]

> I go away for a few weeks, and my housemate, Tim, promises to water my geranium while I'm gone. He forgets, and, as a result, the geranium dies.

In this example, the actual values of the variables involved are: x = Tim does not water my plant, and y = the plant dies. According to Woodward, Tim's failure to water my geranium is a cause of the geranium's death because there are alternative values of these variables—x' = Tim does water my plant, and y' = the plant lives, such that if an intervention were to change the value of X from x' to x, the value of Y would change from y' to y. From now on, I will refer to counterfactuals with this form as 'interventionist counterfactuals'.

Woodward's account of actual causation (like the Menzies–McGrath model) relies heavily on the notion of an *intervention*. Woodward provides a technical definition of this notion, which there is not space to include here.[130] However, as discussed in §1.3, an intervention in a variable can be understood as a manipulation of that variable from

---

[128] The above truth conditions are a simplified version of the truth conditions Woodward provides for token (or actual) causation, given by Peter Menzies: Menzies, "Platitudes and Counterexamples": 357-58. (Italics in the original.) For the detailed version of Woodward's position, see Woodward, *Making Things Happen: A Theory of Causal Explanation*: §2.7.

[129] Woodward does not discuss this example, but another case with the same structure. In Woodward's example, a doctor who neglects to treat his patient with antibiotics is considered the cause of the patient's death. However, another person, who lives a great distance away, has no connection to the patient and is not a doctor, is not considered the cause of the death, despite the interventionist counterfactuals associated with both omissions having the same truth values. Woodward, *Making Things Happen: A Theory of Causal Explanation*: 87-88.

[130] See Woodward, *Making Things Happen: A Theory of Causal Explanation*: §3.1.

outside the system in question. More specifically, an intervention in a variable, *X*, directly fixes the value of *X* to some value, *x*, and does not affect the values of any other variables in the system except through its influence on *X*. Importantly, the definition of 'intervention' does not involve reference to human action, so natural (i.e. non-human) interventions (e.g. earthquakes) are on a par with human actions, and are thus equally valid causes.

Note that the rough version of Woodward's truth conditions for token causation provided above are very similar to those of both Menzies' account of token causation and the Menzies–McGrath account of deviant causal claims, which are discussed in detail in Chapter 1. The reason for this similarity is that Menzies formulates his account of causation in terms Woodward's account of actual causation (at least in the version of Menzies' account that I have relied on most heavily in formulating the truth conditions of the Menzies–McGrath model, which is given in 'Platitudes and Counterexamples').[131] However, Menzies modifies Woodward's account of token causation to take into account the insight that causes and effects are relative to the normal course of evolution of a system.

Although Woodward's account of causation can be seen as a development of Menzies and Price's agency account, Woodward takes pains to distance himself from the subjective conclusions of Menzies and Price's agency theory (i.e. from the claim that causation is a secondary quality), by clearly stating that he takes the relationship between agency and causation to be *epistemological*, rather than metaphysical:

> [C]onsiderations having to do with agency and manipulability help to explain why we developed a notion of causality having the features it does[,] … play a

[131] Menzies, "Platitudes and Counterexamples": 360.

> heuristic role in helping to characterise the meaning of causal claims, and have considerable epistemic relevance when we come to test causal claims.[132]

In this passage, Woodward points out the epistemic links between agency and causation that are emphasised by all agency and manipulability accounts of causation. First, the notion of acting on, or manipulating, a system (i.e. wiggling the value of a variable) is useful as a means of conceptualising causation, and, in particular, the technical notion of an *intervention*. This is because interventions resemble free human actions, in that it is a central feature of both interventions and free human actions that they act on a system from *outside* (i.e. in both cases, a variable is set to a particular value, independently of the other variables in a system). Second, the ability to intervene in the world gives us a means to test causal claims.

However, when it comes to the *metaphysic*s of causation, Woodward is explicitly objectivist. He goes on to claim that: 'agency is not in any way "constitutive" of causality'.[133]

> [Q]uite independently of our experience or perspective as agents, there is a certain kind of relationship with intrinsic features that we exploit or make use of when we bring about *B* by bringing about *A*. Moreover, because this relationship is intrinsic and can exist independently of anyone's experience of agency, it can also be present even when *A* is not in fact manipulable by humans.[134]

Thus, unlike Menzies and Price, Woodward explicitly claims that the causal relation is objective and mind-independent. In other words, causation is a primary quality, both in the sense that the causal relation is constitutionally mind-independent, and in the sense that the extension of the term 'cause' is fixed by mind-independent factors.

---

[132] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 125-26.
[133] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 126.
[134] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 125.

3.3.1 Serious possibilities

Woodward does, however, acknowledge that there is one feature of our causal reasoning that is less than completely objective. To illustrate this feature, let us return to the claim that Tim's failure to water my geranium caused the geranium's death. Recall that in §1.2.2, I compared Tim's failure to water the geranium with Barack Obama's failure to do the same. Although the associated counterfactuals (i.e. if Tim/Obama had watered the plant, it would have survived) are both true, we consider Tim's omission to be the cause of the plant's death, but not Obama's. The Menzies–McGrath model can account for this type of example, because Tim's omission is a deviation from the normative standard of *promise keeping*, whereas Obama's omission is consistent with all the relevant standards.

Woodward has a different way of accounting for this kind of example, involving an appeal to the notion of a *serious possibility*. In discussing this notion, he says:

> [T]he extent to which an outcome represents a serious possibility is a matter of degree, to which a number of considerations are relevant ... the probability of an event's occurring, given the actual obtaining background conditions (or perhaps those that usually or commonly obtain in similar situations) is one relevant consideration. However, it is plainly not the only relevant consideration; considerations having to do with moral requirements, expectations, and custom may matter ... Similarly, considerations having to do with whether an outcome is controllable at all (or easily or cheaply controllable) by current technology may also matter. It seems unlikely that there is any algorithm for determining whether a possibility will be or ought to be taken seriously.[135]

[135] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 88.

For example, Tim's watering my geranium was a serious possibility: he had a responsibility to do so, because of his promise; he was in a position to do so, because he was living in the house with the plant; etc. Obama's watering my plant, on the other hand, was not a serious possibility—there is no reason for us to even entertain the possibility of Obama's flying to Perth and performing that action.

The factors that are relevant to Woodward's notion of a *serious possibility* are the same sorts of factors that I have argued (following McGrath) are relevant to judgements of what is normal, and therefore to the construction of default worlds (in the Menzies–McGrath model). This is unsurprising, because Woodward is invoking the notion of a *serious possibility* to account for the same features of causal discourse that motivate the claim that causes are deviations from the normal. However, the way Woodward integrates the notion of a *serious possibility* into his account of causation is different to the way McGrath's notion of *normal* is incorporated into the Menzies–McGrath model.

Recall that according to causal objectivism, all token causes have an objective, mind-independent feature in common. Woodward's claim is that it is the notion of an *intervention* that is objective—that is, 'intervention' refers to a mind-independent relation, instances of which are picked out by true interventionist counterfactuals. For this to be the case, the truth values of interventionist counterfactuals must not *depend* on which possibilities we take seriously, as the notion of a *serious possibility* is not completely objective. Instead, which possibilities we take seriously must just *restrict* which interventions with the right pattern of counterfactual dependency we judge to be causes of an effect. In other words, whether or not a potential cause, x, is a serious possibility, counterfactuals of the form: 'if an intervention were to change the value of the X variable from x' to x, the value of the Y variable would change from y' to y' must have objective, mind-independent truth values.

Woodward argues that *causal judgements*, as opposed to interventionist counterfactuals:

> [R]eflect both objective patterns of counterfactual dependence and which possibilities are taken seriously; they convey or summarize information about patterns of counterfactual dependence among those possibilities we are willing to take seriously. In other words, to the extent that subjectivity or interest relativity enters into causal judgments, it enters because it influences our judgments about which possibilities are to be taken seriously. However, once the set of serious possibilities is fixed, there is no further element of arbitrariness or subjectivity in our causal judgments; relative to a set of serious possibilities or alternatives, which causal claims are true or false is determined by objective patterns of counterfactual dependence. Thus, relativizing causal judgments to a set of serious possibilities (or, what I take to be the same thing, to the choice of some system of representation that reflects those possibilities) does not introduce subjectivity everywhere or indiscriminately but rather, at most, introduces it in a constrained or limited way.[136]

Woodward's account of the causal reasoning concerning the death of my geranium is therefore as follows: first, we construct a causal model of the situation, which includes only variables representing serious possibilities (e.g. the possibility of Tim's watering the geranium is included in the model, but the possibility of Obama's watering it is not). Then, in order to determine the cause of a certain effect, we consider the counterfactual dependence of the effect on those possibilities we are willing to take seriously, by considering the relevant interventionist counterfactuals. For example, we judge Tim's failure to water my geranium as a cause of its death, because if Tim *had* watered the geranium, it would have survived. (More formally, the counterfactual 'if an intervention

[136] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 90.

were to change the value of the variable representing Tim's watering of the plant from 'watering' to 'not watering', the value of the variable representing the plant's state of health would change from 'alive' to 'dead'.) Although Obama's failure to water the plant counts equally as an intervention on which the plant's death depends, we do not consider Obama's omission to be a cause of the death, because Obama's watering the plant is not a serious possibility.

Woodward is ambiguous as to whether the *truth* of causal claims depends on the notion of a *serious possibility*, or whether we are just more willing to *accept* claims in which the cause is a serious possibility. For example, he says: 'which causal claims we accept as true (or a least *readily* accept) are influenced by what we take to be a "serious possibility".'[137] For this reason, it is not clear whether Woodward thinks that Obama's omission is *literally* a cause of the geranium's death.

Notice that Woodward's response to the plant-watering example is very similar to Schaffer's schema for responsibility (introduced in §2.4.2). Both Woodward and Schaffer admit that normative considerations enter into our causal judgements, but claim, nevertheless, that the *metaphysics* of causation is entirely objective. They achieve this objectivity by dividing our causal reasoning into two steps: *first*, we select the serious possibilities (Woodward) or apply the schema of responsibility (Schaffer) and *then* we evaluate the resulting counterfactual against a mind-independent structure. However, I will now argue that interventionist counterfactuals are not objective in the sense that Woodward and Schaffer intend, because evaluating interventionist counterfactuals often requires holding normative standards fixed.

3.3.2 The subjectivity of interventionist counterfactuals

Consider the following passage from Jennan Ismael:

[137] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 118. (Italics in the original.)

> The *existence* and *strength* of influence depends on what we hold fixed and what we allow to vary. Whether A and B are connected at all and the strength of that connection ... depends on what we hold fixed. Even the most robust local connection like the connection between smoking and cancer, or even the ambient temperature and the level of mercury in a thermometer[,] will disappear if we don't hold fixed a good deal of local infrastructure. These connections are contingent on that local infrastructure and don't hold generally.[138]

If Ismael is right, the truth values of interventionist counterfactuals, and therefore of causal judgements, depend on what we hold fixed.

To examine Ismael's claim in more detail, let us consider the connection between a particular person, Steve, smoking two packs of cigarettes a day for thirty years, and his getting lung cancer. Assume that smoking *was* the cause of Steve's lung cancer. According to Woodward, this causal claim is true because an interventionist counterfactual of the form: 'If an intervention were to change the value of the *X* variable from x' to x, the value of the *Y* variable would change from y' to y' is true.

Here: *X* represents Steve's smoking habit, so:

x = Steve's smoking two packs a day for thirty years, and

x' = Steve's not smoking at all.

Y represents whether or not Steve gets lung cancer, so:

y = Steve has lung cancer

y' = Steve does not have lung cancer.

If we substitute these values of *X* and *Y* into the above counterfactual, the counterfactual comes out true. That is, in saying that smoking caused Steve's lung cancer, we are

[138] Jennan Ismael, "Causation, Free Will, and Naturalism" *Uncorrected Proof* (2012): 225.

claiming that, relative to the situation in which Steve does not smoke and does not have lung cancer, if we intervene to make him smoke two packs a day for thirty years, he will (or is likely to) get lung cancer.

However, this counterfactual is only true *because of what we hold fixed*. In judging the effect of intervening to make Steve smoke, we have to hold fixed numerous variables, including the chemical composition of both cigarettes and the Earth's atmosphere, and countless facts about human biology—the functioning of the human respiratory system, the process of cell division and replication, and so forth. As Ismael notes in the above quote, if we do not hold this 'infrastructure' fixed, the connection between Steve's smoking and his getting lung cancer will disappear.

Of course, it is reasonable to hold this infrastructure fixed. The problem for Woodward is that what we hold fixed depends on what we take to be normal. For example, we hold the chemical composition of the earth's atmosphere fixed at 78% nitrogen, 21% oxygen and 1% argon (plus small amounts of other gases) because it is normal for the atmosphere to be constituted of gases in roughly this ratio. The norm specifying the constitution of the earth's atmosphere is statistical (or descriptive) and is therefore not problematically subjective.[139] However, when evaluating interventionist counterfactuals, we often have to hold *prescriptive* norms fixed.

For example, the sentence 'If an intervention had changed the amount of RAM available from 2GB to 4GB, the computer would have kept functioning and would not have crashed' is an interventionist counterfactual, the truth of which depends on holding fixed certain *artifactual* norms (or standards) governing the construction and functioning of computers. Similarly, the truth value of the sentence 'If the team had fielded better, they would have won' depends on holding the rules of cricket fixed. Just

[139] Recall that 'subjective' is intended to denote a contrast with 'objective', where to be objective is to be independent of human concerns and values (i.e. mind-independent).

as in the smoking example, in these examples there is only a causal connection between the cause and effect if a large amount of local infrastructure is held fixed. However, much of this infrastructure consists of exactly those subjective factors that are involved in the notion of a *serious possibility*—the factors that Woodward wants to *exclude* from the truth conditions of interventionist counterfactuals.

Thus, unless we exclude large numbers of interventionist counterfactuals (and thus swathes of token causal judgements) in advance, which I have argued (in §2.3.4) that we should not do, the separation that Woodward relies on, between the subjectivity of the notion of a *serious possibility*, and the objectivity of interventionist counterfactuals, does not hold up to scrutiny. The truth values of interventionist counterfactuals are not necessarily independent of human concerns—these truth values often depend on subjective, normative factors. Thus, the truth values of causal judgements (or deviant token causal judgements, at least) are not independent of human concerns, either.

The above argument applies to Schaffer's contrastivism as well as to Woodward's interventionism, and therefore explains why the distinction Schaffer relies on between an objective semantics of 'cause' and a subjective, pragmatic 'conceptual filter' that is used to determine the contrast events does not hold up to scrutiny (as suggested in §2.4.2). The problem is that the evaluation of four-place contrastive propositions is not a completely objective matter. Thus, neither Woodward nor Schaffer succeeds in reconciling the subjectivity of aspects of causal discourse with an objective metaphysical image.

In the remainder of this chapter, I will discuss a third account of causation—Price's perspectivalism. Contrary to Woodward, Price claims that interventionist counterfactuals *do not* have objective truth conditions. He argues that neither the term

'intervention' nor 'causation' carves nature at its joints, because the direction of causation is *perspectival*.

3.4 Price's perspectivalism

In 'Causal Perspectivalism', Price defends another ancestor of the agency theory that he and Menzies present in 'Causation as a Secondary Quality'. However, unlike Woodward's, Price's modifications shift this earlier account in the direction of increasing subjectivity. In his later paper, Price explicitly denies that causation is an objective feature of the world, claiming instead that the direction of causation, at least, is a feature of our perspective as agents, and, importantly, could have been different, if we had been constituted differently.

My interest in Price's perspectivalism is not so much in whether he is right that the *direction* of causation is perspectival (I will not adjudicate on this question), but in the relationship between the claim that causation is perspectival *in some sense*, and our experience of agency. I will argue that Price's discussion actually supports a *different* kind of perspectivalism, according to which the truth values of token causal judgements are relative to a *purpose-dependent perspective*.

Price's argument for the lack of an objective direction of *causation* is closely related to arguments he makes elsewhere which conclude that there is also no objective direction of *time*.[140] The objectivity (or lack thereof) of the direction of causation is closely related to that of the direction of time. Although philosophers disagree on exactly *how* these two phenomena are related, they are typically taken to be a joint package, in that the objectivity of one entails the objectivity of the other. To see this, note that there are four ways that the directions of time and causation could be related: first, the direction of causation is a fundamental feature of reality, in which case the direction of causation

[140] Huw Price, "The Flow of Time" *The Oxford Handbook of Time*, ed. C. Callender (Oxford: Oxford University Press, 2001).

would provide the basis of the direction of time; second, the direction of time is a fundamental feature of reality, in which case the direction of time would provide the basis of the direction of causation; third, there is some other fundamental feature of reality which underpins both the direction of causation and the direction of time; or finally, there is no objective direction of either time *or* causation. Price is endorsing this last option. The idea is that the direction of both causation and time arises from the perspective we occupy as beings embedded in time—neither are features of the mind-independent world itself. There could be beings that are oppositely orientated with respect to time, and these beings would also be oppositely directed with respect to causation.

3.4.1 An analogy between 'cause' and 'foreigner'

According to Price, the perspectivalism of causation is analogous to that of the term 'foreigner'. He claims that 'foreigner' is perspectival in that:

> [F]or speakers in different circumstances (in the case of foreigners, belonging to different tribes), the concept picks out something different. *Our* use of the concept picks out *them*, and vice versa, but there's an obvious sense in which it is the same concept in both cases.[141]

From my perspective, as an Australian, someone from Germany counts a foreigner, whereas for German people, it is Australians who are foreign. That is, the term has an asymmetric extension that arises from the viewpoints of different groups of people, but this asymmetry is not a feature of the world itself—from a God's-eye perspective, there are no foreigners, just different groups of people.

[141] Price, "Causation, Intervention and Agency—Woodward on Menzies and Price": §3.1.1. (Italics in the original.)

For Price, then, the term ‘perspectival’ means something like ‘context sensitive’. His claim is that, like the term ‘foreigner’, ‘cause’ is also perspectival, because the temporal asymmetry of causal discourse (i.e. the difference between causes and effects, and the fact that causes typically precede their effects) arises from the human perspective, and in particular, the fact that we are embedded in time. His claim is that for beings *differently* orientated with respect time, the same concept of causation would have a different extension.

The difference between the perspectivalism of the terms ‘foreigner’ and ‘cause’, according to Price, is that the perspective that gives rise to our causal discourse is shared by all humans—it is *intersubjective*—and therefore far harder to discern. We cannot just go on holiday, and observe the term ‘cause’ being used differently by people in different parts of the world. (Notice that Price’s application of the term ‘context sensitive’ to the concept of causation involves using this term differently from the way it is normally used in philosophy. We usually say a concept is context sensitive if it has a different extension amongst *actual* speakers. Price, on the other hand, is extending the use of the term ‘context sensitivity’ so that it applies to differences in extension between humans and merely *possible* beings.)

In claiming that causation is perspectival, Price has explicitly altered his position relative to the agency theory he and Menzies defend in ‘Causation as a Secondary Quality’. In this earlier paper, Menzies and Price respond to the objection that agency theories render causation excessively anthropocentric by claiming that, although the concept of causation is anthropocentric to the extent that possession of the concept of agency is a necessary condition of possession of the concept of causation, *any* degree of agency would be enough to ground a concept of causation with the same extension as our actual causal concept (see §3.2.3). This earlier theory does not allow for the

possibility of the time-reversed agents that Price describes in 'Causal Perspectives', which pick out *different* causal relations to us (or at least causal relations with the opposite directionality). In the next chapter, I will argue that Price should go even further, and allow that the extension of 'cause' (in deviant token causal judgements) varies between *actual* human agents.

Importantly (and unsurprisingly, given that he endorses an agency theory of causation) Price argues that the perspective from which we make causal judgements arises from our experience as *agents*. In particular, our sense of agency includes the idea that our actions are directed towards the future—we are able to bring about effects in the future, but not in the past. He claims that it is a contingent fact about our epistemic access to the world that from our internal, subjective perspective, the past appears to be fixed, whereas the future appears to be open. The notion of intervening in (or manipulating) the world thus inherits this temporal bias.

If Price is right, interventionist counterfactuals do not have objective truth conditions, at least with respect to the direction of causation. It follows that the extension of the term 'cause' is not completely mind-independent either. Consider any two events, c and e, linked by a causal relation. According to Price, the reason it is true that c is a cause of e, but not true that e is a cause of c, is due to features of us, and our perspective as agents embedded in time, rather than features of the external world itself. This is what Price means when he says that the direction of causation is perspectival.

Thus, contrary to Woodward, Price holds that neither 'causation', nor 'intervention', is objective—these terms do not refer to completely mind-independent relations. Rather, the relations picked out by both concepts reflect features of us, and, in particular, of our epistemic position as agents.

In its (implicit) rejection of the claim that interventionist counterfactuals have objective (or mind-independent) truth conditions, Price's perspectivalism is thus similar to the alternative metaphysical picture that will be defended in this thesis. According to both Price's perspectivalism and the Menzies–McGrath model, causation is not a perfectly natural, objective relation.

3.4.2 Price's epistemology of agency

To support the claim that causation is perspectival, Price formulates 'an abstract characterisation of the structure, or functional architecture, of deliberation in general—what is essential to anything that deserves to count as an agent.'[142] That is, he outlines a template of deliberation—the cognitive structure that is necessary for the possession of *any* kind of agency—and then distinguishes the features of this general template from the contingent features of *human* deliberation and agency.

Price points out that in order for any particular instance of deliberation to be possible (and therefore for agency to be possible in general), there must be a number of alternatives that the agent takes herself to be capable of bringing about, and therefore able to choose between. These alternatives form a class of propositions, which he calls OPTIONS. They are 'the propositions the agent takes herself to have the option of "deciding to make true".'[143] OPTIONS can be divided into DIRECT (those alternatives the agent takes herself to be capable of bringing about immediately), and INDIRECT (the propositions she takes herself to be capable of bringing about via one of the alternatives in DIRECT OPTIONS).

All other propositions fall into another class, which Price calls FIXTURES. The propositions in FIXTURES are those that the agent considers to be unalterable, and thus matters of fact. FIXTURES can be further categorised into two subsets: KNOWABLES,

[142] Price, "Causal Perspectivalism": 274.
[143] Price, "Causal Perspectivalism": 275.

which consists of all the matters of fact the agent takes to be knowable, in principle, at the time of deliberation, and KNOWNS, which contains all the matters of fact that the agent takes herself to know at that time. (KNOWNS is thus a subset of KNOWABLES.) The point Price is making is that the possibility of deliberation requires not just that there are alternatives available for the agent to choose between, but that there are also a body of facts (or, at least, propositions that that agent takes to be settled), on which to base the decision. These two categories seem to be mutually exclusive: the possibility of bringing about a particular proposition by acting is not consistent with that proposition being known (or knowable) at the time of the action. This explains why KNOWNS and KNOWABLES are subsets of FIXTURES, and separate from OPTIONS.

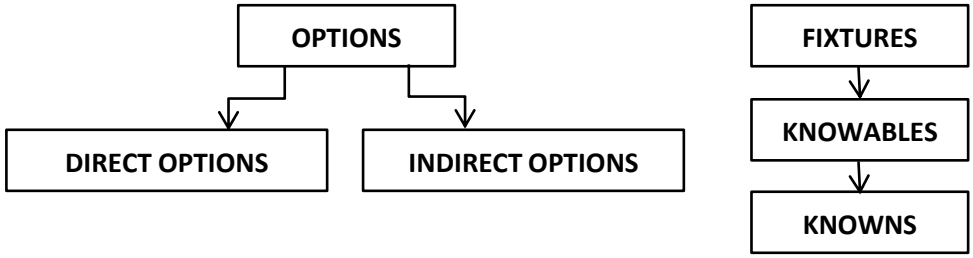

Figure 3.1: Price's architecture of deliberation

The structure of Price's architecture of deliberation is summarised in Figure 3.1. Price intends this structure as a general 'epistemic template', common to all (possible) agents. He notes that '[i]n terms of this template, acting, or intervening, is a matter of fixing something not already fixed—of moving something from OPTIONS to FIXTURES.'[144] This feature of the architecture of deliberation—that actions and interventions have a particular epistemic status—is what marks this structure as pertaining to *agents*. It is what differentiates Price's template from the epistemology of classical empiricism, in which knowledge of the world is gained only through passive *observation*.

[144] Price, "Causal Perspectivalism": 276.

Although a distinction between propositions in OPTIONS and FIXTURES is necessary for the possibility of *any* agency at all, Price argues that it is *not* an essential feature of agency that the division of propositions into these categories falls largely along temporal lines—that propositions about the past (from our perspective) generally fall into FIXTURES, whereas the members of OPTIONS are typically propositions about the future.[145] Price imagines the deliberation of a god who exists outside of the universe, and, in particular, outside time, who is capable of intervening at *any* point in spacetime. He argues that the structure of this god's deliberations would not have the temporal bias that human deliberation does.[146] If Price is correct, it is a contingent fact about humans, and the epistemic access we have to the world, that we (typically) take the past to be fixed, and the future to be open.

As already mentioned, I do not intend to adjudicate on Price's claim that the direction of causation is perspectival. Rather, I will use Price's arguments, including, in particular, his discussion of the architecture of deliberation, to support the claim that causal judgements are made from a context-dependent perspective. First, however, I will discuss two other respects in which Price points out that the concept of agency (and therefore both the notion of an *intervention*, and the extension of the term 'cause') is dependent on features of us, rather than the mind-independent world alone. These both involve the claim that agency requires a degree of *ignorance*.

First, we have seen that it does not seem to be possible for an agent to consider the same proposition to be both known (or knowable), and also able to be made true by acting. (This is why the classes KNOWNS and KNOWABLES are both subsets of FIXTURES.) Consequently, the propositions in OPTIONS are *not* either known or knowable, as far as the agent is concerned. Thus, the possibility of deliberation entails that agents are

[145] Note that in any particular deliberation, many propositions about the future are considered fixed, for the purposes of deliberation. For more on this, see §4.6.
[146] Price, "Causal Perspectivalism": 276.

epistemologically limited.[147] That is, the perspective from which agents deliberate must be a *partial* perspective. Deliberation would not be possible for a God with complete knowledge of the entire universe.

Second, because the agent must take herself to be capable of choosing which member of OPTIONS to make true, she must see herself as *external* to the system on which she is contemplating intervening, and her action as uncaused by anything except her own choice. This is not to say that the agent's actions are, in fact, causally independent of all other factors, just that the possibility of agency requires that, *from the agent's perspectiv*e, she is the cause of her own actions. As Price points out, this constraint on agency also involves '*ignorance* on the agent's part—roughly, ignorance of the causes of her own actions.'[148] Therefore, the possibility of agency also requires that a being possesses a perspective that is partial, in that the causes of the being's own actions are external to the system on which she is considering intervening.

These two constraints on the epistemology of agency provide further reason for doubting the veracity of the natural network model of causation, by revealing a tension between the picture of causation accepted by those who endorse the natural network model, and the picture painted by interventionist accounts of causation. Once again, close examination of the concepts of intervention and causation, and their link to agency, reveals that these concepts are not as objective or mind-independent, as they may first appear. Causal judgements are not made (or made true) from a God's-eye perspective, but from a perspective that is partial—by beings with a certain amount of ignorance, as well as a degree of knowledge.

[147] Price, "Causal Perspectivalism": 280.
[148] Price, "Causal Perspectivalism": 282. (Italics in the original.)

3.4.3 Towards a more comprehensive epistemology of agency

Although Price focuses on the contingency of the *direction* of causation in his discussion of the architecture of deliberation, it is important to note that this is not the only type of contingency involved in deliberation. In particular, the categorisation of propositions into OPTIONS and FIXTURES depends on contextual, as well as temporal, considerations. Price himself says

> As we have just seen, what is accessible to us cannot be something we can take ourselves to control. Yet ... there is much flexibility: something knowable and hence in FIXTURES under some circumstances, may nevertheless be regarded as controllable and hence in OPTIONS in other circumstances.
>
> In practice, for us, this seems especially true of states of affairs in the future. We plan under certain assumptions about what the future will be like, which we take as KNOWNS—for example, normally, that the sun will rise tomorrow. But this seems to be very context-sensitive: if we want to consider an action that involves eliminating the sun, we won't take the fact that it will rise tomorrow as a given—its rising will be in OPTIONS, not FIXTURES.[149]

That is, events in the future can move back and forth between OPTIONS and FIXTURES, depending on the context. In particular, whether a particular future event is part of OPTIONS or FIXTURES will depend on the end an agent is trying to achieve—that is, on the agent's purpose.[150]

Additionally, if we assume that an agent's perspective at any point in time involves a certain *representation* of the world, there must also be *cognitive* preconditions of agency. Any particular being will be able to handle representations involving a certain

[149] Price, "Causal Perspectivalism": 276.
[150] If backwards causation is a conceptual possibility (as it seems to be) then propositions about past events can also (sometimes) be included in OPTIONS.

degree of complexity. This will not affect the categorisation of propositions into OPTIONS and FIXTURES, but will limit the propositions within each of these classes that are actively considered. These two further features of our perspective as agents shed important light on respects (other than the direction of causation) in which the concept of causation is perspectival. This is because these features of our perspective as agents restrict the way that we *model* causal systems. Or so I will argue in the next chapter.

Chapter 4: Causal Perspectives

In the previous chapter, I discussed the metaphysics of three agency, or manipulability, accounts of causation: Menzies and Price's agency theory, Woodward's interventionism, and Price's perspectivalism. I argued, contrary to Woodward, that the truth values of interventionist counterfactuals are not completely objective. Since interventionist counterfactuals feature in the truth conditions of these accounts of causation (and the Menzies–McGrath model), this entails that the truth values of token causal judgements are not completely objective, either.

Throughout this thesis, I have argued that the perspective from which we make and evaluate causal judgements is a partial perspective, from which the world is represented in terms of open systems with a normal course of evolution. The aim of this chapter is to elaborate on this claim in more detail, by telling a story about the way that perspectives contribute to causal judgements—that is, to engage in what Price calls 'philosophical anthropology'. This will involve considering three perspectives from which we could potentially make and evaluate causal judgements, which I will call the 'intersubjective perspective', the 'purpose-dependent perspective', and the 'control-based perspective'.

The first of these perspectives—the *intersubjective perspective*—is shared by all humans, and is determined by our perceptual and cognitive capacities. The second, *purpose-dependent perspective*, is a representation of part of the world that depends on the interests, or purposes, of the person (or group) doing the representing. Finally, the *control-based perspective* represents the alternative courses of action available to a particular individual. The control-based perspective is therefore the perspective that we use to deliberate (and the point of view that Price describes in his epistemology of agency).

According to Price, the perspectivalism of causal judgements is limited to the intersubjective perspective. However, I will argue that, if the Menzies–McGrath model is correct, deviant token causal judgements are made from a *purpose-dependent* perspective. Each purpose-dependent perspective includes a set of default worlds, as well as a set of possible interventions in these default worlds.

In the second half of the chapter, I develop a more detailed version of Price's architecture of deliberation (introduced in §3.4.2), and use the resulting structure to construct a template of the perspective from which we make and evaluate causal judgements—that is, to arrive at a more detailed characterisation of the purpose-dependent perspective. Although the architecture of deliberation is a description of the control-based perspective, rather than the purpose-dependent perspective (and thus not a description of the perspective from which we make causal judgements), it is possible to use the architecture of deliberation to characterise the structure of the purpose-dependent perspective, because the control-based perspective is still a *causal* perspective. In fact, the control-based perspective is a subset of the purpose-dependent perspective, which is subject to the constraint that the possible interventions are all actions that could be performed by the person doing the deliberating.

4.1 What do we mean by 'perspective'?

The term 'perspective' is often used in analytic philosophy, but rarely defined—we are supposed to rely on the vague, metaphorical interpretation of perspectives as 'points of view'. This metaphor can, indeed, be understood intuitively. However, because certain kinds of perspectives play a crucial role in the Menzies–McGrath model, it is necessary to consider the relevant notion of a *perspective* in some detail. The aim of this section is therefore to give a reportive definition of a certain sense of the term 'perspective', and to discuss the features of perspectives that are relevant to causal reasoning.

The sense of 'perspective' defined here is not the sense we apply when we take 'perspective' to be roughly synonymous with 'opinion' (and I am certainly not claiming that all perspectives are equally viable). Rather, in the sense of 'perspective' that is relevant to this thesis, each perspective consists of a representation of the world, made from a particular epistemic position. Roughly, the term describes the way an individual represents the world, at a particular time.

The notion of a *perspective* (or a point of view) entails a distinction—not necessarily sharp—between a viewer (or viewers), and something viewed. (More precisely, since not all perspectives are visual, this distinction is between something that is represented, and someone who does the representing.) For there to be a distinction between viewed and viewer, there must be a boundary between what is viewed and who is doing the viewing. Importantly, because perspectives are *partial*, in that they selectively represent certain parts of the world, this boundary can be drawn in a variety of different places. This partiality affects the resulting perspectives in two ways: first, it results in perspectives with differing *content*, and second, it results in a difference in *which individuals hold* (or have access to) a particular perspective.

4.1.1 Variation in the content of perspectives

Perspectives differ in content because they selectively represent some *part* of the world. There are a number of different ways in which the parts of the world that are represented can be constrained. To take a very basic example, the view from any particular spatiotemporal location is partial, because it includes a representation of only the segment of the world that can be seen from that location. That is, the information included in a particular perspective can be limited by the viewer's access to the world in a spatiotemporal, or geographic/historic sense.

Different *sensory apparatus* (and different scientific instruments) also give rise to perspectives with different content, because they provide access to the world via a different type, or range, of signal. For example, the perspective arising from the human visual system is partial in that our visual system is responsive to a particular *type* of input—in this case, electromagnetic radiation within a certain frequency range.[151] We cannot see radio waves, for example, because they have frequencies outside the range that our eyes are sensitive to. Bees, on the other hand, have access to a different perspective because they have a sensory apparatus that responds to electro-magnetic radiation in the ultraviolet range. In general, the information included in a particular perspective is limited by the type of data, or input, used to generate it.

There are further, less obvious ways in which the partiality of perspectives can generate different content. For example, when we try to imagine the world from the perspective of another person—to put ourselves in someone else's shoes—we do not just consider their capacity to detect certain stimuli, but also other factors, including their desires and purposes. This is because our desires and purposes can influence which aspects of the world we pick out as being *salient*. For example, a chef might walk into a kitchen and see a top of the range oven, a knife that is kept razor sharp, and a fresh barramundi waiting to be filleted. A friend, however, might walk into the same kitchen, note that their host is in a good mood and that a fish is in the early stages of preparation, and look forward to the prospect of a good meal and some interesting conversation. Despite the fact that these two people occupy the same location, and are also in possession of visual systems that are sensitive to the same range of stimuli—that is, although they have the same basic *capacity* to see the kitchen—they represent the kitchen differently. They have access to perspectives with different content, because of their different interests

[151] Giere discusses the effect that the type of input has on the partiality of perspectives in Ronald N. Giere, *Scientific Perspectivism* (Chicago: University of Chicago Press, 2006): 35-36.

and purposes.[152] This final way in which the content of perspectives can differ stems not from our sensory capacities, but from our cognitive capacities. We are forced to simplify the world because we are only capable of conceptualising representations with a certain degree of complexity. As agents, capable of intervening in the world, which features we deem to be salient is often determined by our purpose.

Clearly, perspectives can represent the world with different degrees of accuracy, or objectivity. It is possible to hold a perspective that completely misrepresents the world—a hallucination, for example. However, even if a perspective does succeed in representing reality, the partiality of perspectives ensures that it is possible for someone else to hold a different perspective (i.e. a perspective with different content) that is equally veridical—consider the chef and the friend in the preceding paragraph.

4.1.2 Variation in the individuals that have access to a perspective

As well as varying in content, perspectives also differ along a second dimension, which concerns who holds them, or has access to them. For example, we can talk about what the world looks like from a human (or intersubjective) perspective; from a dolphin's perspective; from a geologist's perspective; or from the perspective of an individual person. It is important to note that to claim that some perspectives can be held by more than one individual is not to claim that these perspectives are held in a collective mind, or that they somehow exist outside of any minds at all. Rather, two individuals who have access to the same perspective each *individually* possess different copies of the same (or similar) representation. (That is, these two individuals possess different *tokens* of the same *type*.)

Finally, talk of perspectives within a realist tradition makes the assumption that perspectives are all points of view *on the same world*. The idea is that there is one

[152] The content of perspectives is probably also influenced by memory and personal history.

external reality, which can be accessed in different ways, by individuals (or groups) with different capacities and interests, giving rise to different representations. To avoid begging the question against anti-realists, however, this presupposition is best thought of as a methodological maxim, rather than a metaphysical truth—it is an assumption that lies at the heart of our everyday understanding of the world, as well as science, and that we would not want to give up without very good reason.[153]

In summary, perspectives are representations of the world, and different perspectives have different content. This content varies both as a result of the epistemic access that a particular individual (or group) has to the world, as well as that individual's (or group's) interests and purposes. Perspectives also vary in terms of who holds them, or has access to them.

In the following sections, I will address the question: in what sense is the concept of causation perspectival? That is, which perspectives are involved in causal reasoning, and what role do these perspectives play? To answer these questions, I will consider three different perspectives that seem to be related to the possession of agency, and thus to potentially figure in our concept of causation: the intersubjective perspective, the purpose-dependent perspective and the control-based perspective.

4.2 The intersubjective perspective

The *intersubjective perspective*, common to all humans, is appealed to in both Menzies and Price's early defence of the agency theory, and Price's later perspectivalism (discussed in §§3.2 and 3.4, respectively). For this reason, I will briefly return to these accounts of causation, to discuss the role that the intersubjective perspective could potentially play in causal reasoning.

[153] In claiming that the uniqueness of the world should be considered as a methodological maxim, I am following Giere: Giere, *Scientific Perspectivism*: 34-35.

According to Menzies and Price, causation is a secondary quality, because the concept of causation makes essential reference to our experience and capacities as agents. Although they do not explicitly make this point, it is clear that when talking about the experience and capacities of agents, Menzies and Price are referring to the *general* capacities of human agents to manipulate events in the world. Thus, on their account, the perspective that is relevant to causation is one that is common to all normal people, and obtains because of our biological and psychological constitution. In other words, the perspective is intersubjective. (To be more precise, because Menzies and Price claim that all *possible* agents would have the same concept of causation, their agency account of causation is actually broader and more objective than the intersubjective perspective. For more discussion, see §3.2.3.)

Price's discussion of the perspectivalism inherent in the direction of causation relies even more obviously on an appeal to an intersubjective perspective. Whether or not he is right in claiming that the direction of causation is a feature of our perspective, embedded as we are in time, it is clear that all humans share the perspective Price refers to—we all take the direction of causation (and of time) to point in the same direction. Again, the relevant perspective is intersubjective.

Notice that if causal judgements are made from a purely intersubjective perspective, they should not have context dependent truth values. In this case, to reach a context according to which there was any variation in the truth value of a particular causal sentence, you would have to venture beyond the perspective of any actual human agent (to Price's imagined time-reversed agent, for example). According to the Menzies–McGrath model, however, the truth value of deviant token causal judgements depends on the system being represented, which is a context sensitive matter. Recall the example of the fire breaking out in the lab (§1.2.1). If it is true that oxygen is a cause of the fire

in situations in which oxygen is not normally present, but is *not* a cause of the fire in situations in which oxygen *is* normally present, the truth values of causal judgements are relative to contextually specified parameters that vary between actual human beings.

Thus, Price's claim that the extension of the concept is partially determined by the intersubjective perspective we share as humans—that is, his claim that if we, as a species, had been different in certain respects, we would have correctly picked out different causal relations—does not go far enough. Price's causal perspectivalism does not account for the actual context sensitivity of causal discourse.

The idea that our conception of ourselves as agents is central to causation (and conceptually prior to causation) suggests another way that perspectives might enter into the causal concept, because the notion of agency itself is linked to that of *purpose*. In §4.1.1, I showed that our individual interests and purposes can influence the way we represent the world, and therefore our perspective at any particular point in time. Perhaps these less general, purpose-dependent representations play an important role in our causal reasoning.

4.3 The purpose-dependent perspective

The best way to illustrate the *purpose-dependent perspective* is by way of example. Consider an ecologist who is studying a particular section of Australian desert—a particular ecosystem. The way she represents this system will largely be determined by objective facts about the plants and animals that live within a certain geographical region, as well as other factors like the climate, soil and atmospheric conditions. To some extent—perhaps when determining the geographical boundaries of the region—she will have to make arbitrary decisions about which variables to include. But her representation will largely be determined by the actual structure of the world—there are

objective reasons why she will not include a species of tree that is endemic to Panama in her model, or rainfall data from London, or the rules of chess.

However, another part of the ecologist's choice regarding which variables to include in her model is determined by her purpose—by what she is interested in. She will divide up the world differently from a botanist or a geologist studying the same part of Australia; and a biochemist studying the photosynthetic apparatus of one particular species of plant would include different variables again. In other words (as argued in §2.3.2), because there are multiple ways of dividing the world into systems, each spatiotemporal region instantiates more than one kind of system. The kind of system any individual picks out as being salient will depend on his interests and purposes.

Notice that these purpose-dependent perspectives are often not *individual* perspectives (although they can be). As a result of their training, ecologists share a common perspective, as do geologists and biochemists. For example, a different ecologist studying the same desert ecosystem would represent the region using (largely) the same variables. The purpose-dependent perspective is, therefore, typically a *group* perspective.

In *Causality*, Judea Pearl provides an influential interventionist account of causation, according to which causal reasoning in science (and elsewhere) relies on the construction of models.[154] In the following passage, he claims that the construction of these models requires that scientists create a boundary between the parts of the world that are included in the model, and the parts that are excluded:

> The scientist carves a piece from the universe and proclaims that piece *in* ... The rest of the universe is considered *out* ... This choice of ins and outs creates an

[154] Pearl, *Causality: Models, Reasoning, and Inference*.

> asymmetry in the way we look at things and it is this asymmetry that permits us to talk about "outside interventions" and hence about causality.[155]

Pearl's talk of 'asymmetry' is another way of claiming that the models scientists use to represent the world are *partial*—if part of the universe is excluded from a model, that model cannot represent the whole universe. Furthermore, to 'carve a piece from the universe and proclaim that piece *in*' is to represent part of the world as a system, subject to interventions. Thus, Pearl's description of scientific causal models is consistent with the Menzies–McGrath model.

Note that Pearl refers to a *choice* that the scientist makes, when deciding how to model the world. Different scientists make different choices with regard to this issue (and the same scientist makes different choices at different times), depending on their purposes. That is, the location of the line drawn between what is included in the model and what is excluded is context dependent (consider the ecologist versus the geologist discussed above).

The scientists' purposes affect not only which variables are included in a model, but also the *accuracy* of the model. For example, the model used by a toy manufacturer to design the firing system of a water pistol would not need to be nearly as accurate as the model used by a weapons manufacturer in the production of a high calibre rifle. The mechanical engineers designing these two systems would therefore base their designs on different causal models. Thus, both the parts of the world represented in any particular scientific model (i.e. the variables included), and the accuracy of that model, depend on the purposes of the scientists (or engineers) who develop and use it.[156]

---

[155] Pearl, *Causality: Models, Reasoning, and Inference*: 350. (Italics in the original.)

[156] Giere makes the point that the accuracy of scientific models depends on the purpose of the scientist doing the modelling whilst defending a perspectival account of scientific knowledge claims. Giere, *Scientific Perspectivism*: 64. Giere's scientific perspectivism is similar to the Menzies–McGrath model in that he holds that scientific judgements are relative to purpose-dependent models.

Assuming that we use the same kind of causal reasoning in everyday life as in science, Pearl's description of the causal models used in science supports the claim that the models used to make everyday causal judgements are purpose-dependent, too. And therefore that the perspective used to make everyday causal judgements is the purpose-dependent perspective described above.

If causal discourse more generally (i.e. not just in science) is based on a purpose-dependent perspective, which is a representation of part of the world as a certain kind of system, it is likely that this is because it is a psychological fact about humans that we tend to conceptualise the world in terms of systems (at least when reasoning about causes). It therefore seems likely that the asymmetry Pearl speaks of, which is essential to our concept of causation, arises from a cognitive ability that humans share, as agents. This ability involves seeing ourselves both as *part* of the world, but also as *external* to a particular system, and thus capable of changing the world around us by intervening. If this is right, the intersubjective perspective that Price alludes to in 'Causal Perspectivalism' is a necessary condition of possession of the concept of causation. However, for the reasons outlined above, this intersubjective perspective does not exhaust the perspectivalism of causal judgments—the truth of each token deviant causal claim is relative to a purpose-dependent perspective, rather than the intersubjective perspective.

The claim that causal judgements are relative to a purpose-dependent perspective is very different from the claim that causal discourse is relative to an intersubjective perspective. For one thing, if causal judgements are relative to perspectives that are determined partially by the interests of groups of agents, there is a sense in which they are significantly *less* objective than the paradigm examples of secondary qualities (e.g. colour properties), which arise from a common intersubjective perspective. It is clear

that different agents can have quite different purposes, and, indeed, the same person can (and does) have different purposes at different points of time. It is important to keep in mind, however, that purpose-dependent perspectives are by no means completely subjective (at least when some effort is made to ensure they are accurate, e.g. in science). As noted above (with regard to the point of view of the desert ecologist), the way we represent systems is constrained, to a large degree, by objective facts about the world.[157]

4.3.1 The normal course of evolution of purpose-dependent perspectives

A further complication is that the perspectives we use to make causal judgements are also influenced by mind-dependent factors *other* than our purposes. Even when different people (or groups of people) represent a causal system using the same variables, people with different interests, or backgrounds, can disagree about the normal course of evolution of that system. Consider this scenario:

> In Tasmania, unlike Western Australia, there are a lot of winding gravel roads. A tourist from Western Australia visits Tasmania, hires a car and, being unused to driving on winding gravel roads, loses control and crashes the car.

In this scenario, the purpose of a causal inquiry is to determine the cause of the accident. The kind of system picked out therefore includes variables such as the road condition, the mechanics of the car, the driver's state of mind, and the driver's experience.

However, if we were to ask the tourist what caused the crash, it is likely that he would give a different answer to that of a local. For the tourist, the condition of the road is a deviation from the actual standard, existing in Western Australia, according to which roads are sealed, straight, and flat. The tourist, then, is likely to cite the condition of the

[157] I will say more about the objectivity of the perspectives used to make deviant token causal judgements in Chapter 5.

road as a cause of his crash. For Tasmanian locals, on the other hand, winding gravel roads are normal, whereas the tourist's lack of experience is a deviation from the standard, existing in Tasmania, according to which drivers gain experience on winding gravel roads, and therefore the ability to drive safely in these conditions. For the locals, the driver's inexperience is likely to be considered the cause of the accident.

Of course, if the tourist and the local were talking to each other, it is likely that they would quickly be able to see each other's perspective, and to agree that both factors can reasonably be cited as causes, depending on which standards, or norms, are taken as relevant. The point is that even after the relevant *system* is determined by the purposes of a causal inquiry, the *normal course of evolution* of that system can still be influenced by additional, context-dependent factors, a point that will come up again in §4.6.

So far, I have discussed the intersubjective perspective we share as humans, and a more localised perspective, determined partly by the purpose, or interests, of groups of people. I will now discuss one final perspective, which is also related to the concept of causation, and which is more localised again. This perspective is based on the notion of *control*.

4.4 The control-based perspective

As agents, we each have a limited sphere of influence in the world, but we nevertheless have a desire to manipulate both the course of our own lives, and, to some extent, the world around us. When we deliberate about acting, we imagine ourselves as intervening in the world. That is, we see ourselves as a source of external intervention in a particular system. To carry out these deliberations, we make use of a *control-based perspective*. The control-based perspective is the point of view that every agent needs to be able to access in order to *be* an agent, and it is the features of this kind of perspective that Price describes in his epistemology of agency (see §3.4.2).

4.4.1 Links between the purpose-dependent and control-based perspectives

If agency and manipulability theories of causation are on the right track, our ability to reason causally is an extension (or generalisation) of our experience of agency—of making things happen in the world. This idea can now be expressed more precisely, by saying that the control-based perspective is a subset of the purpose-dependent perspective. The purpose-dependent perspective is a context sensitive representation of part of the world as a kind of system with a normal course of evolution that is responsive to external interventions. The control-based perspective also consists in a representation of part of the world as a kind of system with a normal course of evolution, but this perspective is subject to the further constraint that the only external interventions included in the representation are the actions of the individual who holds the perspective (i.e. the person who is deliberating). That is, this person sees *himself* as the source of external intervention.

Consider the example in which my car fails to start when I turn the key (introduced in §1.1). To determine the cause of the car's failure to start, we assume a purpose-dependent perspective, according to which the car is represented as a mechanical system with a normal course of evolution, and look for interventions in (or deviations from) this normal course. If I were to take the car to a mechanic, he would also represent the car as a mechanical system with a normal course of evolution, but his goal would not just be to identify the source of the problem, but also to *fix* it. When determining the cause of the car's failure to start, the mechanic would engage in the same causal inquiry that I did (although his reasoning would be far more informed). However, when deliberating about how to fix the car, the only interventions he would consider would be his own potential actions (or those of a team he is part of).

To fix the car, the mechanic needs to see *himself* as the source of external intervention in the system. There is no point in his coming up with a plan to fix the car in which he plays no role at all (not even to convince someone else to do the work), because he only has control over his own actions. Again, when deliberating—that is, when assuming the control-based perspective—the external interventions under consideration are restricted to the agent's own actions.

It is important to distinguish between the perspectives that I have called 'purpose-dependent' and 'control-based', respectively. I will now argue that by failing to make this distinction, psychologists studying causal reasoning have been led to erroneously conclude that counterfactual accounts of causation are mistaken.

4.4.2 Psychological experiments

Psychologists interested in the connection between causation and counterfactuals have carried out experiments in order to investigate this connection. In one type of experiment, psychologists present subjects with a vignette consisting of a sequence of events culminating in an unfavourable outcome, and ask subjects to 'undo reality' by completing the sentence: 'If only …'. For example, David Mandel and Darrin Lehman carried out an experiment using the following vignette:

> Mr. Jones is 47 years old, the father of three and a successful banking executive. His wife has been ill at home for several months.
>
> On the day of his accident, Mr. Jones left his office at his regular time. He occasionally left early to take care of home chores at his wife's request, but this was not necessary on that day. Mr. Jones did not drive home by his regular route. The day was exceptionally clear, so Mr. Jones decided to drive along the shore to enjoy the view.

> The accident occurred at a major intersection. The light turned yellow as Mr. Jones approached. Witnesses noted that he braked hard to stop at the crossing, although he could easily have gone through. His family recognised this as a common occurrence in Mr. Jones's driving. As he began to cross after the light changed, a truck charged into the intersection at high speed and rammed Mr. Jones's car from the left. Mr. Jones was seriously injured.
>
> It was later ascertained that the truck was driven by Mark Smith, a teenager who was under the influence of alcohol. Mark was on his way to a beach party that his friend had told him about earlier that day.[158]

When one group of participants were given this vignette, and were individually asked to imagine they were Jones and to complete the sentence 'If only ...' with Jones' thoughts, they were likely to cite factors that were under Jones' control—most often, his taking an unusual route home. Another group who were directed to answer the same question, except that they were asked to imagine that they were the other driver, Smith, were far more likely to cite Smith's drunkenness and his reckless driving as factors they would 'undo'.[159] This suggests that when we think about how we would have liked a certain course of events to have gone differently, we tend to focus on changing factors that are under our control.

A second set of participants in this study was divided into a further two groups: the members of one group were asked to imagine they were Smith, and the members of the other were asked to imagine they were Jones. Members of both groups were then

[158] David R. Mandel and Darrin R. Lehman, "Counterfactual Thinking and Ascriptions of Cause and Preventability" *Journal of Personality and Social Psychology* 71 (1996): 454. Mandel and Lehman's experiment is an example of the research paradigm that Woodward claims is used to investigate judgements of the actual cause (see §1.1.1). This study is a modification of an experiment that was originally reported in Kahneman and Miller, "Norm Theory: Comparing Reality to Its Alternatives".

[159] Mandel and Lehman, "Counterfactual Thinking and Ascriptions of Cause and Preventability": 454-55. When a third set of participants were asked to think about how the accident could have been *prevented* the corresponding two groups gave a similar distribution of answers to those of the set of subjects who were asked to complete the sentence 'If only …'

individually questioned about the *cause* of the accident. In contrast to the results of the experiment in which the participants were prompted by 'If only …', when asked what caused the accident, both groups (i.e. both those participants who were assuming Jones' perspective and those who were assuming Smith's) were more likely to cite Smith's drunk and reckless driving.

The 'If only …' construction is intended to encourage subjects to engage in counterfactual reasoning (to consider how the world could have unfolded differently). Thus, because subjects who were asked about the *cause* of the accident cited different factors from those cited by subjects asked to complete the sentence *'If only ...'*, psychologists have used the results of this study to argue that counterfactual reasoning and causal reasoning involve quite separate cognitive systems.[160] Assuming that an account of causation is intended to be psychologically realistic, this suggests that counterfactual accounts are on the wrong track.

However, these psychologists are basing their conclusion on the assumption that the kind of counterfactual reasoning induced by the construction 'If only …' is the *only* kind of counterfactual reasoning. If we give up this assumption, an alternative explanation of Mandel and Lehman's results is available: namely that, when reasoning about 'If only …', we base our judgements on a different *perspective* from the perspective used to make deviant token causal judgements. Nevertheless, both kinds of judgements require the evaluation of counterfactuals.

According to this explanation, when prompted by '*If only* ...', subjects typically thought about what the person whose perspective they were asked to assume (i.e. Smith or

[160] For example, Byrne uses the results of Mandel and Lehman's experiment to conclude that causal and counterfactual judgements rely on different psychological processes. Ruth M. J. Byrne, "Counterfactual and Causal Thoughts About Exceptional Events" *Understanding Counterfactuals, Understanding Causation*, eds. C. Hoerl et al. (Oxford: Oxford University Press, 2011). It is worth noting that Mandel himself is more cautious about interpreting the results of his experiment, and the implications of these results in terms of the connection between causal and counterfactual reasoning. See Mandel, "Mental Stimulation and the Nexus of Causal and Counterfactual Explanation".

Jones) could have done differently, to prevent the accident. That is, the participants were engaging in a form of causal reasoning in which the potential causes were restricted to the agent's own actions—they were basing their judgements about on the *control-based perspective*.

When asked about the *cause* of the accident, on the other hand, subjects typically made deviant token causal judgements, from a *purpose-dependent perspective*. In this purpose-dependent perspective, the two drivers were both part of a system governed by certain standards, including the road rules and the physical laws describing the movement of cars. Smith's reckless driving was the most significant deviation from the normal course of evolution of this system, and was therefore judged by most people to be the cause of the accident.

On the interpretation I am suggesting, the results of Mandel and Lehman's experiment entail that we do not use the control-based perspective to make deviant token causal judgements. This conclusion is, of course, unsurprising, because the events we pick out as causes are not just those that *we* could have done differently—the concept of causation extends beyond recognition of the ways that we, as individuals, can manipulate the world. For this reason, the control-based perspective is too restrictive to be used to determine which events count as causes of an effect.

Notice that if the above interpretation is correct, judgements about 'If only …' and judgements about deviant token causes both require the use of counterfactuals.[161] Both to reason about how *we* could have acted differently to prevent an event, and to determine the *cause* of an event, we must consider how a sequence of events could have unfolded differently. It is just the perspective from which these two types of judgements

[161] In the preceding sentence, I am using the term 'counterfactual' in the sense in which it is used in philosophy. Unfortunately, 'counterfactual' is often defined differently in psychology, which tends to create confusion. For an explication of the different uses of 'counterfactual' in these two fields, see Woodward, "Psychological Studies of Causal and Counterfactual Reasoning".

are made—the model of the world used to determine which aspects of reality to undo—that is different.

Thus, once we take into account the difference between the purpose-dependent and control-based perspectives, the connection between the results of experiments like Mandel and Lehman's and very general conclusions about the nature of causal reasoning starts to appear tenuous. What we *can* conclude is that causal reasoning is not based purely on a control-based perspective, and that the human ability to reason counterfactually—to consider ways in which the world might have been different—is a lot more complicated than we might initially have thought.

4.5 Back to the Menzies–McGrath model

So far in this chapter, I have considered three perspectives that could conceivably influence our causal reasoning: first, an *intersubjective* perspective arising from general human biology and psychology; second, a perspective determined by the purpose of a particular inquiry which is usually (but not always) a *group* perspective; and third, a perspective based on an *individual's* ability to manipulate the world. I have argued that the first, intersubjective, perspective is not sufficient to account for the context sensitivity of causation, and also that the final, control-based, perspective is too restrictive to be used to distinguish those events that are causes from those that are not. Rather, the perspective used to make and evaluate causal claims is the purpose-dependent perspective. I will now return to the Menzies–McGrath model, and give an explanation of how the purpose-dependent perspective is applied to deviant token causal judgements.

First, as a result of the epistemically and cognitively limited situation we find ourselves in as human agents (subject to certain biological, psychological and cognitive constraints), we tend to conceptualise the world in terms of kinds of systems with a normal course of evolution. Second, the kind of system we pick out when undertaking a

particular causal inquiry is dependent on the purpose of that inquiry. The causal inquiry thus generates a representation, or model, of a kind of system that is instantiated in the situation being investigated.

Recall that the truth conditions for deviant token causation, according to the Menzies–McGrath model, are as follows:

> $X = x$ causes $Y = y$ relative to the default values $X = x'$ and $Y = y'$ if and only if the following conditions hold:
>
> (i) the actual values of X and Y are x and y respectively; and
>
> (ii) if an intervention were to change the value of the X variable from x' to x, the value of the Y variable would change from y' to y.
>
> (where the default values (x' and y') are the values of X and Y in the default worlds).

The purpose-dependent perspective generates the default values of X and Y (x' and y'), because the content of each purpose-dependent perspective includes a set of default worlds which represent the normal course of evolution of the kind of system in question. Obviously, the actual values of X and Y (x and y) are determined by looking at the actual world.

These truth conditions, and the role of the purpose-dependent perspective in applying them, can be illustrated by returning to the example in which Tim fails to water my geranium after promising to do so while I am on holiday (§1.2.2). In this example, we are interested in determining the cause of the geranium's death. The purpose of the causal inquiry thus specifies a particular purpose-dependent perspective, which is a representation of the geranium's physiological system, including the source of the nutrients and other factors necessary to keep the plant alive. The default values of the variables in this system are the values those variables take when the system follows its

normal course of evolution (i.e. the values of the variables in the default worlds). Because Tim has promised to water the geranium, his watering the plant is the normal source of the geranium's water.[162] Thus, in the default worlds, the values of x' and y' are that Tim waters the plant, and the plant stays alive, respectively. The actual values of the variables (x and y) are that Tim does not water the plant, and it dies. The interventionist counterfactual 'If Tim had watered the geranium, it would have survived', is therefore true. This example shows that the purpose-dependent perspective can supply the information required to fill the truth conditions of the Menzies–McGrath model.

4.6 A more comprehensive epistemology of agency

In the remainder of this chapter, I will reconsider the epistemology of agency proposed by Price (introduced in §3.4.2) in the light of the above conclusions about the perspectives used to make deviant token causal judgements. My intention is to draw some connections between Price's architecture of deliberation and the Menzies–McGrath model, and to use these connections to give a more precise characterisation of the purpose-dependent perspective, including the default worlds.

As noted above (§4.4), Price's architecture of deliberation is a description of the control-based perspective. The purpose of the first half of this chapter was to show that deviant token causal claims are *not* made from the control-based perspective, but from the purpose-dependent perspective. Thus, it may seem that further investigation of Price's architecture of deliberation is unlikely to shed much light on either the Menzies–McGrath model, or deviant token causal claims. However, I also claimed that the control-based perspective is a subset of the purpose-dependent perspective, where the former perspective is subject to the constraint that the agent is, himself, the source of possible interventions in the normal course of evolution (§4.4.1). This restriction on the

[162] 'Normal' is being used here in McGrath's sense, as defined in §1.4.

control-based perspective does not affect the *structure* of the perspective, relative to the purpose-based perspective, but only constrains the source of external interventions. As I have already claimed, and will now illustrate in some detail, both perspectives consist of a representation of part of the world as a kind of system with a normal course of evolution. Illuminating the structure of the architecture of deliberation (i.e. the control-based perspective) is therefore one way of shedding light on the structure of the purpose-based perspective, as well.

4.6.1 Modifications to OPTIONS

Recall that according to Price, the epistemology of agency requires that propositions are classified into two categories: OPTIONS and FIXTURES. The former category contains those alternatives the agent takes to be open when we are making a particular decision, and the latter consists of those propositions the agent holds fixed. Price points out that to *act* is to move a proposition from OPTIONS to FIXTURES. In other words, to act is to make certain propositions true (from the agent's internal perspective). More precisely, Price divides OPTIONS into two further categories: DIRECT (those propositions the agent can make true directly), and INDIRECT (which the agent can make true—or, at least, can increase the likelihood of becoming true—via one of her DIRECT OPTIONS).

However, the philosophy of action teaches us that to act is not just to make certain propositions true, but to do so with *intention*—for some reason, with the aim of achieving some purpose.[163] Thus, OPTIONS must include another subset—ENDS—the class of propositions the agent both *wants* to bring about, and believes herself capable of bringing about.[164] Further, since it is crucial that the agent knows which members of DIRECT OPTIONS are likely to increase the likelihood of particular ENDS (if I am trying to catch a fish, it is important to know that casting a line with bait on it, which I can do

[163] For example, see G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957).
[164] The propositions in ENDS can be members of DIRECT OPTIONS, but will far more often be members of INDIRECT OPTIONS.

directly, is greatly going to increase my chances of catching something relative to just sitting on the shore) the epistemology of agency must be structured as a kind of model, or map, which charts the links between the relevant propositions in OPTIONS. The point of the model is to allow the agent to reason back from ENDS to DIRECT OPTIONS, so she can work out which action to perform.

In more technical terms, the possession of agency requires that a being is capable of representing part of the world as a causal system, in which she can intervene. In particular, the agent's model must include information about the *normal type causal relations* that hold between the members of DIRECT OPTIONS, INDIRECT OPTIONS and ENDS. (The term 'normal type cause' refers to one of the kinds of causes included in the taxonomy of causal judgements introduced in §1.1.1.)[165]

Thus, the architecture of deliberation is not just a collection of propositions, filed into different categories (i.e. OPTIONS and FIXED). Rather, these propositions are *structured*—they are ordered into networks, representing both their relationships to one another, and the type causal relations between kinds of events in the world.

4.6.2 Modifications to FIXTURES

It is also possible to provide more detail regarding the class of propositions Price calls 'FIXTURES'. In his formulation of the architecture of deliberation, Price divides FIXTURES into two sub-categories: KNOWNS and KNOWABLES, where KNOWNS is the class of propositions that the agent takes herself to know at the time of deliberation, and KNOWABLES is the class of propositions that she takes to be knowable, at least on principle. At face value, this terminology suggests that the propositions in FIXTURES

[165] According to the taxonomy introduced in §1.1.1, causal claims can be divided into a four-fold taxonomy based on: first, whether a claim refers to a cause that is a *deviation* from the normal course of evolution of a kind of system, or *part* of the normal course; and second, whether the claim refers to a *type* or *token* cause. The four categories at thus: 'deviant token', 'deviant type', 'normal token', and 'normal type'.

are able to be known precisely because they are fixed and unchangeable. However (as Price acknowledges), most propositions in FIXTURES are not fixed in this sense. It is true that some members of FIXTURES are unchangeable. For example, necessary truths (e.g. 6 + 3 = 9) are permanently fixed. And Price argues that it is a contingent fact about human agency that (from our perspective), propositions about the past are unchangeable. However, contingent facts about the future are *not* fixed in this way—they cannot be, because if facts about the future were fixed (from our perspective), it would not be possible for us to make them true by acting.

Deliberation does require that we hold certain propositions about the future fixed (or at least temporarily settled), though. Otherwise there are just too many possibilities to take into account. In Price's words, '*something* has to be held fixed, for otherwise the question, "What changes, if we change this?" has a trivial answer: "Everything!"'[166] For example, if I am deciding whether to walk to campus, or to drive, there are many factors about the future that I am likely to hold fixed. These include the cost of parking, the time it will take to drive versus the time it will take to walk, what I am intending to do after work and whether I need a car to get there, etc. The important point is that the values of all these variables are changeable. If I were performing a *different* deliberation—for example, if I were trying to decide what to do after work, or if I were on a committee whose task was to consider revising the parking fees—I would allow the values of these variables to vary. That is, in a different context, the relevant propositions would be members of OPTIONS, rather than FIXTURES.[167]

For this reason, the propositions in FIXTURES can be divided into two subsets, the first of which I will call 'PERMANENT FIXTURES'. PERMANENT FIXTURES consists of matters of fact that are unchangeable from the perspective of a deliberating agent. Thus,

[166] Price, "Causal Perspectivalism": 280. (Italics in the original.)
[167] Price also points out that which propositions are in FIXTURES is a context dependent matter. Price, "Causal Perspectivalism": 276.

PERMANENT FIXTURES includes propositions about the past, as well as necessary truths.[168] The second subset consists of contingent facts about the future. These propositions are in fact changeable, but are held fixed for the purpose of deliberation. I will call this subset 'TEMPORARY FIXTURES'. As is evident from the above discussion regarding the choice whether to drive or walk to campus, which propositions are members of the class TEMPORARY FIXTURES is a context sensitive matter.

To fully specify the architecture of deliberation, it is therefore necessary to provide an account of how we decide which propositions about the future to place into the category TEMPORARY FIXTURES. Unfortunately, giving a precise account of which propositions about the future we should hold temporarily fixed is not an easy task. Consider the following example, taken from Ismael:

> If I am wondering whether I should move to Miami for part of the fall, I can't just hold everything that is the case now fixed, I have to know how the weather will be at the time and whether my children will still be in school. And I have to judge how my priorities and feelings will have evolved in the meantime ... There is no simple recipe for making these judgements.[169]

As Ismael's example demonstrates, if we deliberate on the assumption that all variables will stay fixed at their *current* values, our planning will go wildly astray. But we do have to somehow determine, or at least estimate, the future values of certain variables, in order to be able to plan at all.[170]

---

[168] More precisely, propositions about the past are *typically* members of PERMANENT FIXTURES. To hold that propositions about the past are *always* members of PERMANENT FIXTURES is to assume that there is no backwards causation.

[169] Ismael, "Causation, Free Will, and Naturalism": 229.

[170] In this chapter, I use the term 'proposition', and the phrase 'the value of a variable' interchangeably. These two linguistic entities are connected in that a proposition describes a variable taking a certain value.

Ismael is right that there is no simple recipe for working out how to set the values of the variables in the class TEMPORARY FIXTURES. However, the fact that determining the values of these variables is not a simple matter does not imply that it is impossible to say *anything* about how we make these judgements. The discussion in Chapter 1 of this thesis provides a clue as to how we might *start* to put together a recipe. Absent any reason to think that something out of the ordinary will happen, we can assume that the world will follow its normal course of evolution. For this reason, propositions in TEMPORARY FIXTURES often describe the values that variables take when the relevant system follows its normal course of evolution.

For example, Ismael can be fairly sure that the school holidays will occur at roughly the same time from year to year, following their normal pattern. She can get an idea of what the weather will be like by looking at the climate data for Florida in the fall, under the assumption that the weather is likely to follow its normal course. And the only way to judge how her priorities and feelings will have changed is by considering what normally causes them to change, and thinking about how she can expect them to evolve, if they follow this normal pattern. Thus, the propositions Ismael should consider settled—that is, the propositions in the set TEMPORARY FIXTURES—are those that describe the normal course of evolution of the relevant systems, absent any specific reason to think that there is likely to be an intervention in these normal patterns.

The process of deliberation does not require that an agent temporarily settles the values of *all* variables in the future. For example, when deciding whether to go to Miami in the fall, Ismael only needs to actively consider the normal course of evolution of a few *relevant* variables. The school holidays are important to take into account, assuming that she has to teach (or has children). So is the weather, and her own priorities and feelings. So also, presumably, is her work, her financial situation, and any other

commitments she might have. I will call the set of propositions describing the projected values of the variables that are actively considered in a particular deliberation 'RELEVANT FIXTURES'.

There are other factors that we tacitly hold fixed when deliberating about a possible trip. In Ismael's example, these include the assumption that the US will remain politically stable, so that it will be safe to travel to Miami; that the airlines (or other forms of transport) will continue to operate, so that it will be possible to get there; and that no other unforeseen emergencies will crop up, requiring her attention. To hold these factors fixed is to assume that the world will stay within certain normal bounds, although we do not usually bring these variables explicitly to mind. I will call the set of propositions stating the future values of these tacitly fixed variables 'BACKGROUND FIXTURES'. Because both categories consist of future values of variables that we hold fixed for the purposes of deliberation, RELEVANT and BACKGROUND are both subsets of TEMPORARY FIXTURES.

Finally, there is also a set of factors that are likely to be simply irrelevant to Ismael's deliberation about her possible trip to Miami. This set includes propositions about events that are (comparatively) isolated from Ismael's life—for example whether or not there is a thunderstorm in Alice Springs on a particular day in April. It also includes propositions about events that are too causally impotent to have any major effect on her life—for example what she has for breakfast on the 3rd of June. The set of propositions concerning the values of these irrelevant variables forms another class, which I will call 'IRRELEVANTS'. The propositions in IRRELEVANTS are not contained in either OPTIONS or FIXTURES (and therefore not in TEMPORARY FIXTURES, either)—these propositions are simply not part of the deliberative process.

In summary, the more detailed architecture of deliberation outlined in this section consists of three main categories: OPTIONS, FIXTURES and IRRELEVANTS. The category of propositions that an agent considers herself to be capable of making true (i.e. OPTIONS) includes two mutually exclusive sub-categories: DIRECT and INDIRECT, as well as a further category which can include both DIRECT and INDIRECT OPTIONS: ENDS. The category of propositions that an agent considers fixed, for the purpose of the deliberation in question (i.e. the members of FIXTURES), is also divided into two sub-categories: PERMANENT and TEMPORARY. The former includes necessary truths and propositions about the past, whereas the latter consists of changeable facts about the future. TEMPORARY FIXTURES can further be divided into RELEVANT (those propositions about the future that are explicitly considered in the deliberation), and BACKGROUND (propositions that describe normal states of affairs that are taken for granted in the deliberative process). The class IRRELEVANTS consists of those propositions that are simply irrelevant to the deliberative process.

Importantly, the propositions in OPTIONS and FIXTURES are structured as a *model*, which includes a representation of the *normal type* causal relations that hold between the variables included in the model. This structure is summarised in Figure 4.1.

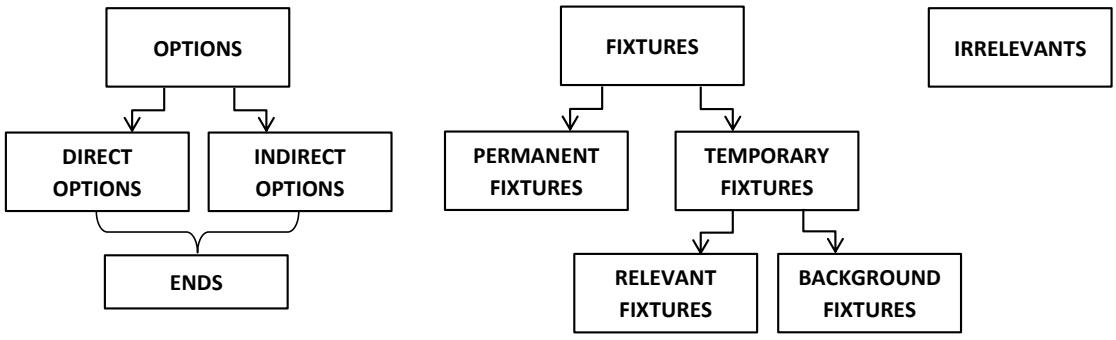

Figure 4.1: A more comprehensive architecture of deliberation

Just as the partiality of perspectives, in general, is dictated by the limited cognitive capacities of beings, the structure and content of the epistemology of agency is also dictated by the cognitive capacities of agents (both actual and possible). The complexity of the model that an agent uses to represent the world at any point in time will necessarily be limited by the cognitive capacities of that agent. That is (assuming the agent is not God), the model must necessarily be a simplification, or abstraction, which represents *part* of the world, including a only fraction of the possible future values of variables. This need for simplified representations gives rise to the need for the class TEMPORARY FIXTURES to be divided into the sub-categories RELEVANT and BACKGROUND, and for the existence of the separate category, IRRELEVANTS.

Finally, just as the categorisation of propositions into OPTIONS, FIXTURES and IRRELEVANTS is highly context dependent, so is the categorisation of the members of TEMPORARY FIXTURES into RELEVANT and BACKGROUND. Which propositions fall into which category is dependent on the agent's desires and purposes at any particular point in time. In other words, the constitution of these classes is partially determined by those propositions in ENDS.

4.7 Linking the epistemology of agency to the Menzies–McGrath model

I have claimed that the architecture of deliberation is a description of the control-based perspective, rather than the purpose-dependent perspective, and that it is the latter that is used to make and evaluate deviant token causal claims. However, these two perspectives have many features in common. In particular, both describe the normal course of evolution of a system. As discussed above, the propositions in TEMPORARY FIXTURES are typically those that describe the normal course of evolution of the relevant system in the control-based perspective. In other words, the class TEMPORARY FIXTURES contains the propositions that describe the default values of the variables—

the values that the relevant variables take in the default worlds. The aim of this section is to show that the division of TEMPORARY FIXTURES into RELEVANT and BACKGROUND maps onto a corresponding division in the way the future values of variables are represented in the default worlds, and that the category IRELEVANTS also has a counterpart in the purpose-dependent perspective.

As discussed above, the propositions in RELEVANT FIXTURES represent the future values of variables (set at their normal value) that are explicitly included in the model that an agent uses to deliberate. The propositions in BACKGROUND FIXTURES also describe the default values of certain variables, but these variables are only *implicitly* included in the causal models. They are the normal events or states of affairs that we just take for granted, in making a particular decision, and never actively attend to. Finally, the variables that are described by the propositions in IRRELEVANTS are not part of the system that is represented in the deliberative process.

The default worlds (which are a component of the purpose-dependent perspective) also represent the normal course of evolution of a system. In addition, the default worlds are subject to the same cognitive constraints as the architecture of deliberation, or control-based perspective. For these reasons, the variables included in the default worlds can also be divided into two categories, corresponding to RELEVANT FIXTURES and BACKGROUND FIXTURES, respectively. The former category consists of the variables that are explicitly contained in the default words, and the latter consists of variables that are tacitly held fixed. The purpose-dependent perspective also contains a third category, corresponding to IRRELEVANTS, which consists of those variables that are not part of the system represented by the default worlds.

The need for a distinction between the variables that are explicitly included in causal models, and a background that is tacitly held fixed, has also been noted by those

working in the causal modelling tradition (or the structural equations framework).[171] This is a system of causal modelling that is becoming increasingly common in the philosophy of causation (having been incorporated from statistics and the philosophy of science).[172]

Pearl, one of the founders of the structural equations framework, writes:

> In most cases the scientist carves a piece from the universe and proclaims that piece *in*—namely, the *focus* of investigation. The rest of the universe is then considered *out* or *background* and is summarised by what we call *boundary conditions*.[173]

The important point is that any causal model of *part* of the world must include a background that is tacitly held fixed, as well as variables that are actively considered. The boundary conditions Pearl refers to in the above quote correspond to the members of the class BACKGROUND FIXTURES in the architecture of deliberation, and the normal values of the variables that are tacitly held fixed in the default worlds. These similarities between the structural equations framework and the Menzies–McGrath model suggest that it would be worth trying to formulate the Menzies–McGrath model in terms of the structural equations framework. However, due to lack of space, such a task will not be undertaken in this thesis.

An important implication of the need to distinguish between the variables that are explicitly included in the default worlds and those variables that are only implicitly included, is that these two types of variables generate two different kinds of *default*

[171] See Pearl, *Causality: Models, Reasoning, and Inference*.
[172] For example, the structural equations framework is used in the following accounts of causation: Woodward, *Making Things Happen: A Theory of Causal Explanation*; Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason"; Christopher Hitchcock, "Structural Equations and Causation: Six Counterexamples" *Philosophical Studies* 144 (2009); Menzies, "Causation in Context"; Joseph Y. Halpern and Judea Pearl, "Causes and Explanations: A Structural-Model Approach. Part I: Causes" *British Journal for the Philosophy of Science* 56 (2005).
[173] Pearl, *Causality: Models, Reasoning, and Inference*: 350. (Italics in the original.)

values, corresponding to variables that are explicitly set at their default values, and variables that are only tacitly held fixed at these default values, respectively. These two types of default values correspond to two different types of *conditions*.

To illustrate, let us return to the example in which my car fails to start when I turn the key. As already discussed, the causal model of this scenario represents the car (or the car's ignition system) as a mechanical system. In the normal course of evolution of this system, the battery is charged, there is petrol in the fuel tank, the spark plugs are working, etc. These states represent default values of variables that are explicitly included in the model of the car's ignition system. These states are also all conditions of the car's starting—it is necessary that these variables take their default values, in order for the car to start. However, there are many other variables that must be set at their normal values if the car is to start. The default values of these variables correspond to the local infrastructure that must be held fixed in order for the evaluation of interventionist counterfactuals to be possible (as discussed in §3.3.2).

For example, the atmosphere must contain a certain amount of oxygen, the combustion of hydrocarbons must have a particular reaction profile, the key must be rigid, etc. The default values of these variables, which are not explicitly included in the causal model of the car's ignition system, are also necessary conditions of the car's starting. The point is that the normal course of evolution of the kinds of systems represented by the purpose-dependent perspectives includes factors that are part of the background, and thus tacitly held fixed, as well as variables that are intentionally held at a particular value. And *both* these kinds of default values are conditions, rather than potential deviant causes (or effects).

There are two important differences between the purpose-dependent and control-based perspectives. First, unlike deliberation, which is necessarily located at a particular point

in time, causal judgements can be made about events in the past, as well as the future. Thus, the purpose-dependent perspective does not include a meaningful division between the past and future values of variables. That is, there is no distinction in the purpose-dependent perspective that maps onto the difference between OPTIONS and FIXTURES. However, the purpose-dependent perspectives do include a type of option. As well as representing the normal course of evolution of a system, modelling (and thus understanding) a causal system requires modelling the kinds of events that could intervene in that system—the kinds of deviations from the normal course of evolution that are likely to have an *effect* on the system. The purpose-dependent perspective must therefore include a set of variables that represent *potential interventions*. These potential interventions are not part of the default worlds—that is, 'possible interventions' is a separate category of propositions, from which deviant causes are selected. *Effects* are also a kind of possible intervention, or deviation from the normal course of evolution, but, like the members of INDIRECT OPTIONS, effects are *brought about* by another external intervention in the system (i.e. a cause). In terms of the taxonomy of causal judgements introduced in §1.1.1, *possible interventions* are *deviant type causes*. The category 'possible interventions' will be more precisely defined in §5.2.

The second difference between the control-based perspective and the purpose-dependent perspective is that the former is a *teleological* structure, whereas the latter is not. The point of deliberating is to achieve a particular outcome, or goal. The options an agent deliberates between are determined by these goals—that is, the members of OPTIONS are constrained by the members of ENDS. The default worlds, on the other hand, are not teleological—they do not include a class of propositions corresponding to the category ENDS. Although the *content* of the default worlds is partially determined by the purposes of the people who construct them, these purposes are not *included* in the default worlds.

In summary, the purpose-dependent perspectives do not just consist of a set of variables representing the normal course of evolution of a system—their structure is more complicated than this basic picture, in a number of ways. First, purpose-based perspectives consist of three main categories of propositions, or values of variables: *possible interventions*, *default worlds*, and *irrelevants*. Second, the perspective is structured as a causal model, in which the values of variables in the default worlds are connected by *normal type* causal relations, and the values of variables that represent possible interventions are *deviant type* causes. Third, the variables included in the default worlds can be divided into two categories, the first of which contains those variables that are actively considered, and the second of which includes variables that form part of a background that is tacitly assumed, but not explicitly considered.

To understand a causal system, and thus to be able to make accurate causal judgements about deviations to that system is to (mentally) have such a causal model available. In other words, to understand a causal system is to have access to a purpose-dependent perspective with the above characteristics, which are summarised in Figure 4.2.

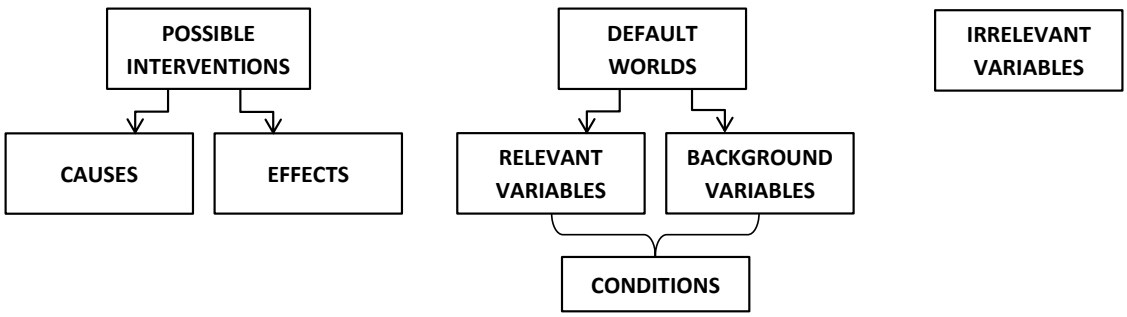

Figure 4.2: The structure of the purpose-dependent perspective

In this chapter, I have used philosophical anthropology to tell a story about the way that perspectives enter into deviant token causal judgements. The conclusion of this investigation is that these judgements are made from a purpose-dependent perspective, rather than an intersubjective or control based-perspective, and that each purpose-

dependent perspective includes a set of default worlds, as well as a set of possible interventions in these default worlds. However, I have yet to say anything about the metaphysical implications of the story I have told—that is, the metaphysical implications of the Menzies–McGrath model of causation. It is to these metaphysical implications that I will turn in Chapter 5.

Chapter 5: The Metaphysics of the Menzies–McGrath Model

In Chapter 3, I discussed three accounts of causation: Menzies and Price's agency theory, Woodward's interventionism and Price's perspectivalism. Like the Menzies–McGrath model, these theories take as their starting point the idea that our causal concept is linked to our status as *agents*. I suggested that Woodward's interventionism and Price's perspectivalism can both be considered as more developed versions of Menzies and Price's earlier agency theory of causation. However, it was also revealed that in metaphysical terms, these authors develop the agency theory in quite different directions—Woodward in the direction of increasing objectivity, and Price in the direction of increasing subjectivity. In this chapter, I will outline a metaphysics of causation (or, at least, deviant token causation) that adopts what I see as the strengths of both Woodward and Price's accounts: Woodward's emphasis on the notion of a *serious possibility*, and Price's contention that the concept of causation is perspectival—that our causal judgements depend on features of us, as well as features of the mind-independent world. In other words, I will defend an account of the metaphysics of causation that attempts to reconcile the subjective and objective features of our causal discourse.

The subjective aspects of deviant token causal claims have been discussed at length in this thesis. First, these causal claims refer to events that are deviations from the normal course of evolution of a kind of system; second, the relevant kinds of systems involve both descriptive and normative standards; and third, *which* kind of system is relevant to a particular inquiry is determined by the purposes of the person undertaking the inquiry. However, despite these subjective or mind-dependent features, it is nevertheless clear that our deviant causal claims are intended to be *about* the world—we think that true causal claims succeed in *describing* happenings in the world. We also know that conceptualising a scenario differently will not change the *effects* of our actions. For

example, I cannot prevent myself from getting a hangover by manipulating the way I represent the physiological system involved in alcohol processing. These subjective and objective features of causal reasoning must both be captured by the metaphysics of causation.

Menzies hints at a metaphysics of causation that reconciles these objective and subjective features, which he calls *perspectival realism*:

> [C]ontextualism about the causal concept does not necessarily support an error theory. For another way to dispute causal realism is to reject at the outset the realist's characterisation of reality in terms of certain privileged mind-independent facts. This relatively unfamiliar position might be called *perspectival realism*. The perspectival realist acknowledges that the truth-value of causal judgements does not depend entirely on the mind-independent structures. The context-sensitive character of causal judgements indicates that their truth value is perspective relative.[174]

Elsewhere, Menzies puts this slightly differently, arguing that the perspectival nature of our contact with reality means that our language does not always map neatly onto a God's-eye perspective of the world:

> [Q]uite often our common sense concepts have a perspectival character that reflects our distinctively human position in the world; and this perspectival character makes it hard to locate the corresponding items in "the view from nowhere" that science aspires to provide.[175]

Menzies has not given a detailed defence of perspectival realism, or indeed any elaboration beyond gesturing at the position in the above passages. My intention in this

[174] Menzies, "Causation in Context": 193. (Italics in the original.)
[175] Peter Menzies, "Peter Menzies" *Metaphysics: 5 Questions*, ed. A. Steglich-Petersen (Copenhagen: Automatic Press, 2010): 49.

chapter is, therefore, to use the conclusions of the previous chapters to formulate this metaphysical position in some detail.

5.1 Perspectival realism and the Menzies–McGrath model

Recall that according to the Menzies–McGrath model, the relationship between the default worlds and the truth values of deviant token causal claims is intended to be understood as a conditional—the idea is that, *if* the given situation instantiates a kind of system, K, governed by standards, S, *then* $X = x$ causes $Y = y$ relative to the default values $X = x'$ and $Y = y'$ if and only if the following conditions hold:

(i) the actual values of X and Y are x and y respectively; and

(ii) if an intervention were to change the value of the X variable from x' to x, the value of the Y variable would change from y' to y.

Conditions (i) and (ii) have already been discussed in some detail (in §§1.5 and 4.5). However, the claim that deviant token causal judgements only hold *if* the situation instantiates a kind of system, K, governed by standards, S, requires further elaboration. This conditional is only true if K is *both* instantiated in the actual world, *and* accurately represented by the purpose-dependent perspective. Thus, this conditional entails constraints on both the system that is being described (i.e. on the structure of a certain part of the world), and on the way these parts of the world are represented (i.e. on the purpose-dependent perspectives). The next section will consist of a discussion of these constraints.

5.2 Kinds of systems

Addressing the constraints on the kinds of systems represented by the purpose-dependent perspectives requires first addressing the question 'What is a kind of system?' In previous chapters, I have argued that the kinds of systems described by the purpose-dependent perspectives are *open*—that is, these systems are subject to

interventions; they consist of *part* of the universe (as opposed to the whole universe); and they have a *normal course of evolution*. However, these specifications are vague, and require further clarification.

Menzies provides a more precise characterisation of the kinds of systems represented by the default worlds in 'Difference-Making in Context'. He claims that:

> [A] system is a set of constituent objects that is *internally organised* in a distinctive fashion; and the properties and relations that configure the objects into a system must be *intrinsic* to the set of constituent objects … A property *F* is *intrinsic to a set of objects* if and only if, possibly, one of its members has *F* although no contingent object wholly distinct from the set exists.[176]

As an illustration of Menzies' definition, consider the system 'Chris's circulatory system'. Chris's circulatory system consists of a set of organs that are organised in a certain way. More precisely, the organs in this system are organised by properties and relations that are *intrinsic* to the objects in the system, in the sense that the fact that this set of objects comprises a system does not depend on any object that is *external* to the system. For example, Chris's heart, his capillaries and his carotid artery are all part of his circulatory system because they share the relational properties of being part of Chris's body, and of transporting Chris's blood. More precisely, the property 'being an organ that transports Chris's blood' is possessed by the objects Chris's heart, Chris's capillaries and Chris's carotid artery, and it is possible that (if a large number of background conditions were held fixed) these objects could possess this property, even if no object external to the Chris's circulatory system were to exist. The organs in Chris's circulatory system organs are *not* grouped together because of *extrinsic*

[176] Menzies, "Difference-Making in Context": 155. (Italics in the original.)

properties and relations—for example, because of the fact that they are all approximately the same distance from Mars.

The systems that are referred to in deviant token causal judgements—that is, the systems that are represented by the purpose-dependent perspectives—are a small subset of the total number of particular systems. The relevant subset consists of those systems that have a normal course of evolution. Chris's circulatory system is one such system—this system has a normal course of evolution that is described by laws, or standards, which specify the normal values (or range of values) of a number of variables, including blood pressure, heart rate, and the oxygen content of each organ.

The purpose-dependent perspectives do not represent *particular* systems, but *kinds* of systems. For example, the default worlds used to represent Chris's circulatory system would represent the normal course of evolution of the human circulatory system *in general*.[177] For this reason, after specifying the *particular* systems that are relevant to causal judgments, Menzies goes on to define a *kind of system*, *K*:

> *Menzies*: A *kind of system K* is a set of particular systems sharing the same intrinsic properties and relations … whose evolution over time conforms to certain laws.[178]

According to the Menzies–McGrath model, the relevant laws are the normative or descriptive *actual standards* described in §1.4, so Menzies' definition can be slightly modified as follows:

[177] In some contexts, the default worlds used to represent Chris's circulatory system might be more specific. For example, if Chris always has an unusual heartbeat, which does not present any health problems, his physician might be interested in deviations from the normal course of evolution of Chris's circulatory system, *in particular* (as opposed to the normal course of evolution of the human circulatory system *in general*). The point, again, is that the content of the default worlds (and thus the purpose-dependent perspectives) is highly context dependent.

[178] Menzies, "Difference-Making in Context": 156. (Italics in the original.)

> *Menzies–McGrath 1*: A *kind of system K* is a set of particular systems sharing the same intrinsic properties and relations whose evolution over time conforms to certain actual standards, *S*.

As it stands, this definition describes a system with a normal course of evolution. However, the systems involved in deviant token causal judgements must also be *subject to intervention*. Recall that in the previous chapter (§4.7), we saw that the structure of the purpose-dependent perspectives (and, therefore, of the systems represented in deviant token causal claims), includes the category *possible interventions*, as well as categories that stipulate the default values of the variables included in the system. The definition of a kind of system must therefore refer to these possible interventions.

In 'Causation, Counterfactuals and the Third Factor', Maudlin gives a description of the kinds of systems that are involved in causal reasoning that is very similar to the definition of a kind of system provided by Menzies, and which includes reference to laws that specify possible intervention.[179]

Maudlin describes causal systems in terms of a kind of law (or lawlike generalisation) which he calls 'quasi-Newtonian'. A system is made up of quasi-Newtonian laws when:

> There are, on the one hand, *inertial* laws that describe how some entities behave when nothing acts on them, and then there are laws of *deviation* that specify in what conditions, and in what ways, the behaviour will deviate from the inertial behaviour. When one conceives of a situation as governed by quasi-Newtonian laws, then typically the primary notion of an effect will be the deviation of the

[179] Menzies also defines an *interfering factor*, which is a possible intervention. Menzies, "Difference-Making in Context": 157.

> behaviour of an object from its inertial behaviour, and the primary notion of a cause will be whatever sort of thing is mentioned in the laws of deviation.[180]

Maudlin's use of the term 'quasi-Newtonian' is a reference to Newton's laws of motion. According to Newton's first law, a body in a state of constant motion (including being at rest) will remain in that state of motion unless an external force is applied to it. Newton's first law thus defines *inertial motion*. Newton's second law states that the force experienced by an object is equal to the object's mass, multiplied by its acceleration ($F = ma$). The second law thus describes the way an object's motion changes on application of an external force. That is, the second law is a law of *deviation*.

In the terminology of the Menzies–McGrath model, Maudlin's *inertial laws* correspond to *actual standards*, which stipulate the normal course of evolution of a particular variable, and *laws of deviation* describe *possible interventions* in the normal course of evolution of a particular variable. The notion of a *law of deviation* can thus be used to give a more precise definition of a possible intervention, as follows:

> Variable X taking the value x is a *possible intervention* in a kind of system, K, if $X = x$ according to one of the laws of deviation of K.

Finally, the notion of a *law of deviation* can also be used to complete the definition of a kind of system:

> *Menzies–McGrath 2*: A *kind of system K* is a set of particular systems sharing the same intrinsic properties and relations whose evolution over time conforms to certain actual standards, and which is subject to certain, specifiable, laws of deviation.

[180] Maudlin, "Causation, Counterfactuals, and the Third Factor": 431. (Italics in the original.)

The requirement that the situation instantiates a kind of system, *K*, governed by standards, *S*, entails that deviant token causal judgements are only true if the cause is a deviation from a kind of system with the internal structure just described. It is, therefore, a condition of each deviant token causal judgement that the part of the world under consideration conforms to this structure. In the next subsection, I will consider this constraint in more detail.

5.2.1 Constraints on the world

After describing the kinds of systems used to make token causal judgements, Maudlin goes on to claim that:

> The special sciences, and plain common sense as well, will seek to carve up the physical world into parts that can, fairly reliably, be described as having inertial states (or inertial motions) that can be expected to obtain unless some specifiable sort of interference or interaction occurs. Or at least, those special sciences that manage to employ taxonomies with quasi-Newtonian lawlike generalisations can be expected to support particularly robust judgments about causes.[181]

Since to 'employ a quasi-Newtonian taxonomy' is just to represent part of the world as an instance of a kind of system with a normal course of evolution, but subject to intervention, Maudlin's claim that the special sciences seek to carve up the world into parts with quasi-Newtonian taxonomies (or that succeeding in this task results in robust judgements about causes) is another way of expressing the idea that causal reasoning (or at least one kind of causal reasoning) requires conceptualising the world in terms of the kinds of systems just defined.[182]

[181] Maudlin, "Causation, Counterfactuals, and the Third Factor": 434.
[182] Hitchcock makes the same point (using very different terminology) in Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason".

However, notice that whether or not any particular special science (and also common sense) *succeeds* in carving up the world into systems with normal courses of evolution will depend partly on the constitution of the world itself. In order for this kind of depiction of the world to be possible, the world (or at least parts of the world) must be structured such that it can be accurately represented by kinds of systems that are open, in that they are responsive to certain kinds of external interventions, but also closed enough to have a normal course of evolution. In other words, the kinds of systems just described must be an appropriate idealisation, or abstraction, of the parts of the world under consideration. When this condition is not satisfied, making causal judgements is vastly more difficult (and making deviant token causal judgements becomes impossible).

As an illustration, consider a world that is completely chaotic. Given the value of a particular variable in this world, X, at time t_0, we cannot say anything about the value of X at a later time t_1—the value of any variable at any time in the future is entirely unpredictable. In this world, there are no systems with a normal course of evolution, and deviant token causal judgements are impossible.

It is possible that parts of the *actual world* are too chaotic to support causal reasoning. For example, consider quantum mechanics. The motion of particles at the quantum level is not Newtonian and predictable, but chaotic and unpredictable. The location and momentum of the electrons in any particular atom, for example, are described by a probability distribution, rather than the differential equations of classical mechanics. For this reason, it is often observed that the concept of causation does not neatly apply to the world of quantum mechanics.[183] Perhaps this is because quantum phenomena are not quasi-Newtonian—they do not form systems with a normal course of evolution. Thus,

[183] For example, see Richard Healey, "Causation in Quantum Mechanics" *The Oxford Handbook of Causation*, eds. H. Beebee et al. (Oxford: Oxford University Press, 2009).

causal reasoning, or at least the form of causal reasoning used to make deviant token causal judgements, breaks down.

Causal judgements are also difficult in situations in which the systems do not have a single, obvious, normal course. This often seems to be the case in economics. As an illustration, consider the 'stimulus package' introduced by the Australian government during the recent financial crisis. In economic terms, this stimulus involved moving from a neutral to an expansionary fiscal policy (i.e. increasing the ratio of government spending to revenue from taxation). The intended effect of this alteration in fiscal policy was to stimulate the economy—that is, to increase consumption. Consumption *did* increase after the stimulus package was introduced. However, because the Australian economy is so complex, and subject to so many interfering factors, there is no consensus on what would have happened if the stimulus package had *not* been introduced. That is, because the system does not have a well-defined normal course of evolution, it is not clear whether the stimulus package actually made a difference and *caused* the increased consumption.[184]

The point of the previous two examples is to illustrate that the difficulties in applying the causal concept to quantum mechanics and economics is due to the nature of the kinds of systems studied in these fields, rather than the nature of the models used to represent these systems. The problem in attributing causes in these fields is due to the organisation of the world itself, and the kinds of systems instantiated in certain parts of the world, and not from human representational practices.

5.2.2 Constraints on the purpose-dependent perspectives

As well as constraining the structure of the parts of the world that can be described by deviant token causal judgements, the requirement that the situation instantiates a kind of

[184] Part of the reason for this lack of consensus is political. But it is not *all* political.

system, *K*, *which is represented by the purpose-dependent perspectives*, also constrains the structure of the purpose-dependent perspectives themselves. This constraint stipulates that the purpose-dependent perspective must be an appropriate model of the relevant kind of system. That is, the structure of the purpose-dependent perspective must correspond to the structure of the kind of system that is represented.

In his paper 'Prevention, Preemption and the Principle of Sufficient Reason', Christopher Hitchcock considers what it takes to be an appropriate causal model of a particular situation.[185] Hitchcock intends the term 'causal model' to refer to a model within the structural equations framework (a system for modelling and evaluating causal networks).[186] For this reason, the precise details of Hitchcock's causal models are different from those of the purpose-dependent perspectives. However, these details do not affect the question of what it takes for the *variable choice* in a model (or purpose-dependent perspective) to be appropriate for a particular situation. The constraints on causal models that Hitchcock discusses are, therefore, also constraints on the purpose-dependent perspectives.

The restrictions that Hitchcock places on appropriate variable choice are as follows: first,

> [I]t is important to choose the variables so that different values of the same variable correspond to events (or versions of events) that are incompatible on broadly logical or conceptual grounds; typically, they will represent incompatible states of a system at the same time.[187]

[185] Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason". In this paper, Hitchcock proposes an account of token causation along the lines of the account suggested above. According to Hitchcock, an event, *c*, is a cause of another event, *e*, if the value of the variable representing *c* is a token cause of the value of the variable representing *e* in a causal model that *appropriately* represents the situation: 503.
[186] See Chapter 4, footnote 172 for a list of citations.
[187] Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason": 502.

For example, when representing Chris's circulatory system as a causal model, it would be appropriate to choose Chris's heart rate as one of the variables, because the different values of this variable (e.g. 60 beats per minute, 100 beats per minute, etc.) are logically incompatible: it is not possible for Chris's heart rate to be 60 bpm and 100 bpm simultaneously.

Hitchcock also points out that '[m]ore nebulously, [a causal model] must include enough variables to capture the essential structure of the situation being modelled. What counts as an appropriate model may depend at least in part on pragmatic factors.'[188] In the terminology of this thesis, Hitchcock's claim here is just that the purpose-dependent perspective must capture the structure of *K*, the kind of system instantiated in the situation being modelled. As has already been argued in this thesis, both the system picked out, and the extent of this system (i.e. where the line is drawn between what is included and what is excluded), will partly depend on pragmatic factors, including the purpose of the enquiry and the cognitive limitations of the being doing the representing.

The most important constraint on a causal model is that it must correctly predict the effects of interventions.[189] That is, an appropriate causal model must entail only true interventionist counterfactuals. Meeting this condition requires that the particular system being represented is an instance of a kind of system, *K*, with a normal course of evolution; that there is a restricted class of possible interventions (specified by the laws of deviation) that can impede this normal course; and that the structure of the purpose-dependent perspective accurately represents both the internal structure of *K,* and the possible interventions in *K*.

In order to successfully represent a kind of system, *K*, the purpose-dependent perspectives must have the structure that was outlined in §4.7. That is, each purpose-

[188] Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason": 503.
[189] Hitchcock, "Prevention, Preemption, and the Principle of Sufficient Reason": 503.

dependent perspective must consist of three categories of variables: first, variables that are explicitly included in the model; second, background variables that are not explicitly included, but are tacitly held fixed; and third, variables that represent possible interventions. As explained in Chapter 4, the first two categories of variables represent the normal course of evolution of *K*, and the latter represent possible interventions in *K*.

Notice that the requirement that causal models generate only true counterfactuals entails that these models can be *tested*. In everyday life, the testing of causal models is informal. For example, a child might learn about the type causal connection between flipping a light switch and the light's turning on by repeatedly flipping the switch, and observing the effects of his interventions. The process of testing causal models has been formalised in science—this, of course, is the role of scientific experiments. The requirement that a causal model entails only true counterfactuals ensures that we have some way of determining whether a model accurately represents the relevant system—we know that a model appropriately represents a causal system if it correctly predicts the effects of interventions. When this condition is fulfilled, we have good reason to believe that our causal models accurately represent the world—that the default worlds are appropriate models of a kind of system, *K*.

The above conditions do not amount to a precise answer to the questions of what it takes to be a kind of system, *K*, or to provide an appropriate causal model of a particular system. This is partly because these notions themselves are vague—whether or not a system has a normal course of evolution is not completely determinate, and there may not always be one *best* model of a particular system, even once the purpose of the enquiry (and thus the relevant purpose-dependent perspective) is specified. However, this section has gone a long way towards specifying the conditions (both of the relevant part of the world and of the model used to represent this part of the world) in which

deviant token causal judgements are successful. In the reminder of the chapter, I will consider the metaphysical implications of these truth conditions, beginning with a response to an argument made by Nancy Cartwright, according to which the mind-independence of causation mirrors the mind-independence of the difference between effective and ineffective strategies.

5.3 Effective versus ineffective strategies

In her paper 'Causal Laws and Effective Strategies', Nancy Cartwright famously counters Russell's suggestion that the concept of causation is 'a relic of a bygone age',[190] by showing that the concept of causation is required to distinguish between effective and ineffective strategies.[191] One of Cartwright's examples concerns a discovery made by the French, when they were building the canal in Nicaragua.

> [T]he French discovered that spraying oil on the swamps is a good strategy for stopping the spread of malaria, whereas burying contaminated blankets is useless. What they discovered was true, independent of their theories, of their desire to control malaria, or of the cost of doing so.[192]

Cartwright's point here is clearly correct—changing the way we model a situation, or our purposes in constructing a model, will not change the fact that some interventions are effective, and others are not. In the terminology of Menzies and Price's agency theory, representing a situation using a different model will not change whether a bringing about a certain event, *A*, is an effective means of bringing about another event, *B* (see §3.2 for a discussion of Menzies and Price's theory).

[190] Bertrand Russell, "On the Notion of Cause" *Proceedings of the Aristotelian Society* 13 (1912-1913): 1.
[191] Nancy Cartwright, "Causal Laws and Effective Strategies" *Noûs* 13 (1979).
[192] Cartwright, "Causal Laws and Effective Strategies": 420.

Any successful account of the causal concept must be consistent with Cartwright's response to Russell. That is, a theory of causation must be able to account for the mind-independence of the difference between effective and ineffective strategies.

Notice that the actions that correspond to effective strategies are precisely those actions that are specified by laws of deviation—that is, the actions that represent possible interventions in the normal course of evolution of the relevant kind of system. For example, spraying oil on the swamps was a good way of preventing the spread of malaria because malaria is spread by mosquitoes, which breed in pools of water. Spraying the swamps results in a film of oil forming over these pools of water, and this film of oil prevents mosquitos from breeding. The spraying thus adheres to a specifiable law of deviation into the normal course of evolution of the life cycle of the mosquito, which reduces the spread of malaria, and accordingly explains why this strategy is effective. Contaminated blankets, on the other hand, are not involved in the life cycle of mosquitos. Burying contaminated blankets, therefore, does not represent a possible intervention in this system, and this explains why such burying is useless. Thus, the Menzies–McGrath model can account for the discovery that spraying the swamps is an effective strategy, whereas burying blankets is not.

More generally, although *which* system is relevant to a particular causal judgement is determined by the purpose of the person making the judgement, this decision is subject to two restrictions. First, the system must be an instance of the kind of system defined in §5.2, and second, the purpose-dependent perspective must be an appropriate model of this kind of system. These two restrictions guarantee that the truth of a deviant token causal judgement is largely mind-independent. In particular, whether or not a strategy is effective or ineffective is determined by the causal structure of the world, not by human representational practices.

5.4 Perspectival realism reconsidered

Recall Price's claim that the concept of causation is perspectival in the same way that the term 'foreigner' is perspectival—the extension of both 'foreigner' and 'cause' is determined at least partly by the perspective of the beings who use the concept (see §3.4.1). Viewed from a God's-eye perspective, there are no foreigners, and, Price claims, no causation.

It is in Price's sense that the Menzies–McGrath model is a form of perspectivalism. From different purpose-dependent perspectives, we carve up the world differently—we pick out different kinds of systems as being salient. Because deviant token causes are deviations from the normal course of evolution of a particular kind of system, which events are causes depends partly on which kind of system is represented. That is, the term 'cause' has a different extension relative to different purpose-dependent perspectives.

To return to the analogy with the term 'foreigner', from a God's-eye perspective, there are groups of people who have arranged themselves into societies. The term 'foreigner' is useful only to a being that sees himself as *within* one of these groups of people. Although God can determine which individuals count as foreigners relative to any particular person, relative to God himself, there are no foreigners, just different societies.

Similarly, from a God's-eye perspective, there are objects and events organised in various ways into systems with different degrees of complexity and susceptibility to intervention. There are also human beings, who conceptualise the world in terms of kinds of systems that can be relatively reliably expected to follow a normal course of evolution. The distinction between events that are part of the normal course of evolution of a system and events that represent deviations from these systems is useful to humans

because we see ourselves as capable of intervening in the world, and thus altering the course of evolution of these systems.

Here, the analogy with the term 'foreigner' begins to break down, because it is possible that God sees the *whole universe* as a system with a normal course of evolution, subject to manipulation by him. It is therefore possible that the concept of causation is useful to God.

However, just as the extension of the term 'foreigner' is relative to a particular group of people, so is the extension of the word 'cause'. In order to determine whether a particular deviant token causal claim is true, God has to assume the perspective of the individual making the claim. For example, to evaluate the truth of an utterance of the sentence 'The presence of oxygen caused the fire', God has to assume the perspective of the person (or people) undertaking the causal inquiry, and determine whether the presence of oxygen is a deviation from the normal course of evolution of the system represented in their purpose-dependent perspective. God cannot just look down on the world and pick out the causes, independent of all human representation.

In terms of the debate between Woodward's objectivist interventionism and Price's more subjectivist perspectivalism (both discussed in Chapter 3), the perspectival realism of the Menzies–McGrath model is broadly in agreement with Price. According to the Menzies–McGrath model, for beings that carve the world into different causal systems, and thus assume different purpose-dependent perspectives, the concept of deviant token causation will have a different extension.

However, as discussed in §3.4, Price claims that our causal judgements are made from an *intersubjective* perspective, rather than a purpose-dependent perspective. His aim is to explain how creatures like us came to have the causal concept that we do (in contrast with other possible beings that are differently embedded in time), whereas the aim of

the Menzies–McGrath model is to account for the *actual* context sensitivity of our causal discourse.

5.4.1 The Menzies–McGrath model as a form of causal realism

According to the Menzies–McGrath model, the truth values of deviant token causal claims are relative to the perspective from which those claims are made. However, the kinds of systems (including the potential interventions in these systems) that are represented by these purpose-dependent perspectives are *constitutionally* mind-independent. Although many of these kinds of systems contain normative, as well as descriptive, standards, and are thus not perfectly *natural*, the question of whether a particular set of norms is *instantiated* in a given spatiotemporal region has a mind-independent answer. For example, although the rules of cricket are normative, and thus dependent on human concerns and values, whether or not a particular game of cricket is governed by these rules is a mind-independent matter. Another way to make this point is to note that whether any particular purpose-dependent perspective is an appropriate model of a certain kind of system is independent of human representational practices.

Deviant token causes are thus features of the world that exist independently of the person making a particular causal claim (except in the case that a person cites her own mental state as a cause). For this reason, the Menzies–McGrath model is a form of causal realism, rather than antirealism. Importantly, however, the form of causal realism embodied in the Menzies–McGrath model (i.e. perspectival realism) is different from the natural network model in two respects.

First, according to the Menzies–McGrath model, causation is not a *natural* relation (in the sense of 'natural relation' defined in the Introduction), because the laws governing the evolution of a kind of system can be normative as well as descriptive. Second, the metaphysics of the Menzies–McGrath model is not *network-like*, because deviant token

causal judgements are relative to a multitude of *different* causal systems, rather than just one. Thus, although causes are constitutionally mind-independent, the *truth values* of causal judgements (or at least deviant token causal judgements) are not completely mind-independent.

5.4.2 The Menzies–McGrath model as a form of contextualism

Because the truth values of deviant token causal judgements are context dependent according to the Menzies–McGrath model, this account is a form of *contextualism* about causation (see Chapter 2). One way of construing the relationship between the semantics and the metaphysics of causation encoded in the Menzies–McGrath model (although probably not the only way) is to extend the model of the semantics of causation defended by another group of contextualists: contrastivists. According to contrastivist accounts of causation (discussed in §2.4), although causal sentences are typically two-place, the full grammar of causal propositions is three- or four-place. The context acts to determine which three- or four-place proposition is specified by a particular causal sentence, and these three- or four-place propositions correspond to a mind-independent causal relation, which is also three- or four-place. Thus, by claiming that both the semantics and the metaphysics of causation are three- or four-place, rather than two-place, contrastivists reconcile semantic contextualism with metaphysical realism.

To apply this account of the relationship between the semantics and metaphysics of causation to the Menzies–McGrath model would be to hold that the fully specified semantics of deviant token causal claims contains a purpose-dependent perspective as part of the causal proposition. That is, the semantics of any particular deviant token causal claim includes a model of a kind of system (including possible interventions), as well as the particular values of variables that represent the cause and effect (x and y).

Because the systems instantiated in the situations described by deviant token causal claims are constitutionally mind-independent, whether a token of the kind of system represented in a particular causal proposition actually exists in a given situation is a mind-independent matter. (More precisely, whether the system exists is independent of the mind of the person (or people) making the corresponding causal claim. However, minds (or mental states) may be included in these systems.) Further, the fact that *x* is an intervention that makes a difference to the normal course of evolution of the relevant kind of system is also mind-independent. That is, the causal propositions just described represent mind-independent states of affairs. Thus, by adopting this approach, the Menzies–McGrath model can also plausibly reconcile semantic contextualism with causal realism.

However, the Menzies–McGrath model deviates from the version of causal realism endorsed by contrastivists in that, according to the latter, causation is typically held to be a *natural relation*. For example, we have seen that both Woodward (§3.3.1) and Schaffer (§2.4.2) are careful to separate the subjective and normative features of causal discourse from the metaphysics of causation. According to Woodward, subjective and normative factors enter into the determination of which events are considered to be 'serious possibilities'. However, the notion of a *serious possibility* functions as a kind of add-on to Woodward's overall theory, which does not affect the objectivity of the underlying causal structure. He claims that the truth values of interventionist counterfactuals are objective, and the distinction between serious and non-serious possibilities just restricts which objective *interventions* we are willing to call 'causes'.

As discussed above, according to the Menzies–McGrath model, the truth values of interventionist counterfactuals are partly mind-dependent, because the kind of system represented by the purpose-dependent perspectives depends partly on the purpose of the

person making the causal judgement, and the truth values of interventionist counterfactuals depend on which kind of system is picked out. In addition, the kinds of systems represented can be governed by prescriptive, as well as descriptive norms. This entails that the evaluation of interventionist counterfactuals often requires holding normative factors fixed (see §3.3.2). Thus, according to the Menzies–McGrath model, causation is not a natural relation. The causal concept is infused with subjectivity and normativity all the way through.

This difference between Woodward's interventionism and the perspectival realism of the Menzies–McGrath model is related to the response provided by each account to the problem of unmanipulable causes.

5.4.3 Back to the problem of unmanipulable causes

In response to the problem of unmanipulable causes, Woodward claims that agency accounts of causation can be extended to causes that are not possible human interventions only if there is a 'certain kind of relationship with intrinsic features that we exploit or make use of'[193] when we make causal judgements. That is, he argues that the agency theory collapses into causal objectivism, according to which all causes have some mind-independent, intrinsic feature in common (§3.2.4). According to Woodward, if the sentence 'A meteor strike caused the extinction of the dinosaurs' is true, this is because the sentence succeeds in referring to an instance of an objective, intrinsic causal relation. However, in §3.2.4, I suggested another option.

According to the Menzies–McGrath model, deviant causes do not have an intrinsic, mind-independent feature in common. Instead, these causes all share the *relational* property of being an intervention in the normal course of evolution of a particular kind of system (defined in §5.2). On this model, if the sentence 'A meteor strike caused the

[193] Woodward, *Making Things Happen: A Theory of Causal Explanation*: 125.

extinction of the dinosaurs' is true, this is because the meteor collision was a deviation from the normal course of evolution of the earth at the time of the dinosaurs, which made a difference to that normal course. Thus, the Menzies–McGrath model can respond to the problem of unmanipulable causes by focusing on the shared characteristics of a certain kind of system, rather than on an intrinsic feature of the causal relation itself.

Finally, notice that the Menzies–McGrath model involves reference to human capacities and responses, in that the extension of the term 'cause' is partially determined by the way groups of humans represent the world. Thus, on Menzies and Price's definition of a secondary quality (given in §3.2.2), the Menzies–McGrath model entails that causation is a secondary quality. However, as causes are not *constitutionally* mind-dependent according to the Menzies–McGrath model, on a traditional interpretation of the primary/secondary quality distinction, the Menzies–McGrath model entails that causation is a primary quality.

In this chapter, I have formulated the metaphysical implications of the Menzies–McGrath model, by developing perspectival realism, a metaphysical picture that is hinted at, but not elaborated on, by Menzies. According to perspectival realism, the truth of deviant token causal claims is relative to a purpose-dependent perspective, which represents part of the world as a system with a normal course of evolution. However, due to constraints on both the kinds of systems represented, and the structure of the representations themselves, the result is a form of causal realism, rather than anti-realism, but a form, nevertheless, which is importantly different from the realism of the natural network model of causation.

Conclusion

In this thesis, I have defended the Menzies–McGrath model as an account of deviant token causes. According to the Menzies–McGrath model, these causes are interventions in the normal course of evolution of a system, which make a difference to that normal course. Although each *particular*, instantiated system is partial, open and constitutionally mind-independent, each system is also an instance of a *kind of system* governed by standards that can be either descriptive or normative. Further, because any given spatiotemporal region instantiates multiple systems, the system included in a particular deviant token causal claim is determined by the purposes of the individual (or group) making the claim. More specifically, the truth values of deviant token causal claims are relative to a kind of system represented by a particular *purpose-dependent perspective*. These truth values are therefore not completely mind-independent—they are determined by features of us, and our perspective on the world, as well as by features of the mind-independent world itself.

Although the perspectival realism of the Menzies–McGrath model entails that the concept of deviant token causation partly reflects features of us, and the way we represent the world, we can still learn something about the world by studying this concept. We can ask the question 'What does the world have to be like for our concept of causation to play the role it does in our lives?' In answer to this question, we know that the world must contain creatures with limited knowledge and limited cognitive abilities that possess the concept of agency. We also know that the world itself must contain reasonably predictable, open systems, with a certain degree of complexity. We know this because in a world that was either completely chaotic, or which consisted purely of closed systems (i.e. systems that were not subject to intervention), the actual

concept of causation would not enable us to make successful predictions or give explanations—it would not be of any practical use.

I will end with three suggestions for future research. First, the account of deviant token causal judgements presented in this thesis does not include a determinate means of specifying *which* variables are included in the kind of system represented by a particular purpose-dependent perspective, or which values of these variables are *normal*—that is, the values the variables take in the *default worlds*. As noted in §4.6.2, providing an exact specification of an appropriate purpose-dependent perspective is a difficult (and perhaps impossible) task. However, the most promising means of developing a more specific account of the construction of purpose-dependent perspectives (and evaluating the account of deviant token causal reasoning presented in this thesis) would be to examine the psychological literature, and, in particular, the results of experiments in social psychology, which attempt to determine which factors we pick out as 'the cause' in complex scenarios.[194] To carefully survey this literature, then, is the first suggestion for future research.

Second, in §5.4.2, I briefly outlined a possible account of the connection between the semantic and metaphysical positions defended in this thesis (i.e. the Menzies–McGrath model and perspectival realism). Another suggestion for future research is to more thoroughly explore both the option discussed (a form of contextualism), and other, similar positions that are defended in the literature, including works within the causal modelling tradition, Giere's scientific perspectivism, and Price's subject naturalism.[195]

[194] These experiments constitute the psychological research paradigm that, according to Woodward, is focused on the concept of actual causation (see §1.1.1). The study by Mandel and Lehman discussed in §4.4.2 is an example of this research.

[195] For a list of works within the causal modelling tradition, see Chapter 4, footnote 172. For the most developed versions of Giere's perspectivism and Price's subject naturalism, see Giere, *Scientific Perspectivism*; Huw Price, *Naturalism without Mirrors* (Oxford: Oxford University Press, 2011).

Finally, and most importantly, there is a need to investigate the connections between the concept of deviant token causation described in this thesis, and the other three concepts included in the taxonomy of causal judgements outlined in §1.1.1.

I have suggested that *normal type causes* form the structure of the normal course of evolution of the kinds of systems described by the purpose-dependent perspectives (i.e. that these causes make up the default worlds); that *deviant type causes* are possible interventions in the normal course of evolution of a kind of system, which are specified by laws of deviation; and that *normal token causes* are particular instances of causation in cases in which a system follows its normal course. However, this is only a very rough overview of the relationships amongst a complicated set of concepts—an overview that leaves many questions unanswered. For example: is there a sharp demarcation between normal and deviant causes, or, as seems more likely, do these positions represent two extremes of a spectrum? Does *all* causal reasoning require that we conceptualise the world in terms of the open systems described in this thesis? And is causal reasoning changing, as developments in both computational power and scientific knowledge allow us to represent the world using increasingly complex models? To answer these questions, it will be necessary to study both human representational practices, and the structure of the world itself.

References

Anscombe, G. E. M. *Intention*. Oxford: Blackwell, 1957.

Beebee, Helen. "Causing and Nothingness." *Causation and Counterfactuals*. Collins, J., Hall, N. and Paul, L. A., eds. Cambridge, MA: MIT Press, 2004. 291-308.

Bernstein, Douglas A., Penner, Louis A., Clarke-Stewart, Alison, and Roy, Edward J., eds. *Psychology*. VIII ed: Houghton Mifflin, 2008.

Byrne, Alex, and Hilbert, David R., eds. *Readings on Color*. Vol. 1: The Philosophy of Color. Cambridge, MA: MIT Press, 1997.

Byrne, Ruth M. J. "Counterfactual and Causal Thoughts About Exceptional Events." *Understanding Counterfactuals, Understanding Causation*. Hoerl, C., McCormack, T. and Beck, S. R., eds. Oxford: Oxford University Press, 2011. 208-29.

Cartwright, Nancy. "Causal Laws and Effective Strategies." *Noûs* 13 (1979): 419-37.

Cohen, Jonathan. "Contextualism, Skepticism, and the Structure of Reasons." *Philosophical Perspectives 13: Epistemology*. Tomberlin, J. E., ed. Malden: Blackwell Publishers, 1999. 57-89.

Collingwood, R. G. *An Essay on Metaphysics*. Oxford: Clarendon Press, 1940.

Collins, John, Hall, Ned, and Paul, L. A., eds. *Causation and Counterfactuals*. Cambridge, MA: MIT Press, 2004.

Davidson, Donald. "Causal Relations." *Causation*. Sosa, E. and Tooley, M., eds. Oxford: Oxford University Press, 1993. 75-87.

DeRose, Keith. "Contextualism and Knowledge Attributions." *Philosophy and Phenomenological Research* 52 (1992): 913-29.

Edgington, Dorothy. "Conditionals." *Stanford Encyclopedia of Philosophy* (Winter 2008 Edition).

Eells, Ellery. *Probabilistic Causality*. Cambridge: Cambridge University Press, 1991.

Fodor, Jerry. "Special Sciences, or the Disunity of Science as a Working Hypothesis." *Synthese* 28 (1974): 77-115.

Gasking, Douglas. "Causation and Recipes." *Mind* 64 (1955): 479-87.

Giere, Ronald N. *Scientific Perspectivism*. Chicago: University of Chicago Press, 2006.

Gopnik, Alison, and Schulz, Laura. *Causal Learning: Psychology, Philosophy and Computation*. Oxford: Oxford University Press, 2007.

Grice, Paul. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 1989.

Hall, Ned. "Structural Equations and Causation." *Philosophical Studies* 132 (2007): 109-36.

Halpern, Joseph Y., and Pearl, Judea. "Causes and Explanations: A Structural-Model Approach. Part I: Causes." *British Journal for the Philosophy of Science* 56 (2005): 843-87.

Hart, H. L. A., and Honoré, Tony. *Causation in the Law*. Oxford: Clarendon Press, 1959.

Hausman, Daniel M. "Causation and Experimentation." *American Philosophical Quarterly* 23 (1986): 143-54.

Healey, Richard. "Causation in Quantum Mechanics." *The Oxford Handbook of Causation*. Beebee, H., Hitchcock, C. and Menzies, P., eds. Oxford: Oxford University Press, 2009. 673-86.

Hitchcock, Christopher. "Farewell to Binary Causation." *Canadian Journal of Philosophy* 26 (1996): 267-82.

---. "Of Humean Bondage." *The British Journal for the Philosophy of Science* 54 (2003): 1-25.

---. "Prevention, Preemption, and the Principle of Sufficient Reason." *Philosophical Review* 116 (2007): 495-532.

---. "Structural Equations and Causation: Six Counterexamples." *Philosophical Studies* 144 (2009): 391-401.

Hitchcock, Christopher, and Knobe, Joshua. "Cause and Norm." *Journal of Philosophy* 106 (2009): 587-612.

Ismael, Jennan. "Causation, Free Will, and Naturalism." *Uncorrected Proof*. 2012.

Jackson, Frank. *Conditionals*. Oxford: Basil Blackwell, 1987.

Jackson, Frank, and Pargetter, Robert. "An Objectivist's Guide to Subjectivism About Colour." *Readings on Color*. Byrne, A. and Hilbert, D. R., eds. Vol. 1: The Philosophy of Color. Cambridge, MA: MIT Press, 1997. 67-80.

Kahneman, Daniel, and Miller, Dale T. "Norm Theory: Comparing Reality to Its Alternatives." *Psychological Review* 93 (1986): 136-53.

Knobe, Joshua. "Person as Scientist, Person as Moralist." *Behavioral and Brain Sciences* 33 (2010): 315-29.

Knobe, Joshua, and Fraser, Ben. "Causal Judgement and Moral Judgement: Two Experiments." *Moral Psychology: The Cognitive Science of Morality*. Sinnott-Armstrong, W., ed. Vol. 2. Cambridge, MA: MIT Press, 2008. 441-47.

Lewis, David. *Counterfactuals*. Malden, MA: Basil Blackwell, 1973.

---. "Causal Explanation." *Philosophical Papers*. Vol. 2. Oxford: Oxford University Press, 1986. 214-40.

---. "Causation." *Philosophical Papers*. Vol. 2. Oxford: Oxford University Press, 1986. 159-213.

---. "Postscripts to 'Causation'." *Philosophical Papers* Vol. 2. Oxford: Oxford University Press, 1986. 172-213.

---. "Causation as Influence." *Causation and Counterfactuals*. Collins, J., Hall, N. and Paul, L. A., eds. Cambridge, MA: MIT Press, 2004. 75-106.

Mackie, J. L. "Review of *Causality and Determinism*." *Journal of Philosophy* 73 (1976): 213-18.

---. *The Cement of the Universe*. Oxford: Oxford University Press, 1980.

Mandel, David R. "Mental Stimulation and the Nexus of Causal and Counterfactual Explanation." *Understanding Counterfactuals, Understanding Causation*. Hoerl, C., McCormack, T. and Beck, S. R., eds. Oxford: Oxford University Press, 2011. 147-70.

Mandel, David R., and Lehman, Darrin R. "Counterfactual Thinking and Ascriptions of Cause and Preventability." *Journal of Personality and Social Psychology* 71 (1996): 450-63.

Maslen, Cei. "Causes, Contrasts, and the Nontransitivity of Causation." *Causation and Counterfactuals*. Collins, J., Hall, N. and Paul, L. A., eds. Cambridge, MA: MIT Press, 2004. 341-58.

Maudlin, Tim. "Causation, Counterfactuals, and the Third Factor." *Causation and Counterfactuals*. Collins, J., Hall, N. and Paul, L. A., eds. Cambridge, MA, 2004. 419-43.

McGrath, Sarah. "Causation by Omission: A Dilemma." *Philosophical Studies* 123 (2005).

Mellor, D. H. "For Facts as Causes and Effects." *Causation and Counterfactuals*. Collins, J., Hall, N. and Paul, L. A., eds. Cambridge, MA: MIT Press, 2004. 309-24.

Menzies, Peter. "Difference-Making in Context." *Causation and Counterfactuals*. Collins, J., Hall, N. and Paul, L. A., eds. Cambridge, MA: MIT Press, 2004. 139-80.

---. "Causation in Context." *Causation, Physics, and the Constitution of Reality*. Price, H. and Corry, R., eds. Oxford: Clarendon Press, 2007. 191-223.

---. "Platitudes and Counterexamples." *The Oxford Handbook of Causation*. Beebee, H., Hitchcock, C. and Menzies, P., eds. Oxford: Oxford University Press, 2009. 341-67.

---. "Peter Menzies." *Metaphysics: 5 Questions*. Steglich-Petersen, A., ed. Copenhagen: Automatic Press, 2010. 39-52.

Menzies, Peter, and Price, Huw. "Causation as a Secondary Quality." *The British Journal for the Philosophy of Science* 44 (1993): 187-203.

Pearl, Judea. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2000.

Price, Huw. "The Flow of Time." *The Oxford Handbook of Time*. Callender, C., ed. Oxford: Oxford University Press, 2001. 276-311.

---. "Causal Perspectivalism." *Causation, Physics and the Constitution of Reality*. Price, H. and Corry, R., eds. Oxford: Clarendon Press, 2007. 250-92.

---. *Naturalism without Mirrors*. Oxford: Oxford University Press, 2011.

---. "Causation, Intervention and Agency—Woodward on Menzies and Price." 2012. ms.

Russell, Bertrand. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13 (1912-1913): 1-26.

Salmon, Wesley C. "Causation without Counterfactuals." *Philosophy of Science* 61 (1994): 297-312.

Schaffer, Jonathan. "The Metaphysics of Causation." *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition).

---. "Contrastive Causation." *Philosophical Review* 114 (2005): 327-58.

---. "Contrastive Causation in the Law." *Legal Theory* 16 (2010): 257-97.

---. "Causal Contextualism." *Contrastivism in Philosophy*. Blaauw, M., ed. Hoboken: Taylor and Francis, 2013. 35-63.

Steward, Helen. *The Ontology of Mind*. Oxford: Clarendon Press, 1997.

Von Wright, Georg Henrik. *Causality and Determinism*. New York and London: Columbia University Press, 1974.

Woodward, James. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press, 2003.

---. "Psychological Studies of Causal and Counterfactual Reasoning." *Understanding Counterfactuals, Understanding Causation*. Hoerl, C., McCormack, T. and Beck, S. R., eds. Oxford: Oxford University Press, 2011. 16-53.