

A New Normal for Accounts of Causation

A thesis submitted to The University of
Manchester for the degree of
Doctor of Philosophy
in the Faculty of Humanities

2021

Jonathan H Bebb
School of Social Sciences/Philosophy

Contents

Abstract.....	5
Declaration.....	6
Copyright Statement.....	6
Acknowledgements.....	7
Introduction	11
Chapter 1 – The Contextual Question	22
1. The Core of Contextualism.....	27
2. The Core of Contrastivism.....	32
3. Schaffer’s Points of Difference.....	37
4. Demarcating Contextualism and Contrastivism	45
5. The Implications of the Contextual Question	55
Chapter 2 – The Normative Question	57
1. What is a Normative Account of Causation?	59
2. The Puzzle Cases	65
3. Two Competing Strategies	76

Chapter 3 – The Pragmatics-and-Biases Strategy	80
1. The First Branch: Pragmatics	81
2. The Second Branch: Biases	104
3. Moving to the Competence Strategy	127
Chapter 4 – Developing the Competence Strategy	128
1. The Relevance of Counterfactual Scenarios and Possible Worlds	129
2. Normal Courses and Difference-Making.....	140
3. Causal Strength	147
4. The Unclarities Objection.....	158
Chapter 5 – Arguing for a Univocal Account of Normality	164
1. The Orthodox View	166
2. Reasons For and Against the Orthodoxy	172
3. Ambiguity Tests.....	174
4. A Moral Reason to Prefer an Alternative to Orthodoxy	181
5. Explaining the Bear & Knobe Data	188
6. Moving Past the Orthodox View	200
Chapter 6 – A Normative Account of Causation	202

1. The Dual-Axis View of Normality	203
2. Returning to Causation	233
3. Answering the Big Questions	252
Chapter 7 – Objections and Replies	254
1. The Natural Relation Objection	254
2. Normal Events as Causes and Effects	269
3. Is this all Completely Unfalsifiable?	273
4. The Indeterminacy Objection	277
Conclusion.....	283
Bibliography	285

Word Count – 79,950

Abstract

The aim of this thesis is to address a gap within many recent accounts of causation that take the truth-values of causal statements to be sensitive to normative considerations such as whether or not an event is normal. These accounts are intended to provide a solution to the problem of the apparent arbitrariness of our selecting some events as causes while others are relegated to being mere background conditions. This problem arises when we consider the intuitions we have about certain puzzle cases, which seem like they can be explained if we adopt a normative account of causation by holding that the truth-values of our causal statements are sensitive to what is and is not normal (Hitchcock and Knobe 2009) (Halpern and Hitchcock 2015) (Menzies 2007) (McGrath 2005). There are other ways of explaining these intuitions that do not commit us to the view that the truth-values of our causal statements are sensitive to facts about normality, however those alternative explanations generally do not succeed at providing a *better* explanation than the one offered by the normative account. We therefore have good reason to prefer a normative account of causation.

Unfortunately, although these normative accounts of causation do provide some explanation for our intuitions about the puzzle cases, they also suffer from a number of problematic unclarities that result from those accounts not containing any kind of conceptual analysis of normality. I will address this issue by developing my own original analysis of normality that I then go on to apply to causation. The end product of this will be a new normative constraint on accounts of causation.

Declaration

I declare that that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

- I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- IV. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=2442> 0), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see <http://www.library.manchester.ac.uk/about/regulations/>) and in The University’s policy on Presentation of Theses.

Acknowledgements

Given the all the discussion of counterfactual reasoning in what follows, I think it is only right to acknowledge the people without whom my thesis would be either non-existent or a much worse and less interesting version of itself. Firstly, I would like to thank my supervisor, Helen Beebee. Helen has, over the last four years, offered me mountains of helpful and well-thought-out comments and suggestions. Whenever I was having trouble with answering a particular objection or articulating an important point, I could always rely on her to provide some kind of illuminating explanation of what I might want to say and how best to say it. I have no doubt that all of her input has considerably improved the overall quality of my thesis. I would also like to thank Ann Whittle for her help as my second supervisor and for previously supervising my Masters dissertation. Her feedback was always extremely helpful, and she was frequently able to direct me towards parts of the literature that I had somehow managed to overlook. My thesis would certainly be worse off without her.

Next I would like to thank my parents, Celia and Garry Bebb. Growing up, they were always very keen to encourage me to think critically about the world and to make me feel confident in expressing my own ideas. There was never any question that a PhD was something that I could do and that they would support me completely if that were something I wanted to do. It is only fairly recently that I have come to appreciate how truly lucky I am to have had them as parents, and without them, there is no way I would have been able to achieve all that I have.

Finally, I would like to thank Leonie Smith, for helping me to recognise and understand the impact that I have made in the world and that I could make in future. I know that without her I would not have had the confidence or the will to stand by some of the claims that I wanted to make as part of this thesis. Her love and support for the last two years have allowed me to appreciate the value I bring to the world when I work hard at something, and that is something I will treasure for the rest of my life.

In memory of Gus Lloyd

Blank Page

Introduction

When common sense delivers a firm and uncontroversial answer about a not-too-far-fetched case, theory had better agree. If an analysis of causation does not deliver the common-sense answer, that is bad trouble. (Lewis 1986a, 194)

This remark from David Lewis serves as a clear reminder of the importance that adhering to common sense intuitions has for metaphysical theories, as well as for philosophy in general. Of course, we cannot possibly account for *all* the intuitions that we may have about different cases, especially when they contradict one another. It is with these instances in mind that Lewis continues by saying:

But when common sense falls into indecision or controversy, or when it is reasonable to suspect that far-fetched cases are being judged by false analogy to commonplace ones, then theory may safely say what it likes. (*ibid.*)

In this thesis I address one specific set of intuitions that we have towards certain cases involving causation. These intuitions are frequently pointed to as evidence that the truth-values of certain causal statements are sensitive to contextual factors and certain normative considerations, such as whether or not the events referred to by the statement are normal. This then raises the question of whether these are intuitions that our theories of causation ought to find a way to agree with, or if our theories can safely say what they like about them. Many, including Lewis himself, have argued for the latter, claiming that these intuitions originate from some mistake or from some pragmatic consideration regarding the utterance of causal statements. Therefore, these intuitions do not and cannot tell us anything about what it means to say that one thing caused another.

My thesis will, in part, be a response to this view. I will argue against the attempt to characterise these intuitions as a mistake or to explain them away by appeal to pragmatics, and I will offer an alternative understanding that will allow our theories to deliver a common-sense answer.

The analysis I offer here is certainly not the first attempt at trying to bring our theories into agreement with these intuitions. However, my analysis does make a significant contribution by addressing a gap found in the existing accounts. Specifically, this gap comes from the fact that many of these accounts try to explain the intuitions in question by appealing to some normative concept – the most common being normality or defaults – without first providing any kind of analysis of whatever normative concept they invoke.

I will remedy this by offering an analysis of normality, which I will then go on to apply to causation, resulting in a theory that can accommodate the intuitions we are interested in. I thus aim to give a complete analysis of how normality shapes our causal judgements, and to argue that our judgements being shaped in this way is not a result of some misunderstanding of how the concept of causation ought to be applied. Rather, such judgements are instances of the concept being applied competently.

Most of the authors who have written about normality's effect on our causal judgements have done so within the realm of 'token' causation, also known as 'actual' or 'singular' causation. That is causal statements of the form 'c caused e' where 'c' and 'e' refer to specific events (or whatever else we take the relata of causation to be, which is an issue I will address shortly). These are the kinds of causal statements that we typically have in mind when discussing causation. Examples of these statements include things like 'the door slamming shut caused

me to jump', 'the bird flapping its wings caused it to be lifted off the ground', or 'the assassination of Archduke Franz Ferdinand caused the First World War'. What all these statements have in common is that they refer directly to singular, token events. This is often contrasted with statements of 'type' or 'general' causation, which take the form 'C causes E', where 'C' and 'E' refer to event-types. A paradigmatic example of this is the statement 'smoking causes cancer'.

Any talk of general/type causation would involve failing to engage with the problem in the manner that it is currently understood and would introduce many additional complexities. For this reason, as a means of ensuring that the scope of the thesis remains appropriately narrow and directed at the area where discussions of this issue most often take place, I will be limiting my focus to singular/token causation and will henceforth be using 'causation' to mean this singular/token version of causation. In addition to this, I will also use the phrase 'causal statements' to refer to the statements of the form 'c caused e' (unless otherwise specified) that make use of this singular/token concept.

When it comes to understanding this singular/token kind of causation, I will also be assuming that a sufficient (though possibly not necessary) condition for the truth of 'c caused e' is for e to counterfactually depend on c. To claim that e counterfactually depends on c is to assert that the following counterfactual holds:

If c had not occurred then e would not have occurred.

So, under the assumption that counterfactual dependence is sufficient for causation, this counterfactual being true would mean that 'c caused e' is also true.

By making this assumption, I find myself in agreement with the counterfactual account of causation, which is the view that causal relations are reducible to relations of counterfactual dependence. Lewis (1973) provides the most influential description of such an account, where he also makes the claim that e 's being counterfactually dependent on c is a sufficient (but definitely not necessary) condition for the truth of ' c caused e '. An advantage of such a view is that it allows us to do justice to the intuition that a cause somehow makes the difference between the occurrence and non-occurrence of its effect. However, there are some problems with this approach, most notably when it comes to how it handles cases of preemption, where there is causation without counterfactual dependence, and transitivity, where there is counterfactual dependence without causation (Menzies and Beebe 2020). More recently, proponents of the counterfactual view have sought to address some of these issues using a causal modelling framework, though it is unclear whether or not this has been achieved. I should be clear that I am not aiming to do the same and to offer a defence of the view that causation can be *reduced* to counterfactual dependence. Instead, I will be proceeding on the assumption that counterfactual dependence is sufficient for causation because it will enable us to better explore and respond to a particular problem that can be raised against any theory of causation that shares in this assumption.

The problem that I will address concerns the apparent arbitrariness of our selecting some events as causes while others are relegated to being mere background conditions. This problem can be found in cases where some effect counterfactually depends on two or more other events but not all of these events can be acceptably described as having caused the effect. In a case where e counterfactually depends on both c and d , we might assent to the claim that ' c caused e ' but refrain from the claim that ' d caused e '. The challenge posed by

this kind of case is to provide an explanation for why our intuitions about the acceptability of these causal statements can differ in this way. I will provide a more detailed discussion of these cases in subsequent chapters but suffice it to say many have argued that our intuitions about these cases are best explained by holding that our causal selection is (at least partly) determined by what is and isn't normal. My aim is to defend a version of this solution to the problem.

It should also be noted that the solution I will provide presents us with two plausible reasons for preferring a broadly counterfactual analysis. One is that some of the cases that best exemplify the problem are cases of causation by omission/absence. We have good reason to think that such cases do exist (Schaffer 2000) (Lewis 2004), yet their existence presents various problems, including the one discussed here, that are best solved by a counterfactual account. The second reason is that, for reasons that will emerge in Chapter 7, the kind of contextualist and normative analysis I will be defending here fits in much better with a counterfactual account than an account that comes with much stronger ontological commitments.

One final preliminary point concerns the issue of what the relata of causation are. So far, I have been presupposing that causal relations hold between events. However, this assumption is far from uncontroversial, with many theorists arguing for some alternative kind of relata. The suggested alternatives have included things like facts (Bennett 1988) (Mellor 1995), states of affairs (Armstrong 1997), features (Dretske 1977), tropes (Campbell 1990), aspects (Paul 2000), and many more besides. In what follows, I will remain neutral on the question of which of these views are the correct one. For the sake of simplicity, I will refer to the relata of

causation as 'events', although the analysis I offer should also be compatible with alternative accounts with the appropriate modifications. The only assumption I will be making regarding the relation is that I will allow for the possibility that omissions/absences can be causes since, as noted above, some of the problem cases I will examine assume that this is the case.

The thesis is divided up into 7 chapters. Chapters 1 and 2 will be focused on providing some of the necessary background for my analysis. I will begin by identifying two key questions that I will attempt to answer over the course of the thesis. One question concerns whether or not the truth-values of causal statements are sensitive to context, while the other concerns whether or not they are sensitive to normative considerations, like whether or not a given event is normal. Explaining what is meant by the former question will be the goal of Chapter 1, while Chapter 2 will do the same for the second question.

To address the context-sensitivity question in Chapter 1 I will be examining the philosophical approach to conceptual analysis known as contextualism. According to this approach, statements that refer to certain target concepts, including the concept of causation, are context-sensitive. The aim of this chapter is to identify the core claims of contextualism, as well as the main motivation for adopting such an approach. I will achieve this by distinguishing it from contrastivism: another approach to conceptual analysis that is often mistakenly viewed as a mere variant of contextualism. To that end I will begin by identifying the core claims of contextualism and how they are cashed out in certain domains, including that of causation. I will then do the same for contrastivism.

I shall then examine an existing attempt to demarcate contextualism and contrastivism provided by Jonathan Schaffer (2004). I will show that, although his argument points us in the

right direction, it somewhat misses the mark when it comes to identifying a metaphysically significant distinction between the two approaches. I then go on to present my own distinction, which is one that does carry with it some metaphysically significant consequences. Specifically, I will argue that contrastivism is compatible with the intuition that causation is a natural relation – it is a relation that is ‘out there’ in the world – while contextualism is incompatible with that intuition. Much of the opposition to contextualism about causation, I shall claim, can be seen as motivated by the wish to uphold this intuition.

In Chapter 2 I will then turn my attention to the second key question, concerning whether or not causal statements are sensitive to normative considerations. I will begin by explaining what is meant by this question and why I take the relevant considerations to be those of normality. From there, I will lay out the case for why we might think that the truth-values of causal statements are sensitive to what is and isn’t normal. For this, I shall lay out a series of four puzzle cases that have all previously been used to argue that causal statements are sensitive to normality in this way. All these puzzle cases involve a pair of causal statements, of the form ‘c caused e’ and ‘d caused e’, such that e counterfactually depends on both c and d but intuitively one of the statements is acceptable while the other is not. For someone who argues that such statements are sensitive to claims about normality, these intuitions are best explained by appealing to that sensitivity. This will also be tied in with the first question since something’s being judged as normal is often taken to be a feature of the context in which that judgement is made. Thus, we have an argument for a contextualist account of causation where the truth-values of causal statements are sensitive to claims about normality.

To conclude Chapter 2, drawing on Blanchard & Schaffer (2017), I shall identify two distinct strategies that could be used to explain our intuitive responses to the puzzle cases: the Competence Strategy and the Pragmatics-and-Biases Strategy. The former strategy accepts the argument that the concept of causation is both contextual and normative; hence the intuitions we have about the puzzle cases are the result of a competent application of the concept of causation. Someone who pursues this strategy must then explain the mechanism by which claims about normality come to determine the truth-values of causal statements. By contrast, according to the Pragmatics-and-Biases Strategy our intuitions about the puzzle cases involve some sort of confusion or only occur at the level of pragmatics. This means that someone who follows this strategy will have to provide an explanation for how we come to have these intuitions without also claiming that these intuitions reflect the actual truth-values of the statements in question. Such explanations tend to appeal either to some cognitive bias or to some pragmatic story. In Chapter 3 I will argue that we ought to avoid the Pragmatics-and-Biases Strategy, while in Chapter 4 I will examine how the existing Competence Strategy accounts could be improved upon.

In Chapter 3, I will show how the Pragmatics-and-Biases Strategy has so far failed to provide a compelling reason to prefer it to the Competence Strategy by examining three Pragmatics-and-Biases accounts and showing how they fail to explain our intuitions about the puzzle cases. The first two accounts, offered by Lewis (2000) and Montminy & Russo (2016), fall into the 'pragmatics' camp. I shall argue that Lewis's fails to capture the strength of our intuitions, while Montminy & Russo fail to offer a unified account that encompasses all of our intuitions about the various cases. In the biases camp we have the explanation offered by Blanchard & Schaffer (2017), which appeals to Kahneman & Tversky's work on the availability heuristic

(1973). I argue that Blanchard & Schaffer's argument fails because they do not provide proper support for their two premises. I conclude that the followers of the Pragmatics-and-Biases Strategy have failed to provide a reason to prefer it over the Competence Strategy.

Having made the case against the Pragmatics-and-Biases Strategy, in Chapter 4 I turn my attention to the Competence Strategy. I examine three different attempts to explain how claims about normality come to determine the truth-values of causal statements. I begin by examining Hitchcock & Knobe's (2009) view that normality determines which counterfactual scenarios are relevant when coming to a causal judgement. I then look at Menzies' account (2007) (2009), according to which an event is taken to have caused some effect only if it makes the difference between the actual course of events for a given causal system and a normal course of events for that system. Finally, I examine a view suggested by Icard et al. (2017), according to which the normality of a given event determines how much causal strength that event is taken to have with respect to some effect. As I will demonstrate, a flaw that all of these accounts share is that they do not attempt to offer any kind of analysis of normality. This is a problem because, as argued by Blanchard & Schaffer (2017, 192), it opens the door to several problematic unclarities. From this I conclude that there is a need to provide an analysis of normality that can be deployed in response to the selection problem raised in Chapter 2.

With the need for an analysis of normality clearly demonstrated, in Chapter 5 I begin to offer such an analysis. The aim of this chapter will be to argue against the commonly held view that 'normal' is an ambiguous term, and instead provide some motivation for a univocal account of normality. I will argue that there are three reasons why it is a mistake to prefer the view

that 'normal' is ambiguous over the alternative. Firstly, existing ambiguity tests do not provide any evidence to support the claim that 'normal' is ambiguous. Secondly, there are plausible moral reasons for preferring a univocal account. Finally, recent experimental data offered by Bear & Knobe (2017) seems to suggest the existence of a generalised sense of 'normal' that would be difficult to account for on the view that the term is ambiguous. I conclude that there are clear grounds for preferring a univocal account of 'normal'. Such an account also has the advantage that it allows us to eliminate the unclarities from the existing Competence Strategy accounts.

Given that we now have good reason to develop a univocal account of normality, in the first half of Chapter 6 I will do just that. I call the account I offer a 'dual-axis' view. Here, what is normal will be shaped along two competing dimensions of prescriptive and descriptive considerations. These dimensions will be given different weights depending on the context, so statements about normality will be context-sensitive. Having developed this account, I will then use it to generate a plausible condition on causation that can be used to solve the puzzle cases from Chapter 2. The end result will be a contextual constraint on causation, and specifically an account of how contextually-determined facts about normality determine the truth-values of causal statements.

In the final chapter, Chapter 7, I defend this analysis by considering and responding to four distinct (but also related) possible objections. The first objection is that by abandoning the intuition that causation is a natural relation, the cost of entry for my account is far too high. I respond that this is not a serious problem if we consider the role that our concept of causation is supposed to play. The second objection is that my account does not seem to allow for the

possibility that normal events can cause other events or be caused themselves. Here I contend that this problem can be avoided by paying close attention to whatever reference class is salient in a context where it seems to be true that a normal event caused/was caused by something else. The third objection concerns the worry that the account is unfalsifiable in the sense that it allows an advocate who is arguing in bad faith to wave away any potential counterexample by simply appealing to some feature of the context at play in that example. In response to this I demonstrate how such responses are avoidable, as long as any putative counterexamples raised against the theory are clear about which contextual features are to be held fixed. Finally, there may also be a worry that the account admits too much indeterminacy. Here I offer some ways that any indeterminacy could be avoided, but also note that this is to be expected when our aim is to provide an account that agrees with folk intuitions.

I will then end by concluding that we ought to adopt the normative constraint on causation provided in this thesis. Doing so will help to address the problem of explaining the role that normality plays in our causal judgements.

Chapter 1 – The Contextual Question

To give a complete analysis of how normality shapes our causal judgements, I will need to answer two specific questions:

- I. Is causation normative?
- II. Is causation contextual?

With the question I, I mean something like ‘are the truth-values of causal statements sensitive to normative considerations?’, where the phrase ‘normative considerations’ refers to claims or facts involving normative concepts like normality or defaults. Meanwhile, question II asks us to consider whether or not the truth-values of causal statements are sensitive to the contexts within which they are used. My answer to I is that causation is normative because the truth-values of causal statements are sensitive to claims about what is and is not normal. This then leads me to my answer for II, which is that causal statements are context-sensitive in virtue of the fact that what’s normal is also context-sensitive. The precise details of these two answers will form the core of my analysis into the role that normality plays in an account of causation.

These two questions are closely related and have, unfortunately, often been understood as being one and the same. As such, a necessary first step in my analysis of the way that normality shapes our causal judgements will be to separate them out, whilst also taking the time to examine what exactly these questions are asking, what answers we can give to them, and why we may prefer one answer over another. To this end, I will consider each question individually, with this chapter focusing on question II, and Chapter 2 addressing question I.

My reason for beginning with question II is that it will allow me to introduce some of the key positions that will need to be understood before I am able to proceed with my argument.

With this in mind, we can start by asking what it means to say that a statement is context-sensitive. To answer this, it will be instructive to consider how this question is tackled in the literature. Generally speaking, talk of context-sensitivity usually centres on discussion of a particular approach to conceptual analysis within philosophy: contextualism. This approach involves regarding the statements that express certain target concepts as being context-sensitive, and many have applied this approach to the concept of causation. This approach is opposed by invariantism, which takes the view that the statements in question are not context-sensitive. This means that an invariantist about causation would answer question (II) with a resounding 'no', which is a complete rejection of the contextualist position. So, it should hopefully be clear by now that to give an informed answer to question (II), we need to understand what it means to be a contextualist about causation. An effective way to accomplish this would be to consider what it means to be a contextualist in general.

The problem with this is that there is currently a great deal of confusion around contextualism, specifically regarding its relationship to another major approach to the conceptual analysis of certain concepts, including causation: contrastivism. Indeed, contrastivism is regularly and, I contend, inaccurately presented as a variant of contextualism, rather than a distinct position. Contrastivist accounts are regularly viewed as proposing the same core claims as any contextualist account that appeals to a notion of contextually relevant alternatives. This presents a problem with understanding contextualism about causation (or contextualism within any domain for that matter) because it dilutes the core

contextualist claim and makes it more difficult to see why we would want to be a contextualist in the first place. As such, this chapter will focus on identifying both the contextualist and contrastivist positions, before explaining why these two approaches ought to be understood as distinct from one another, in that they carry significantly different metaphysical consequences.

It is worth noting that I am not the first person to try to distinguish between these two approaches. In his 2004 paper 'From Contextualism to Contrastivism', Jonathan Schaffer attempts to argue for such a distinction, albeit with regards to knowledge rather than causation. In the paper, Schaffer presents his correspondence with Lewis, who argues for a contextualist view about knowledge that appeals to relevant alternatives and quips that:

The only thing we disagree about is whether we disagree. (Schaffer 2004, 97)

This scepticism towards the contextualist/contrastivist distinction was shared by Neta, who described Schaffer's account as:

... one of the most explicit, comprehensive, and thoroughly defended *contextualist* theories of knowledge to date. (Schaffer 2004, 97)

Despite this resistance, in his 2004 paper Schaffer perseveres with the distinction, outlining three alleged differences between his contrastivist account and the accounts proposed by the contextualists. He concludes that, while the two views are 'sibling theories', contrastivism provides a 'more suitable ... model [of the context dependence of knowledge statements], which preserves the core contextualist insights while resolving the main objections to contextualism' (2004, 97). From this, it is clear that in 2004, Schaffer saw contrastivism as a

position that was distinct from and superior to the kind of contextualist account advanced by Lewis and his ilk, despite there being some shared elements between the two views.

Yet in his more recent works on the topic, we can see that even Schaffer has abandoned this commitment to the distinction between contextualism and contrastivism. His 2012 co-authored paper with Knobe now refers to a more developed version of his 2004 account of knowledge as a ‘form of contextualism’ (2012, 676). We see a similar dialling back in Schaffer’s work on causation. In ‘Contrastive Causation’ (2005), we are presented with an explicitly contrastivist account of causation, with little mention being made of contextualism¹. But by ‘Causal Contextualism’ (2012), the same contrastivist account is explicitly presented as being a kind of contextualism.

This move by Schaffer away from his 2004 view was, I suggest, a mistake – one which has not been identified as such, including by Schaffer himself. There *is* a metaphysically significant difference between the kind of contextualism that Lewis argues for and the contrastivism that Schaffer argues for, and so it is a mistake to paint contrastivism as a mere brand of contextualism. The difference is that the contextualism is incompatible with certain naturalist intuitions about the metaphysical nature of the referent of the relevant concept, while

¹ Contextualism is, in fact, only mentioned twice. The first is when Schaffer notes that the framework provided by his contrastive account of causation is ‘roughly analogous to that of epistemic contextualism’ (2005, 315) in virtue of the fact that they both integrate ‘relevant alternatives into the relation’ (*ibid.*). Here Schaffer is referring to epistemic contextualist accounts like Lewis’s, where the contextual salience of different epistemic possibilities will affect whether or not a given knowledge claim is true. The fact that Schaffer considers these views as analogous suggests that he recognises there are some similarities between the two views. However, he also states that epistemic contrastivism would make for a better analogy with causal contrastivism, which clearly shows that he still maintains a distinction between contextualism and contrastivism. The second time that contextualism is mentioned in this paper is in an endnote where he notes that the causal contextualist claim that ‘cause’ is context-sensitive is more well supported than the epistemic contextualist claim that ‘knows’ is context-sensitive. He does not say any more than this, so it is unclear from this what he takes the relationship between contextualism and contrastivism to be.

contrastivism *is* compatible with these intuitions. The intuitions in question concern the referent of the concept's existence being an ontological matter and something that is independent of context. This difference in compatibility with these intuitions fundamentally affects our motivations for choosing one position over the other. Those who wish to uphold these naturalist intuitions will thus have a good reason to prefer a contrastivist account over a contextualist account. So, by understanding and accepting that contrastivism is a distinct approach from contextualism, we can achieve a new understanding of what it is that we are buying into when we decide to endorse one approach rather than the other.

In what follows I will aim to demonstrate the existence of this difference between contextualism and contrastivism, and to argue that a demarcation should be made between the two approaches. This chapter will be divided into four sections and will proceed as follows. In section one I will provide a brief overview of the contextualist approaches to various different areas of philosophy and identify a core set of commitments that are common to all these approaches, but which have largely been left unclear in previous literature. Particular attention will be paid to the contextualist accounts that rely on a notion of relevant alternatives. In section two I shall provide the same overview for contrastivism. In section three I will begin to explain how these two views differ by drawing on the three differences that Schaffer provided in his 2004 paper, arguing that Schaffer's distinctions do not quite capture the extent of these differences. Finally, in section four, I will develop my positive account of the distinction further and demonstrate that there is an important metaphysical distinction to be made between the two views – one that has important implications for contextualism about causation.

1. The Core of Contextualism

In the literature, the term ‘contextualist’ denotes a particular approach to conceptual analysis, according to which the truth of statements involving terms that express the target concept are context-sensitive. What is meant by ‘context-sensitive’ here is that these statements could, in one context, express a proposition that is true, and in a different context they could express a proposition that is false. The concepts that this approach has been applied to are both numerous and varied. As I have already mentioned, there have been contextualist accounts of causation, with the most notable being found in the works of Peter Menzies (2004a) (2004b) (2007) (2009). Additionally, there have also been contextualist accounts of knowledge, modality (Kratzer 1977), quantifiers (Westerståhl 1989) (Stanley and Gendler Szabó 2000), gradable adjectives (Kennedy, Vagueness and Grammer: The Semantics of Relative and Absolute Gradable Adjectives 2007), morality (MacFarlane 2007), vague expressions (Kamp 1981), and many more.

Contextualists generally limit their focus to a single concept, so they are only interested in demonstrating that statements containing terms that are an expression of that concept are context-sensitive. To give some examples of this: contextualist about causation argue that statements containing the term ‘caused’ are context-sensitive; contextualists about knowledge argue that statements containing the term ‘knows’ are context-sensitive; contextualists about morality do the same with statements containing terms like ‘ought’, ‘good’, or ‘virtue’; contextualists about gradable adjectives focus on terms like ‘tall’, ‘smooth’, or ‘hot’; and for contextualists about modality, it’s terms like ‘possible’ and ‘necessary’. These are terms that are importantly related to the target concept and are typically used to

predicate something about the subject(s) of a statement. I will henceforth refer to such terms as 'c-terms'. The contextualist about concept x is thus attempting to argue that statements containing c-terms that express x are context-sensitive.

There are many different ways of cashing out this claim that c-term statements are context-sensitive. For my purposes, I will only need to focus on those accounts that are likely to be viewed as examples of both contextualism and contrastivism. These are accounts where the context-sensitivity of statements containing c-terms is explained by appeal to the contextual relevance of possible alternatives to one or more of those statements' subjects. To make this clearer, let's examine some actual examples of contextualist accounts. I will of course discuss Menzies' contextualist account of causation, but to begin with, we shall examine how contextualism has been applied to the case of knowledge, as this is much more straightforward.

Contextualism about knowledge is a fairly common position, and includes amongst its supporters the likes of Williams (1991), DeRose (1992) (1995) (1999), Lewis (1996), Cohen (1988) (1998) (1999), Heller (1999), and Neta (2003a) (2003b). All these theorists endorse the claim that statements containing the c-term 'knows' are context-sensitive. In other words, they hold that a statement of the form 'S knows that p ' can express a true proposition in one context and a false proposition in a different context.

Lewis's contextualist analysis of knowledge is a paradigm example of a contextualist account that appeals to a notion of relevant alternatives. On his account, 'S knows that p ' is true only if S can use the evidence available to them in order to rule out all relevant alternatives to p (1996, 551-553). To see how this works, consider the following statement:

1) Moore knows that he has hands².

The traditional invariantist view of statements such as these is that they express the same proposition with the same truth-value in all contexts. Contextualists usually respond by stating that (1) expresses something true when uttered in an everyday context, but in the context of an epistemology seminar, for example – where the bar for what counts as knowledge is much higher and sceptical scenarios are taken more seriously – it expresses something false.

On Lewis's contextualist account, (1) is true only if Moore is able to eliminate all *relevant alternatives* where he does not have hands. Alternatives include possibilities such as Moore's having stumps instead of hands and Moore's being a brain in a vat. The relevance of an alternative is determined by contextual factors. In an everyday context, the more straightforward alternatives like Moore's having stumps will be relevant, while the more extreme alternatives, like his being a brain in a vat, will not be relevant. So, in this everyday context, Moore only needs to eliminate the possibility that he has stumps for (1) to be true. Since Moore can clearly see that he does not have stumps, this possibility is eliminated by the evidence. The possibility that he is a brain in a vat cannot be eliminated through the use of perceptual evidence, but this has no effect on the truth-value of (1) because this alternative is not contextually relevant. However, it *is* a relevant alternative in the context of the epistemology seminar. As a result, the fact that this alternative cannot be eliminated by the available evidence renders (1) false in this context. So, on Lewis's account, whether or not a

² Example taken from (Schaffer 2004, 73).

given knowledge statement expresses something true will (at least partly) depend on which alternatives are relevant in the context of utterance.

Contextualist accounts that appeal to a notion of contextually relevant alternatives are not just limited to the domain of knowledge. So, let's now turn to the contextualist account of causation offered by Peter Menzies (2007) (2009), who argues that the truth-values of causal statements are relative to contextual orderings of possible worlds based on normality (2007, 826-828). Put simply, the truth or falsity of a given causal statement will depend on what we consider to be the 'normal state' of the causal system that we are interested in, as this will be the relevant counterfactual scenario/alternative state that we appeal to when evaluating a given causal claim. A cause is understood as the event(s) that makes the difference between the system being in its normal state and its actual state.

To illustrate, imagine a lightning strike occurs in the middle of a forest, leading to a devastating forest fire³. The forest fire is counterfactually dependent on the lightning strike, but it is also counterfactually dependent on the presence of oxygen in the atmosphere. Assuming a counterfactual account of causation, we are provided with two candidate causes, each expressible as one of the following causal statements:

- 1) The lightning strike caused the forest fire.
- 2) The presence of oxygen caused the forest fire.

A group of forest rangers are attempting to figure out what caused *this* fire in order to prevent future fires. In this context, (2) seems intuitively acceptable, while (3) seems intuitively

³ This version of the example is from Schaffer (2012, 42).

unacceptable. So, the lightning strike gets to count as having caused the forest fire and the presence of oxygen is relegated to just being a mere condition for the occurrence of the fire.

Now consider a second context of assessment, where a group of Venusians observe both the lightning strike and the subsequent forest fire from space⁴. Lightning strikes do in fact occur on Venus, and for the purposes of the example we can suppose that they are far more common than on Earth. Another important feature of Venus is that it has no oxygen in its atmosphere, and consequently no fires occur on its surface. So, the Venusians will be surprised at the sight of the forest fire and will want to know how it happened, since all previously observed (by them) lightning strikes produced no such thing. In this context, (3) seems to be far more acceptable than it was in the context of the forest rangers, and (2) seems to be far less acceptable.

Contextualists like Menzies argue their account best explains the difference in the intuitive acceptability of causal statements that we find in cases like this. (2) is acceptable in the context of the forest rangers and unacceptable in the context of the Venusians because it is true in the former and false in the latter. Likewise, (3) is true in the context of the Venusians and false in the context of the forest rangers. On Menzies account⁵, these differences in the truth-values of (2) and (3) are a result of the fact that in different contexts, a different alternative state of the causal system is relevant. In the context of the forest rangers, the normal/relevant alternative state is one where oxygen is present, but there is no lightning

⁴ Inspired by (Putnam 1982, 150).

⁵ Menzies himself discusses a structurally isomorphic case in (Menzies 2007, 209-211).

strike and consequently no forest fire. Here, the lightning strike makes the difference between the relevant alternative state of the system and the actual state of the system, while the presence of oxygen does not. So, in this context, (2) is true and (3) is false. Yet in the context of the Venusians, the normal/relevant alternative state of the system is one where there *is* a lightning strike but no oxygen, resulting in no forest fire. Here the presence of oxygen *does* make the difference between the relevant alternative state of the system and the actual state of the system, while the lightning strike does not. Consequently, in this context, (3) is true and (2) is false. From this it should be clear that, on Menzies' account, the truth or falsity of any given causal statement will depend on contextually relevant alternative states of the causal system in question, with the relevant states being those that are the most normal.

The lesson here is that all contextualists agree on one thing: that whether or not a statement containing a c-term expresses something true will always be context-sensitive. This context-sensitivity can be explained in many different ways, with one being that it is the result of whether or not non-actual alternatives to the subject of those c-term statements are, in the context, *relevant* alternatives. Contrastivism also invokes alternatives to explain the context-sensitivity of certain c-term statements. As a result of this, contrastivist accounts are often *mistakenly* labelled as contextualist. Before examining where the error lies, we first need to be clear on how contrastivist accounts themselves make use of alternatives in their analyses.

2. The Core of Contrastivism

I will begin by examining the ways in which contrastivism is uncontroversially similar to contextualism. The most apparent point of similarity is that contrastivism is also largely used

as a way of analysing the c-term statements of a target concept that involves a kind of context-sensitivity. As with contextualism, this approach has been applied to a wide range of concepts, including causation (Hitchcock 1996a) (Schaffer 2005) (Schaffer 2012) (Northcott 2008) (Maslen 2004), knowledge (Schaffer 2004) (Schaffer and Knobe 2012), explanation (van Fraassen 1980) (Garfinkel 1981) (Lipton 1990) (Hitchcock 1996b), grounding (Schaffer 2016), obligation (Snedegar 2017), moral luck (Driver 2015), and more. The contrastivist's main claim is that the c-term statements that we typically use are context-sensitive. Specifically, they are sensitive to the different alternatives that may or may not be contextually salient. With this in mind, it becomes easy to see how the likes of Lewis and Neta might assume that contrastivism is merely a form of contextualism that appeals to relevant alternatives.

What makes a contrastivist approach distinct from the relevant-alternatives contextualist accounts we have looked at so far is in the explanation that they give for this context-sensitivity. An important feature⁶ of this explanation is the claim that c-term statements have one or more extra argument places that are typically hidden or suppressed (Schaffer 2004, 87) (Schaffer 2005, 308) (Blaauw 2013, 90). What this means is that the kinds of c-term statements that we use most often are, when taken on their own, semantically incomplete. That is, when considered outside of a context, these statements don't express a complete proposition, and only do so when used within a context. So, it is the context that completes

⁶ As pointed out by an anonymous referee, there are some examples of theorists providing accounts that are widely recognised as contrastivist and that do not have this feature. Notable examples of this include Sinnott-Armstrong's work on contrastivism, as well as the account presented in Menzies' 'Platitudes and Counterexamples (2009). For the sake of simplicity, I will avoid discussion of these accounts until section 4, where I will address their existence as a potential objection to my analysis.

the proposition, and different contexts will complete the proposition in different ways, producing different propositions. Thus, the relevant c-term statements are context-sensitive.

This then raises the question: what exactly it is that context supplies to complete these propositions; what fills in these hidden argument places? The contrastivist's eponymous answer is that the argument places are filled in with a set of contrasts. This means that a completed c-term proposition will express a multi-part relation that holds between the statement's original referents and this contrast set. The members of this contrast set will be a contextually salient set of possible alternatives to the referents of the c-term statement in question. Therefore, the proposition that is expressed by a typical c-term statement will be partly determined by which possible alternatives are salient within the context of utterance for that statement.

To help make it clear what this all means, it will be useful to examine a specific contrastivist account. I will focus primarily on the contrastivist account of knowledge given by Schaffer (2004, 101-112) (2008, 235-245) to allow for a comparison with Lewis's account from the previous section. On Schaffer's account, binary knowledge statements of the form 'S knows that p' are semantically incomplete and contain a hidden argument place that is to be filled by a salient alternative to p. Schaffer refers to this as the mechanism of 'ternicity', as it means that all knowledge relations are three-place relations. A complete knowledge statement expresses a ternary relation between a subject, a proposition, and a contrast to that proposition. A statement expressing such a relation will be of the form 'S knows that p rather than p*', where 'p*' stands for at least one contextually salient alternative proposition to p. Statements like this are acceptable for English speakers to use. However, people typically stick

to the incomplete versions of the statements, and so those are the statements that the contrastivist focuses on.

Let's return to statement (1). Since this is a binary knowledge statement, the contrastivist claims that it has a hidden argument place that needs to be filled by a set of contextually salient alternatives to the proposition <Moore has hands>. In an everyday context, the alternatives we *would* consider salient are close possible alternatives like <Moore has stumps>, while distant alternatives like <Moore is a brain in a vat> would *not* be salient. So, an utterance of (1) in an everyday context will, according to the contrastivist, express something like the following proposition:

1~) <Moore knows that he has hands rather than stumps>.⁷

For this to be true, it must be the case that Moore is able to rule out the possibility that he has stumps. Fortunately, he *is* able to do this, as he can look and see that he has hands and not stumps.

Things are different in the context of a philosophy seminar, where the alternative <Moore is a brain in a vat> *is* salient. Therefore, an utterance of (1) in this context will express something like:

1#) <Moore knows that he has hands rather than that he is a brain in a vat>.

⁷ Of course, there may be more alternatives that are salient in an everyday context and will consequently be a part of the contrast set. However, for the sake of readability, the examples I provide here will only contain a single alternative.

This is false because Moore is unable to rule out the possibility that he is a brain in a vat using the evidence that is available to him.

It is at this point that we can clearly see how someone might conclude that contrastivism is merely a form of contextualism. Schaffer's account fits in perfectly with the contextualist accounts discussed in the previous section as he argues for the context-sensitivity of a set of c-term statements that is cashed out in terms of the contextual relevance of alternatives. This is equally true of other contrastivist accounts. It therefore seems entirely reasonable to conclude that contrastivism is merely a brand of contextualism that is extremely similar to (or even identical with) the kind of contextualist accounts that appeal to relevant alternatives.

I believe that this conclusion is mistaken, as there is an important point of difference that can be easily overlooked: contrastivists are committed to the view that the relations they are interested in have relata that are not specified by typical c-term statements about those relations. However, those relata can be specified by c-term statements ((1~) and (1#) are examples of how this can be done), and this entails that there are (at least in principle) complete c-term statements with no role for context to play. For example, 'Holmes knows that Mary stole the bicycle rather than the wagon' (Schaffer 2004, 78) seems to be an acceptable statement in English that explicitly specifies a contrast. On Schaffer's account, this statement is semantically complete: there are no empty argument places that need to be

filled in by the context. Consequently, this statement, along with many others like it, is *not* context-sensitive⁸.

The fact that contrastivism allows for c-term statements that are not context-sensitive will be important when we come to discuss the difference between contextualism and contrastivism in section four. In the next section I will lay the groundwork for that distinction by examining Schaffer's attempt to do the same. Ultimately, I will argue that his account misses the mark by failing to offer a substantial difference between the two approaches, but it does send us off in the right direction.

3. Schaffer's Points of Difference

The picture that has so far emerged from the prior discussion is that contrastivism is strikingly similar to the contextualist accounts that appeal to relevant alternatives. Indeed, the two positions are so similar that one could be forgiven for thinking that they are identical. The only notable difference that we can point to is that the contrastivist argues that certain c-term statements contain a hidden argument place that is typically filled in by the context of utterance. Yet for those who maintain that contrastivism is just a form of contextualism, this

⁸ It may be argued that the contextualist could attempt a similar move to eliminate the context-sensitivity from a given c-term statement by specifying the contextual features that the statement is sensitive to within the statement itself. As an example, we could rewrite (1) so that it says, 'Moore knows that he has hands relative to the standards of assessment S_1 ', where S_1 is a placeholder for the standards that are in place in a specific context. This statement will be invariant because we no longer have to appeal to the context to find out what standards we should be assessing the knowledge claim against. However, this is not analogous to the move made in the contrastivist case because this statement is not a completed version of (1). This would only be the case if the contextualist were to claim that (1) has a hidden, empty argument place that needed to be filled by a set of standards. As things stand, the contextualist does not make this claim, and so this statement refers to a different relation than the one referred to by (1). Likewise, the contextualist would also have to hold that the relation that is referred to by (1 \sim) and (1 $\#$) is different to the relation that is referred to by (1). To argue otherwise would be to endorse a theory that is not contextualism.

can simply be dismissed as a minor linguistic point that does not make for a significant difference between the two approaches. However, this way of understanding these approaches is mistaken. This linguistic point does in fact point towards a metaphysically significant distinction that reflects their differing commitments to certain naturalist intuitions.

In this section, I will begin to spell out this distinction by returning to Schaffer's 2004 paper, where he outlines three points of difference between his contrastivist account of knowledge and the corresponding contextualist accounts. What I will demonstrate is that although these points of difference are a useful step in the right direction, when taken on their own they fail to pick out a difference that is significant enough to ground a substantial demarcation between contextualism and contrastivism.

Schaffer's three points of difference between contextualism and contrastivism about knowledge are as follows:

- i) They employ distinct linguistic models for generating context-dependent truth-conditions of knowledge statements.
- ii) They have distinct linguistic mechanisms for factoring alternatives into the truth conditions of knowledge statements.
- iii) They offer distinct philosophical accounts of how ordinary knowledge is shielded from sceptical doubt. (2004, 82)

We can immediately rule out difference (iii) as a means of providing a general distinction between contextualism and contrastivism, as it only applies within the domain of knowledge.

The points of difference identified by (i) and (ii) are more promising, although they ultimately do not get at the heart of the distinction. In his discussion of (i), Schaffer distinguishes between the models of 'indexicality' and 'ternicity' (2004, 82-87). On the model of

indexicality, the context-dependence of knowledge statements results from the fact that the same statement can express different propositions in a different context. As an analogy, a statement containing an indexical like 'I' (e.g. 'I am in Paris') will express many different propositions with different truth-values depending on who uses it in a given context.

It is important to note, however, that although many contextualist accounts do employ this model, this is not true of *all* contextualist accounts⁹. This is evident in the work of MacFarlane¹⁰, who distinguishes between indexical and non-indexical forms of contextualism (2009, 231-250), where 'indexicality' is used to mean the same thing as it does on Schaffer's understanding. Non-indexical forms of contextualism model the context-sensitivity of c-term statements (in this case knowledge statements) as being a result of the same proposition being judged by contextually varying standards. That is, the same statement will express the same proposition in all contexts and any difference in the truth-values of those statements across different contexts is the result of that proposition being evaluated according to different standards in those contexts. MacFarlane's contextualist account of knowledge is an example of a view that models the context-sensitivity of knowledge statements in this way (2009, 236-237). This model is distinct from the one offered by the indexicalist, but it still belongs in a contextualist account because the claim being made is that the c-term statements of our target concept are all context-sensitive: they can express a true proposition in one context and an (identical) false proposition in another. What this means for our inquiry

⁹ It is also not always made clear which model is being used in any given contextualist account.

¹⁰ As noted by an anonymous reviewer, this distinction may not be exhaustive. However, the fact that it can even be made, clearly suggests that there are forms of contextualism that do not employ the model of indexicalism.

is that (i) cannot be used as a point of difference between contextualism and contrastivism, because, to put it bluntly, non-indexical forms of contextualism exist. However, it can help us to understand what indexical and non-indexical forms of contextualism have in common that they do not share with contrastivism.

To see this, let's compare the model of indexicality with Schaffer's competing model of context-dependence, ternicity. On this model 'context-dependence is generated by the absence of an explicit setting for $\{p^*\}$, which is then implicitly saturated by different alternatives in different contexts' (Schaffer 2004, 82). In other words, the context-sensitivity in contrastivism comes from the fact that there is an argument place that can be filled in a number of different ways depending on the context.

Schaffer further distinguishes the two models when he states that 'by indexicality what is variable is the relation denoted by 'knows', whereas by ternicity what is variable is not the relation denoted by 'knows' but rather the value of the relatum $\{p^*\}$ when left implicit' (Schaffer 2004, 83). What this means is that, on the indexical contextualist account, the relation that the term 'knows' refers to will vary across different contexts. So, in one context (C_1), the statement 'S knows that p' will express a proposition that refers to the relation '...knows- C_1 ...', while in a different context (C_2) the same statement will express a *different* proposition in virtue of the fact that the relation referred to by 'knows' will be a different relation, '...knows- C_2 ...'. Applying this to statement (1), we find that in an everyday context (C_E) it expresses something like the following:

3) <Moore knows- C_E that he has hands>

While in the context of an epistemology seminar (C_S), it will express something like:

4) <Moore knows- C_S that he has hands>

Here 'knows- C_E ' and 'knows- C_S ' are two distinct relations, so it is possible for both propositions to have a different truth-value. We therefore have a model of context-sensitivity where the relation expressed by 'knows' – and consequently the propositions expressed by the whole knowledge statement – varies across contexts. As a consequence, there is no one single knowledge relation, rather a set of differing knowledge relations that all obtain in different contexts¹¹.

This differs from the model used by non-indexicalists like MacFarlane. On this model, the relation that is referred to by 'knows' remains the same in all contexts. The relation that is referred to simply *is* the knowledge relation – a two-place relation that holds between a subject and a proposition – and whether or not it obtains depends on some feature of the context where the knowledge statement is used. As an example, on a non-indexicalist version of Lewis's account, the statement 'S knows that P' will always mean something like, 'S believes that p and can rule out all the alternatives that are relevant in the context of utterance'. Whether or not this is true will depend partly on the context, and which alternatives are relevant there. So, this is still a model for context-sensitivity, but unlike the indexicalist model, 'knows' will always refer to the same two-place relation in all contexts.

This is distinguished from the contrastivist account, where the relation picked out by 'knows' is always the same *three-place* relation, regardless of context. What does vary across contexts

¹¹ To avoid any additional complexity, I will refrain from offering any more details regarding what these relations are and how they differ. These details are not needed for my argument.

is its relata, as the context determines what belongs in the contrast set. So, a statement of the form 'S knows that p' will always refer to the same three-place relation, but in different contexts the relation may have a different contrast set as a relatum. If the relatum differs across contexts, then so does the overall proposition that is expressed by the statement. This model of context-dependence clearly differs from the model of indexicality, but then so does the model of non-indexicality. So right now, all we appear to have are three different kinds of contextualist account that all use different models for context-dependence.

If anything, at this point, contrastivism looks like something of a compromise between indexical and non-indexical contextualism. Like the indexicalist, the contrastivist does not claim that the same proposition is expressed in all contexts, but like the non-indexicalist they also claim that the same (three-place) relation is expressed in all contexts. What makes the proposition expressed different in different contexts is a change in the (hidden) relatum, as different contrast sets can be salient in different contexts. With this in mind it seems entirely reasonable to conclude that contrastivism is just another breed of contextualism. Therefore, point (i) on its own is not enough to motivate a general demarcation between contextualism and contrastivism. Furthermore, as I will go on to show, it is also not sufficient when taken in conjunction with point (ii).

In his discussion of (ii), Schaffer distinguishes between two different linguistic mechanisms for factoring alternatives into the truth conditions of knowledge statements: 'relevance' and 'saturation' (2004, 87-90). The kind of alternatives that Schaffer is talking about here are the non-actual propositions that, if contextually salient, will make up the contrast set for a completed knowledge claim. The mechanism of saturation consists in those alternatives being

used to fill in the empty argument place of an incomplete, binary knowledge statement. That is, the mechanism takes us from 'S knows that p' to 'S knows that p rather than {p*}'. So, on this mechanism, alternatives are factored into the truth conditions of knowledge statements by filling in the gaps of binary knowledge statements. In ternary knowledge statements, this mechanism is not needed since there are no empty argument places.

Schaffer distinguishes this from the mechanism of relevance. He states that on this mechanism 'alternatives enter into the truth-conditions via the semantical rule of relevant alternatives (as triggered by the 'knows' indexical)' (2004, 87). There are two points worth highlighting in regard to this mechanism. The first, is that not every contextualist account argues that the relevance of alternatives matter for the truth-values of knowledge statements. For instance, MacFarlane's account of knowledge holds that the truth-values of knowledge statements are sensitive to the relevance of epistemic standards and not alternatives. The reason that Schaffer chooses to focus on alternatives here is, I believe, that he is looking to distinguish his contrastivist account from contextualist accounts like Lewis's, where the truth-values of knowledge statements *are* sensitive to the contextual relevance of alternatives. Recall that on Lewis's account, someone can only be truly said to know that p if they are able to rule out all and only the *relevant* alternatives to p. So, alternatives are only factored into the truth-conditions of knowledge statements if they are relevant, making this an example of the mechanism of relevance.

The second point concerns how this mechanism actually functions, and specifically relates to Schaffer's use of the term 'indexical' when he states that the mechanism is 'triggered by the "knows" indexical' (Schaffer 2004, 87). As discussed earlier, the use of 'indexical' here is

inappropriate because there are non-indexical versions of contextualism, and the mechanism seems to function differently depending on which kind of contextualist account is being offered. On the indexical account, different alternatives are factored in by the term 'knows' being used to pick out different relations. To illustrate, we can use an indexicalist version of Lewis's account. In the everyday context, (1) expresses a proposition containing the '...knows- C_E ...' relation, which can only obtain if Moore can eliminate relevant alternatives, like his having stumps. The relation that (1) refers to in the epistemology seminar, '...knows- C_S ...', is much less likely to obtain because it relies on Moore also being able to eliminate the alternative that he is a brain in a vat. What this means is that, on the indexicalist account, alternatives are incorporated into the truth-conditions of knowledge statements by being relevant to whether or not the relation picked out by the term 'knows' actually obtains.

Things are slightly different on the non-indexicalist account. On this account the relation that is picked out by 'knows' will always be the same, but the relation is characterised broadly enough that whether or not it obtains depends on specific features of the context. On the non-indexicalist version of Lewis's account, the relation referred to by 'knows' will be something like '...believes that... and can rule out all the alternatives that are relevant in the context of utterance'. So, this knowledge relation will obtain if the contextually relevant alternatives have been eliminated. In an everyday context, only alternatives like Moore's having stumps are relevant. Since this alternative can be ruled out, the relation will obtain in this context, and (1) will be true. Yet in the context of a philosophy seminar, alternatives like Moore being a brain in a vat become relevant. Since this alternative cannot be ruled out, the relation does not obtain and (1) is false. So, as with the indexicalist account, alternatives are incorporated into the truth-conditions of knowledge statements by being relevant to whether

or not the relation picked out by the term 'knows' actually obtains. The difference for the non-indexicalist is that 'knows' will always pick out the same specific relation. What we can conclude from this is that the indexical and non-indexical views use very similar mechanisms for factoring alternatives into the truth-values of knowledge statements, and that this mechanism differs significantly from the one used by the contrastivist for the same end.

This point of difference is a more promising line of inquiry, particularly when taken in conjunction with the first point. However, all that has really been demonstrated so far is that contextualism and contrastivism use different linguistic mechanisms to achieve the same end of having the contextual relevance/salience of alternatives affect the truth-values of knowledge statements. There is definitely a point to be made here, as it seems like there is a very real difference between an account that characterises knowledge statements as making claims about contextually-sensitive relations and an account that characterises knowledge as a three-place relation where one of those places is usually filled in by the context. Yet this difference does not seem metaphysically significant enough to motivate a demarcation between contextualism and contrastivism. This means that in the next section, I will need to take this point much further to demonstrate that there *is* a metaphysically significant difference between these two views that goes beyond a minor difference in linguistic mechanisms.

4. Demarcating Contextualism and Contrastivism

An advantage of approaching this topic from the perspective of causation, is that this is an area where the metaphysical significance of my distinction can be easily understood. As such, in this section I will mainly focus on showing how the distinction applies in the domain of

causation. Having said that, however, this demarcation should also be applicable in all domains that contain both contextualist and contrastivist approaches, and to demonstrate this I will also be applying it in the domain of knowledge.

The invariantist/contextualist/contrastivist debate within the domain of causation is best understood as the result of two conflicting intuitions:

The Natural Relation Intuition: The intuition that causal relations exist out there in the world in a way that is independent of our thinking about causal relations. Strawson refers to this as the intuition that causation is a natural relation (Strawson 1992, 109).

The Contextual Intuition: The set of intuitions we have about cases which entail that contextual factors affect our evaluations of causal statements.

The source of the first intuition is fairly straightforward. We tend to see causal relations as being a natural feature of the universe that would exist even if we did not. They are a part of a real causal structure that is the object of scientific inquiry and other forms of empirical investigation. Such investigations are conducted with the aim of uncovering the underlying causal structure of the universe. So, when we state that 'c caused e' we are making a claim about how that underlying structure is arranged, and the truth of that statement will depend on whether or not the structure is as the statement describes, which is something that will be in no way affected by the context of assertion. This intuition is widespread, and there are many theorists who believe that a theory of causation must be able to accommodate it.

Unfortunately, this intuition is challenged by a set of intuitions that we get from cases that seem to suggest that our evaluations of causal statements are (at least partly) determined by

contextual factors. For an example, recall the forest fire case from section one, where we had two competing causal statements that were being assessed in two different contexts:

- 2) The lightning strike caused the forest fire.
- 3) The presence of oxygen caused the forest fire.

Our initial intuitions about this case were that (2) is acceptable in the context of the forest rangers and unacceptable in the context of the Venusians, while (3) is acceptable in the context of the Venusians and unacceptable in the context of the forest rangers. Many philosophers take these intuitions to suggest that contextual factors can determine whether or not it is true to say that one event caused another.

This runs counter to our intuition that causation is a natural relation. On *that* view, the universe has an underlying causal structure that would exist even if we were not around to study it. A consequence of this would be that the truth or falsity statements of statements about that structure would be entirely dependent on the nature of that structure, which is something that is an objective feature of reality. It simply would not be possible for the truth or falsity of those causal statements to be affected by the contexts that those statements are used in. Yet, cases like the forest fire seem to suggest that the truth-values of those statements *are* affected by contextual factors, which would suggest that there is no underlying causal structure to the universe and that causation is something we have constructed through our own inquiries into the nature of the universe. We thus have two competing intuitions about the metaphysical nature of causation: one where causation is an objective, underlying feature of the universe and one where facts about causation are only to be had when embedded in a context of enquiry.

This metaphysical question about the nature of causation is deeply connected to the linguistic question of whether or not causal statements are context-sensitive. If we accept that causal statements *are* context-sensitive, then we will have to deny the intuition the causation is a natural relation. This means that contextualism is incompatible with any view on the nature of causation that can accommodate the natural relation intuition. To be a contextualist is to effectively deny that intuition. A clear illustration of this point can be seen from the fact that Menzies refers to the intuition as a ‘philosopher’s myth’ (2009, 355). This has also meant that those who have sought to uphold this intuition have often found themselves drawn towards invariantism, where they affirm the natural relation intuition by denying the contextual intuition – e.g. by arguing that our contextual intuitions are merely the result of our being unwilling to assert certain causal statements for pragmatic reasons, or of our being misled by some cognitive bias. None of this is to say that all invariantists seek to affirm the natural relation intuition and deny the intuitions that we have about examples like the forest fire case¹². What it *does* mean is that someone cannot assent to the natural relation intuition whilst also being a contextualist, as the two are incompatible.

We have now arrived at the difference between contextualism and contrastivism about causation, as the latter is *not* incompatible with the natural relation intuition. The reason why is connected to the conclusion of the previous section: that the two approaches utilise different linguistic mechanisms to achieve the same end of having the contextual relevance/salience of alternatives affect the truth-values of statements. For the

¹² For an invariantist approach that tries to accommodate both of these intuitions, see (Montminy and Russo 2016).

contextualists, these alternatives are not considered to be part of a semantically complete version of the statement.

Compare this to the contrastivist account of causation that is offered by Schaffer (2005) (2012). Here causation is characterised as a four-place relation that holds between a cause, an effect, a causal contrast (a contrast set of alternatives to the cause), and an effectual contrast (a contrast set of alternatives to the effect). This means that an utterance of 'c caused e' expresses something like 'c rather than c* caused e rather than e*', where c* stands for the causal contrast and e* stands for the effectual contrast. As with all contrastivist accounts, these contrasts are usually supplied by the context the statement is used in, but they don't have to be. If we were to simply use a complete, quaternary causal statement, then the context of use is no longer needed to complete the proposition. An example of such a statement would be:

- 5) The lightning strike rather than the absence of a lightning strike caused the forest fire rather than the absence of a forest fire.

This statement will express the same proposition with the same truth-value regardless of the context where it is used, since there are no hidden argument places that need to be filled by that context. The same holds true for the other statement that we may want to make about this case:

- 6) The presence of oxygen rather than the absence of oxygen caused the forest fire rather than the absence of a forest fire.

There may be pragmatic reasons for the forest rangers to refrain from uttering (6) and for the Venusians to refrain from uttering (5), but both statements will still be true in those

contexts¹³. Therefore, unlike the contextualist account, the contrastivist account holds that complete causal statements like (5) and (6) are not context-sensitive.

So far, this is just a linguistic point about the behaviour of causal statements on the contrastivist account. However, this point has important implications for the metaphysical nature of causation if we were to accept such an account. These implications can be brought out when consider the question of what ‘caused’ is referring to in (5) and (6). It is entirely consistent and entirely natural for the contrastivist to answer that the thing being referred to is a four-place causal relation that exists as an objective part of the underlying causal structure of the universe. The existence of this four-place relation is entirely consistent with the intuition that causation is a natural relation: it is something out there in the world for us to discover, and there are there independently of any conversational context. This means that the truth of all causal statements will depend on the obtaining of these four-place relations, even the truth of those statements that are semantically incomplete. The contrastivist does not *have* to go down this route, but it *is* preferable for them to do so since it allows them to

¹³ One issue that may be raised here is that Schaffer’s account does not adequately accommodate the contextual intuition in this particular case, as it does not provide an obvious mechanism for saying that the truth-values of statements (6) and (7) can vary across different contexts. However, this is only a problem if we believe that contrastivism has to be able accommodate this intuition as successfully as contextualism. As I suggest on the next page, contrastivism is best thought of as a middle-ground position between contextualism and invariantism. So, contrastivism only needs to do a better job at accommodating the contextual intuition than invariantism does. I believe that it achieves this by including contrast sets in the formulations of complete causal statements. The inclusion of these contrast sets provides us with more reasons for finding a causal statement to be acceptable in one context and unacceptable in another. As an example, the forest rangers may find (6) more acceptable than (7) because the causal contrast of there not being any oxygen present is a more remote possibility than there not being a lightning strike. Explanations like this are much more difficult to provide on the invariantist approach, and so the contrastivist is able to offer more resources for handling the contextual intuition.

do justice to the natural relation intuition. The contrastivist *can* accommodate the natural relation intuition, while the contextualist *cannot*.

It is important to note that this does not mean that the contextualist view is incompatible with there being other relations out there in the world that can ground causal relations. For instance, there may still be relations of counterfactual dependence that obtain independently of context as part of the structure of the universe. These relations could plausibly be described as natural. However, for those relations to be properly described as causal on the contextualist account, certain contextual factors must be in place. If there can be no *causal* facts without these contextual factors, then causation cannot be a natural relation. On the contrastivist account, causal relations *can* obtain without these contextual factors, and so these *are* natural relations.

Thus, in the domain of causation, contrastivism is metaphysically distinct from contextualism by virtue of being able to accommodate the natural relation intuition. Contrastivism is also able to account for the contextual intuition, as the contrastivist can argue that any influence that contextual factors have on the truth-values of binary causal statements are the result of those statements being completed differently in different contexts through the selection of different contrasts. On this understanding, contrastivism is something of a middle-ground between contextualism and invariantism, since it is able to accommodate both of the intuitions I have discussed here at the cost of a different (and also very plausible) intuition:

The Two-Place Relation Intuition: Causation is a two-place relation.

This gives us three distinct approaches to analysing the concept of causation, and our reasons for preferring one approach to another will depend on which intuition we are willing to

sacrifice: the natural relation intuition, the contextual intuition, and the two-place relation intuition. Therefore, we have a clear demarcation between contextualism and contrastivism that is of real metaphysical significance.

We must now turn to the question of how this demarcation can be applied more generally by looking at how it applies in the domain of knowledge. As with causation, contextualism about knowledge involves the denial of a metaphysical claim about knowledge: that knowledge is a relation that is out there in the world and can be studied through empirical investigation. This idea leads us to the widely defended claim that knowledge is a natural kind. What this claim means is that knowledge is a distinctive psychological state that can be studied empirically in the same way that other psychological states are. This intuition is certainly not as widespread as the natural relation intuition is in causation, but there are many that endorse it. The claim was first explicitly proposed by Kornblith in 2002¹⁴, but, as Kumar (2014, 440) notes, the guiding intuition has also been identified in earlier works from Dretske (1981) and Millikan (1993). This intuition is ruled out by contextualism, as someone could be in the exact same psychological state in two different conversational contexts and be truly said to know that *p* in one and to not know that *p* in another. Of course, an invariantist account *is* compatible with the natural kind intuition, but it does not have the advantage of being able to account for the intuition that many philosophers have that the truth-values of knowledge statements are affected by the contextual relevance of sceptical scenarios (for an example of this, refer back to the discussion of Moore's hands in section one). So, choosing between contextualism

¹⁴ See (Kornblith 2002).

and invariantism in this domain will ultimately come down to deciding which of these two competing intuitions to reject: the natural kind intuition or the contextual intuition.

Unsurprisingly contrastivism is able to accommodate both of these intuitions at the cost of the intuition that knowledge is a two-place relation. In section two we saw how the account handles the contextual intuition by applying it to Moore's hands. It is also compatible with the intuition that knowledge is a natural kind, as complete knowledge statements could be understood as indicating that a subject is in a particular psychological state – albeit one that is more complex than the state they are taken to be in on the invariantist account. Therefore, we again have three distinct approaches to analysing the target concept and our reasons for preferring one approach to another will again depend on which intuition we are willing to sacrifice: the natural kind intuition, the contextual intuition, and the two-place relation intuition. This gives us another domain where the demarcation between contextualism and contrastivism is metaphysically significant in a manner that should affect our reasons for choosing one approach over the other. I also see no reason why we could not continue and apply the demarcation to other domains that permit both contextualist and contrastivist approaches.

At this point it may be objected that my analysis has failed to properly identify a feature that marks contrastivism out as distinct from contextualism, as there are accounts that seem to not conform to my definitions. The demarcation rests on the point that contrastivists are committed to providing analyses of relational concepts, according to which those concepts refer to relations with more relata than we might have typically expected them to have. Yet

there are examples of theorists offering contrastivist accounts that don't make such a claim¹⁵. As an example, Sinnott-Armstrong provides a contrastive account of knowledge without endorsing the claim that knowledge statements refer to a relation with more than two places (2004) (2006) (2008). In his 'A Contrastivist Manifesto' he even quotes Schaffer saying that knowledge is a ternary relation (2008, 257) but notably refrains from making the same claim himself. Likewise, Menzies, in his 'Platitudes and Counterexamples', presents his same contextualist account of causation as a 'contrastive' account (2009, 342). Both these authors label themselves as presenting a contrastive account but avoid postulating additional places in the relations that they talk about. Without any kind of claim about additional argument places, these accounts are more or less identical to the kind of contextualist that appeal to relevant alternatives. Therefore, if we accept that these views are contrastivist, then my analysis has failed to identify a point of difference between contextualism and contrastivism.

My response to this is to deny that we ought to think of these accounts as contrastivist – the labelling of them as such by their proponents notwithstanding. That is, when deciding which theories should be included within the umbrella term of contrastivism, there is good reason for excluding those that don't involve a claim about additional argument places. If we want to be consistent with how we label our theories, then we really only have two options available to us: one is to follow the understanding I have provided in this chapter and the other is to characterise contrastivism broadly enough as to include all of the accounts that are labelled as such. If we take the latter approach then the term 'contrastivism' will simply refer to any view that talks about contextually relevant alternatives, which would include just

¹⁵ Thanks to an anonymous referee for raising this issue.

about every account I have referenced thus far (several of which have never been described as contrastivist). This approach would have the advantage of allowing all the accounts that have been labelled (by others) as contrastive to be properly described as such. However, it would erase the important metaphysical distinction that I have highlighted between the accounts that do and do not claim additional argument places¹⁶. Therefore, I believe that it is far more useful to characterise ‘contrastivism’ in the way that I have done so in the course of this chapter, meaning that we ought to consider contextualism and contrastivism to be two separate views.

5. The Implications of the Contextual Question

Here I have argued that both contextualism and contrastivism offer different prospects for how we understand the metaphysical natures of our target concepts. Contextualism requires us to deny any naturalist intuitions that we may have in favour of our contextual intuitions. This is opposed by the invariantist, who denies those contextual intuitions in order to uphold their naturalist intuitions. Contrastivism then offers a middle-ground position between the two, where we can accommodate both kinds of intuition, at the cost of making our analysis more complex by specifying additional relata. We thus have three metaphysically distinct

¹⁶ It may be objected that this response is question-begging because I am using the fact that there is a difference between contextualism and contrastivism to motivate the claim that there is a difference between contextualism and contrastivism. However, that is a misunderstanding of my argument. I have already demonstrated that there is a metaphysically significant difference between the kinds of account that do argue for extra-argument places that are filled in by the context, and those views that do not make this claim. As the boundaries of different views ought to generally respect these differences, it makes sense to label them differently. The labels of ‘contextualism’ and ‘contrastivism’ are the most apt for this, given how those terms are already used. The difference that I identify here exists prior to any labelling of views as ‘contextualist’ or ‘contrastivist’, so it is not question-begging to then argue that our use of these labels ought to respect this difference.

approaches to understanding a target concept, with our motivations for choosing one over the other being partly determined by our own prior intuitions about that concept.

All of this carries some important consequences for how we think about contextualist and contrastivist accounts of causation. One of the main questions that I hope to address with this thesis is:

II. Is causation contextual?

It should now be very clear what this question is asking. On top of this, it should also be apparent that the only account that can answer it with a strong and resounding ‘yes’ is contextualism. Contrastivism, despite appearances to the contrary, is still compatible with the natural relation intuition, and will therefore allow for the possibility that causal relations exist independently of context as part of reality. This makes the position far less appealing to those who wish to fully embrace the kind of context-sensitivity that seems to be implied by examples like the forest fire case. For those people, causal contextualism should carry a much greater appeal, as there is no attempt made to incorporate the natural relation intuition.

This distinction will also affect how we approach the normative question, which concerns whether or not the truth-values of causal statements are sensitive to normative considerations. Examining what this question means will be the focus on the following chapter. Beyond that, the incompatibility of contextualism with the natural relation intuition will also be important when I come to defend my analysis in Chapter 7, since many who would argue against the constraint that I provide will do so to uphold that intuition.

Chapter 2 – The Normative Question

In the previous chapter I identified the two central questions that I will answer in this thesis:

- I. Is causation normative?
- II. Is causation contextual?

Having stated these questions I then clarified question II, distinguishing between three separate positions one can take regarding whether or not a given concept – in this case the concept of causation – is contextual: contextualism, invariantism, and contrastivism. The contextualist about causation will answer question II with a firm ‘yes’, while the invariantist will answer with an equally firm ‘no’. Meanwhile, the contrastivist will occupy the middle-ground position that incomplete causal statements are context-sensitive, while causation itself is not. Leaving aside the contrastivist for now, the disagreement between contextualist and the invariantist can be understood as a dispute over whether we should uphold the intuition that causation is a natural relation, or the set of intuitions that we have regarding certain puzzle cases that suggest our evaluations of causal statements vary across different contexts.

The aim of this chapter is to introduce some of the important considerations for answering question I. This question asks whether or not the truth-values of causal statements are sensitive to facts involving some specific normative concept like normality or defaults. Those who answer ‘yes’ endorse a normative account of causation, while those who answer ‘no’ reject such an account. This disagreement can be understood as motivated by similar concerns to those that gave rise to the contextualist/invariantist debate. Theorists who argue

against the view that causation is normative do so due to a commitment to the intuition that causation is a natural relation; while those who argue *for* a normative account of causation are motivated by upholding the intuitions that we have about certain puzzle cases. As such, most of this chapter will be focused on describing some of the various puzzle cases that have been previously offered in support of a normative account of causation.

What all of these puzzle cases have in common is that they present us with different causal statements that vary in how intuitively acceptable they are. The proponent of the normative account argues that these variations in our intuitions are best explained by claiming that they result from the influence of some normative notion. So, the main argument in favour of a normative account is an inference to the best explanation. That is, the reason given for endorsing the normative account is that it best explains the intuitions that we have about the puzzle cases. For this argument to be successful it must therefore be the case that the normative account's explanation is better than any alternative we could give. In Chapter 3 I argue against several existing alternative explanations. However, in Chapter 4 I will also show how the explanation given by the normative account is incomplete by virtue of not including any kind of analysis of whatever normative notion the explanation appeals to. *This* chapter will provide the necessary background for making that argument and for understanding why we might want to endorse a normative account of causation more generally. As such, in what follows I will focus on setting out the puzzle cases and examining how they are to be explained by the normative account.

This chapter will be divided into three sections. In section one I will provide a brief summary of the normative account's main claim and address the question of which normative concept

ought to be employed here. In section two I will provide my list of puzzle cases that are used to motivate a normative account of causation. All told, I will present four different cases. For each example, I will outline the case as it is presented in the literature, and then I will explain how it can be used to support a normative account of causation. Then in the final section I will consider some of the strategies that are available to us in responding to our intuitions about these puzzle cases.

1. What is a Normative Account of Causation?

As stated earlier, a normative account of causation is one that takes the truth conditions of causal statements to be sensitive to claims involving some specific normative concept. Two key concepts that are often invoked here are normality and defaults. What both of these concepts have in common is that they refer to ways that the world ought to be, albeit in a fairly weak sense. That is, they tell us how we can reasonably expect the world to be. By introducing such concepts into our understanding of causation, we start to see our process of causal reasoning as a means of determining how an unexpected actual course of events came to be. We go through a process of recognising that things are not the way they ought to be (in this weak sense of 'ought'), and so look to try and identify the abnormal or deviant event that brought us here.

This is important because, as we will see in the following section, many of the puzzle cases are instances of the more general problem of selection. As stated earlier, this problem consists in trying to find an answer to the question of why we pick out some events as causes, while others are relegated to mere background conditions. For any given effect, there will be a considerable number of candidate causes that the event counterfactually depends upon.

Yet when asked to identify the event that caused the effect in question, we tend to respond with one or two events that we select from the almost endless set of candidates. The remaining members of this set are almost never mentioned, and when they are, they are seen as mere background conditions for the occurrence of the effect we are investigating. The problem comes from trying to answer two important questions about this phenomenon:

- i) On what basis do we distinguish between causes and background conditions?
- ii) Why do we make such a distinction in the first place?

So, the problem is to explain how and why we pick out some events as causes of an effect while ignoring others, even when all the events in question stand in the same kind of counterfactual dependence relation with that effect.

Adopting a normative account allows us to answer question (i), which then helps us to make progress towards answering question (ii). For (i) it can explain how we come to decide whether an event is a cause or a mere condition by appealing to whatever normative notion that the theory considers causal statements to be sensitive to. In this thesis, I will take the relevant normative concept to be that of normality, as I take this concept to be broad enough to do the work that we require of it. Another concept that is often employed towards the same end is that of a default. I will not argue against accounts that use this concept here, but I will avoid doing so myself. My reasoning here is that it seems like anything that is the default for something will be normal, yet there are things that are normal that are not defaults¹⁷. It

¹⁷ A straightforward example of this is that Google Chrome is not the default internet browser for computers that use the Windows operating system, yet it would be absurd to say that using Chrome on those computers is not normal.

therefore seems that normality is more broadly applicable than the notion of a default, and thus an account that appeals to normality will potentially be more versatile than one that appeals to defaults.

By taking normality to be the relevant concept in our normative account, we can answer question (i) by stating that the events we select as causes are chosen because they are not *normal*. Meanwhile, the events that *are* normal are relegated to being mere background conditions. With this explanation in place, it then becomes easier to address question (ii): if an event's being normal or abnormal is what determines whether or not it is selected as having causal status, then the reason for our selecting causes in this way must have something to do with normality. One possibility is presented by Hitchcock and Knobe, who draw on the interventionist idea that the concept of causation enables us to pick out appropriate targets for intervening to ensure that our interests are met (2009, 591). We tend to pick out abnormal events because those events will be either much easier to prevent than the normal alternatives, or more likely to be missing from a set of events that would otherwise lead to our desired effect. Therefore, such events are more suitable targets for intervention, which explains why we select *them* to be causes instead of the more normal alternatives. This shows how a normative account of causation can be used to solve the problem of selection.

One further complication here is that the puzzle cases that are used to support a normative account of causation are often also used to support a contextual one. Contextualists about causation will typically present such cases and argue that our intuitions about them are sensitive to contextual factors. We saw this in the previous chapter with the forest fire case. There our intuitions about the acceptability of the two main causal statements varied across

the contexts of the forest rangers and the Venusians. This suggests that context plays some role in determining whether or not these statements are true. If we accept this then we will end up endorsing a contextualist account of causation. However, even if we are perfectly happy to sign up to contextualism, there is a further question of what specific feature of the context these causal statements are sensitive to.

A normative account can help us here by identifying what this feature is. On a normality-based account of causation, the feature will be what is normal in the context where a causal statement is being assessed. This presupposes that what is normal will itself be context-sensitive, such that something can be normal in one context and not normal in a different context. However, such an assumption is relatively uncontroversial and is intuitively very plausible. Specifically, it seems that, at the very least, statements about normality are sensitive to whatever reference class is salient in the context where those statements are uttered. We can therefore accept the claim that normality is contextual without too much trouble, leading us to an account of causation where the truth-values of causal statements are (at least partly) determined by claims about what is normal in the context where those statements are being assessed. As we will see in the following section, this is what allows the proponent of a normative and contextual account of causation to solve the various puzzle cases associated with the problem of selection.

At this point it should be clear why, as stated in the previous chapter, the normative question and the contextual question (questions I and II) tend to be treated as though they are asking the same thing. In principle, it is definitely possible to provide a normative account that is not contextualist: we just simply stipulate that the truth-values of causal statements are sensitive

to some normative notion that is context independent. This will include any account that appeals to normality as the relevant notion but understands statements about what is 'normal' as referring to context independent facts. Additionally, it is also possible to offer a contextualist account that is not a normative account by holding that causal statements are sensitive to some non-normative feature of the context. Yet despite this, there is little reason to adopt either one of these views. A normative invariantist account of causation will have to contend with our intuition that normality is contextual. Meanwhile, a contextualist account that is not normative will struggle to identify a feature of the context that explains our intuitions about the puzzle cases that is not ultimately reducible to a normative notion. For example, we might appeal to something like salience to explain the intuition, but then it seems like what makes an event salient is the fact that it is not normal. It therefore makes the most sense to adopt an account that is both contextualist and normative, rather than an account that is just one but not the other. In Chapter 6 I will provide such an account.

Here it might be tempting to ask what exactly is meant by 'normal'. Unfortunately, as noted earlier, there has been surprisingly little philosophical analysis done into the concept of normality, especially outside of the philosophies of biology and medicine. This may be surprising to some, as philosophers frequently use a term that is obviously very closely related: 'norm'. We very often hear talk of 'moral norms', 'social norms', 'statistical norms', 'legal norms', and more, so how can I claim that normality has not been given a considerable amount of philosophical attention? Well, this is because I take this notion of a 'norm' to be distinct from normality. I cannot offer a full discussion of this distinction, as this would fall outside of the scope of this thesis. I can however offer some more general remarks of what I take the difference between the two notions to be.

On the one hand, norms tend to be fairly precise but also somewhat universal rules that tell us how we ought to behave. This idea that ‘norm’ is a term used to denote a prescriptive rule is echoed by Halpern & Hitchcock, who state that, ‘To conform with a norm is to follow a prescriptive rule’¹⁸ (2015, 430). Understanding norms as rules further suggests that their scope is limited to human behaviour and they don’t apply outside of that. Additionally, violating one of these norms tends to entail that you have done something wrong, further supporting this idea that they are rules.

Statistical norms are an exception to this understanding of norms as rules. These norms don’t seem to be in any way prescriptive and they obviously apply outside the domain of human behaviour. I take this as fairly clear evidence that the use of ‘norm’ here means something different than it does when talking about other kinds of norm. Indeed, this view seems to be shared by many other philosophers, who often talk about ‘statistical *regularities*’, or ‘statistical *frequencies*’, rather than ‘statistical *norms*’.

Turning now to normality, it seems to me that referring to something as ‘normal’ is more descriptive than saying that it is a ‘norm’. Statements about normality seem to inform us (at least to some degree) about how the world is, rather than just telling us how it ought to be (though they may also do this as well). Unlike norms, they also seem to apply outside of the domain of human behaviour. As well as this, going against what is normal does not guarantee that you have done something wrong: e.g. extreme philanthropy may not be normal, but it is

¹⁸ Although it is important to note that Halpern & Hitchcock do not take ‘norm’ and ‘normal’ to have distinct meanings as I do here. Rather, they take ‘norm’ to be synonymous with a prescriptive sense of ‘normal’. In Chapter 5 I discuss this view that ‘normal’ is ambiguous in more detail, and then eventually go on to argue against it.

also certainly not wrong. Interestingly, people often talk about something's being 'the norm', which can be understood as being the same as saying that something is 'normal' rather than saying that it is '*a* norm'. In what follows, I will largely be focusing on normality, and will avoid any use of 'norm' wherever possible to remove any potential sources for confusion.

It should now be apparent what a normative account of causation is, and roughly how it can be used to address some of the puzzle cases that arise from the problem of selection. Additionally, it should also be broadly clear what I take normality to be and what its role is in the account. To illustrate all this, it is now time to turn our attention to the puzzle cases themselves, and to see how we might be able to solve them through a normative account of causation.

2. The Puzzle Cases

In this section I will simply list some of the puzzle cases that are the most helpful for understanding the argument in favour of a normative account of causation. Before I begin, there are two points that I want to emphasise. The first is that this list of examples is not intended to be comprehensive. I will consider four cases that have previously been used to motivate a normative account of causation. This does not mean that the cases I am considering here are the only ones that can be used to support a normative account.

The second point is to reiterate that the main argument for the normative account is an inference to the best explanation. If we take these puzzle cases on their own, they don't prove anything regarding whether or not causation is normative. They only generate intuitions about the acceptability of causal statements, not their truth-values. The acceptability of an utterance *can* be a result of its having a certain truth-value. For example, an utterance of

‘there are eight days in a week’ would be unacceptable (at least in most contexts) because it is clearly false. However, acceptability and truth-values can also come apart. A statement can be true, but unacceptable due to it conveying some false implicature, by being mistakenly disbelieved by its audience, or by simply being inappropriate to say in the context of utterance.

This distinction is important, because the proponent of the normative account of causation is trying to offer the best explanation of our intuitions about the acceptability of the causal statements in these puzzle cases by claiming that these intuitions reflect the real truth-values of those statements. Yet since truth and acceptability can come apart, it may be the case that our intuitions about the acceptability of the statements could be explained without having to hold that they reflect the truth-values of those statements. Such an explanation may also be a better explanation than the one offered by the normative account, which would undermine the argument for that account. In Chapter 3 I will consider some of the attempts to provide such an explanation, but for now I will simply lay out the puzzle cases and provide some indication of how they are to be explained on the normative account.

The Venusians and the Forest Fire

The first case that I will address is the forest fire case from the previous chapter. This case involved a lightning strike that led to a forest fire. We are then asked to assess two causal statements from the context of a group of forest rangers. In this context, the following causal statement seems acceptable:

- 1) The lightning strike caused the forest fire.

This intuition can be justified by appealing to the fact that the forest fire counterfactually depends on the lightning strike. Yet in the very same context, an utterance of the following would not be acceptable:

2) The presence of oxygen caused the forest fire.

This raises the question of why (1) is acceptable, whilst (2) is unacceptable. The normative account can explain this difference in acceptability by appealing to the fact that, on Earth, the presence of oxygen is normal whilst the occurrence of a lightning strike is not – and the forest rangers know this. So, the forest rangers ascribe the lightning strike the status of a cause and the presence of oxygen is judged to be a mere background condition. This solves the puzzle by explaining the difference in acceptability between (1) and (2).

A further problem then comes when we are asked to instead imagine that we are assessing statements (1) and (2) in the context of some spacefaring Venusians who have observed the fire from a distance. It was previously stipulated that lightning strikes are more common on Venus than they are here on Earth. To make the example more vivid, let's suppose that they are so common on Venus that the Venusians regard lightning as just a background feature of their world, just as we view the presence of oxygen here on Earth. As such, when the Venusians see the lightning strike they are surprised that it is followed by a fire, since the same thing does not happen on their world. They then set out to try and understand how the forest fire came about. Eventually, they discover that the forest fire came about as a result of the lightning strike interacting with a gas that is not present in the atmosphere of Venus: oxygen.

Now in this context it seems that (1) is intuitively unacceptable, in the same way that (2) was unacceptable in the context of the forest rangers. Likewise, it seems that (2) is intuitively acceptable, in the same way that (1) was acceptable to the forest rangers. This can be explained by the fact that, for the Venusians, the lightning strike is normal, while the presence of oxygen is not normal. As a result, upon seeing the forest fire the Venusians selected the presence of oxygen as the thing that caused the forest fire. This demonstrates how a normative (and contextualist) account can also explain the variations in the acceptability of (1) and (2) across contexts. Explanations like this can also be found in the literature, as Menzies gives one just like it when discussing a case with the same kind of causal structure (2007, 209-211).

To summarise, we have two causal statements that both refer to events that stand in the same kind of counterfactual dependence relation, yet one statement is acceptable, and the other is not. In addition, our intuitions the acceptability of these statements seem to vary based on the context within which we are assessing them. A proponent of a normative account would argue that these variations in acceptability can be understood as variations in truth-values that are in turn determined by what is and is not normal within each context. Alternative explanations are available, but I will save discussion of them until the next chapter. For now, I will turn to the second puzzle case.

Causation by Omission/The Gardener and the Queen

One of the first areas within causation where this idea of normality determining the truth-values of causal statements was raised was as a response to a particular instance of the problem of selection that occurs in cases of causation by omission. Before stating the

problem, it is worth acknowledging that there is some debate over whether or not there can even be causation by omission. In particular, there are those who follow the Davidsonian view by insisting that causation is a relation that holds between events, and since omissions are not events (they are indeed the absence of an event) they cannot be causes (Davidson 1967). For the purposes of this thesis, I will proceed under the assumption that there *is* causation by omission, as my aim is to account for the folk uses of causation and there is ample evidence to show that the folk often judge that omissions are causes¹⁹.

Yet if we accept that omissions can be causes, then we quickly run into a problem. This problem is best summarised by Lewis, who says of causation by omission that ‘if there is any of it at all, there is a lot of it – far more of it than we would normally want to mention’ (2000, 196). The claim here is that, for any effect, there are many possible events that could have prevented that effect from occurring but in fact did not. This means that for all effects, there will be an almost unlimited number of prior omissions upon which the effect counterfactually depends. The problem is to explain why we select some of these omissions as causes and not others.

To better understand this problem, it will be helpful to consider an example of causation by omission that Blanchard & Schaffer dub ‘the gardener and the Queen’ (2017, 184). This case is a version of the classic plant-watering case that is often used when discussing causation by absence. In these kinds of cases, someone has an obligation to perform a certain task (often watering a plant), but they fail to do so. As a result, they are judged to be the cause of

¹⁹ Most notably, Livengood and Machery (2007) directly tested the intuitions of the folk and found strong support for the claim that they attribute causal status to omissions.

whatever effect resulted from the task not being done. Different versions of this kind of case have been discussed by Hart & Honoré (1985, 38), Stapleton (1994, 122), Beebee (2004, 295-296), Menzies (2004a, 145), McGrath (2005, 126-127), and Sartorio (2010, 262-263). The version presented here originates in Schaffer (2000, 295) and is discussed further by Montminy & Russo (2016, 69-70) as well as Blanchard & Schaffer (2017, 184-185).

Suppose that I own a flowerbed and employ a gardener to regularly water my flowers for me. One day, while I'm on holiday, the gardener decides to not go to work and thus does not water the flowers. Sometime time later the flowers wilt. Intuitively, the gardener's not watering the flowers caused their wilting. On Lewis's account, this is true because the flowers' wilting is counterfactually dependent on the gardener's not watering them: they would not have wilted had she watered them, just as they would not have wilted had the gardener watered them. Yet it seems unacceptable to claim that the Queen's not watering the flowers caused them to wilt. So, we have differing intuitions about the acceptability of the following two causal statements:

- 3) The gardener's not watering the flowers caused them to wilt.
- 4) The Queen's not watering the flowers caused them to wilt.

On its own, the counterfactual account cannot explain this difference in the acceptability of these statements because the flowers' wilting is counterfactually dependent on both the gardener and the Queen not watering them. This raises the question of why our intuitions about the acceptability of (3) and (4) differ.

One answer comes from McGrath, who offers a normative account of causation by omission. She begins with reference to Hart & Honoré's (1985, 38) observation that an omission is

usually ascribed causal status when it deviates from 'man-made norms'. In the case of the gardener, they claim that 'the 'failure' on the part of persons other than the gardener to water the flowers would...be a normal though negative condition [so] no mention of them would be made. The gardener's failure to water the flowers...is not merely a breach of duty on his part but also a deviation from a system or routine' (1985, 38). McGrath then takes this further by explicitly invoking the concept of normality and building it into the truth conditions of causal statements involving omissions (2005, 138-144). On her account, (3) is true because the gardener's omission was not normal, while (4) is false because the Queen's omission was normal. This explains the difference in acceptability between (3) and (4).

Knobe Effects/The Pen Vignette

At this point, it may be tempting to question whether or not the intuitions that I have been describing thus far are as universally held as the proponent of a normative account might have us believe. Fortunately, these intuitions have since been put to the test, with a recent flurry of experimental research into the ways that norms affect our causal judgements. The focus of much of this research has been on a phenomenon that has become known as the Knobe effect (as much of the work on this has been done by Joshua Knobe²⁰). Put simply, this is the effect of moral and social norms on the extent to which we agree with a given causal statement. The most widely known case where this effect has been shown to occur is Knobe

²⁰ See (Knobe and Fraser 2008) (Knobe 2009) (Hitchcock and Knobe 2009) (Knobe 2010).

& Fraser's (2008) pen vignette. This case was presented to a group of subjects²¹ as part of a study into the effects of norms on our causal judgements. The vignette is as follows:

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 repeatedly e-mails 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. (2008, 143-144)

Having been shown the vignette, the subjects were then asked whether they agreed or disagreed with the following statements:

- 5) Professor Smith caused the problem.
- 6) The administrative assistant caused the problem.

On average, the subjects were more likely to agree with (5) than (6). This shows that (5) is more acceptable than (6). This can be explained by appealing to the fact that it normal for the administrative assistants to take the pens, but it is not normal for the faculty to do the same. This suggests that the acceptability of statements (5) and (6) are being affected by

²¹18 students in an introductory philosophy class at University of North Carolina-Chapel Hill (Knobe and Fraser 2008, 144). A fairly small sample size, but many more recent studies have replicated the results with far more subjects.

considerations of what is normal in that context. These results are easily explained if we endorse a normative account of causation.

It should also be noted that these results have since been replicated in multiple studies (Alicke, Rose and Bloom 2011) (Clarke, et al. 2015) (Hitchcock and Knobe 2009) (Phillips, Luguri and Knobe 2015) (Icard, Kominsky and Knobe 2017) (Samland and Waldmann 2016), and the Knobe effect has also been shown to occur in cases of causation by omission (Clarke, et al. 2015). These cases can also be explained if we endorse a normative account of causation, lending further support to the theory.

Causal Superseding

The fourth and final example comes from a study on causal superseding by Kominsky et al. (2015). This phenomenon is related to Knobe effects, except instead of being about the causal status we ascribe to abnormal actions, it concerns how the norm violation of one event affects the causality attributed to other events. This phenomenon is described well by Icard et al. when they state, 'suppose an outcome depends on a causal factor C as well as an alternative causal factor A, such that the outcome will only occur if both C and A occur. Then people will be less inclined to say that C caused the outcome if A is abnormal than if A is normal' (2017, 81-82).

The study conducted by Kominsky et al. provides us with a good case for illustrating this phenomenon. In the experiment, the subjects were divided up into two groups. One group was presented with a morally good version of the scenario and the other was presented with

a morally bad version of the scenario. Both versions of the scenario began with the following background:

Bill's wife, Sue, is out of town for the weekend. She leaves Bill a message that says, 'I just saw this marvellous bookend. It's called a Bartlett bookend. So pretty! I'm going to go back tomorrow and get one. It will be perfect for the left side of our bookshelf' (2015, 200)

The next part of the case involves Bill visiting a friend of his who has a right-side Bartlett bookend. If Bill is able to acquire this bookend and Sue buys the one she was describing, then they will together be in possession of the paired set of bookends. At this point one group was presented with the morally good version of the case:

Bill goes and visits his friend. Bill and his friend talk for a while. Bill asks if his friend is willing to sell the bookend and his friend is happy to sell it. Bill makes an offer, but his friend insists on him not paying so much. Finally, Bill buys the right-side Bartlett bookend from his friend and goes home (*ibid.*)

Meanwhile, the other group were presented with the morally bad version of the case:

Bill goes and visits his friend. Bill and his friend talk for a while. Bill asks if his friend is willing to sell the bookend. His friend tells him it's a precious heirloom and it is not for sale. Bill waits until later when his friend is in the bathroom and slips the bookend into his bag. Finally, Bill leaves his friend's house with the stolen right-side Bartlett bookend in his bag (*ibid.*)

Then both groups were presented with this outcome:

Then the next day, Sue goes and buys the left-side Bartlett bookend. So, when Sue got home, they had the paired set of bookends (*ibid.*)

The subjects in both groups were then asked to what extent they agreed with each of the following two statements:

- 7) Sue caused them to possess the paired set of bookends.
- 8) Bill caused them to possess the paired set of bookends.

The subjects were then asked to rate the valence of both Sue's and Bill's actions. As expected, they viewed Sue's actions as morally neutral and judged Bill's actions in the morally bad example to be much worse than his actions in the morally good example.

On average the subjects who were presented with the morally bad example were less willing to agree with (7) than the subjects who were presented with the morally good example. In other words, the subjects were less inclined to select Sue's action as having caused the outcome when Bill behaved in a way that violated a moral norm. This provides us with another case that can be explained by appealing to a normative account. The explanation would be that the subjects found (8) to be more acceptable than (7) in the morally bad version of the case because Bill's action of stealing from his friend is not normal. In the morally good version of the case, neither persons actions strike us as abnormal, so neither one is selected over the other when we are assessing (7) and (8). Here we have another instance where the acceptability of causal statements are sensitive to what is and is not normal. Again, this can be explained by a normative account of causation.

As a bonus, the study also replicated the Knobe effect, with the subjects being more inclined to agree with (8) in the morally bad version of the case than in the morally good version of the case. As stated earlier, this effect can be easily captured by a normative account of causation.

3. Two Competing Strategies

At this point it should be clear that the normative account is able to provide an explanation for our intuitions about all of these puzzle cases. The question we now need to ask is whether or not this is the *best* explanation. So far, all we really have is a list of cases where it seems that our causal judgements about the acceptability of causal statements are sensitive to certain other claims about what is and is not normal. There are many different ways that these results could be explained, and so it would be far too hasty to conclude from all of this that we ought to adopt a normative account of causation. The reason for this is that, as noted in section 1 of this chapter, acceptability and truth can come apart. The proponent of the normative account wants to argue that our intuitions about the acceptability of these statements reflect those same statement's truth-values. However, this ignores the fact that we can find a particular statement to be acceptable or unacceptable for reasons that are entirely unrelated to its truth value. We can therefore appeal to those reasons to provide an alternative explanation of our intuitions. It is thus worth taking a step back, and considering some of the options that are available when it comes for developing such an explanation.

To this end, it will be helpful to introduce a distinction made by Blanchard & Schaffer (2017, 207) between two competing strategies for explaining our intuitions about the puzzle cases. The first is known as the 'Competence Strategy'. The strategy here is to argue that the variations in acceptability we have identified in the causal statements presented in this chapter, are the result of a competent use of the concept of causation. This is what the proponent of a normative account is doing when they claim that these differences in acceptability are the result of differences in truth-values. The upshot of this strategy is that it

means viewing all of the intuitions we have looked at thus far as being entirely accurate assessments of the truth-values of the relevant causal statements; no mistakes have been made and no biases have affected the outcome of this process. Our folk intuitions about these puzzle cases are true and arrived at competently. Naturally, treating these intuitions this way will lead to us producing a normative and contextualist account of causation.

The second strategy that Blanchard & Schaffer identify is what they refer to as the ‘Heuristics-and-Biases Strategy’, although I will refer to it as the ‘Pragmatics-and-Biases Strategy’, as it better captures the accounts that follow this line of thinking. This strategy begins with the denial that the variations in acceptability we have been studying in this chapter are the result of competently using the concept of causation. Our intuitions about the causal statements in the puzzle cases don’t reflect the truth-values of those statements. It is entirely true to say that the Queen’s omission caused the plants to wilt or that the presence of oxygen caused the forest fire (even in the context of the forest rangers), so our sense that these statements are unacceptable must come from something other than their truth-value.

What that thing may be will depend on the theorist in question. Generally, those who have previously pursued the Pragmatics-and-Biases strategy have tended to argue that the variations in acceptability are the result of either some kind of pragmatic process or a specific cognitive bias (hence the name). Going the pragmatic route means holding that a causal statement could be true but unacceptable for pragmatic reasons, such as it being irrelevant. An example of this is the way in which Lewis appeals to pragmatics to explain the cause/condition distinction and how he does the same when faced with cases of causation by

omission (2000, 196). He cites Grice's (1975) theory of conversational implicature as a way of explaining why claims like (4) seem unacceptable.

An alternative option is to try and explain the variations in acceptability in terms of some cognitive bias. Here the idea is that a given causal statement will be true, but we don't recognise it as such as a result of being misled by some bias. One such explanation is the motivational-bias hypothesis, which is intended as a response to the cases offered by Knobe (Alicke 2008) (Alicke, Rose and Bloom 2011) (Sytsma, Livengood and Rose 2012) (Rose 2017). Here the claim is that our causal judgements are being affected by a bias to confirm an earlier judgement of blame. Another explanation is Blanchard & Schaffer's appeal to Kahneman & Tversky's heuristics-and-biases framework²², and specifically to their discussion of the availability heuristic.

What both pragmatics route and the biases route have in common is that they both lead us to the conclusion that causation is not normative. Any intuitions that we have about the acceptability of causal statements that seem to suggest a normative account, will instead be dismissed as not reflecting the real truth-values of those statements.

The main advantage of adopting this strategy is that it allows us to uphold the intuition that causation is a natural relation. After all, causation can hardly be a relation that is 'out there' in the universe if it is also determined by our constructed views on what is and is not normal.

What we can draw from this is that the same kinds of motivations are involved in both the

²² Hence why Blanchard & Schaffer refer to their approach as the Heuristics-and-Biases Strategy. I will be using the term Pragmatics-and-Biases Strategy because it more broadly encompasses the accounts that I want to discuss in this thesis.

rejection of normative accounts of causation and the rejection of contextualist accounts of causation. There is a strong intuition that causation is a relation that is 'out there' in the world, and the Competence Strategy challenges this idea in the same way that contextualist accounts do.

Despite this, my overall goal with this thesis is to develop and defend a normative constraint on an account of causation that follows the Competence Strategy.

However, I first need to demonstrate why such a constraint is necessary. This will be the focus of the following two chapters. In Chapter 3, I will look at how the existing views that follow the Pragmatics-and-Biases Strategy have so far failed to provide any compelling solutions to the puzzle cases outlined in this chapter. Then in Chapter 4 I will show how the existing views that follow the Competence Strategy tend to present unclear and incomplete accounts by not offering any kind of analysis of normality.

Chapter 3 – The Pragmatics-and-Biases

Strategy

We have now seen that there are two distinct strategies for responding to the puzzle cases that are used to motivate a normative account of causation: the Competence Strategy and the Pragmatics-and-Biases Strategy. So, when we are faced with an example like the gardener and the Queen case, we must decide which strategy to adopt. On the one hand, we could hold that our intuitive judgements about the case are true because they come from a competent application of the concept: the gardener's omission caused the flowers to wilt, whilst the Queen's did not. On the other hand, we could take the view that one of the intuitive judgements is false: either both omissions caused the flowers to wilt or neither of them did. We would then argue that this error is either a result of the fact that we have some pragmatic reason to refrain from asserting that the Queen's omission caused the flowers to wilt, or the fact that we have some bias which prevents us from accurately judging the truth-values of such causal statements.

My view is that the Competence Strategy is the one we should prefer when it comes to explaining our intuitions about the puzzle cases outlined in the previous chapter. In this chapter I will examine how the Pragmatics-and-Biases Strategy has failed to generate any explanations that are better than those offered by a normative account. To do this, I will consider accounts from both branches of this strategy: the appeal to pragmatics and the appeal to biases. Specifically, for pragmatics I will examine Lewis's (2000) Gricean view and Montminy & Russo's (2016) usage of various different pragmatic phenomena (amongst other

things) as part of a pick-and-mix account; then for biases I will examine a recent account from Blanchard & Schaffer (2017) which is based on Kahneman & Tversky's notion of an availability heuristic. For each of these accounts I will show how it fails to successfully explain our intuitions regarding the puzzle cases that I outlined in the previous chapter. This will then allow me to conclude that we have no reason to prefer the Pragmatics-and-Biases Strategy over the Competence Strategy.

What follows will be divided up into two sections, one for each branch of the Pragmatics-and-Biases Strategy. In the first I will examine two accounts that appeal to pragmatics, and in the second I will examine Blanchard & Schaffer's account, which appeals to biases.

1. The First Branch: Pragmatics

I will begin my argument against the Pragmatics-and-Biases view by focusing on those accounts that attempt to explain our intuitions about the puzzle cases by appealing to some pragmatic phenomenon. I will look at two accounts that adopt this strategy. Firstly, I will consider is Lewis's (2000) appeal to the Gricean notion of conversational implicature. Secondly, I will look at the various ways in which Montminy & Russo tackle this problem. For each account I will argue that they fail to explain our intuitions about the puzzle cases. This will allow me to conclude that the existing accounts fail to justify the claim that causation is non-normative, and so the pragmatics branch of the Pragmatics-and-Biases Strategy has not given us a reason to prefer it over the Competence Strategy.

The Appeal to Conversational Implicature

The earliest attempt to provide an account which uses a pragmatic phenomenon in support of the view that causation is non-normative is in Lewis's defence of his counterfactual account of causation (1973) (2000). One objection that is often raised against Lewis's counterfactual account is that it delivers counterintuitive results in the causation by omission puzzle cases. Recall the gardener and the Queen case. There we were asked to consider two causal statements:

- 1) The gardener's not watering the flowers caused them to wilt.
- 2) The Queen's not watering the flowers caused them to wilt.

On Lewis's counterfactual account, both of these statements are true because the flower's wilting is counterfactually dependent on both the gardener and the Queen not watering them. Obviously, this runs counter to our intuition that (1) is acceptable, while (2) is unacceptable. The normative account explains this difference by holding that (1) is true while (2) is false. Lewis explicitly rejects this position when he states that he is trying to capture a 'nondiscriminatory' (read: non-normative and invariant) concept of causation (Lewis 1973, 559). So, he needs to explain why (1) is acceptable and (2) is unacceptable, whilst still maintaining that they are both true.

To achieve this, Lewis cites Grice's account of conversational implicature (Grice 1975). This account is meant to explain how we come to understand the meanings of utterances. Grice claims that, in general, participants in a conversation are expected to observe the following principle:

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (1975, 45)

Grice calls this the 'Cooperative Principle'. The thought is that there is a purpose or a goal to any conversation and each participant in the conversation assumes that all the other participants are aiming to achieve that goal. This means that any utterance a participant makes as part of a conversation is understood as contributing toward the realisation of that goal. Of course, there are many utterances that, when interpreted literally, don't seem to contribute to the realisation of the goal of the conversation. An example of this is sarcasm, where we say something that is literally false and thus is unlikely to help in cases where the goal of the conversation is to provide accurate information. This means that such utterances need to be understood non-literally.

To understand how this works, we need to introduce the Gricean notion of conversational implicature. This notion reflects the idea that what a speaker *means* by an utterance can differ from what they actually say. Every speaker is expected to be following the Cooperative Principle. So, when an utterance is made, we try to understand it as though it were contributing towards achieving the goal of the conversation. If an utterance appears to not be contributing, then it is understood as expressing some implicature which *does* contribute. The content of that implicature will depend on how the utterance fails to contribute. This is where Grice's four maxims come in. These maxims provide us with guidance for how best to follow the Cooperative Principle. They are as follows²³:

²³ The maxims are found in (Grice 1975, 45-46), while the wording I use here is taken from (Davis 2014).

- i) **Quality** - Make your contribution true; so, do not convey what you believe false or unjustified.
- ii) **Quantity** - Be as informative as required and be no more informative than is required.
- iii) **Relation** – Be relevant.
- iv) **Manner** - Be perspicuous; so, avoid obscurity and ambiguity, and strive for brevity and order

If the literal meaning of an utterance fails to fulfil one of these maxims, then the speaker is understood as meaning something other than what they literally said. Such an utterance generates an implicature which both differs from what was literally said and conforms to the Cooperative Principle. For example, suppose that I enter a room clutching a soaking wet umbrella. One of the people in the room asks me what the weather is like and I sarcastically respond, 'it's lovely'. The purpose of this conversation is partly to inform my audience about the fact that it is raining outside. What I have literally said is false and I clearly know that it is false. So, my utterance violates the maxim of quality and thus appears to violate the Cooperative Principle. When my audience, who expect me to act in accordance with the Cooperative Principle, realise that I have violated one the maxims they infer that I must mean something other than what I literally said. They understand my utterance to be generating an implicature which does conform to the Cooperative Principle. In this case it would be the true claim that the weather is terrible today because this moves the conversation towards fulfilling its purpose. To summarise what is being said here, we interpret utterances which don't fulfil Grice's maxims as generating implicatures that conform to the Cooperative Principle.

This now gives us a better understanding of how we can have an unacceptable utterance which is still true. Imagine a different version of the case where the weather really is lovely, and I know that this is the case. My umbrella is still wet, but rather than being the result of

the weather, it is instead the result of my accidentally dropping it in a pond earlier that day. The person I am talking to is unaware that the weather is lovely, and so thinks my umbrella is wet from rain. When they see me, they gesture towards the umbrella and ask what the weather is like. In this case if I were to respond by simply saying 'it's lovely' and nothing else, my audience would understand my utterance as being sarcastic, since I have said something that, to them, is plainly false. My true utterance would therefore generate the false implicature that the weather is terrible. The fact that this utterance generates a false implicature makes it unacceptable to say because it would be as if I were actually saying something false. So, we have a case where a true statement is uttered, but it is judged to be unacceptable because it generates a false implicature.

We can now see how this applies to the gardener and the Queen case. To begin with, it is important that we are clear on the conversational context. So, suppose that I am talking to the owner of the flowerbed, who is asking me to identify what caused the flowers to wilt. It is natural to assume that the goal of this conversation is to identify who or what is to blame for the wilting of the flowers and not to provide a complete account of the causal structure of this case. Taken at face value, then, neither claim (1) nor (2) moves the conversation towards achieving this goal, since neither explicitly says anything about who is to blame. Thus, they both violate the maxim of Relation. This means that an utterance of either must be understood as generating an implicature about who is to blame for the flower's wilting. It would therefore be misleading to claim that the Queen's omission caused the flowers to wilt because it would generate the false implicature that the Queen is to *blame* for the flower's wilting. So, an utterance of (2) would hinder our pursuit of the goal of the conversation. This gives us a pragmatic reason to refrain from an utterance of (2). There are no such reasons to

refrain from uttering (1) since it generates the true implicature that the gardener is to blame. This difference between the truth-values of the implicatures generated by utterances of (1) and (2) is meant to account for the difference in the acceptability of these claims.

This appears to solve the problem raised by this case, and we can use a similar explanation with the other cases. As an example, consider Knobe & Fraser's (2008) pen vignette. Here, as with the gardener and the Queen, we had two causal claims to consider:

- 3) Professor Smith caused the problem.
- 4) The administrative assistant caused the problem.

On average, the subjects judged (3) to be more acceptable than (4). On the Gricean account, the reason for this is, again, that (4) generates a false implicature while (3) does not. Here the conversational context is that of the subjects being asked whether they agree or disagree with (3) and (4). It seems plausible to claim that the subjects took the goal of this conversation to be to identify who is to blame for the problem. This means that an utterance of (4) will generate the false implicature that the administrative assistant is to blame for the problem and would thus hinder the realisation of the conversation's goal. As a result, (4) is unacceptable in this context even though, according to Lewis, it's true.

By now it should be clear how Lewis's appeal to Grice is meant to solve the problems raised by the puzzle cases in the previous chapter. I will now argue that the account fails to explain our intuitions about the puzzle cases, for three reasons. The first is that the implicatures which the Gricean account claims are generated by utterances of (2) and (4) don't appear to be cancellable. A general feature of implicatures is that they can be cancelled by adding a certain

type of conjunct to the end of the utterance that generated them. For example, suppose that Alice and Bob have the following exchange (Grice 1975, 51):

Alice: 'I am out of petrol.'

Bob: 'There's a garage around the corner.'

Here Bob's utterance violates the maxim of Relation which generates the implicature that Alice can get petrol at the garage around the corner. But Bob could have cancelled this implicature with the following:

Bob: 'There's a garage around the corner, but it's closed today.'

This utterance no longer generates the implicature that Alice can get petrol at the garage around the corner. It is considered to be a feature of most (if not all) implicatures that they can be cancelled in this manner.

The problem for the Gricean solution is that the causal claims that we have been considering don't seem to be cancellable in this way. In the Gardener and the Queen case we could try to cancel the implicature by uttering the following:

The Queen's not watering the flowers caused them to wilt, but she is not to blame for the flower's wilting.

The first conjunct in this claim seems just as unacceptable as when it was on its own. This suggests that the unacceptability of the claim is not a result of it generating a false implicature.

The second reason is that our intuitions about the acceptability of (1) and the unacceptability of (2) (ditto for (3) and (4)) remain the same even when we change the conversational context

so that we are not looking for someone to blame. Suppose that instead of flowers, I am growing a poisonous plant which I intend to use as part of a plot to assassinate the Queen. As with before, I go on holiday and the gardener (who is unaware of my plot) neglects to water the plant which then dies. Thus, my assassination plot is foiled. Now imagine that the police discover what I was planning, and I am arrested. News of my arrest is then received by a group of journalists working at a national newspaper. When discussing the case, their aim is to piece together the facts about what happened to cause the plot to fail. Since the outcome was positive they will not be looking for someone to blame. Likewise, we can also suppose that they are not looking to praise the gardener, because he was unaware of my plot. So, it is plausible to hold that the journalists are only interested in the causal structure of the case and not in making judgements of praise/blame. Yet even in this context, they would be unwilling to accept the claim that by not watering the plant the Queen has (unknowingly) prevented her own assassination.

Even so, it might still be objected that, strictly speaking, it is true that the Queen prevented her own assassination, the journalists just refrain from saying it because doing so would violate the maxim of relevance. The problem here though is that it is not clear why this violation would make a true statement unacceptable. In the original version of the case, violating the maxim was bad because it generated the false implicature that the Queen was to blame for flowers wilting. This cannot be the case in the new version of the case, because the outcome is good. Could it then be that the false implicature is that the Queen is praiseworthy for what happened? It could, but it does seem rather odd to say that the natural way of interpreting (2) in this context is as an attempt on my part to communicate that I think the Queen was praiseworthy for preventing an assassination attempt on her own life. It seems

far more likely that the only maxim I would be violating is the maxim of quality, because what I have said is just false.

This problem also does not just arise for cases of causation by omission, as it can also be seen with the pen vignette. Recall that the conversational context was that the subjects were being asked to what extent they agreed/disagreed with (3) and (4), with the presumed goal of identifying who was to blame for the receptionist not being able to write down a message. So, the subjects refrained from agreeing with (4) because it generated the false implicature that the administrative assistant was to blame for the bad outcome. This seems implausible, as it has been shown that the effect remains even when the outcome is good (Hitchcock and Knobe 2009, 603-604). Suppose that, rather than taking a message, the receptionist wanted to use a pen to stab the department chair's eye out²⁴ and she was unable to do so because there were none left. In this context, the participants are unlikely to interpret the question as asking who ought to be assigned blame, yet they still view Professor Smith's taking a pen as causing the problem.

The third reason is, quite simply, that our intuitions about these cases seem to be rooted in more than just an unwillingness to assert a pragmatically inappropriate truth. When considering the gardener and the Queen case, it doesn't feel like (2) is an obvious truth that we just refrain from saying because we don't want to blame the Queen for what happened. Rather, it feels like (2) really is just false, and that is why we don't want to assert it. Indeed, if we were to be asked directly if the Queen caused the flowers to wilt, it seems that our

²⁴ This example is taken from (Knobe 2010, 323).

answer would be 'no'. This has even been tested by Livengood & Machery, who found that (at least in some cases) people will, when asked, disagree with statements where a seemingly irrelevant omission is claimed to have caused some effect (2007, 123). This can also be seen outside of causation by omission, as noted by Hitchcock & Knobe subjects actively disagreed with (4) when presented with the pen vignette (2009, 594). This suggests that the unacceptability of statements like (2) and (4) is a result of them being true but misleading, but rather a result of their being false.

All of this demonstrates that the Gricean account which is suggested by Lewis cannot provide a successful explanation for our intuitions regarding the puzzle cases. This means that another explanation is needed. In the next sub-section, I will consider a different possible explanation suggested by Montminy & Russo. This explanation appeals to the notion of pragmatic presupposition.

Montminy & Russo's Pick-and-Mix Approach

The second account which I will consider comes from Montminy & Russo in their recent paper, 'A Defense of Causal Invariantism' (2016). In this paper they respond to several different puzzle cases that have previously been used to argue for either or contrastivism. Specifically, they focus on the cases given by Schaffer in 'Causal Contextualism' (2012), as well as one case from Northcott's 'Causation and Contrast Class' (2008). Both Schaffer and Northcott use these cases to motivate a contrastive account of causation. As noted in the first two chapters, the contextualist/contrastivist/invariantist debate is closely related to the question of whether or not causation is normative, as many tend to approach that question with the assumption that statements of normality are context-sensitive. This means that Montminy & Russo end up

responding to most of the puzzle cases listed in the previous chapter, although not every case they discuss is a part of this group. For each of the cases, Montminy & Russo begin by outlining how our intuitions about those cases are meant to support contextualism/contrastivism, before demonstrating how the invariantist can account for those intuitions. They then go on to show that the contextualist/contrastivist explanation runs into problems when responding to some of the cases. This allows them to conclude that invariantism is the superior view of causation.

The most important thing to understand when discussing Montminy & Russo's answer to the question of whether causation is normative is that the answer they give varies depending on the case which they are responding to. Broadly speaking, these cases, and Montminy & Russo's responses to them, can be divided up into three groups. The first group includes the cases where Montminy & Russo respond using the Pragmatics-and-Biases Strategy. The responses they give all fall on the pragmatics branch, which is why I am discussing them in this section. The second group includes a different set of cases where Montminy & Russo respond using the Competence Strategy. Finally, there is the group that includes the cases that aren't relevant to the question of whether causation is normative. Since my aim with this chapter is to argue against the Pragmatics-and-Biases Strategy, I will focus on Montminy & Russo's responses to the cases in the first group. However, I shall first comment on their responses to the cases in the other two groups and how these responses relate to the question of whether or not causation is normative.

To begin with, I note that Montminy & Russo seem to be at least somewhat in favour of a normative account of causation. This is shown early on when they claim that the relevance of

a given counterfactual alternative partly 'depends on norms, including statistical, functional, as well as social, moral, and prudential norms' (2016, 52). They then go on to sketch their own account of causation, which is similar to Schaffer's in some ways, but completely rejects the claim that contextual factors are what determine whether or not a possible alternative is salient. Instead, they argue that a possible alternative is relevant 'if it would obtain in virtue of some norm or other' (2016, 70). The important thing to note here is that Montminy & Russo take norms to be a part of the circumstances of a case and not the conversational context. Therefore, the fact that norms affect the truth-values of causal statements does not, on their view, entail that causation is contextual, because the norms themselves are not contextual.

Since this distinction between circumstance and context plays an important role in how Montminy & Russo respond to several of the cases they consider, it will be helpful to examine this in more detail. This point is first made in response to a case from Northcott (2008, 112-114), which he calls the 'two assassins puzzle'. This is an extended version of an example given by Hitchcock (2003, 9-11), which is meant to show that there are cases where we are unclear about what causes what. Hitchcock presents the case as follows:

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. (2003, 10)

Hitchcock concludes that it is unclear whether or not Captain's yelling fire caused Victim to survive, because the threat to Victim's life only came about as a result of Captain's yell. This kind of example is known as a short-circuit case. This is where some event causes a threat to some future event but also causes something which defeats that threat. In this case the threat

was Assistant's firing, and the defeater was Victim's hearing Captain's yell. The fact that Captain's yell caused both the threat and the defeater is what makes it unclear whether or not it is a cause of Victim's survival.

Northcott develops this example by asking us to imagine two distinct versions of this case (2008, 112-113). The first version is as follows:

Captain and Assistant are pursuing Victim through a crowded marketplace. A surge of people causes Captain to become separated from Assistant and threatens to carry Victim off to safety. Captain realises that he needs to do something before Victim escapes and so he yells 'fire!'. Assistant hears the yell and fires, but Victim also hears the yell and ducks. As a result, Victim survives the shot and escapes with the crowd.

According to Northcott, in this version of the case it would be unacceptable to claim that Captain's yell caused Victim's survival. This is because Victim would have still gotten away if Captain hadn't yelled. There is no salient alternative to Captain's yell where Victim dies. Therefore, Captain's yell didn't make any difference to the outcome. Compare this with Northcott's second version of the case:

Captain and Assistant are standing on a balcony overlooking Victim with plenty of time to fire a shot. Captain and Assistant have previously agreed that Captain's raising his finger is the signal for Assistant to fire. However, just as he was about to raise his finger, Captain impulsively yells 'fire!'. Assistant fires and Victim, having heard Captain's yell, ducks. As a result, Victim survives the shot and escapes.

Here Northcott claims that in this version of the case it is acceptable to say that Captain's yell caused Victim's survival. This is because in this case there is a salient alternative to Captain's yell that would have made a difference to whether or not Victim survives. Captain could have raised his finger which would have led to Assistant firing without alerting Victim. Northcott

concludes that we have a case where the following causal statement seems to be either true or false depending on the context:

5) 'Captain's yell caused Victim's survival'

Obviously, Montminy & Russo wish to reject this conclusion by arguing that the truth of (5) doesn't vary with the context that it is made in.

The most important thing to note about this is that Northcott doesn't attempt to argue that the difference between versions in our judgements of the acceptability of (5) are the result of normative considerations. Instead, he argues that the difference can be explained in terms of contrasts (2008, 113-114). In the first version, there was no alternative action that Captain could have performed that would have resulted in Victim's death. On the other hand, in the second version there is a salient alternative to Captain's yelling which would have resulted in Victim's death (namely, Captain raising his finger). This means that the difference in the acceptability of (5) between the two versions of the case can be explained by the difference in the causal contrast in both versions.

What this demonstrates is that Northcott's use of the two assassins case is meant to motivate his contrastive account of causation and not to support the view that causation is normative. This means that the example belongs in the third group of cases that aren't relevant to motivating a normative account. This means that I do not need to respond to what Montminy & Russo have to say about this case, because it has no bearing on the question of whether causation is normative. However, looking at their response will help to clarify their use of the distinction between circumstance and context.

Montminy & Russo begin their response to Northcott by drawing on a point made by Steglich-Petersen. He notes that 'the details added to the two assassins case ... do not provide us with clear causal intuitions by virtue of 'subtle contextual cues' ... but simply by virtue of providing details about the scenario that are relevant to whether [5] is true' (Steglich-Petersen 2012, 134). In other words, the difference in our intuitions regarding the acceptability of (5) between Northcott's two versions of the case is not the result of there being a difference in conversational context between the two versions. Rather, the difference is a result of Northcott 'varying the circumstances surrounding the alleged cause and effect' (Montminy and Russo 2016, 53). Montminy & Russo then conclude that this example doesn't support contextualism because 'a good argument for causal contextualism should appeal to variations in intuitive truth-values that are affected solely by variations in the conversational context' (2016, 53). However, what is important for our purposes is that this case does not present a puzzle that arises from normative factors. As such, it belongs in the third group of cases that are irrelevant to the question of whether or not causation is normative.

The same can also be said of Montminy & Russo's response to Schaffer's (2005, 309-311) use of McDermott's dog-bite case (1995, 531-532), which I will not discuss here for the sake of brevity.

Importantly though, Montminy & Russo use the context/circumstance distinction in response to two cases that *are* used to motivate a normative account of causation. One such case is Schaffer's version of the classic switching case. Imagine a set of train tracks that split up and then reconverge just before a station. Now suppose that there is a switch at the point where the track splits up. This switch controls which route the train takes. Flipping the switch causes

the next train to take a different route than it otherwise would have but doesn't cause the train to arrive at the station, because that would have happened anyway.

Schaffer's version of this case involves a set of tracks that split off in *three* different directions and a switch with three corresponding settings (2012, 38-40). The three settings are named: *broken*, *local*, and *express*. When the switch is set to *broken*, the train turns left and derails. If the switch is set to *express*, the train follows the middle route and arrives at the station quickly. If the switch is set to *local*, then the train goes right and arrives at the station slowly. Schaffer then asks us to consider whether or not the switch being set to *local* caused the train to arrive at the station. He claims that the answer seems to depend on the context. In a context where there is a 'background assumption' (2012, 38) that the switch was set to *broken*, we find it acceptable to claim that the switch being set to *local* caused the train to arrive at the station. However, in a context where there is a 'background assumption' that the switch is set to *express*, it would be unacceptable to claim that the switch being set to *local* caused the train to arrive at the station, since we are assuming it would have arrived there anyway.

According to Schaffer, this means that the following causal statement will be more acceptable in the first context than in the second:

- 6) The switch being set to local caused the passengers to arrive at the station.

Montminy & Russo's response is to again argue that this difference in acceptability is a result of the fact that 'the *circumstances surrounding the switch and the train* change from the first case to the second' (2016, 67). For there to be a difference in the 'background assumption', it must be the case that there is also some difference in where the switch is set most of the

time. This means that the example doesn't support contextualism because the variation in the acceptability of (6) is a result of a change in circumstance and not context. This means that this case cannot be used in support of a contextualist account of causation. However, it may still be used to support a normative account, since a thing being statistically typical means that it will be normal, so here it seems like our intuitions can be explained by appealing to normality's effect on the truth-values of causal statements.

Having stated that norms affect the truth-values of causal statements, Montminy & Russo then go on to examine how this happens in their response to the gardener and the Queen case. As mentioned earlier, Montminy & Russo claim that a possible alternative is relevant for counterfactual analysis if it would obtain 'in virtue of some norm or other' (2016, 70). To see what this means, let's apply it to the following statement:

- 1) The gardener's not watering the flowers caused them to wilt.

On Montminy & Russo's account this comes out as true because the relevant alternative that we appeal to in our counterfactual analysis will be one where the gardener waters the flowers, and in this counterfactual scenario the plants don't wilt because they have been watered. The reason that this alternative is relevant is that there is a norm for the gardener to water the plants. So, the alternative where the gardener waters the flowers is relevant, because it obtains in virtue of a norm. Now compare this to the other statement:

- 2) The Queen's not watering the flowers caused them to wilt.

This statement comes out as false because there is no relevant alternative where the plants are watered. This is because there is no norm for the Queen to water the flowers. This means

that the alternative where the Queen waters the flowers is not relevant, because it doesn't obtain in virtue of a norm.

Montminy & Russo claim that this appeal to norms solves the problems raised by these cases without inviting a contextualist account because these norms are understood to be part of the circumstance rather than the context. Interestingly, this response follows the Competence Strategy because it holds that our judgements about the puzzle cases are the result of a competent use of the concept of causation and not the result of a pragmatic phenomenon or a cognitive bias. This suggests a normative account of causation, but one that is invariantist rather than contextualist, as whether or not something is normal will be a feature of the circumstance rather than the context. For reasons that I will outline in Chapter 6, I support a contextualist view of normality, and so I reject the invariantist view that Montminy & Russo are suggesting here. However, the fact that they are following the Competence Strategy here means that I don't need to argue against this response to show that we shouldn't follow the Pragmatics-and-Biases Strategy.

Of course, Montminy & Russo respond to more cases than the ones which I have examined so far. Their responses to these other cases differ significantly because they appeal to pragmatic phenomena to attempt to explain our intuitions about the cases. The problem with these responses is that the solutions simply cannot be generalised, and Montminy & Russo make no effort to do so.

To illustrate this point, I will examine Montminy & Russo's appeal to the notion of pragmatic presupposition. But before I do that, I will need to first explain what is meant here by the term 'pragmatic presupposition'. This term typically refers to a particular phenomenon that was

discussed by Stalnaker (1999). A pragmatic presupposition is something that a speaker can, under normal conditions, be expected to hold in common with all the other participants in a conversation. In other words, a pragmatic presupposition is a belief in some proposition that the participants in a conversation can reasonably expect to be taken for granted.

Montminy & Russo's use of the term is clearly related, but they mean something slightly more specific. They introduce the term when responding to an example discussed by Schaffer (2012, 40) that is originally from McDermott (1995, 540). The example is intended to illustrate how the way in which we describe events can affect our judgements about the acceptability of certain causal statements. One important thing to note is that this example falls into the third group of cases that are not used to motivate a normative account. Rather, it is intended to motivate Schaffer's contrastive account and nothing more. With that in mind, let's consider the following two statements (2012, 40):

- 7) McEnroe's tension caused him to serve awkwardly.
- 8) McEnroe's tension caused him to serve.

Schaffer's point is that both statements refer to the same events²⁵. Yet (7) seems acceptable, while (8) seems unacceptable. The only difference between the two statements is in how McEnroe's serve is described. For Schaffer, this difference can be explained by there being a difference in the effectual contrast for (7) and (8). With (7), the description of McEnroe's serve as being awkward suggests that the contrast is his serving normally. In (8), the serve is just

²⁵ This point is somewhat controversial and presupposes a view of events as being coarse-grained enough to allow these two events to be identical. Someone like Lewis would disagree with this way of understanding events (1986b).

presented as is, and so we take the contrast to be McEnroe not serving at all. Therefore, with these contrasts, (7) is true and (8) is false. This explains the difference in acceptability between the two statements.

Montminy & Russo reject this and argue that the difference in acceptability can instead be explained using the notion of pragmatic presupposition. They claim that the reason we find (8) unacceptable is that it is 'pragmatically associated' with the proposition that 'McEnroe's being tensed is causally relevant to his serving' (Montminy and Russo 2016, 59), in the sense that if the tension were not there then he wouldn't have served at all. This proposition is an instance of pragmatic presupposition. A presupposition that is associated with a sentence is 'a condition that someone uttering the sentence would typically assume to hold' (Montminy and Russo 2016, 59-60). In other words, a presupposition is a proposition that we can expect someone to believe when they utter a particular sentence. To illustrate this, Montminy & Russo ask us to consider the following sentence:

9) Homer does not know that the earth is flat.

They claim that this sentence is true because one cannot know something that is false. However, it also seems unacceptable. Montminy & Russo argue that the reason for this unacceptability is that (9) is associated with the presupposition that 'the Earth is flat' is a proposition that can be known and therefore must be true. This presupposition is false because, again, one cannot know something that is false. So, a true statement can be made unacceptable when it is associated with a false presupposition.

Likewise, with the McEnroe case, we have a true statement that is unacceptable because it is associated with a false presupposition. Both (7) and (8) are true, but (8) is unacceptable

because it is associated with the presupposition that McEnroe's tension is causally (or perhaps explanatorily) relevant to his serving. Since this presupposition is false, the associated statement is judged to be unacceptable.

This response from Montminy & Russo seems plausible when it is applied to the McEnroe case, but it faces the problem that it doesn't generalise to the puzzle cases that are used to motivate a normative account. This is evidenced by the fact that Montminy & Russo don't attempt to apply the response in this way. We can see why it doesn't generalise when we consider how it might be applied to the pen vignette. To do this, we would need to answer the question of which false proposition is associated with the following statement such that it gives us a reason to not assert it:

- 4) The administrative assistant caused the problem.

The problematic associated proposition is clearly not that the administrative assistant's action is causally relevant (in the sense defined by Montminy & Russo), because their action definitely *is* causally relevant: if the assistant hadn't taken a pen, then there wouldn't be a problem. An alternative might be to say that (4) presupposes that the assistant is to blame for the problem, which is false. However, this runs into the same problem faced by the Gricean solution, which is that if we alter the situation to eliminate any judgements of responsibility, the intuition still remains. All of this suggests that the appeal to pragmatic presuppositions can't be generalised to provide a successful account on the Pragmatics-and-Biases Strategy, even though it may be successful in responding to the McEnroe case.

Of the cases that used in support of a normative account, the only one where Montminy & Russo's response involves an appeal to pragmatics is the case of the Venusians. The puzzle here concerned the following causal statement:

2) The presence of oxygen caused the forest fire.

In the context of the Venusians this statement is acceptable, whereas in the context of the forest rangers this statement is unacceptable. For Montminy & Russo's account to be successful it needs to explain this difference in acceptability.

To accomplish this, they begin by noting that 'the forest rangers are *presupposing* that oxygen is present and wondering about *how the fire started*' (Montminy and Russo 2016, 65). This means that an utterance of (2) in this context 'fails to speak to the question under discussion, and thus violates the maxim of relation' (2016, 65). Additionally, they state that such an utterance 'is arguably inappropriate because it is uninformative, given that the forest rangers already know that oxygen is needed for fire' (2016, 65). So, to summarise, the forest rangers have a pragmatic reason against uttering (2) because it tells them what they already know and thus does not help them realise the goal of their conversation. Likewise, the Venusians are in a similar situation, except they have a reason against ascribing a causal status to the lightning strike. This is, according to Montminy & Russo, all supported by the fact that the forest rangers probably wouldn't deny (2) if it were presented to them, they would merely refrain from saying it themselves.

An immediate response we can make here is to reiterate the third reason given for rejecting Lewis's explanation. Specifically, the available evidence we have suggests that the forest

rangers *would* actually deny (2) if it was presented to them, just as Knobe & Fraser's subjects deny (4) when it is presented to them.

This account also has the problem of not being generalisable. Consider, for example, how this might be applied to the pen vignette. Here the subjects were asked whether they agreed or disagreed with two different causal claims. The goal of this 'conversation' was to gauge their intuitions about the case and they had an equal amount of information about both of the candidate causes. As a result, it can be difficult to see what kind of pragmatic reason the subjects would have against an utterance of (4). Therefore, we again have a problem of generalising the account.

At this point I want to make it clear that my aim here is not to completely refute Montminy & Russo's entire account. Instead, I want to highlight a particular issue with the pragmatics branch of the Pragmatics-and-Biases Strategy. The problem is that many of the accounts that are developed in this branch tend to only apply to a very limited set of cases, and thus don't seem to provide much insight on what appears to be a systematic feature of people's causal judgements. If we compare the responses that Montminy & Russo give on the Pragmatics-and-Biases Strategy to the response they give on the Competence Strategy, there is a clear difference in how successful they are. They offer a plausible response to both the switching case and to the gardener and the Queen case. The responses they give on the Pragmatics-and-Biases Strategy are much less convincing. Of the three that they give, two of them concern cases that aren't relevant to the current discussion of normality, and none of them can be easily generalised to the other cases. As a result, the Competence Strategy comes off

looking like a much more attractive option because it can allow us to respond to a wider range of cases in a unified manner.

Cutting the First Branch

In this section I have examined two accounts which attempt to explain the intuitions we have about the puzzle cases by appealing to two different pragmatic phenomena. With both accounts, I demonstrated that they are unsuccessful in providing such an explanation. Therefore, when it comes to the pragmatics branch of the Pragmatics-and-Biases Strategy, we have not been given a reason to prefer it over the Competence Strategy. This means that the first branch of the Pragmatics-and-Biases Strategy has failed to give us a reason to believe that causation is not normative.

I will now turn to the second branch of the strategy.

2. The Second Branch: Biases

In the previous section I demonstrated that the pragmatics branch of the Pragmatics-and-Biases Strategy failed to give us a reason to prefer it over the Competence Strategy. In this section I will address the second branch, the appeal to biases. A recent account comes from Blanchard & Schaffer's 'Cause without Default' (2017). They offer three arguments against the idea of a normative account of causation. The first argument is that these are heterogeneous concepts whose inclusion in our understanding of causation generates unclarity (2017, 192-195). The second is that many of the puzzle cases that are given to motivate a normative account of causation rely on non-apt causal models and when we amend the examples so that the models *are* apt, we find that they no longer support the view

that causation is normative (2017, 195-206). Finally, the third argument provides an alternative explanation for the psychological evidence that is often given in support of the normative view. This evidence includes the results of studies like Knobe & Fraser's (2008) use of the pen vignette. Blanchard & Schaffer argue that these results can be better explained by appealing to a general cognitive bias concerning the availability of alternatives (2017, 175). They conclude that the results from the psychological studies don't give us any reason to accept the view that causation is normative.

In this chapter I will be responding to this third argument. I will provide a more detailed explanation of the first argument at the end of Chapter 4, where the challenge it raises will be more relevant, as it effectively highlights the shortcomings of the current Competence Strategy accounts. The problem comes from the fact that these accounts fail to offer an analysis of normality, which would then allow them to address some of the unclarities that Blanchard & Schaffer draw attention to. Since I go on to provide such an analysis over in Chapter 6, that can be considered my response to the objection. When it comes to the second argument I am not in a position to give such a developed response, as any discussion of apt causal models will fall outside the scope of this thesis. So, for now, I will focus on just their third argument, with the aim of demonstrating that it fails to give us sufficient reason to prefer the Pragmatics-and-Biases Strategy over the Competence Strategy.

My response will be divided up into two sub-sections. In the first sub-section, I will look at Blanchard & Schaffer's argument. Specifically, I will examine how they use the notion of a cognitive heuristic to explain the psychological results from studies like Knobe & Fraser's in a manner that avoids the claim that causation is normative. I will then go on to show how

Blanchard & Schaffer argue that this is a better way to explain the results than the equivalent explanation from the Competence Strategy.

In the second sub-section I will offer my response to Blanchard & Schaffer's argument. To do this I will examine the two main premises that Blanchard & Schaffer use to justify their conclusion. The first premise is that their account explains the results of studies like the pen vignette without having to hold that our concept of causation is normative. The second premise is that not only is this a successful explanation, but it is also the best available explanation. I will demonstrate that both these premises are false, and therefore Blanchard & Schaffer's conclusion, that we ought to prefer their account, is unsupported.

The Appeal to the Availability Heuristic

For Blanchard & Schaffer to provide a successful account on the Pragmatics-and-Biases Strategy, they need to show that our intuitive judgements about cases like the pen vignette are not instances of a competent use of the concept. Specifically, they have to show that our impulse to disagree that the administrative assistant caused the problem is the result of some mistake or bias. To do this, they draw on Kahneman & Tversky's (1974) heuristic framework, which was later developed by Kahneman & Miller (1986). So, before I look at Blanchard & Schaffer's argument, it will be helpful to understand what this framework is.

Kahneman & Tversky (1974) aimed to understand how it is that we make judgements about probability and frequency. Their general answer to this question is to claim that people use a limited number of heuristics to reduce these complicated judgements of probability and frequency into much simpler judgements (Kahneman and Tversky 1973, 207). Put simply, they

are 'cheap and dirty' (Blanchard and Schaffer 2017, 208) ways of arriving at beliefs about things that are uncertain.

The heuristic that is central to Blanchard & Schaffer's account is the availability heuristic (2017, 208-209). This is one that is identified by Kahneman & Tversky (1973) (1974) as something that we use to determine the likelihood of possible future events. The term 'availability' refers to the ease with which possible events can be brought to mind (1974, 11-14). A central part of this idea is how associations are strengthened through repetition (1973, 208-209). For example, I have had many experiences of rain in Manchester, which means that, for me, Manchester and rain have a fairly strong association. On the other hand, I have no experiences of people being abducted by aliens, so for me there is no association between people and alien abduction. This makes it easy for me to bring to mind the idea of it raining in Manchester and comparatively difficult for me to bring to mind the idea of someone being abducted by aliens. The former possibility is more available to me than the latter. Therefore, I consider it more likely that it will rain in Manchester today than that someone will be abducted by aliens today.

Heuristics like this are useful because they allow us to quickly and easily make judgements that are at least somewhat accurate (after all, it is indeed more likely that it will rain in Manchester than that I will get abducted by aliens). However, they are also prone to biases that can result in false beliefs. For instance, dramatic images can affect the availability of certain possible events. This is because such images are very memorable, so they can be brought to mind more easily than more mundane images, especially if we have recently been exposed to them. An example that Kahneman & Tversky use is that people who have recently

been shown a photo of a car accident will tend to judge the likelihood of such an accident to be higher than what someone who has not been shown that image judges it to be (1973, 230). This is because the subjects who were presented with the photo could more easily bring the possibility of a car accident to mind because the image stuck out in their memories. This shows a bias with the availability heuristic that can result in erroneous beliefs: the people who have been shown the photo tend to overestimate the likelihood of a car accident.

Having explored the role heuristics play in our cognition and seeing how they can sometimes lead to error, it is now time to examine how Blanchard & Schaffer apply this to our concept of causation. The basic idea is that we also employ a heuristic when making causal judgements. To understand how this heuristic works, we first need to look at how Blanchard & Schaffer understand the concept of causation. In past works, Schaffer has defended a contrastive account of causation (2005) (2012). Here that commitment to contrastivism is reaffirmed when Blanchard & Schaffer state that ‘the concept of actual causation is contrastive, with the structure “c rather than c* causes e rather than e*”’ (2017, 209). As is the case with Schaffer’s earlier accounts, the contrasts are selected based on contextual factors, though Blanchard & Schaffer are careful to point out that ‘there is nothing normative in this contrastive approach’ (2017, 210). They do acknowledge in a footnote that normative considerations might be involved in contrast selection, which may initially seem contradictory. However, as I identified in Chapter 1, contrastivist accounts come with the option to eliminate the muddying effects of context by explicitly referring to the contrasts for a given causal statement when that statement is uttered (i.e. by saying the ‘rather than...’ parts of the structure that are normally kept hidden). This will also ensure that no normative concepts can get involved either. If the context plays no role in selecting the contrasts, then,

assuming that these normative concepts are part of the context (as I have been doing here with normality), there will be no mechanism for those normative concepts to determine the truth-values of the resulting causal statements.

Now given that, according to Blanchard & Schaffer, causation has this contrastive nature, the heuristic²⁶ we use to make causal judgements must be a specific kind of counterfactual judgement. Specifically, when trying to determine whether *c* caused *e*, we consider a counterfactual scenario where some causal contrast *c** obtains and then we try to determine whether or not some effectual contrast *e** would have also obtained (2017, 210). If it would have done so, then we conclude that '*c* caused *e*' is true. If it would not have done so, then we conclude that '*c* caused *e*' is false. So, by using the heuristic of counterfactual judgement, we can quickly make judgements about causation.

One important thing to note about this heuristic is that it is a two-step process. When attempting to judge whether or not *c* causes *e*, the first thing we do is select our causal and effectual contrasts. Only once those contrasts have been selected are we in a position to employ our counterfactual reasoning. This is different to the kind of counterfactual account given by Lewis because we aren't merely considering a scenario where the alleged cause

²⁶ I am using the term 'heuristic' here because that is the term that Blanchard & Schaffer use to describe this process of counterfactual reasoning, however I disagree that this process is a heuristic. The purpose of a heuristic is to provide us with a cheap and dirty way of reaching judgements about things that are uncertain. They are an imperfect but practical alternative to a more ideal methodology that would guarantee true beliefs. The process of counterfactual reasoning outlined by Blanchard & Schaffer is not an imperfect alternative to an ideal process, it just is the process we go through if we wish to determine whether or not a given causal statement is true. Blanchard & Schaffer have just given us the truth conditions for causal statements, rather than a cheap and dirty heuristic for determining whether those statements are true. In Chapter 7 I raise a similar point against Sytsma et al. (2012), where what they describe as a mistake is, I argue, actually an example of the concept of causation being applied competently.

doesn't occur, but rather one where a specific counterfactual alternative occurs instead. This means that the selection of alternatives is important for what our causal judgements will end up being.

It is at this point where we can start to see how bias can creep into the use of this heuristic. With the availability heuristic we saw that presenting a dramatic image to someone tends to lead to them making mistaken judgements about the likelihood of possible future events. This was because the dramatic images were more available to the subjects when it came time to make their judgements. Likewise, Blanchard & Schaffer claim that the process of contrast selection is partly affected by availability. They state that, "'Deviant' events tend to leap out as especially salient to people and tend to trigger thoughts of the more normal alternative, while 'default' events tend to duck out of view and not trigger thoughts about alternatives at all' (2017, 210). This means that when we are attempting to determine the cause of some effect, abnormal events tend to leap out at us because it is easier to imagine a more normal alternative to those events. In other words, we have a bias towards assigning causal status to abnormal events and ignoring other more normal events because it is easier to imagine how those abnormal events could have been otherwise. This bias can then lead to mistaken causal judgements.

To illustrate how we make these mistaken judgements, Blanchard & Schaffer use Knobe & Fraser's pen vignette (Blanchard and Schaffer 2017, 211). As a reminder, the puzzle in this case is to explain the difference in acceptability between:

- 3) Professor Smith caused the problem.
- 4) The administrative assistant caused the problem.

So, Blanchard & Schaffer need to explain why the subjects found (3) acceptable and (4) unacceptable, without claiming that this is the result of a competent use of the concept of causation. They state that when the subjects were asked whether Professor Smith caused the problem ‘the contrastive view of causation has it that [they] need to consider a salient event involving Professor Smith, and a salient contrast event, as well as a salient event involving the problem ... and a salient contrast to that’ (2017, 211). Then once the contrasts have been selected, the subjects can be ‘expected to use the heuristic: “if the salient contrast event to the salient event involving Professor Smith ... had occurred, would things have then run smoothly?”’ (*ibid.*). The same process is also involved when we are attempting to determine whether the administrative assistant caused the problem.

This means that to understand where the subjects are supposed to have gone wrong, we first need to select the causal and effectual contrasts for both (3) and (4). Since the alleged effect is the same in both (3) and (4), and they are presented in the same context, it stands to reason that the effectual contrast would be the same for both. Specifically, we are looking for a salient alternative to the receptionist being unable to take a message. Given the context of the study, it seems reasonable to say that the subjects would consider the salient alternative to be the receptionist’s being able to take a message and everything running smoothly. So, this is the effectual contrast for both (3) and (4).

It is with the causal contrast that we start to see the effect of our bias. According to Blanchard & Schaffer, the difference in acceptability between (3) and (4) can be explained by the ‘differences in the cognitive availability of the salient events as well as their salient contrasts, for the professor as opposed to the assistant’. Professor Smith’s taking a pen is an

abnormal/deviant event. So, when we are making a judgement about what caused the problem, Professor Smith's taking a pen 'will tend to leap to mind and will tend to trigger thoughts of the default alternative of Professor Smith's not taking any pen' (Blanchard and Schaffer 2017, 211). By comparison, the normal/default event of the administrative assistant's taking a pen 'will not tend to leap to mind so readily and will not so readily tend to trigger thoughts of the deviant alternative of the assistant's not taking any pen' (2017, 211). What this means is that, when we are considering the question of what caused the problem, Professor Smith's action is more available to us because it tends to trigger thoughts of the more normal alternative, whilst the administrative assistant's actions are less available because they will not trigger thoughts of the default alternative. The result of this is that we are more willing to accept the claim that Professor Smith caused the problem, than the claim that the administrative assistant caused the problem. All this, despite the fact that, according to Blanchard & Schaffer's contrastivist account, both claims are equally true.

Blanchard & Schaffer claim that this demonstrates how our counterfactual heuristic contains a bias that has led us to a mistaken causal judgement, that Professor Smith's action is a cause, but the administrative assistant's action is not. This follows the Pragmatics-and-Biases Strategy because it explains our intuitions about the case without claiming that they are the result of a competent use of the concept of causation. However, this on its own doesn't get us to the conclusion that we ought to prefer their version of the Pragmatics-and-Biases Strategy over an account that follows the Competence Strategy. It is entirely consistent to say that Blanchard & Schaffer have provided a successful explanation, but an account which follows the Competence Strategy provides a better explanation and so we should prefer that account.

This gives us the second premise of Blanchard & Schaffer's argument: that their explanation is better than any account which follows the Competence Strategy. Their argument for this premise is very brief and appears only in the penultimate paragraph of the paper. As such, the entire argument is contained within the following quotation:

...the rationale for pursuing some form of Heuristics-and-Biases Strategy²⁷ is not just to keep the probability/causal calculus pure, and not just to fit the structure of existing psychological theorizing, but to best explain the generality and systematicity of availability effects. On the rival Competence Strategy ... in order to account for the generality of availability effects, a theorist extending the ... strategy would presumably wind up positing lots of individual concepts (the concept of frequency, the concept of probability, the concept of representativeness, etc.) that just so happened to each make reference to norms, and just so happened to each do so in similar ways, so as to generate similar availability effects. From the perspective of this sort approach it just looks like a pure accident that these many concepts each just so happen to make reference to norms in similar ways. A theorist invoking the [Heuristics-and-Biases] strategy instead has a ready explanation for the generality and systematicity of normative influences. After all, she is positing a single background feature of cognitive performance playing a role in all of these domains. (Blanchard and Schaffer 2017, 211)

The first sentence of this quotation offers us three reasons why we might prefer Blanchard & Schaffer's account to one that follows the Competence Strategy:

- i) It keeps the 'causal calculus pure'.
- ii) It 'fit[s] the structure of existing psychological theorizing'.

²⁷ As I stated earlier, Blanchard & Schaffer use the term 'Heuristics-and-Biases strategy' to refer to accounts like theirs, which appeal to Kahneman & Tversky's heuristics framework. This shouldn't be confused with the term 'Pragmatics-and-Biases Strategy', which I am using as more of an umbrella term for all of the views which are opposed to the Competence Strategy.

iii) It 'best explain[s] the generality and systematicity of availability effects'.

Regarding (i), the claim here is that Blanchard & Schaffer's account is preferable because it allows us to construct causal models which don't have to incorporate normative concepts. I can think of two reasons why this would be a good thing. One is that it makes it simpler to model systems because we don't have to consider unclear normative concepts. I suspect that this is not what Blanchard & Schaffer had in mind, as it is more or less just a restatement of their unclarities objection (the first argument I mentioned in the beginning of this section). The second reason is that it makes causal models more scientific, in the sense that it allows them to represent causation as some mind-independent feature of the world. Fundamentally, this is an appeal to the intuition that causation is a natural relation, with the argument being that there is something wrong with our causal models if they don't respect this intuition. Blanchard & Schaffer's use of the term 'pure', suggests that this is what they have in mind when they give (i) as a reason to prefer their account.

Of course, I don't want to overstate the importance of (i) in Blanchard & Schaffer's argument, as it only receives an offhand mention. I believe that this is because the intuition that causation is a natural relation supports all accounts that don't follow the Competence Strategy, not just Blanchard & Schaffer's. Of course, the intuition presents a serious challenge that any normative account of causation will need to overcome, but it does not give us a reason to prefer Blanchard & Schaffer's account over any other account on the Pragmatics-and-Biases Strategy. As such, here is not the appropriate place to respond to this objection, so I will postpone discussion of this point until Chapter 7.

This leaves us with reasons (ii) and (iii), which are closely related. With (ii), the claim is that we ought to prefer Blanchard & Schaffer's account because it coheres nicely with the existing psychological literature. More specifically, it coheres with Kahneman & Tversky's work because it uses their heuristics framework. This is taken further with reason (iii). Here the claim is that we ought to prefer Blanchard & Schaffer's account because it explains why we see the same pattern of availability influencing our judgements in various different domains. We have seen already that availability affects our judgements of probability, as demonstrated by the effect that showing images of car accidents had on Kahneman & Tversky's subjects. As noted by Blanchard & Schaffer in the quoted paragraph above, similar effects involving availability have been observed in our judgements of frequency (Kahneman and Tversky 1973) and representativeness (Kahneman and Tversky 1972). These are what Blanchard & Schaffer are referring to when they use the term 'availability effects', and the occurrence of these effects has been well-documented in the psychological literature. This gives us a phenomenon that requires a psychological explanation.

Within the non-causal domains, this explanation is given by Kahneman & Tversky's use of the availability heuristic. In the case of probability, the fact that more available events are judged to be more likely is explained by the fact that we use the availability heuristic when making those judgements. Since the availability of an event will be partly affected by how normal it is (an event that is more normal will be more available than one that is less normal), the result is that our judgments of probability will be affected by facts about normality.

Blanchard & Schaffer's contribution is to simply extend this account to the domain of causation, where our judgements also seem to exhibit availability effects. Their claim here is

that our judgements about the cases like the pen vignette are instances of this effect. These are cases where events that have more available alternatives are favoured when it comes to attributing causal status. The fact that these events are favoured can, according to Blanchard & Schaffer, be explained by the fact that we use the availability heuristic when making causal judgements. This also explains why causal judgements are affected by norms, because norms affect the availability of events. This means that the occurrence of availability effects in our causal judgements can be explained using the same framework as our judgements in the non-causal domains. The result of this is that we have a unified explanation for the occurrence of availability effects in several different domains. So, not only should we prefer Blanchard & Schaffer's account because it coheres with existing work in psychology, but it also contributes to a more general existing account that explains the occurrence of a well-documented psychological phenomenon.

In the quoted passage above, Blanchard & Schaffer compare how their account handles the availability effects to how the Competence Strategy would do the same. They claim that a follower of the Competence Strategy will need to give a different explanation of the occurrence of availability effects for each domain that they occur in. In other words, they will need to give one explanation for the availability effects in causation, another for probability, and another for frequency (as well as any other domain where availability effects occur). A consequence of this is that we will end up with several distinct concepts which all seem to accidentally generate the same availability effects. In other words, we would have to hold that the concepts of both causation and probability (alongside several others) just happen to be such that they both contain the same kinds of references to norms that produce the same effect. For Blanchard & Schaffer, the fact that their account avoids this implausible result is

the main reason why they believe that we ought to prefer it to an account that follows the Competence Strategy.

This completes my summary of Blanchard & Schaffer argument for the Pragmatics-and-Biases Strategy. In the next sub-section, I will argue that they have failed to demonstrate that this strategy is superior to the Competence Strategy.

Responding to Blanchard & Schaffer's Argument

Based on what we saw in the previous sub-section, Blanchard & Schaffer's argument seems to argument seems to rely on two key premises:

- I) The cases where it seems that our causal judgements are affected by normative considerations can be explained using the heuristics framework outlined in the previous sub-section.
- II) That explanation is better than the explanation given by the Competence Strategy.

With these two premises, Blanchard & Schaffer are then able to conclude that we ought to prefer their account over any account that follows the Competence Strategy. In this sub-section, I will respond to Blanchard & Schaffer's argument by criticising both premises in turn. However, before I offer my objections, I will first need to offer some general remarks about how contrastivism and the heuristics framework interact, as well as the role that availability is playing in the account.

The contrastivist argues that causation is a quaternary relation that holds between a cause, an effect, and one or more contrasts. This is a claim about the nature of causation itself. Now the heuristics framework, as presented by Kahneman & Tversky, only tells us about how people come to make judgements about things that are uncertain. For instance, in the domain

of probability the framework explains how people come to make particular judgements, but it tells us nothing about the nature of probability itself. Talk of heuristics tells us nothing about whether or not 'x is likely' is true; all it tells us is how someone could come to believe that statement, regardless of whether it is true or false. Likewise, talk of heuristics won't tell us anything about the truth-values of causal statements.

This then raises the question of how the two views interact. The question here concerns the precise role that the heuristics are playing in Blanchard & Schaffer's account. To answer this question, we should take a closer look at the analogy they draw between the domains of causation and probability. In the domain of probability, availability is one of the heuristics that we use to judge how likely a given possible future event is. When asked whether or not a given possible future event is likely, our answer will be correlated with the degree to which that event is available to us. With this in mind, it is plausible to suppose that if we were presented with two possible future events and asked which one was the most likely, we would, all things being equal, choose the event that was the most available. So, availability directly affects our judgements of probability by enforcing a correlation between the availability of a possible event and how likely we judge it to be.

This suggests a clear point of disanalogy with the role that the heuristic plays in the domain of causation. If it were the case that availability was directly affecting which event we ascribe causal status to, then it is difficult to see why Knobe & Fraser's subjects would choose Professor Smith's action over the administrative assistant's action. After all, both events were mentioned in the vignette and so we would think that they are equally available to the subjects. Alternatively, we might think that the more normal event will be more available; in

which case the subjects would have picked out the normal action of the assistant's taking a pen over the abnormal action of Professor Smith doing the same. It does not seem that there is any plausible understanding of availability that would direct the subjects to assign causal status to Professor Smith's action and not the assistant's action. What this shows is that availability cannot be directly affecting our selection of causes. This means that the role availability plays in our causal judgements will have to be different to the role it plays in our probability judgements.

In the previous section I noted that, according to Blanchard & Schaffer, we arrive at our causal judgements via a two-step process: first we select the contrasts (c^* and e^*), then we infer 'c caused e' from 'if c^* had occurred, e^* would have occurred'. Since heuristics aren't involved with the second step, it must be the case that they are used in the process of contrast selection. In other words, the heuristic is somehow involved in the process of determining which possible alternative to the cause/effect is salient.

With this disanalogy in place, I return to the question of the role that availability plays in Blanchard & Schaffer's account and to raise my objection against the first premise in their argument. My objection is that Blanchard & Schaffer's account is unsuccessful at explaining the results of the pen vignette, and this is because they fail to properly define the role that availability is playing in the selection of contrasts. Recall that Blanchard & Schaffer explain the results by claiming that the subjects were more willing to accept the claim that Professor Smith caused the problem because the alternative where he doesn't take a pen is more available than the alternative where the administrative assistant doesn't take a pen. The problem with this is that it misrepresents how the study was carried out. It is not the case

that the subjects were presented with the vignette and then asked the open-ended question of ‘who caused the problem?’²⁸. Rather, they were explicitly asked whether they agreed or disagreed with the following two statements:

- 3) Professor Smith caused the problem.
- 4) The administrative assistant caused the problem.

This difference affects the role that availability can plausibly occupy in the account. If the subjects were asked an open-ended question, then it would be plausible to say that their answer was determined by which event had the more available alternative. The problem for Blanchard & Schaffer is that the subjects were not asked an open-ended question; they were asked whether they agreed/disagreed with both (3) and (4). This means that Blanchard & Schaffer’s explanation that the subjects simply *prefer* (3) due to it having a more available causal contrast than (4) doesn’t hold. They need to explain how availability affects the judgements of the subjects with the result that they agree with (3) and disagree with (4).

Based on what we have seen thus far, it seems that the only role availability must play in our assessments of causal statements is in the process of contrast selection. In other words, we use the availability heuristic when selecting the contrasts for a given binary causal statement. So, the alternatives that we deem salient will be the ones that are the most available. This means that any mistaken causal judgements can be explained by the fact that a cheap and dirty heuristic was used to make those judgements – one which, on this occasion, led to error.

²⁸ Blanchard & Schaffer do, in fact, begin by describing the subjects as responding to the question of whether or not Professor Smith/the administrative assistant caused the problem. However, the way that they describe the situation changes such that the question becomes more open-ended.

The problem with this can be seen by revisiting the pen vignette and consider how the subjects would reason about statements (3) and (4) when asked the closed questions about whether they agreed or disagreed with each statement. If Blanchard & Schaffer are right, then the subjects would approach the question by first selecting the contrasts for the statement indicated by the question. The most available alternative to the effect in both statements is everything running smoothly due to the pens not running out. The reason for this is that this contrast is more normal than what actually happened. Professor Smith's not taking a pen is the causal contrast for (3), again for the reason that it is more normal than his taking the pen in the actual world. So, the complete version of (3) is:

3*) Professor Smith's taking the pen rather than his not taking the pen cause the problem rather than everything running smoothly.

This is clearly true, which explains why the subjects were willing to agree with (3). The problem comes with trying to identify the causal contrast for (4), which is not something we need to do when trying to answer the open question of 'who caused the problem?'. Here Blanchard & Schaffer simply claim that the administrative assistant's taking the pen 'will not tend to leap to mind so readily and will not so readily tend to trigger thoughts of the deviant alternative of the assistant's not taking any pen' (2017, 211). We could try applying this to the subject's thinking about the closed question by arguing that there is *no* available alternative to the assistant's taking a pen. The problem is that this seems to be a rather odd thing to say. Clearly there is an available alternative to the assistant's taking a pen: the assistant's *not* taking a pen. This alternative may be normal, but there is more to availability than normality. After all, the subjects are trying to find an alternative to the assistant's taking a pen, so surely

the assistant's not taking a pen would be the thing that leaps immediately to mind. It is hard to see how, if the subjects really are using the availability heuristic, they would select anything other than the assistant's not taking a pen as the causal contrast. This would mean that the complete version of (4) is:

4*) The administrative assistant's taking a pen rather than not taking a pen caused the problem rather than everything running smoothly.

The problem for Blanchard & Schaffer is that this is true, and so it doesn't explain why the subjects disagreed with (4).

To respond to this, Blanchard & Schaffer have to show how using the availability heuristic does not, in fact, lead us to select the assistant's not taking a pen as the causal contrast for (4). There are two ways for them to do this: they can either argue that there is some other alternative that is *more* available to the subjects, or they can argue that there are no available alternatives at all.

If Blanchard & Schaffer take the first option, then they could argue that the available alternative is something like – for example – the assistant's taking two pens. If we take this as the causal contrast, then the complete version of (4) is:

4') The administrative assistant's taking a pen rather than taking two pens caused the problem rather than everything running smoothly.

This is false because if the assistant had taken two pens then there still wouldn't have been any pens for the receptionist to take the message. This would explain why the subjects tended to disagree with (4).

However, this response seems ad hoc. There doesn't seem to be any reason to think that the assistant's taking two pens is more available to the subjects than the assistant's not taking any pens. It is not the case that the former is more normal than the latter, since the norm in the department is that the assistant *can* take the pens, not that they *must* take the pens. There also doesn't seem to be any other alternatives that are more easily brought to mind in the context of the subjects than the assistant's not taking a pen. So, it is unlikely that the availability heuristic is producing a mistaken judgement by picking out an alternative other than the assistant's not taking a pen.

This means that Blanchard & Schaffer are forced to hold that there are *no* available alternatives to the assistant taking a pen. So, the subjects are unable to construct a complete causal statement because there is no causal contrast available to them. This then leads to their refusal to agree with (4).

The problem with this is that it seems absurd to hold that when the subjects were trying to identify a salient alternative to the assistant's taking a pen, the possibility of the assistant not taking a pen was completely unavailable to them. The availability of an alternative is the ease at which it can be brought to mind, and in the context of the subject it is hard to see how the possibility of the assistant not taking a pen cannot be easily brought to mind in this context. Admittedly, it might be a bit harder to bring it to mind than it is to bring the professor's not taking a pen to mind. But, in figuring out whether they agree or disagree with (2), what the professor does or should do plays no role.

All of this demonstrates that Blanchard & Schaffer's account fails to explain our intuitions about cases like the pen vignette. The reason for this failure is that the account only seems to

apply to cases where we are asked to pick out one event as a cause, and it completely breaks down in cases where we are expected to assess specific causal statements. This means that Blanchard & Schaffer have failed to provide sufficient reason to accept premise (I) of their argument.

One response that Blanchard & Schaffer might give at this point is to draw attention to the fact that the subjects in this study were asked whether they agreed/disagreed with both (3) and (4) at the same time. In this context the subjects might have thought that they were being asked to pick just one statement to agree with and so they chose the one with the most available alternative to the cause. If this is the case, then Blanchard & Schaffer's account can successfully explain the results.

This response implies a prediction that if the subjects were presented with only one of the statements, then their judgements would have reflected the actual truth-values of the statements (i.e. they would have agreed with both). Such a test has already been carried out by Clarke et al. (2015). They presented an analogous case to their subjects, who were then asked whether they agreed or disagreed with exactly one statement (Clarke et al., 2015, p. 282). The results of this study support the view that causation is normative: the subjects tended to agree with the statement that assigned causal status to an abnormal event and didn't agree with the statement that assigned causal status to a normal event. Therefore, this response fails, and we still have no grounds for accepting premise (I) in Blanchard & Schaffer's argument.

This gives us grounds to reject Blanchard & Schaffer's conclusion that we ought to prefer an account that follows the Heuristics-and-Biases Strategy over an account that follows the

Competence Strategy. However, I also believe that Blanchard & Schaffer fail to give us sufficient reason to accept premise (II).

Recall from the beginning of this section that premise (II) is the claim that Blanchard & Schaffer's account, if true, provides a better explanation of the results of the studies than an account that follows the Competence Strategy. Their argument for this premise focuses primarily on the fact that adopting the Heuristics-and-Biases Strategy allows us to explain the occurrence of availability effects across a wide variety of domains as being the result of a bias in the use of the availability heuristic. This compares favourably with the explanation offered by the Competence Strategy, where it is just a mere accident that we find availability effects within all of these different domains. Since Blanchard & Schaffer's account doesn't entail this kind of accidental uniformity, it is a better explanation than the one offered by the Competence Strategy.

Against this premise I dispute the claim that the Competence Strategy can only offer a mere accidental uniformity that is inferior to the uniformity offered by the Heuristics-and-Biases Strategy. Specifically, this 'accidental uniformity' is not as obviously problematic as Blanchard & Schaffer make it seem. Their description gives the impression that, if the Competence Strategy were right, we would be surprised to discover that the content of all these different concepts (causation, probability, frequency, etc.) are such that they all produce availability effects. Yet this is exactly what we ought to expect when we consider the role that these concepts play. Specifically, they all play an important role in our decision-making process. For instance, when deciding what actions to take, it is vital that we know how likely certain possible future events are, so we make use of the concept of probability. Likewise, causation

is useful because it enables us to intervene in the course of events to achieve a desired outcome²⁹. This shows that these different concepts are all related by the fact that they all play a role in our decision-making process.

With this in mind, it is less surprising that all of these concepts just happen to produce availability effects. In life, we are often forced to make decisions quickly. This means that a great deal of importance will be given to how quickly information can be brought to mind. After all, if we are in a situation where we need to make a decision quickly, we might not be able to afford to stand around and try to bring less available things to mind. So, we should expect that all the concepts involved in this process would be sensitive to availability in some way. It is not a mere accident that we find availability effects in all these domains.

What all of this shows is that Blanchard & Schaffer have failed to defend the claim that their account offers a properly unified explanation for availability effects, and that the Competence Strategy can only provide a mere accidental unity. From this, we can conclude that Blanchard & Schaffer have not given sufficient reason for us to accept premise (II), in addition to premise (I).

Cutting the Second Branch

These objections demonstrate the Blanchard & Schaffer have failed to provide a successful argument against the Competence Strategy. Therefore, we don't currently have any reason

²⁹ The fact that causation can help us to intervene in this way forms the basis for interventionist accounts of causation like Woodward's (2004).

to prefer an account on the biases branch of the Pragmatics-and-Biases Strategy over an account on the Competence Strategy.

3. Moving to the Competence Strategy

In this chapter I have argued against the Pragmatics-and-Biases Strategy as a means of responding to the problem raised by the puzzle cases from the previous chapter. In the first section I examined two accounts that appeal to pragmatics, and in the second section I examined one that appeals to biases. In both cases I demonstrated that the accounts fail to properly explain our intuitions about the puzzle cases, and consequently they fail to provide a better explanation than the one offered by the normative account.

This means that the field is clear for us to pursue an account that follows the Competence Strategy. In the following chapter I will analyse the existing attempts to do this and argue that these accounts do not succeed because they fail to offer a precise or complete analysis of normality. Then in subsequent chapters, I will fill this gap by providing my own analysis of normality.

Chapter 4 – Developing the Competence Strategy

At this point in my thesis, we have seen that there is a problem for invariantist and non-normative accounts of causation. This is the problem of selection, and it can be seen in the puzzle cases presented in Chapter 2. There are two strategies we could use in response to this problem: the Competence Strategy and the Pragmatics-and-Biases Strategy. In the previous chapter I demonstrated that the Pragmatics-and-Biases Strategy has so far been unsuccessful and so we should instead consider an account that follows the Competence Strategy. This means arguing that the intuitions we have about the puzzle cases are not the result of some mistake or pragmatic consideration. Instead, they are instances of a competent use of the concept of causation. By arguing for this claim we would be committing ourselves to the view that causation is normative. For the remainder of this thesis, I will focus on developing and defending such an account.

The best place to start will be to look at some of the existing attempts to provide an account that follows the Competence Strategy by incorporating normative notions into the truth-conditions of causal statements. To do this, I will follow Menzies by distinguishing between two different ways of applying these normative notions to existing counterfactual accounts of causation (2010, 347). The first is the view proposed by Hitchcock & Knobe (2009) where normative notions determine that counterfactual alternatives to the actual world are relevant to our causal judgements. The second view is the one that Menzies adopts, although it was originally proposed by Hart & Honoré (1985). On this view, a cause is something that makes

the difference between the normal course of events and the actual course of events. So, when making causal judgements we compare the actual course of events with what would 'normally' have occurred. Additionally, I will also be looking at a third, more recent account from Icard, Kominsky & Knobe (2017) where normality affects people's judgements of causal strength.

I will clarify these three positions and the differences between them over the course of this chapter. My aim will be to show that all these views are unsuccessful because they fail to clarify certain key normative notions and to fully explain how those notions are integrated into their accounts. My argument for this conclusion will be divided into four sections. In the first three sections I will examine the accounts of Hitchcock & Knobe, Menzies, and Icard et.al. respectively. In the final section I identify a gap that all these accounts share in not providing an analysis of normality. I will then show how this gap leads to some problematic unclarities that are best articulated by Blanchard & Schaffer (2017, 192-195) I will then argue that, although this objection is not fatal to the Competence Strategy, it does demonstrate the need for these accounts to provide a conceptual analysis of normality. Providing such an analysis will be the focus of the following two chapters.

1. The Relevance of Counterfactual Scenarios and Possible Worlds

It should be clear at this point that the aim of these accounts is to explain how normative considerations affect the truth-values of causal statements. The most straightforward way to do this from the perspective of a counterfactual account is to argue that these normative considerations determine which counterfactual scenarios are relevant when evaluating a

given causal statement. This is the view that Hitchcock & Knobe (2009) defend, and it will be the focus of this section.

The easiest way of understanding this account is to compare it to the original Lewisian view. According to Lewis (1973), the causal statement 'c caused e' is true if the following counterfactual holds³⁰:

If c had not occurred then e would not have occurred.

So, for us to understand the truth conditions of causal statements, we first need to identify the truth conditions of these kinds of counterfactuals. For Lewis, the truth-conditions of counterfactuals are understood in terms of similarity relations between possible worlds. He offers the following analysis:

A counterfactual 'If it were that A, then it would be that C' is (non-vacuously) true if and only if some (accessible) world where both A and C are true is more similar to our actual world, overall, than is any world where A is true but C is false. (Lewis 1979, 465)

Put simply, for a given counterfactual 'A > C' to be true it must be the case that of all the worlds where A is true, the one that is most similar to the actual world is one where C is also true.

It should be clear from this that the notion of similarity plays an important role in the account.

This suggests an ordering of possible worlds that is often understood in terms of closeness.

³⁰ Strictly speaking this doesn't give us Lewis's complete account of causation, since he considers counterfactual dependence to be sufficient but not necessary for causation. What is necessary (and also sufficient) for causation is that there is a chain of counterfactual dependence relations that link the cause to the effect. This extra detail is important to Lewis's account, but it would overcomplicate things to introduce it here.

Roughly speaking, worlds that are more similar to the actual world in terms of natural laws and matters of fact are considered to be closer to the actual world than the worlds that are less similar in this regard. So, the closest possible worlds will be those that differ to the smallest degree in terms of matters of fact and the natural laws. This means that the actual world will be the closest possible world to itself, as it does not differ at all from itself in terms of the laws and matters of fact. The more a world differs from the actual world in its matters of facts and natural laws, the more distant it will be from the actual world. This is how similarity orders possible worlds. The worlds that are relevant for the evaluation of counterfactuals are those that are the most similar to the actual world in terms of the matters of facts and natural laws.

It should be clear how all of this can be applied to causation. The counterfactual we are interested in has the antecedent 'if c had not occurred', so the world that we are interested in is the closest possible world at which, through some small miracle, c did not occur. If this world is one where the consequent is also true (e doesn't occur), then the counterfactual is true and thus the statement 'c caused e' is also true. What this shows is that, on Lewis's account, similarity is the thing that determines which counterfactual scenarios/possible worlds are relevant when making causal judgements.

Hitchcock & Knobe's account resembles Lewis's, with one key difference. The difference is that rather than the relevance of counterfactual scenarios being determined by similarity, it is instead determined by normality (at least in some cases). Their account begins with the following claim:

[P]eople's judgments about the relevance of counterfactuals depend in an essential way on norms. The basic suggestion is that people classify events on a scale from 'normal' to 'abnormal.' Then, when something abnormal occurs, they regard as relevant counterfactuals those that involve something more normal having occurred instead. (Hitchcock and Knobe 2009, 597)

In other words, when we are attempting to discern the cause of some abnormal event, we use a counterfactual analysis where we appeal to possible scenarios/worlds that are more normal than the actual world. This is a descriptive claim about the kinds of non-actual scenarios that people consider to be relevant when using counterfactual reasoning. As an example, recall the gardener and the Queen case. When trying to judge what caused the flowers to wilt, we appeal to the world where the gardener actually waters the flowers, because that world is more normal than the actual world. We don't appeal to the world where the Queen waters the flowers because that world is considerably less normal than the actual world and is therefore considered to be irrelevant.

Crucially though, this way of ordering worlds remains the same even if we change the distance between these worlds and the actual world. For instance, suppose we said that the gardener chose to take an unscheduled holiday to Australia on the day that he was supposed to water the flowers. Meanwhile the Queen just so happened to be walking by the flowers when there was a full watering can on the ground near the flowerbed. In this version of the case, the world where the gardener waters the flowers would be quite far from the actual world: a lot of matters of fact would need to be different for him to be present to water the flowers. On the other hand, the world where the Queen waters the flowers is fairly close: all she has to do is spontaneously decide to water the flowers and it will be done very quickly. Yet even Hitchcock & Knobe would claim that the world we appeal to as part of our counterfactual

reasoning is one where the gardener did not go on holiday and instead did his job by watering the flowers. This world is still the more normal world, and so we appeal to it over the far closer but less normal world where the Queen waters the flowers.

In support of their claim, Hitchcock & Knobe appeal to the psychological literature on counterfactual reasoning. Kahneman & Tversky (1982) and Kahneman & Miller (1986) have both demonstrated that norms affect our counterfactual reasoning in the way described by Hitchcock & Knobe. Kahneman & Miller state that '[e]xceptions tend to evoke contrasting normal alternatives, but not vice versa' (1986, 143). In other words, abnormal events tend to trigger thoughts of more normal counterfactual scenarios, whilst normal events don't trigger thoughts of abnormal scenarios. So, when we reason counterfactually about abnormal events, we mentally undo it by changing it to a more normal alternative. Yet we won't change a normal event to an abnormal event unless prompted to do so. These mental changes to actual events are referred to as downhill, uphill, and horizontal changes. A downhill change is where we replace an actual event with a more normal alternative. An uphill change is where we replace an actual event with a less normal alternative. A horizontal change is where the actual event is replaced with an equally normal alternative. As the names would suggest, a downhill change is cognitively easier than an uphill change.

This is related to Blanchard & Schaffer's account, which I discussed in the previous chapter. As we saw, they argue that we pick out events as causes based on the availability of the alternatives to those events. Hitchcock & Knobe state that '[i]f some abnormal event occurred, the hypothetical situation in which it does not occur, or some more normal event occurs instead, will typically be considered relevant', and 'if some normal event occurs, we

may never get around to considering the counterfactual situation in which some more abnormal alternative occurs instead' (2009, 601). In other words, we pick out abnormal events as causes because we can more easily imagine how they could be otherwise. We don't pick out normal events as causes because there is no relevant alternative to appeal to as part of our counterfactual reasoning. The main difference between this view and the one proposed by Blanchard & Schaffer, is that Hitchcock & Knobe argue that it is part of our competent use of the concept of causation to pick out abnormal events. In other words, according to Hitchcock & Knobe it is false that, for example, the Queen's failure to water the flowers caused them to wilt, and that the administrative assistant's taking a pen caused the problem; according to Blanchard and Schaffer, by contrast, we are induced by the lack of cognitively available alternatives into mistakenly believing that those causal claims are false

The basis for this account of the truth-conditions of causal statements comes from the idea that there are worlds that are more or less normal than the actual world to various degrees. This suggests an ordering of possible worlds based on the degree to which those worlds can be considered normal. At the top of this ordering, we have the worlds that are maximally normal, relative to the actual world. Unlike with similarity, the actual world will not be the most normal relative to itself. Rather, the worlds that are the most normal relative to the actual world will be those where all events follow the normal course and thus there are no events that deviate from that course. Further down that ordering we have the worlds that are less normal because they contain a few minor deviations from the normal course. Then even further down we have the worlds that contain more significant deviations from the normal course. This continues until we reach the bottom of the ordering, where we find all the worlds that are the most abnormal. When making causal judgements, the worlds that are relevant

are those at the top of this normality ordering. Here we see a clear analogy with Lewis's account, where worlds are ordered according to their similarity to the actual world, and those that are the most similar to the actual world are the worlds that are relevant to causal judgements.

This idea of a normality ordering of worlds is further developed by Halpern & Hitchcock (2015). Their account builds on the existing work on causal models done by Pearl (2000) and others. Specifically, they provide an account of extended causal models that include an ordering of worlds based on the degree to which they can be considered normal (2015, 434-435). They then go on to add that we 'can use normality to rank actual causes' and '[d]oing so lets us explain the responses that people make to queries regarding actual causation' (2015, 436). These queries they have in mind here are the judgements that we have about cases like the ones I discussed in Chapter 2. Halpern & Hitchcock state their method for ranking causes in formal terms, but for the sake of brevity I will just use plain English. No doubt some nuance and complexity will be lost in translation, but I believe that I can still capture the spirit of the account.

Halpern & Hitchcock claim that we 'grade candidate causes according to the normality of their best witnesses' (2015, 436). The term 'witness' is being used to refer to those worlds at which neither the candidate cause nor the effect occurs. The 'best witness' is the world that is most normal out of these. This method of grading candidate causes is meant to explain why some events are favoured over others when it comes to assigning causal status. The cases they have in mind are those like the gardener and the Queen, or the pen vignette. In both these cases we are presented with two candidate causes that the effect is counterfactually dependent on.

Yet our intuitions are to select one event or omission (the gardener's not watering the flowers or Professor Smith's taking a pen) over the other (the Queen's not watering the flowers or the assistant's taking a pen). They argue that what we are doing when presented with these cases is 'picking the 'best' cause, where best is judged in terms of normality' (*ibid.*). The 'best' cause, according to Halpern & Hitchcock, is the one with the most normal witness.

To see how this works, we can examine how it applies to one of the puzzle cases. I will go with the pen vignette, as it is one of the cases that Halpern & Hitchcock explicitly discuss (2015, 439-441). The puzzle here is to explain the difference in acceptability between:

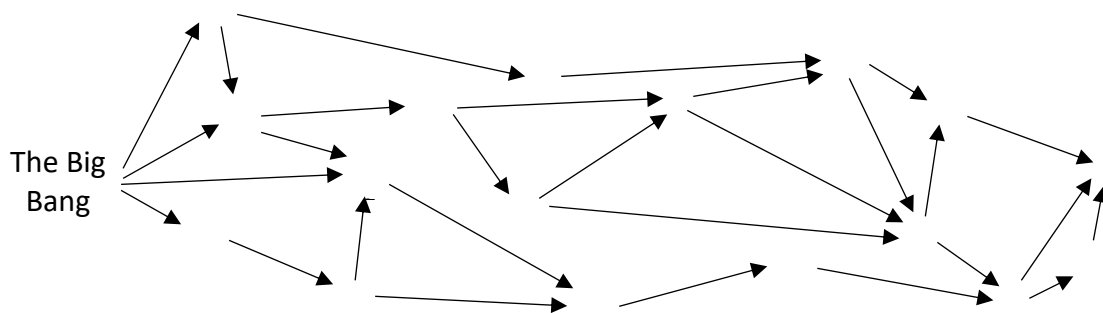
- 1) Professor Smith's taking a pen caused the problem.
- 2) The administrative assistant's taking a pen caused the problem.

(1) is considered acceptable, whilst (2) is unacceptable. Halpern & Hitchcock can explain this difference by claiming that the best witness for (1) is more normal than the best witness for (2) (2015, 441). The best witness for (1) is the world where Professor Smith doesn't take a pen, the assistant does take a pen, and there is no problem (because there is still a pen left for the receptionist). The best witness for (2) is the world where Professor Smith takes a pen, the assistant doesn't take a pen, and there is no problem. The best witness for (1) is more normal than the best witness for (2) because the world does not contain any violations of the norm of only the assistants being allowed to take pens. This is why the subjects are more willing to attribute causation to Professor Smith's action than the assistant's action.

This demonstrates how Hitchcock & Knobe's account can be used to solve the problem cases discussed in Chapter 2. Of course, everything that has been said so far is consistent with the view that our intuitions about those cases are mistaken and that competent users of the

concept of causation do not appeal to a normality ordering of possible worlds when making causal judgements. However, Hitchcock and Knobe go on to claim that following this process *does* represent a competent application of the concept of causation because of the purpose that concept serves. Specifically, its purpose is to help us to identify ‘the most suitable targets for intervention’ (2009, 597), a function that is best served by having our judgements be sensitive to a normality ordering of worlds.

To see why this is the case, we need to consider how this normative concept of causation might generally be used. At the centre of Hitchcock and Knobe’s discussion of this issue is the intuition that for any effect (e) there is a complex web of earlier events that e is counterfactually dependent upon. This web will include events from the distant past, up to and including the Big Bang. These events will also stand in counterfactual dependence relations with each other. If we were to model this causal structure, then we would end up with a more complicated version of the following:



Here the letters represent events, and the arrows represent counterfactual dependence relations.

Now obviously, if we were questioned about what caused *e*, we would not respond by listing all the events featured in this web. Rather, when faced with such a question we tend to respond by picking out a small number of specific events (usually just one) that are relevant. So, when asked 'what caused *e*?' we would provide a straightforward answer like '*c* caused *e*'. This raises the issue of what it is that distinguishes *c* from all the other events that *e* counterfactually depends upon. In other words, what is it about the counterfactual dependence relation that obtains between *c* and *e* that makes us select *c* over all the other events that *e* is counterfactually dependent upon?

As stated earlier, Hitchcock & Knobe's answer is to claim that the purpose of our concept of causation is to identify suitable targets for intervening to achieve a desired outcome (2009, 606-612). If we wanted to prevent or encourage events like *c*, then it would be more useful to know the abnormal events that *c* is dependent upon, as those events can be much more easily manipulated. If our aim is to prevent fires, it is much easier to stop short circuits from occurring than it is to remove all oxygen. Indeed, norms are often in place for a good reason, and so trying to alter those norms could be difficult to implement and potentially harmful. In the pen vignette, it is implied that the faculty are in a much better position to buy their own pens than the administrative assistants. This means that it would be more effective and less objectionable to make the faculty stop taking pens.

From this it should be clear how Hitchcock & Knobe's account follows the Competence Strategy. They identify the 'legitimate purpose' (2009, 612) of causation as being to identify targets for intervention, and then note that this is best achieved by picking out abnormal events as causes. This means that when we do the same in response to the puzzle cases in

Chapter 2, it is an instance of the competent application of the concept of causation in a manner that serves its legitimate purpose.

One problem that might be raised at this point is that the account falls victim to the same problem I raised against Blanchard & Schaffer's account in the previous chapter. The problem lies in explaining why Knobe & Fraser's subjects found (2) unacceptable, even when they were asked a direct and closed question about whether or not the assistant caused the problem. This proved challenging for Blanchard & Schaffer because they had no way to explain why the subjects don't select the assistant's not taking a pen as the relevant alternative to consider in their counterfactual reasoning. In their current forms, both Hitchcock & Knobe and Halpern & Hitchcock's versions of the account fall prey to this objection as well. To go with Halpern & Hitchcock's formulation, their claim is that the subjects made their judgement by comparing the normality of the best witnesses for the candidate causes. Yet this explanation only works if we assume that the subjects are answering an open question (who caused the problem?) that prompts them to weigh up each candidate cause against the other. The explanation seems to falter when trying to explain why the subjects disagreed that the assistant's taking a pen caused the problem, since there was no need for them to compare different witnesses.

There are two ways for Halpern & Hitchcock to respond here. One would be to say that the subjects always considered all the candidate causes, even when they were not specifically asked to do so. We might think that this is implausible in cases where the causal structure is very complex and so there are many candidate causes, since it would mean that every judgement involves comparing the relative normalities of multiple different witnesses to answer even the most basic of causal questions.

Hence, a second option might be preferable. This way of responding would be to argue that the subjects aren't comparing the best witnesses of the candidate causes to each other, but rather to the actual world. In other words, if the best witness of a candidate cause is more normal than the actual world, then it is considered an actual cause. On the other hand, if the best witness of a candidate cause is not more normal than the actual world, then it is judged to not be an actual cause. This explains why the subjects disagreed with (2), because the best witness is a world that contains the same number of norm violations as the actual world, and so the two worlds are equally as normal. Of the two options available here, this is the one that I believe a proponent of this account should choose.

This gives us the first of the two normative accounts that I will be examining in this chapter. A promising start, but as I will explain in the final section, not a perfect one. In the next section, I will look at Menzies' account of causes as the things that make the difference between the normal course of events and the actual course of events.

2. Normal Courses and Difference-Making

The guiding idea behind Menzies' account of causation can be found in Hart & Honoré's *Causation in the Law* (1985). Their view is given in the following passage:

Human action in the simple cases, where we produce some desired effect by the manipulation of an object in our environment, is an interference in the natural course of events which *makes a difference* in the way these develop. ... 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 commonsense concept of cause [W]hat is identified as the cause of some occurrence ... is a difference to the

normal course which accounts for the difference in the outcome. (Hart and Honoré 1985, 29)

The key points Menzies takes from this are: that there is a normal course of events that will occur naturally in the absence of any external interference; that human actions and other events can interfere with the course of events such that it no longer follows the normal course; and that the event that makes the difference between events following and not following the normal course is a cause of whatever abnormal event follows. Put simply, a cause is something that makes a difference between events following the normal course – the course that would happen without intervention – and events not following the normal course. This idea is at the core of Menzies' account.

Menzies then goes on to further elucidate this guiding idea from Hart & Honoré. He understands the normal course of events as the default state of the causal system under consideration (2009, 359). This is where all the events that are contained within the system are default/normal events. To illustrate what this means we can use the gardener and the Queen case. There the default state of the system is where the gardener waters the plants, the Queen doesn't water the plants, and the plants don't wilt. This is the default state of the system because everything is the way that it is supposed to be: the gardener is fulfilling his obligation to water the plants, and so isn't violating any moral or social norms; the Queen is doing her queenly things and isn't doing any work that might be seen as beneath her; and the plants are alive and flourishing. This means that the normal course of events is one where the plants are watered by the gardener and only the gardener, which results in them not wilting.

Now that we have a clear understanding of a normal course of events/default state of a system, we can examine how this course/system is interfered with. It is here that Menzies

appeals to the notion of an intervention. This use of the term 'intervention' refers to 'a source of causal influence that is exogenous to the causal set-up being considered' (Menzies 2011, 199). In other words, an intervention is a causal factor that comes from outside the system under consideration. It should be noted that Menzies' use of the term differs from the more precise, technical definition offered by Woodward (2004). Rather, it serves a similar purpose to Lewis's use of the notion of a miracle, in that it serves to rule out backtracking counterfactuals. It does this by allowing us to imagine a counterfactual scenario in a deterministic world, without requiring that the facts about the past be changed to make the scenario possible.

With this notion of an intervention, Menzies then introduces a distinction between default and deviant counterfactuals (2011, 199). The former are counterfactuals where the 'closest antecedent-worlds are ones in which the antecedent is realized in the normal course of events in the absence of any intervention' (*ibid.*); while the latter are counterfactuals where the 'closest antecedent-worlds are ones in which the antecedent is realized by an exogenous intervention' (*ibid.*). In other words, the counterfactual 'if a had occurred then c would have occurred' is a default counterfactual if the closest world at which a is true is one where a is part of the default state of the system under consideration. The same counterfactual is deviant if the closest world at which a is true is one where a comes about through an intervention. This means that typically the antecedents of default counterfactuals will be normal, and the antecedents of deviant counterfactuals will be abnormal.

As with both Lewis and Hitchcock & Knobe, Menzies understands the truth-conditions of counterfactuals in terms of the closeness of possible worlds. Menzies offers the following characterisation for the worlds that are relevant to evaluating counterfactuals:

Consider a counterfactual 'If P were the case, Q would be the case' where P is about states of affairs at a time t. Consider all those worlds w such that:

- a) P is true at w;
- b) w is exactly like the actual world at all times before a transition period beginning shortly before t;
- c) w conforms to the norms of the actual world at all times after t; and
- d) during the transition period w departs from the norms of the actual world no more than it must to permit P to hold.

The counterfactual is true if and only if Q holds in every such world w. (Menzies 2011, 199)

For the sake of simplicity, I will be assuming that there is only a single relevant antecedent world.

We are now in a position to examine Menzies' account of causation. He offers the following truth-conditions for causal statements:

A state of affairs c causes a wholly distinct state of affairs e if and only if (i) if c were to obtain, e would obtain; and (ii) if c were not to obtain, e would not obtain, where (i) is a deviant counterfactual and (ii) is a default counterfactual. (Menzies 2011, 200)

Put simply, c must make the difference for e's occurrence, and c must also be an abnormal event.

As with Hitchcock & Knobe's account, we can illustrate this by applying it to the pen vignette. Recall that the account must show that (1) is true and (2) is not true. With statement (1), there are two counterfactuals that we need to consider:

- i') If Professor Smith had taken a pen, there would have been a problem.
- ii') If Professor Smith hadn't taken a pen, there wouldn't have been a problem.

For (1) to be true, it must be the case that both (i') and (ii') are true (on Menzies' account of the truth-conditions of counterfactuals), and it must also be the case that (i') is a deviant counterfactual and (ii') is a default counterfactual.

Let's start by considering (i'). The relevant world for the truth-value of this counterfactual is one where: (a) Professor Smith took a pen, (b) everything is exactly like the actual world until a transition period just before Professor Smith took a pen, (c) there are no further norm violations after Professor Smith took a pen, and (d) any norm violations in the transition period are kept to a minimum. Under these conditions, the relevant world would be one that is just like the actual world up to the transition period. During that period is when the norm-violating act of Professor Smith taking the pen takes place (so far, we still haven't deviated from the actual course of events). Since we need to keep norm violations to a minimum, the rest of the transition period will play out as normal and conform to the actual course of events (any deviation from the actual course will be a lawbreaking, and thus norm-violating event). This means that the assistant will still take a pen. The result of this is that the receptionist will have a problem because there are no pens left. This means that the relevant world is just the actual world. As a result, the counterfactual (i') comes out as true, since Professor Smith did take a pen and there was a problem. Additionally, Professor Smith's taking a pen is not a default/normal event, and so (i') must be a deviant counterfactual.

With (ii'), the relevant world is one where: (a) Professor Smith didn't take a pen, (b) same as (i'), (c) there are no further norm violations after Professor Smith failed to take a pen, and (d) any norm violations in the transition period are kept to a minimum. So, the relevant world is one where Professor Smith doesn't take a pen, the assistant does take a pen (since that would ensure minimal deviation from the norms of the actual world), and there is no problem. This means that (ii') is true. Additionally, since the Professor Smith's not taking a pen is part of the default state of the system, (ii') is a default counterfactual. This means that both (i') and (ii') are true, and the former is a deviant counterfactual, while the latter is a default counterfactual. Therefore, (1) comes out true on Menzies' account.

Now we can see how the account handles (2). Here the counterfactuals we are interested in are:

- i*) If the administrative assistant had taken a pen, then there would have been a problem.
- ii*) If the administrative assistant hadn't taken a pen, then there wouldn't have been a problem.

To deliver the correct result (on the Competence Strategy) of (2) being false, it must be the case that at least one of these counterfactuals are false. Menzies states that the relevant world for (i*) will be one where Professor Smith doesn't take a pen, and so the counterfactual is false (Menzies 2011, 201). The falsity of (i*) entails the falsity of (2), thereby explaining our intuitions about the case.

However, it is worth noting that there are some unclarities here. With both (i*) and (ii*) there is a problem with determining exactly which world is relevant for determining their truth-value. The wording of the example suggests that Professor Smith's taking a pen occurs before

the assistant's taking a pen. If this is the case, then the former ought to be held fixed in order to maximise the similarity to the actual world prior to the antecedent. However, if Professor Smith's taking the pen occurred at the same time or after the assistant's taking a pen, then the need to minimise norm violations kicks in and we have to change it to Professor Smith not taking a pen. If the relevant world is one where Professor Smith takes a pen, then both (i*) and (ii*) will be true. On the other hand, if the relevant worlds are ones where Professor Smith doesn't take a pen, then (i*) is false and (ii*) is true. Menzies doesn't acknowledge this lack of clarity when he applies his account to the case. He does not even consider what the relevant world would be for (ii*), as the falsity of (i*) is enough to show the falsity of (2). This means that his account gets the right result, but also shows a real lack of clarity in how that result was reached.

A further source of unclarity for Menzies' account can be seen when we consider whether (i*) and (ii*) are default or deviant counterfactuals. The relevant world for (i*) will be one where the assistant takes a pen, which is part of the normal course. As a result, (i*) is a default counterfactual. So far so good, but the problem appears when we turn our attention to (ii*). The relevant world will be one where the assistant didn't take a pen, and it isn't clear whether this can be considered part of the default state of the system, as no norms are being violated. This is a point that I will return to in the final section. For now, it is enough to note that Menzies' account gets the correct result, but also suffers from some serious unclarity.

This completes my explanation of the second account that follows the Competence Strategy. Overall, Menzies' account seems to deliver the correct result for the pen vignette, but this relies on an assumption which he may not be entitled to. Additionally, there is some unclarity

regarding whether (i*) and (ii*) are default or deviant counterfactuals. In the next section, I will examine a final account from Icard et al.

3. Causal Strength

The final account that I wish to examine is presented in Icard, Kominsky, and Knobe's 'Normality and Actual Causal Strength' (2017). The central claim of the account is that normality affects our judgements of causal strength, with events that are abnormal having more causal strength attributed to them. This explains our intuitions about the puzzle cases from Chapter 2.

Before I begin explaining how this account is supposed to work, it should be noted that this account differs from the accounts I discussed earlier because it doesn't involve an explicit commitment to the Competence Strategy. In fact, what follows is consistent with the claim that the influence of norms on our causal judgements are the result of a bias and not a competent use of the concept of causation. The account is therefore neutral between the Competence Strategy and the Pragmatics-and-Biases Strategy. However, it could potentially be employed as a means of delivering on the Competence Strategy if we apply what is said here about causal judgements to the truth-conditions of causal statements. As such, it is worth considering whether or not such an approach would be successful.

In trying to understand Icard et al.'s account, the best place to start is to examine what they mean by the term 'causal strength'. The way that they introduce this notion is to go through some of the existing measures of causal strength that have been discussed in the literature (Icard, Kominsky and Knobe 2017, 83). The accounts they consider have no bearing on what we are discussing in this chapter, but the first account they consider will be useful in

illustrating what is meant by the notion of 'causal strength'. They state '[o]ne prominent proposal is aimed at capturing the intuition that a cause C should raise the probability of the effect E above its unconditional value' (*ibid.*). This view is then attributed to Reichenbach (1956), Spellman (1997), and Suppes (1970). Put simply, the occurrence of c should make the occurrence of e more likely, and the greater the effect that c has on the probability of e, the greater its causal strength.

Now that we understand what is meant by the term 'causal strength', we can examine how Icard et al. use it to solve the problem raised by the cases given in Chapter 2. One important thing to note is that, of the accounts discussed so far, Icard et al. are the only ones to address the phenomenon of causal superseding. Recall from Chapter 2 that causal superseding occurs in cases where some effect is counterfactually dependent upon the conjunction of two or more earlier events. In such cases, if one of those earlier events is abnormal, then we are typically less willing to attribute causal status to any normal events that the effect is counterfactually dependent upon. As Icard et al. put it, 'suppose an outcome depends on a causal factor C as well as an alternative causal factor A, such that the outcome will only occur if both C and A occur. Then people will be less inclined to say that C caused the outcome if A is abnormal than if A is normal' (2017, 81-82). Of course, it would be slightly unfair to criticise the earlier accounts for not directly addressing the phenomenon of causal superseding, as it was only identified by Kominsky et al. in 2015. However, it is certainly an advantage of Icard et al.'s account that it can explain a broader range of phenomena than the earlier accounts.

Another important point that needs to be considered before discussing the account is that the way Icard et al. present their account involves a high degree of formalisation (even more

than the accounts discussed earlier in this chapter). Since including all of this would massively increase the complexity and length of the following exposition, I will again stick to plain English. Hopefully, this will not undermine the force of my criticisms in the next section.

Icard et al. begin the explanation of their account by offering their own measure of causal strength. They are quick to point out that their measure differs from the earlier measures (such as the probabilistic account I mentioned at the beginning of this section) because it is 'not motivated by an intuition about what causal strength really is' (Icard, Kominsky and Knobe 2017, 84). Rather, 'it is motivated by an intuition about the psychological processes people go through when they are making causal judgments' (*ibid.*). They then go on to add 'we will be drawing on the idea that people make causal judgments through a process of probabilistic sampling' (*ibid.*). Specifically, this involves the sampling of counterfactuals through a process that converges on a measure of causal strength. Icard et al. then state that 'this measure accurately captures the patterns in the impact of normality on people's judgments of actual causation' (*ibid.*).

It is at this point that Icard et al. introduce two counterfactual notions that are relevant to how we make causal judgements. The first is the kind of straightforward negative counterfactual that can be found in all the accounts I have discussed thus far (including Lewis's). The kind of counterfactual we are concerned with here is as follows:

If C had not occurred, E also would not have occurred [sic]. (Icard, Kominsky and Knobe 2017, 84)

Icard et al. refer to this notion as 'actual necessity'. This is how counterfactual dependence is traditionally understood. They then distinguish this from another notion that several

researchers³¹ have claimed is involved in our process of making causal judgements. They refer to this notion as 'robust sufficiency', and it concerns the following counterfactual:

Given that C occurred, E would have occurred even if background conditions had been slightly different. (*ibid.*)

This notion is just like the first condition that Menzies proposed in the previous section, since for Menzies (unlike Lewis), the truth of 'had c occurred, e would have occurred' is not guaranteed by the fact that both c and e do, in fact, occur. Icard et al. suggest that since both notions are involved in our causal judgements, a measure of causal strength will in some way combine the two. This then raises the question of 'how people assess each of these factors and how they combine the two' (*ibid.*).

Icard et al.'s answer to this question is to appeal to the process of probabilistic sampling. The thought is that, when making causal judgements, people consider a wide variety of counterfactual scenarios that are then evaluated probabilistically rather than deterministically (2017, 84). Higher probabilities will be assigned to counterfactual scenarios that are normal, than to those that are abnormal (2017, 85). From this process of probabilistic sampling, we are then able to derive a measure of causal strength.

This measure of causal strength comes from the how we measure actual necessity and robust sufficiency. When measuring the former, we imagine a scenario where, before some transition period, c doesn't occur but everything else is held fixed. If it follows from that that

³¹ Here Icard et al. (2017, 84) cite Hitchcock (2012), Lombrozo (2010), and Woodward (2006).

e also doesn't occur, then c has a necessity strength of 1. If *e* does occur at such a world, then c has a necessity strength of 0 (*ibid.*).

To illustrate this, it might be helpful to consider one of the cases that Icard et al. provide (2017, 87). In the case, Billy and Suzy are colleagues who are working on a project together. Both arrive in their office building at 9am and set off a motion detector. Four different versions of the case are presented that vary whether or not Billy was supposed to be there (and so it varies whether or not his action was normal), and whether the case has a conjunctive or disjunctive structure. For now, let's consider the two versions where Billy was not supposed to be there and so his arrival was abnormal. In the conjunctive version, the motion detector is set to go off when it detects *more than* one person arriving at the same time. In the disjunctive version the motion detector is set to go off when it detects *at least* one person arriving. So, in the former, each of Billy's and Suzy's arrival at nine is necessary for the motion detector to go off; whereas, in the latter, the motion detector going off is causally overdetermined.

In the conjunctive case, Billy's arriving at nine has a necessity strength of 1 because if he had not arrived at that time, the motion detector would not have gone off. The same is true of Suzy's action, so it also has a necessity strength of 1. However, this differs in the disjunctive case. If Billy hadn't turned up, Suzy would have still arrived and set off the motion detector. Therefore, the necessity strength of Billy's action in this case is 0. Likewise, if Suzy hadn't turned up, the motion detector would have still been set off by Billy, and so her action also has a necessity strength of 0.

That explains how we measure actual necessity; now it is time to examine how we measure robust sufficiency. Here we are attempting to test whether *e* will still occur as a result of *c* when we vary the other conditions. This means that the counterfactual scenarios we are interested in are those where *c* is held fixed and the background conditions are varied. If *e* still occurs because of *c* in all such minor variations of the background conditions, then *c* has a sufficiency strength of 1. If not, then the sufficiency strength of *c* will be the probability that we assign to the occurrence of the actual background conditions (Icard, Kominsky and Knobe 2017, 85). The reason for this is that in such a case, the background conditions make a difference between the occurrence and non-occurrence of *e*, and so how likely we judge them to be will be relevant in assessing the causal strength of *c*.

To illustrate this, let's return to the motion detector case. In the disjunctive version of the case, Billy's action would have a sufficiency score of 1. This is because, the scenario where he still arrived, and Suzy didn't, would still result in the motion detector going off. Ditto for Suzy's action. However, the conjunctive case is more complicated. There the scenario where Billy arrived, and Suzy didn't, would not result in the motion detector going off. Therefore, the sufficiency strength of Billy's action will be the probability that we assign to Suzy arriving at nine. Since this is what would normally happen, the probability will be quite high. This means that the sufficiency strength of Billy's action would also be quite high. On the other hand, the sufficiency strength of Suzy's action will be the probability that we assign to Billy's arriving at nine. Since this is abnormal, the probability will be low, which means that the sufficiency strength of Suzy's action will also be low.

At this point we now understand how actual necessity and robust sufficiency are measured. This means that we now have all the pieces we need to understand Icard et al.'s measure of causal strength. Put simply, the causal strength of *c* is the weighted sum of 'its necessity strength and sufficiency strength' (Icard, Kominsky and Knobe 2017, 86). This sum is weighted by the probabilities assigned to both the non-occurrence and occurrence of *c* respectively. So, in simple terms, the following equation can be used to calculate the causal strength of *a*:

$$Ca = ((Na \times Prob: \sim a) + (Sa \times Prob: a))$$

With the key:

Ca = The causal strength of *a*

Na = The necessity strength of *a*

Sa = The sufficiency strength of *a*

$Prob:a/\sim a$ = The probability assigned to *a*/ $\sim a$

Recall from earlier in this section, that the probabilities assigned to the occurrence/non-occurrence of *c* are determined by how normal *c* is judged to be. If *c* is an abnormal event, there will be a low degree of probability that it will occur, and a high degree of probability that it will not. Likewise, if *c* is normal, then its occurrence will have a high degree of probability, and its non-occurrence will have a low degree of probability. It is in this way that normality influences our judgements of causal strength.

To show how this all works, we can now apply this to the motion detector case. Of the four versions of the case that Icard et.al. present, I will only talk about the two that feature a norm-

violating agent within a conjunctive and disjunctive causal structure. As a reminder, in both these cases, Billy is not supposed to be there, whilst Suzy *is* supposed to be there.

I will begin with the conjunctive version of the case. Recall that in this version, the motion detector will only be set off if both Billy *and* Suzy arrive at the same time. We have already established that the necessity strength for each of them arriving at nine is 1. Additionally, we also know that the sufficiency strength for each of them arriving at nine will be the probability that the other will arrive at nine. It is normal for Suzy to arrive at nine because it is where she is supposed to be. Consequently, the probability of her arriving at nine will be high (0.8), whilst the probability of her not arriving at nine will be low (0.2). On the other hand, Billy has been told not to come in at nine, and so it would not be normal for him to do so. So, the probability of him arriving at nine will be low (0.1), whilst the probability of him not arriving will be high (0.9). This tells us that the sufficiency strength of Billy's arriving at nine will be 0.8 (the probability of Suzy arriving), while the sufficiency strength of Suzy's arriving then will be 0.1 (the probability of Billy arriving).

We now have all the numbers we need to fill in the equation and calculate the causal strength for both Billy and Suzy's arriving at nine. For the former, we would get the following sum:

$$(1 \times 0.9) + (0.8 \times 0.1) = 0.98$$

This means that Billy's action has a causal strength of 0.98, which is very high. Now let's compare it to Suzy's action:

$$(1 \times 0.2) + (0.1 \times 0.8) = 0.28$$

So, the causal strength of Suzy's action is 0.28, which is much lower than Billy's. Note that the factor that makes all the difference here is the probability that we assign to the event's non-occurrence (0.9 for Billy and 0.2 for Suzy). Abnormal events will have a higher probability of non-occurrence, so, in conjunctive cases, they will have a much higher degree of causal strength.

Now let's look at the disjunctive version of the case. Recall that in this version, the motion detector will go off if either Billy *or* Suzy arrives. Earlier we established that, for each action, the necessity strength was 0 and the sufficiency strength was 1. The probabilities for the occurrence/non-occurrence of Billy and Suzy's actions can remain the same as in the previous version. So, the causal strength of Billy's action can be worked out as follows:

$$(0 \times 0.9) + (1 \times 0.1) = 0.1$$

This means that the causal strength of Billy's arriving at nine in the disjunctive case is 0.1, a lot lower than in the conjunctive case. In doing the same for Suzy we get the following:

$$(0 \times 0.2) + (1 \times 0.8) = 0.8$$

Here, Suzy's causal strength is 0.8, which is much higher than Billy's 0.1. Notice that with this causal structure, the thing that makes the difference is the probability that the event occurs (0.1 for Billy, 0.8 for Suzy). Since normal events will have a higher probability of occurrence, they will also have a greater degree of causal strength in disjunctive cases. This seems to run contrary to what is suggested by the earlier accounts and it is a point that I will return to shortly.

This completes my exposition of Icard et al.'s account of causal strength. We have now seen how it is that causal strength is derived from actual necessity, robust sufficiency, and the probabilities assigned to the occurrence/non-occurrence of events. The final point to consider is why this notion affects our causal judgements in the way that Icard et al. say that it does. As with Hitchcock & Knobe's account, the thought here is that our judgements of actual causation are a way to identify suitable targets for intervention. To that end, it makes sense that we would pick out the events that exhibit the greatest degree of causal strength, because intervening on those events would likely make the greatest difference in whether or not the effect occurs. This means that, when we are picking out an event as the cause of some event, we tend to go for the event with the highest degree of causal strength.

We are now able to examine how this account solves the problem presented by the puzzle cases in Chapter 2. With the pen vignette, the puzzle was to explain why the subjects agreed with (1) but disagreed with (2). This can be explained by first recognising that the case is isomorphic with the conjunctive version of the motion detector case. This means that the abnormal event will have a greater degree of causal strength than the normal event, because it is more probable that the event will not occur. This means that Professor Smith's taking a pen will have a greater degree of causal strength than the assistant's taking a pen. So, when the subjects are asked to make a judgement about the case, they consider the causal strength of all the candidate causes, and then pick out the strongest one to be the actual cause. In this case, the strongest event will be Professor Smith's taking the pen, so that is the event to which we attribute causal status. This explains the difference in acceptability between (1) and (2), and thus solves the problem.

However, just like Halpern & Hitchcock's account, there is also the potential worry that this account falls victim to the objection that I raised against Blanchard & Schaffer's account in the previous chapter. The account performs fine when trying to explain the answers that people give to open questions: we calculate the causal strength of all the candidate causes and then select the strongest. The problem arises when we are asked to explain what happens when we are given a closed question about whether we agree or disagree with a particular causal statement. When we are weighing up if we agree with (2) it certainly seems like we are just thinking about the assistant's taking a pen, rather than comparing its causal strength to other relevant events like Professor Smith's taking a pen.

As with Halpern & Hitchcock, Icard et al. have two ways to respond. One is to hold that we always consider every candidate cause, even when not prompted to do so. Again though, this might seem implausible with complex causal structures. The second option is to claim that there is some threshold on causal strength, where anything higher is considered to be a cause. So, we could say that any event with a causal strength above 0.5 counts as a cause. Again, this second option seems more promising.

Overall, I find Icard et al.'s view the most plausible account of the three discussed in this chapter. The reason for this is that it manages to explain our judgements in the broadest range of cases, including the recently discovered effect of causal superseding. Additionally, the account also makes a novel prediction. Earlier I noted that, in cases with a disjunctive causal structure – that is, cases of overdetermination – normal events will have a greater degree of causal strength than abnormal events. This leads Icard et al. to predict that the subjects who are presented with this case will attribute causal status to the more normal

event (2017, 86). This is not something that the accounts I have discussed earlier (including those that follow the Pragmatics-and-Biases Strategy) would predict. Yet, this prediction is then tested and confirmed (2017, 86-87). This is another advantage that Icard et al.'s account has over the alternatives and shows why I consider it to be the best of the accounts discussed in this chapter.

To turn this into a view that belongs within the Competence Strategy is a fairly simple matter, as all we need to do is stipulate that the measures of causal strength offered here are part of what it means to say that 'c caused e'. That is, a condition for the truth of a given causal statement is that the strength of the cause that the statement refers to must be greater than all other relevant events, or greater than some threshold. Such an account would then provide a plausible Competence Strategy view, since our intuitions about the puzzle cases are the result of us correctly identifying the relative causal strengths of the events referred to by the causal statements.

That completes my summary of the three accounts that follow the Competence Strategy. In the next and final section, I will discuss the areas in which these accounts can be developed, specifically when it comes to the notion of normality.

4. The Unclarities Objection

In the previous chapter I mentioned that Blanchard & Schaffer offer three arguments for their conclusion that causation is not normative: the first is that normality is an unclear and heterogenous concept; the second is that many of the puzzle cases used to support normative accounts are based on non-apt causal models; and the third is that the effects that normality has on our causal judgements is best explained by appealing to heuristics and biases. I

responded to the third argument by demonstrating how it fails to properly motivate its conclusion. However, I believe the first argument is far more successful, and it raises an issue that the accounts I have been considering in this chapter fail to address.

Blanchard & Schaffer summarise their argument as follows:

What concerns us most about default-relativity is the unclarity it generates... Normality judgements seem to draw on diverse and typically competing factors, in a highly context sensitive way. As such default-relativity often seems to us to come close to a free parameter in an otherwise so precise and objectively constrained formalism, which basically gives the theorist leeway to hand-write the result she wants. So we think that default-relativity generates complicating and under-constrained unclarity. (Blanchard and Schaffer 2017, 192)

In other words, normality is a broad concept that seems to be shaped by many different sorts of considerations. As a result, a lot of theorists have accounts that are practically unfalsifiable, since what can count as normal or abnormal can be whatever is needed for the theory to come out as true.

Now while I don't believe that this is true, I do believe that the followers of the Competence Strategy have done very little to persuade people that it's not. When normality is spoken of in the literature, it is done in a way that is purposefully meant to include a broad range of cases. The reason for this is that various different kinds of 'norm' have been shown to have an effect on our causal judgements, from statistical norms (Hilton and Slugoski 1986) (Cheng and Novick 1991), to moral/social norms (Knobe and Fraser, Causal Judgment and Moral Judgment: Two Experiments 2008) (Samland and Waldmann 2016), to norms of proper functioning (Hitchcock and Knobe 2009) (Clarke, et al. 2015). This means that theorists have

had to talk about a general and heterogenous conception of normality so that their account can explain all these instances of norms affecting our causal judgements. The problem is that this raises some important questions that have yet to be answered.

Firstly, what reason do we have to believe that there is this unified notion of normality in the first place? After all, when we say, 'it is not normal for it to rain in the Atacama Desert', we seem to mean something entirely different to 'Donald Trump is not a normal president'. The former is entirely descriptive and is based on statistical facts about the prevalence of rainfall in a particular part of the world. The latter, by contrast, is based on moral or social norms about how a president ought to behave and suggests that Trump is doing something wrong with him when he acts contrary to those norms. I will discuss this issue further in the next chapter, but for now it's enough to note that we appear to have two very different senses to the term 'normal'. Yet we are expected to believe that they have the same effects on our causal judgements.

Of course, many theorists recognise that there are different kinds of 'norm'. For instance, Halpern & Hitchcock provide the following list:

- Statistical norms concerning what happens most frequently
- Moral norms concerning what is permissible
- Social norms concerning what policies are in force
- Functional norms concerning how systems are supposed to operate (2010, 402-403)

Yet, as Blanchard & Schaffer point out (2017, 193), they also claim that this list is not exhaustive, which means they haven't provided a complete account for what gets to count as normal.

This then leads onto the next question, which is also raised by Blanchard & Schaffer (2017, 193), concerning how these norms interact, especially in cases where they come into conflict. Does one kind of norm trump another, or do they interact in some other way? Sytsma et al. (2012) tested this by adding in information about statistical norms to the pen vignette and presenting it to a new group of subjects. The kind of statistical norms they added were both population-level (e.g. the faculty always takes pens) and agent-level (e.g. Professor Smith always takes a pen). They found that population-level norms had no effect in this case, but that agent-level norms did. However, contrary to what we might expect, the subjects were more likely to consider an action a cause when it was agent-level normal, rather than agent-level abnormal. None of the accounts I have discussed in this chapter have the resources to explain these results because they don't offer an account of normality.

Additionally, Blanchard & Schaffer identify another source of unclarity (2017, 193-195). They argue that 'default status seems underdetermined in many cases'. In other words, it is not always clear whether something is a default or deviant event. For instance, would a coin landing on heads be a default or deviant event? It certainly isn't abnormal for a coin to land on heads, so we might be tempted to say that it's a default event. But this then raises the question of what we can say about cases where normal events can plausibly be said to cause other normal events. Icard et al.'s account fares better than the others in this regard because a normal event can still be said to have some causal strength. However, it's not clear what we should say in cases where we have multiple normal events with equal causal strength.

One final question concerns Icard et al.'s account specifically. The question returns to a point made in the last section about how we can assign numerical values to events based on their

normality. This seems plausible when we are considering statistical norms, but not when we are thinking about moral or social norms. What probability should we assign to an act that violates a moral norm? Should that probability be sensitive to the severity of the norm violation? If so, by how much? Also, the question of how to weight conflicting norms becomes even more pressing here. Should we ascribe a higher probability to an act that violates a moral norm, when we know that the person doing the act frequently violates moral norms? If so, to what degree? Icard et al. don't seem to have any answers to these questions.

All of this demonstrates that the accounts discussed in this chapter all fail in one key respect: they don't provide any analysis of normality. This then leads to questions that the accounts are ill-equipped to answer, and cases where it is unclear what our causal judgements ought to be. Of course, this isn't a fatal objection to these accounts. They have just failed to provide enough information to adequately solve the problem raised in Chapter 2. Icard et al. even acknowledge that there is more work to do by stating, '[d]ifficult questions arise when we inquire into the exact nature of this amalgamation of statistical and prescriptive normality' (2017, 85). Yet they also fail to address these questions themselves³².

These questions need to be answered if we hope to provide a complete answer to the problem posed in Chapter 2. Therefore, in the final chapters of this thesis I will attempt to answer them by providing a new conceptual analysis of normality, which I will then apply to causation in order to solve the problem. To start with, I will begin my analysis of normality by arguing against the orthodox view that 'normal' is an ambiguous term. This will help to

³² Instead, they cite another paper by Bear & Knobe (2017) that I will discuss in the next chapter.

address the first question by motivating a univocal (and therefore unified) account of normality. In Chapter 6 I will then go some way to addressing the second question by showing how these prescriptive and descriptive considerations interact to shape what is normal. I will not tackle the third question directly, since I am not interested in providing a probabilistic analysis of causation, but hopefully the remarks I offer will be helpful to others in this endeavour.

Chapter 5 – Arguing for a Univocal Account of Normality

In the previous chapter, we saw how many of the existing attempts to provide a normative account of causation are incomplete due to the fact that they do not provide an analysis of the normative concepts they employ. The accounts all discuss things like ‘normal states of causal systems’ or ‘default and deviant events’, but these notions remain frustratingly undefined. It seems to be widely assumed that we always have an immediate, intuitive grasp of what the normal state of a system will be, which we then use when reasoning to our causal judgements. If we didn’t have such an intuitive understanding, the normative account would be unable to explain how we can so easily form clear and definitive judgements about the causal status of so many events. The reason that the unclarities objection is so forceful is that it directly challenges the assumption that we *have* such an intuitive understanding. It draws attention to the fact that the class of factors that determine whether or not something is normal seems to be very heterogenous: encompassing statistical facts, moral judgements, social rules, and claims about the thing’s proper form or function. It is thus very difficult to actually make progress in developing a clear and precise normative account of causation without first specifying what normality actually is.

Of course, the heterogeneity of normality is likely the reason that the existing normative accounts of causation have refrained from offering such an analysis. What’s normal seems to be subject to so many different competing considerations that it would be very difficult to provide an analysis of such a concept. Discussion of the concept outside of causation has also

been fairly limited, with the only serious attempts at analysis coming from the philosophy of biology.

In the absence of much philosophical scrutiny, a particular orthodoxy has formed around how we are to understand the term 'normal'. The heterogeneity of normality has led to the widespread assumption that 'normal' is an ambiguous term that carries at least two different senses. It is generally agreed that one of those senses is one where the use of 'normal' is taken to express something wholly descriptive and reducible to statistical facts. Other senses tend to incorporate some of the other factors that seem to shape what is normal: paradigmatic forms, proper functions, moral norms, etc. Authors who adhere to this orthodox view typically begin by simply stating that 'normal' is ambiguous and will then go on to list the different senses that they take the term to have. This is typically understood as something that is so obviously true that it does not need to be established through philosophical argument.

Challenging this orthodox view and providing motivation for adopting a univocal account of normality will be the focus of this chapter. My discussion of this topic will be divided into five sections. Section 1 will be a brief overview of the orthodox view and some of the different senses of 'normal' that have previously been identified. This will largely be based on the existing work on normality that can be found in the philosophy of biology. In section 2 I will examine the existing motivations for adopting the orthodox view, before arguing that they do not provide sufficient grounds for preferring the orthodox view to a univocal account. Then in section 3 I will consider whether the proponent of the orthodox view can use ambiguity tests to provide support for their position. I will then discuss two additional reasons for avoiding the orthodox view and instead preferring a univocal account. The first reason will be

discussed in section 4, which is that there are moral reasons for such a preference. In section 5 I will provide the second reason, which is that the results of a recent work of experimental philosophy (Bear and Knobe 2017) are more easily explained by a univocal account than by the orthodox view. I will then conclude by arguing that we ought to provide a univocal analysis of normality, which will be what I go on to do in Chapter 6.

1. The Orthodox View

The orthodox view of normality is that the term ‘normal’ is ambiguous between at least two different senses. One of those senses is always taken to be entirely descriptive, with what is normal usually being reducible to statistical facts. What the other senses are taken to be tends to vary depending on the author. Generally, there is an evaluative sense, where what is normal can be reduced to certain prescriptive norms (e.g. moral, social, and legal norms). This sense of ‘normal’ is distinct from the statistical sense, as there is an implied value judgement that comes with the use of this sense: saying that something is ‘not normal’ on this sense entails that it is bad or wrong in some way. The statistical sense is, by contrast, taken to be completely value-neutral.

Further senses of ‘normal’ that have also been suggested – either instead of or in addition to the evaluative sense – are prototypical and teleological senses. In the former sense, something is normal if it closely resembles the typical form taken by other things that of the same type. In the latter sense, something is normal if it is able to perform the function that that type of object is supposed to perform. It is clear that these senses differ from the statistical sense because they in no way supervene on statistical facts, instead making claims about the forms that things ought to take or the functions that they ought to be performing.

It is perhaps less clear that they are distinct from the evaluative sense, which also makes claims about how the world ought to be. Those who postulate a prototypical or teleological sense either do not identify a distinct evaluative sense at all, or they argue that their prototypical/teleological sense differs from the evaluative sense by being value neutral. The point here is that the orthodox view is to understand 'normal' as ambiguous between two or more senses, where one of those senses is entirely descriptive and statistical, while the remaining senses are either evaluative, prototypical, teleological, or some other sense that I have not considered here.

It may help to offer some examples of the different senses that I have mentioned so far. The statistical sense is typically understood as being used in statements like the following:

- 1) It's normal for it to rain at some point during the day in the Amazon rainforest.

Now compare this to a statement that would be typically understood as using the evaluative sense of 'normal':

- 2) It's not normal to lie to your partner.

What makes (1) true or false are statistical facts about how often it rains in the Amazon, while the truth or falsity of (2) seems largely independent of any statistical facts, as even if it were true that most people lie to their partners it would still be plausible to claim that (2) is true. Instead, it seems that the truth-value of (2) is dependent on the existence of certain moral norms regarding the wrongness of lying to one's partner. This is why the orthodox understanding considers (2) to be value-laden and (1) to be value-neutral.

Examples of the prototypical and teleological senses of 'normal' are commonly found in biology and medicine, which is why there has been some discussion of the concept of normality in the philosophy of both those areas. An example of a statement that uses the prototypical sense is:

3) Birds normally have feathers.

The truth of this statement depends on what we take to be a prototypical characteristic of birds. If feathers are prototypical then (3) is true. This does not entail that *all* birds have feathers, just that feathers are a feature of the typical form of birds. This also does not entail that most or even some birds will have feathers. If some disease caused every bird to lose its feathers, it would still be true that feathers are a prototypical feature of birds, and so (3) would also still be true. This is what makes the prototypical sense distinct from the statistical sense, as the truth-values of statements that use the former sense do not supervene on statistical facts. The prototypical sense also differs from the evaluative sense, as it can be used to make claims that are value-neutral.

This final example can be understood as using the teleological sense:

4) The heart normally pumps blood around the body.

Here the truth-value of the statement depends on what we take to be the proper function of the heart. As with the prototypical sense, this does not supervene on statistical facts, as the heart's function would still be to pump blood, even if every single heart in the universe failed at the same time. As a result, the teleological sense is distinct from the statistical sense. It is

also distinct from the evaluative sense because it can be used to express claims that are value-neutral.

It should now be clear what the orthodox view is and what some of the different senses are that the proponents of this view take 'normal' to have. As we have already seen, this view has been widely assumed by those involved in discussing whether or not causation is normative. It tends to just be accepted that something can be normal in the sense of being statistically regular, in the sense that it conforms to moral norms, or in some different sense entirely. The unclarity objection simply draws attention to this fact and argues that this makes it unclear what the proponent of a normative account means when they talk about a 'normal event' or 'normal state of the system'. Indeed, Blanchard & Schaffer seem to suggest that this ambiguity allows the proponent of a normative account to cheat by picking whichever sense best serves their purpose at any given moment (Blanchard and Schaffer 2017, 192). Such an argument strongly suggests that Blanchard & Schaffer are assuming the orthodox view.

Outside of causation, we can also see the orthodox view being assumed in other fields. Most notable is in how 'normal' has been understood within the philosophy of biology. For instance, Vácha (1978) is an early example of a proponent of the orthodox view. He distinguishes between a statistical and 'value' sense of 'normal', where the value sense just is the evaluative sense that I have explicated. This is taken further by Wachbroit, who claims that 'normal' is ambiguous between *three* different senses: statistical, evaluative, and biological (1994, 580-581). This biological sense is a teleological sense and is, unsurprisingly, intended to capture the sense of 'normal' that is used in biology.

Schurz (2001) seems to disagree that the sense used in biology is teleological. He distinguishes two kinds of normality that any use of 'normal' could be referring to: prototypical and statistical. On his understanding, a feature is a prototypical trait of a certain biological class if that trait was selected for as part of that class's evolutionary history. This analysis can equally be applied to non-biological classes that also have an evolutionary history (Schurz 2001, 494). Schurz then goes on to argue that prototypical normality implies statistical normality, since a trait's being evolutionarily selected for within a certain class will likely result in most of the members of that class having that trait (Schurz 2001, 495). Schurz does not postulate an evaluative sense, or any other sense of 'normal' that is value-laden. The result is a view that can plausibly be understood as holding that 'normal' is ambiguous between a statistical and a prototypical sense.

Some authors have sought to challenge the notion that the sense of 'normal' used by biologists is entirely descriptive. For instance, Dupré (1998) argues that the use of the term when talking about people will always express an evaluative sense. Chadwick also recognises that there are many cases where the use of 'normal' can have ethical ramifications, though she still maintains a distinction between the 'scientific' definitions of 'normal' (which include the statistical and teleological senses) and the 'normative' definitions (Chadwick 2017).

We can also see the orthodox view being assumed in other areas where normality is relevant. In philosophy of medicine, it is generally agreed that the main goal of healthcare is to restore 'normal functioning', but there is widespread disagreement over what this means. Naturalists like Boorse understand normal functioning though statistical typicality (Boorse 2014, 684), while normativists like Sedgewick argue that the failure of normal functioning can be

understood as a failure to conform to some value-laden norm (Sedgwick 1982, 32). Neither side argues that their opponent has failed to provide a correct analysis of normality; they merely disagree over which sense of 'normal' is appropriate here. Likewise, in epistemology we have Martin Smith's work on justification, where he assumes that there are different senses of 'normal', with one being statistical and another being a sense that seems to represent what we expect to happen, such that any deviation from these expectations needs to be explained (2016, 39-40).

Outside the realms of biology and medicine, normality has received far less philosophical attention, though we can still see the orthodox view being presupposed. Within the causation debate this comes through in the accounts that were discussed in the previous chapter. Hitchcock and Knobe note that a potential objection to their account would be that it 'suffers from a fundamental ambiguity' (2009, 597) because 'it might be suggested that terms like 'norm' and 'normal' can be used in a number of different senses' (*ibid.*): 'statistical norms' and 'moral norms' (*ibid.*). At no point do they challenge this claim, instead proceeding with the somewhat underdeveloped claim that the relevance of different counterfactual scenarios in causal reasoning are determined by 'a single overall degree of "normality"' (2009, 598). Later Halpern and Hitchcock state that '[t]he word 'normal' is interestingly ambiguous', seeming to 'have both a descriptive and a prescriptive dimension' (2015, 429).

This should all go some way towards demonstrating how widespread the orthodox view is, despite the fact that it has not, to my knowledge, been explicitly argued for. Therefore, in the following section I will attempt to provide a candidate explanation for why we should believe it. I will then explain why that explanation is unsuccessful.

2. Reasons For and Against the Orthodoxy

When it comes to accounts that postulate an ambiguity, it is important to keep in mind Grice's razor: that senses are not to be multiplied beyond necessity (Grice 1989, 47). In other words, we should generally avoid claiming that a term is ambiguous unless we absolutely have to. Obviously, this does not mean that we ought to immediately discount the orthodox view, but it does mean that the proponents of this view need provide us with some reason for thinking that it is necessary to claim that 'normal' is ambiguous. That is, we need a reason to believe that it is not possible to provide a successful account of normality that does not make this claim. Otherwise, we have no reason to accept the orthodox view, before considering the alternative, namely that 'normal' is not ambiguous.

In the absence of a full philosophical argument for the ambiguity claim, the only positive reason we might have to accept the orthodox view is that it best explains our intuitions about how 'normal' is actually used. To this end, the view's proponents can point to the fact that we seem to use 'normal' in a wide variety of different ways. Statements (1)-(4) in section one of this chapter provide some good examples of this. Each of these statements seems to use 'normal' to mean something very different to what it is being used to mean in the other statements. This would seem to entail that 'normal' is ambiguous.

Unfortunately for the proponent of the orthodox view, this is not enough to provide a reason to believe that 'normal' *must* be ambiguous, as it could be that the term is merely underspecified in a way that allows a plurality of different possible uses that are still consistent with a singular meaning. As an example, the term 'large' is underspecified in a couple of ways. The term generally means something like 'of considerable size', but the

precise conditions that a thing needs to fulfil in order to meet this description can vary depending on certain features of the context that the term is used in. One feature is the reference class that is contextually salient: a bus may be considered large relative to other road vehicles but not relative to all human-made artefacts. Additionally, our judgements about whether or not something is large are graded along multiple axes (height, width, and depth) and one axis may be more relevant in a particular context. As an example, a bus may be too large to fit through a tunnel by virtue of being too tall, while a much longer and wider train is not too large to fit through the same tunnel.

It might be that we could say something similar about 'normal'. In a recent work of experimental philosophy, Bear & Knobe (2017) present the findings from their study into the use of 'normal' and use it to argue for the existence of an 'undifferentiated representation of what is normal' that is shaped by both what is 'average' (though it is left unspecified which sense of 'average' they mean) and what is 'ideal'. This kind of normality does not fit neatly into any of the senses of 'normal' that we have discussed so far, so it will be difficult for the orthodox view to account for it. We may have more success imagining an account where what's normal is represented across two axes: conformity to the average, and conformity to the ideal. As with the bus case, we can say that our normality judgements will be influenced by the relative weighting we give to one axis over the other, as well as by the contextual salience of different reference classes. Such an account has (to my knowledge) not been previously explored in the literature.

I will expand further on this dual-axis view in the following chapter. All that matters for my present purposes is to note that this view is not obviously implausible and benefits from the

fact that it does not require us to postulate more than one sense of 'normal'. On its own this is not sufficient reason to conclude that we ought to prefer such an account to the orthodox view. To do this, we will also need to show that there are no obvious reasons that could be offered in support of the orthodox view, and we also need to provide some reasons for preferring the alternative. This will be the focus of the remainder of this chapter.

In the next section I will explore how the proponent of the orthodox view might try to appeal to the results of ambiguity tests to support their position, before demonstrating that the tests not only offer very little support to the orthodox view; they also suggest that defending the view will be much harder than we might have initially supposed.

3. Ambiguity Tests

A widely accepted test for ambiguity comes from Zwicky & Sadock's 'Ambiguity Tests and How to Fail Them', where they provide a number of tests that are meant to identify ambiguous terms as opposed to terms that are merely underspecified (Zwicky and Sadock 1975). If 'normal' passes this test, then the proponent of the orthodox view will have a compelling case for the claim that 'normal' is ambiguous.

The test that I shall focus on is known as 'conjunction reduction' (Zwicky and Sadock 1975, 17-20). The test focuses on trying to recognise whether or not specific conjunctions exhibit zeugma. This is a feeling of absurdity that we have when we are forced to read two or more different senses of the same term within the same sentence. The conjunction reduction test has us construct sentences that force multiple readings of the term we are testing. We do this by taking two statements that each contain a use of the term we are testing and then putting them together to form a conjunction. The resulting sentence is then reduced so that the term

we are testing only occurs once. If the resulting sentence feels zeugmatic, then the term is ambiguous.

To illustrate how this works, let's apply the test to the genuinely ambiguous term, 'bank'. Our test sentence is:

5) Adam and Charlotte both live near a bank.

Since 'bank' is a genuinely ambiguous term, each conjunct should have two different readings.

These can be represented in the following table:

	1st Conjunct	2nd Conjunct
	Adam lives near a bank...	...Charlotte lives near a bank.
Sense A	Adam lives near a financial institution.	Charlotte lives near a financial institution.
Sense B	Adam lives near a riverbank.	Charlotte lives near a riverbank.

This means that when both conjuncts are taken together, there are a total of four possible readings (senses A+A, B+B, A+B, and B+A). For the test, we are only interested in the crossed readings (A+B and B+A). To force those reading we can imagine that Adam lives near Barclays, Charlotte lives by a river, and the speaker of (5) is aware of both these facts. In such a context, we are forced into the A+B reading of (5), which is clearly zeugmatic. So, 'bank' passes the test, giving us clear evidence that it is ambiguous.

The question now is whether or not ‘normal’ passes the test. If the sentences we apply the test to have crossed readings that are zeugmatic, then we will have evidence for the orthodox view. However, as I will demonstrate, these sentences do not seem to exhibit zeugma on any reading. As an example, consider the following sentence:

6) Finding a four-leaf clover is not normal, and Trump’s presidency isn’t either.

According to the orthodox view, normal has at least two senses, with one being a statistical sense and the other most likely being an evaluative sense. So, as with the first two examples, there are four possible readings here:

	1st Conjunct	2nd Conjunct
	Finding a four-leaf clover is not normal...	...Trump’s presidency isn’t normal.
Sense A	Finding a four-leaf clover is statistically irregular.	Trump’s presidency is statistically irregular.
Sense B	Finding a four-leaf clover violates some evaluative norm.	Trump’s presidency violates some evaluative norm.

My intuitions regarding this case are that there is nothing zeugmatic about either crossed reading. B+A does seem to be rather obviously false in virtue of the fact that it seems implausible to say that finding a four-leafed clover violates an evaluative norm. However, the B+B reading also seems false in this way and thus any weirdness felt on the B+A reading is

likely not the result of this reading being zeugmatic. As such, it seems reasonable to conclude that (6) provides no evidence for 'normal' being ambiguous: 'normal' fails the conjunction reduction test.

It is important to note that these results *do not* show that 'normal' definitely is not ambiguous. The reason for this is discussed by Viebahn (2018), who argues that failing the test should not be used to argue for the non-ambiguity of a term (as Stanley (2005) does with the term 'knows'). Viebahn notes that there are genuinely ambiguous terms that seem to fail the test in certain instances (2018, 753-754). He then goes on to contend that the distinction between the ambiguous terms that reliably pass the test and those that don't maps onto an existing distinction between two different kinds of ambiguous term: terms that are homonymous and terms that are polysemous. Terms of the former kind will have multiple senses that are completely unrelated. 'Bark' is a good example of a homonymous expression, as it can be used to refer to the sound a dog makes or a part of a tree. The fact that we can use the same word to refer to both of these things is entirely coincidental, as there is no connection between the two senses of the term. Terms that are polysemous are distinguished by the fact that they have multiple senses that *are* related in some way. 'Book' is an example of this, as the term can be used to refer to either an abstract work or a singular physical copy of that work. These two senses differ in the kind of thing that is being referred to, but it is clear that the two are related by the fact that they both refer to a work of literature in some form. Hence 'book' is polysemous.

As Viebahn demonstrates, polysemous terms, while ambiguous, do not reliably pass the conjunction reduction test. That is, the kinds of sentences we generate when testing

polysemous terms will sometimes have non-zeugmatic readings, despite the fact that such terms *are* ambiguous. As an example, Viebahn offers the following:

- 7) 'I heard of the book and picked it up at the library a few hours later.' (Viebahn 2018, 754)

Intuitively this sentence is not zeugmatic – it fails the conjunction reduction test – despite the fact that the context pushes us towards a crossed reading (we don't hear about specific physical copies of books, and we don't pick up abstract works). If failing the conjunction reduction test were proof that a term is not ambiguous, then we would have to conclude that 'book' is not ambiguous. Yet this is not the case. Therefore, failing the conjunction reduction test cannot be used to argue that a term is not ambiguous, because polysemous terms are ambiguous and yet do not reliably pass the test. So, the fact that 'normal' does not pass the test only gives us reason for thinking that the term is not homonymous; it leaves open the possibility that it could be polysemous.

Unfortunately for the proponent of the orthodox view, the project to provide a polysemous account of 'normal' will be complicated by the fact that, as demonstrated by Moldovan (2019), it is not the case that *all* polysemous terms reliably fail the test; only the members of a certain subset do. This distinction between polysemous terms that do and do not reliably pass the test corresponds to an existing distinction between two different forms of polysemy: regular and irregular³³. Regular polysemy 'is obtained in a rule-governed manner and is characterised as systematic or productive' (Moldovan 2019, 12). What it means to say that

³³ For an early discussion of this distinction, see (Apresjan 1974).

regular polysemy is systematic/productive is that it is both predictable and consistent. So, for any regular polyseme, we would expect there to be other words that have senses which will be distinguished from one another and related to each other in a manner that corresponds to the polyseme in question. As an example, the distinction between the two senses of 'book' as referring to either an abstract work or a physical copy is replicated in other terms like 'film' or 'magazine'. Therefore, all of these terms can be considered examples of regular polysemy because the distinctions between their senses are systematic.

Irregular polysemy differs in that it has 'less predictable connections between the different senses' (Eddington and Tokowicz 2015, 14). Moldovan uses the term 'idiosyncratic' to describe the ways that the senses of irregular polysemes are distinguished, to contrast them with the predictability and consistency of regular polysemy. 'Bank' is an example of this, as the two senses are both derived from a common source, but in a way that is not obvious or systematic: the term was originally used to refer to both the slopes of a riverbank and the sloped tables used by moneylenders (OED Online 2020a) (2020b). It is important to note that more recent works characterise this distinction as being between two different parts of a spectrum that runs from homonymy to regular polysemy, with irregular polysemy being somewhere in the middle (Rabagliati and Snedeker 2013) (Vicente and Falkum 2017). On the regular polysemy end of the spectrum, the different senses of a term are connected by a well-defined and systematic relation that also obtains between other similar senses (such as 'book'). Then on the other end, we have terms with senses that are not connected in any way, other than being expressed by the same word (such as 'bark'). Then in the middle we have irregular polysemy, where the connection is less predictable and more idiosyncratic (such as 'bank'). So, all ambiguous terms fall somewhere on this spectrum.

Moldovan's main claim is that the closer an ambiguous term is to the regular polysemy end of the spectrum, the more likely it is to fail the conjunction reduction test for ambiguity. Therefore, there are only two possible explanations for a term's failing to pass the test. One is that the term is closer to the regular polysemic end of the spectrum. The other is that the term is simply not ambiguous at all. As things currently stand, it is an open question as to which of these explanations best applies in the case of 'normal'. However, taking the view that 'normal' is a regular polyseme requires demonstrating that the relationship between its different senses is systematic, and can consequently be seen in other regular polysemes.

One analogy that may be given is with the term 'possible'. This term is ambiguous between a few different senses, including a logical sense, a nomological sense, and an epistemic sense. This is an example of regular polysemy because all these senses express something related to the notion of consistency. If we describe something as 'possible' in the logical sense, what we mean is that it is consistent within itself. When we describe something as 'possible' in the nomological sense, we are saying that it is consistent with the laws of nature. Finally, when we say that something is 'possible' in the epistemic sense, then we mean that it is consistent with what we already know to be the case.

With 'normal' we could potentially offer a similar explanation by holding that the statistical and evaluative senses are related by the fact that they express some variation on the idea of meeting a standard. That is, when something is described as being 'normal' in the statistical sense, we take it to mean that it meets some kind of statistical standard, while describing something as 'normal' in the evaluative sense means that it meets one kind of moral or social standard. However, the problem here is that the relation between the two senses is not

explained by appealing to a notion that is clearly distinct from the term we are investigating. Explaining the relation between the difference senses of 'possible' in terms of consistency is successful because those notions can be defined independently from possibility. When it comes to the notion of a standard, it's less clear that the explanation works because it is difficult to see how we could define it without some reference to normality. As a result, all this explanation for the relation between the senses of 'normal' can offer us is an insistence that they both tell us something about normality. This is a poor explanation because it a) does not provide us with any insight on how the senses of 'normal' differ from one another, and b) does not provide any reason to prefer the orthodox view over a univocal account.

All this demonstrates how difficult it would be to provide an account of 'normal' as a regular polyseme, and so, as things stand, we have no positive reason to endorse the orthodox view.

4. A Moral Reason to Prefer an Alternative to Orthodoxy

So far we have seen how, beyond our immediate intuitive responses to certain examples, there are no positive reasons for thinking that 'normal' is ambiguous. In this section and the next I will outline a pair of reasons to prefer an alternative account, such as the dual axis view that I hinted at earlier.

The reason offered in this section is a moral one, and it stems from a problem with the orthodox claim that there is a distinct and entirely descriptive sense of 'normal'. The problem is that this claim seems to come with the unappealing consequence that we cannot effectively challenge cases where 'normal' is used to express bigoted claims.

To illustrate this point, imagine the following exchange taking place on social media:

Brad: 'Gay people aren't normal.'

Ellie: 'That's a horrible thing to say, there is absolutely nothing wrong with being gay. Of course it's normal!'

Brad: 'But gay people make up less than 5% of the population. I'm not saying that there's anything wrong with being gay, but out of all the people in the world they are a very small minority. Of course they aren't normal!'

Ellie: 'No you're wrong. They *are* normal.'

This case is very unlike an ordinary case where a listener is confused about what sense is being expressed by an ambiguous term. To see how, compare it to the following exchange:

Alice: 'Charlotte lives near a bank.'

Emma: 'No she doesn't. The nearest bank is Lloyds, and that's miles away from where she lives.'

Alice: 'Sorry, I meant that she lives next to a riverbank. Her house sits right on the bank of the Thames, so she does live near a bank.'

Emma: 'No you're wrong. Lloyds is miles away.'

In the first exchange, Ellie's final utterance seems entirely appropriate (as well as also being entirely right), while Emma's final utterance in the second exchange seems entirely inappropriate. It is difficult to see how this could be the case if 'normal' were ambiguous because we would have to tell the same story about both.

Rejecting the orthodox view provides us with a very straightforward explanation for the difference in the appropriateness of Ellie and Emma's responses. The reason for the difference is that the first exchange is not a genuine instance of confusion regarding what

sense of an ambiguous is being expressed. Indeed, such confusion is not possible because 'normal' is univocal and not ambiguous. There is no confusion about what sense of the term Brad is using because there is only one sense of the term that he *could* be using. There is no distinct, value-free sense of 'normal'. The sense he is using must be an evaluative sense. Therefore, Brad's initial utterance expresses a negative value judgement, and Ellie is right to challenge him on this.

Brad's subsequent attempts to defend his view by supplying statistical facts are a means to disguise his initial bigoted utterance as a merely descriptive claim that is entirely value neutral. If we are committed to the orthodox view, we have to allow that such a move is possible. We then have to concede that it is just as wrong for Ellie to continue to deny what Brad is saying as it is for Emma to continue to deny what Alice is saying. It is only by denying the orthodox view that we can do justice to the intuition that Ellie is right to continue to challenge Brad's utterances, as well as to deny him the chance to disguise his bigotry as something else. This gives us a compelling reason to reject the orthodox view.

At this point the proponent of the orthodox view has several possible responses to this argument available to them. For the remainder of this section I will consider three responses and demonstrate how they fail to provide reasons to reject my argument. We will begin with the most straightforward response, which is that Brad is just being dishonest when he tries to appeal to a descriptive reading of his first utterance. That is, Brad's initial use of 'normal' was always intended to express the evaluative sense and his later attempt to justify it using statistics is a disingenuous attempt to deflect any accusations of homophobia. On this reading of the case, such accusations are entirely justified, since Brad's initial utterance entails that

being gay is not the way that people ought to be. Consequently, the attempt to appeal to a descriptive sense of 'normal' is an example of what Saul (2017) refers to as a 'figleaf': a linguistic device that is used to provide cover for overt bigotry³⁴. In this case, Brad's appeal to statistics is meant to give the impression that his initial use of 'normal' was intended to express an entirely descriptive sense of the term and thus it cannot be understood as making any negative value judgements about people who are gay. This means that the appeal to a descriptive sense functions as an effective figleaf because it provides a cover for Brad's homophobia.

If this reading of the case is correct, then the argument I have presented in this section will be unsuccessful. The argument rests on the intuition that Ellie's responses to Brad are entirely appropriate and not an instance of someone failing to identify the correct sense that is expressed by the use of an ambiguous term. Yet if we understand Brad's appeal to a descriptive sense as a mere figleaf, then the proponent of the orthodox view is no longer forced to argue that Ellie has misunderstood Brad's initial utterance. Instead, they can say that Ellie was right to understand this use of 'normal' as expressing an evaluative sense and that she was also right to stick with that understanding even after Brad's attempt to provide a cover for the homophobic consequence of that utterance. Therefore, the proponent of the orthodox view can continue to endorse the claim that 'normal' is ambiguous, without having to argue that Ellie's response rests on a misunderstanding.

³⁴ Saul's original paper focuses specifically on the use of figleaves as a cover for racism, but the framework can also be applied to cases of homophobia.

In defence of my argument, I will make one specific change to how the initial exchange is framed. Specifically, I want to stipulate that Brad never intended for his utterance to carry any homophobic meaning or suggest any kind of homophobic conclusion. His utterance is solely motivated by his belief that only a minority of people are gay and does not represent any kind of negative value judgement he has made. His intentions with that utterance are entirely pure. We can even suppose that Brad is blamelessly unaware that homophobia exists and is a persistent social ill. Yet even with all these stipulations, Ellie's response still seems entirely appropriate and does not appear to rest on a misunderstanding. This can only seem to be explained if we hold that 'normal' is not ambiguous and thus always comes with some kind of value judgement.

Of course, this is not to suggest that Brad is deserving of blame in this instance, as if all of these extra stipulations hold, he does not seem to have done anything wrong in this amended version of the case. Rather, he just seems to have misunderstood what 'normal' means, which is something that Ellie's response will help him to recognise. Yet despite this fact, his utterance has still communicated a negative value judgement that warrants being challenged by Ellie. So, the first response is unsuccessful.

A second possible response that is available to the proponent of the orthodox view is to argue that Ellie's response is appropriate because of the pragmatics – as opposed to the semantics – of Brad's initial utterance. The thought behind this objection is that, if we just take the literal meanings of the words used (granting that this use of 'normal' really does have a descriptive meaning), Brad's initial utterance expresses something that is entirely unobjectionable: that being gay is statistically irregular. Strictly speaking, this is true, and so it would be

inappropriate to respond to this utterance in the way that Ellie does based on its semantic meaning alone. However, her response was not aimed at challenging to the *semantic* meaning of the utterance. Rather, its target was the *pragmatic* implication of the utterance. On this way of understanding the exchange, Brad's utterance of 'Gay people aren't normal' in a context where homophobia exists in the background comes with the false implicature that being gay violates some prescriptive norm. The fact that this utterance comes with a false and harmful implicature provides an overwhelming reason against its assertion, even if we accept that it is strictly speaking true. As such, Ellie's response is entirely appropriate – despite the fact that what she has said is strictly speaking false – because the implicature from Brad's utterance is both false and potentially harmful if left unchallenged. Therefore, the proponent of the orthodox view is not forced to argue that Ellie's response rests on a misunderstanding, thus avoiding the force of my argument.

My reply appeals to Haslanger's notion of an ameliorative inquiry (2000). That is, when providing an analysis of a term or concept, we should also consider what the term can do for us, and the kinds of effect that different ways of understanding this term can have. In this case, if we accept the above response to my argument, then we are forced to admit that everything Brad said is true and the problem lies in the fact that he used it in the context that he did. Any attempt to rebuke him would have to come with the caveat that what he said was, strictly speaking, true, and so his utterance would be entirely appropriate if the context was sufficiently different to prevent the generation of the false implicature. This is a problem if we are committed to an ameliorative analysis of 'normal', because it allows bigoted speakers like Brad to hide what they're doing and deflect any reasonable criticism by insisting that their utterances are taking place in a context that renders them acceptable. Therefore,

there are clear ameliorative reasons for wanting to avoid any response that holds Brad's utterance to be true. This means avoiding any view that allows for there to be an entirely descriptive sense of 'normal', including the orthodox view.

The third and final possible response that I will consider is perhaps the most convincing, as it does not allow for the possibility that Brad has said something true with his initial utterance. This response is to follow Dupré (1998) in arguing that there is something special about the term 'normal' such that when it is predicated of a person (or at the very least, a person who is a member of a marginalised group), it will always express an evaluative sense, regardless of any other features of the utterance or the context surrounding it. This means that, even with Brad's subsequent appeal to a descriptive sense and despite any stipulations about his intentions to the contrary, the utterance will always express the claim that being gay violates a prescriptive norm. Therefore, Ellie's response is entirely appropriate, as the utterance expresses something that is both false and harmful. As a result, the proponent of the orthodox view is able to avoid having to argue that Ellie's response rests on a misunderstanding. In effect, then, this response agrees with the univocal account that in this particular case 'normal' cannot be taken to have two different senses, and hence that what Brad says is evaluative; but – unlike the univocal account – it takes this particular case (and others that involve making normality claims about persons) to be a special case where the usual ambiguity of 'normal' does not apply.

I see this as the most plausible response because it does not commit us to the claim that Brad's utterance was true, nor does it make any assumptions about his intentions or psychological states that could be undermined by simply changing the example. That said, it

does require us to accept that we can never describe persons as being ‘normal’ in a statistical sense, which many who support the orthodox view will be reluctant to do.

Additionally, the proponent of the orthodox view to provide some kind of story to explain why ‘normal’ behaves this way. To see why, consider the analogous case of the ambiguous term ‘hard’. This term is ambiguous between meaning ‘difficult’ or ‘very physically solid’. Yet it seems that when it is predicated of abstract entities, the ambiguity vanishes as we are forced to read it in the difficulty sense (e.g. ‘this is a hard problem’). The reason for this is that the physical sense of ‘hard’ is just completely inconsistent with the non-physical nature of abstract entities. The same does not seem to be the case when it comes to predicating a statistical ‘normal’ of persons. What is it about persons that would render it inconsistent to describe them as ‘normal’ in a statistical sense? After all, there are many terms that are entirely (and uncontroversially) descriptive that we can predicate of persons without issue. This is a question that the proponent of the orthodox view will need to answer if they wish to follow the route suggested by this third response.

Let’s now turn to the second reason for preferring a univocal account to the orthodox view.

5. Explaining the Bear & Knobe Data

The second reason that we ought to prefer an alternative, univocal account to the orthodox view is that such a view will likely be more successful at explaining the results of a recent set of studies conducted by Bear & Knobe (2017). Bear & Knobe aim to demonstrate that we employ a general representation of what is normal when making judgements about normality. This general concept of normality captures – and is influenced by – both descriptive and prescriptive considerations but is also distinct from both the alleged statistical and

evaluative senses of 'normal' that we have been discussing this far. In what follows I will summarise and explore each of three studies that Bear & Knobe offer up in support of this conclusion³⁵.

Study 1: Normal Amounts

The first study concerns people's judgements about what counts as a normal amount of something (Bear and Knobe 2017, 27-28). This study had ninety-two participants, divided into three roughly equal-sized groups, each presented with variations on questions about the amounts of things in twenty different domains. As an example, take the domain 'number of hours of TV watched in a day'. The first group were asked what they would consider to be *average* amounts of TV watching, the second were asked about *ideal* amounts, and the third were asked about *normal* amounts, with specific respective questions as below:

- (i) What would you guess is the *average* (emphasis added) number of hours of TV that a person watches in a day?
- (ii) What do you think is the *ideal* number of hours of TV for a person to watch in a day?
- (iii) What is a *normal* amount of hours of TV for a person to watch in a day? (Bear and Knobe 2017, 27)

Bear & Knobe then calculated the mean answer to each question by each group, giving us the mean for the average, ideal, and normal amounts provided in each domain.

The results from this study are presented in the following table:

³⁵ Bear and Knobe in fact present four studies in the paper, however, the first three will provide sufficient material for the purpose of this chapter.

Domain	Average	Ideal	Normal
Books read/yr	10.07	26.15	9.9
Percent middle school students bullied	27.59	2.31	27.26
Mins doctor is late/appointment	17.78	3.97	18.47
Money cheated on taxes	604.56	136.45	636.6
Romantic partners/lifetime	8.04	4.25	8.47
Percent high school dropouts	12.64	3.82	11.13
Drinks of frat brother/weekend	16.79	5.91	14.3
Calories consumed/day	2159.26	1757.84	2063.33
Mins waiting for customer service	15.04	5.78	12.73
Times checking phone/day	45.33	13.12	37.17
Sugary drinks/wk	9.67	3.52	7.3
Servings of vegetables/mnth	34.81	67.67	51.97
Hrs TV watched/day	4	2.34	3.03
Percent students cheat on exam	34.64	3.5	15.97
Times cleaning home/mnth	5.57	6.75	4.72
Hrs exercising/wk	5.37	7.31	6.77
Lies told/wk	24.25	2.75	8.43
Computer crashes/mnth	4.78	0.5	1.6
International conflicts/decade	19.3	1.59	4.82
Times calling parents/mnth	6.04	6	5.23

Figure 1

Bear & Knobe argue that this data is evidence that our understanding of what is normal is determined by our beliefs about both what is average and what is ideal. Out of twenty different domains, fourteen are such that the normal value falls in between the average and ideal values. This is significant, Bear & Knobe explain, because it means that the normal value falls in between the other two values more often than would be the case as the result of chance³⁶. It is also worth noting that in most of the cases where the normal value didn't fall between the average and the ideal figures provided, it still ended up very close to one of them (usually the average figure).

This study was later repeated with an additional twenty domains (Bensinger, Bear and Knobe 2016). This time, twenty-five out of the forty normal values fell in between both the average and the ideal values. Once again, this is much higher than what we would expect from chance³⁷. As with the first iteration of the study, most of the normal values that fell outside the range between the average and the ideal were still very close to one of the other values (again, usually the average).

From all of this, Bear & Knobe conclude that our judgements about what counts as a normal amount of something are determined by our beliefs about what are average *and* ideal

³⁶ There are three values in play, so the odds of any one of them falling in between the other two is roughly 1/3. So, we would expect the normal value to fall in between the average and the ideal values in roughly six or seven of the domains, much fewer than the fourteen domains where this actually happened.

³⁷ Again, if it was random, we would have expected around 1/3 of the normal values to fall between the average and ideal values. This means that we would have seen this happen in roughly thirteen or fourteen domains, rather than twenty-five.

amounts of that thing, and so our concept of normality is affected by both prescriptive and descriptive considerations.

Study 2: Gradable adjectives

Let's now look at the second study (Bear and Knobe 2017, 28-29), which is very similar to the first in that it concerns judgements of amounts. The main difference is that this study was intended to examine the participants' use of gradable adjectives, such as 'large', 'hot', 'loud', 'fast', and 'difficult', instead of their use of 'normal'. These are all terms that allow us to understand entities along a scale. For example, when it comes to size, we can have entities that are described as 'small', 'medium', or 'large'. Kennedy (1999) has demonstrated that whether or not a particular use of one of these terms will count as accurate will depend on what entity we are referring to and what class of things it is being compared to. However, once those things are fixed, our judgements then concern how the object compares to a *standard* object of the same type. As an example, if I were handed a bowl and then asked to judge whether or not it was a large bowl, I would reason by comparing the bowl than I am holding with some standard sized bowl that I am imagining. If the bowl is sufficiently greater in size than the standard bowl, then I will judge that it is a large bowl. Put simply, the use of gradable adjectives involves an implicit comparison to some kind of implicit standard.

Bear & Knobe argue that this behaviour can be understood in terms of judgements about normality (Bear and Knobe 2017, 27). They point out how earlier studies have demonstrated that our intuitions about the relevant standard are influenced by both descriptive (Barner and Snedeker 2008) and prescriptive (Egré and Cova 2015) considerations. Bear & Knobe propose that 'when people are trying to determine which point along the scale counts as the standard,

they are influenced by intuitions about which point is the *normal* one' (Bear and Knobe 2017, 27). In other words, a *standard* amount of something is just a *normal* amount of something. Therefore, if Bear & Knobe can demonstrate that we form our judgements of standard amounts from our beliefs about average and ideal amounts, then that will provide additional support for their conclusion that we have a general understanding of normality.

To do this, they took 101 new participants and presented each of them with a single question about each of the domains from the first study. The questions were all structured as follows:

Imagine that a person watches y hours of TV in a day. Please rate the extent to which you think this is a large or small number of hours of TV for a person to watch in a day (Bear and Knobe 2017, 28)

Here, ' y ' stands for a randomly selected integer that fell between 50% and 150% of the average value for the corresponding domain from study one (in this case, the number of hours spent watching TV). The participants were then asked to indicate on a seven-point scale to what extent they thought that ' y ' was a large or small amount, with linear regression used to estimate the standard point for each domain from these responses³⁸ (Bear and Knobe 2017, 29). The results are below, alongside the average and ideal values from study one for reference:

³⁸ One problem with this method is that it can, under certain circumstances, produce unusual results, as seen in "Percent high school dropouts" and "International conflicts/decade". These results are likely due to subjects consistently picking the highest values on the seven-point scale, regardless of what number took the place of " y ".

Domain	Average	Ideal	Standard
Books read/yr	10.07	26.15	17.79
Percent middle school students bullied	27.59	2.31	10
Mins doctor is late/appointment	17.78	3.97	8.02
Money cheated on taxes	604.56	136.45	35.54
Romantic partners/lifetime	8.04	4.25	12.38
Percent high school dropouts	12.64	3.82	624.82
Drinks of frat brother/weekend	16.79	5.91	18.43
Calories consumed/day	2159.26	1757.84	23.03
Mins waiting for customer service	15.04	5.78	14.51
Times checking phone/day	45.33	13.12	5.49
Sugary drinks/wk	9.67	3.52	10.08
Servings of vegetables/mnth	34.81	67.67	32.46
Hrs TV watched/day	4	2.34	5.45
Percent students cheat on exam	34.64	3.5	3.83
Times cleaning home/mnth	5.57	6.75	40.95
Hrs exercising/wk	5.37	7.31	23.7
Lies told/wk	24.25	2.75	3.82
Computer crashes/mnth	4.78	0.5	13.79
International conflicts/decade	19.3	1.59	2007.57
Times calling parents/mnth	6.04	6	6.29

Figure 2

Again, these results support Bear & Knobe's conclusion, with the standard values falling between the average and ideal values in fourteen out of the twenty domains: a much higher rate than we would expect to happen by chance.

Study 3: Normality as Prototypicality

Study three differs significantly from the first two, and the results that it produces are far more complex. However, we will not need to go through all the results of this experiment here, as they are not required for my argument.

The third study concerns people's judgements regarding what counts as a *prototypical* example of a given category. It has previously been demonstrated that such judgements are influenced by descriptive (Rosch and Mervis 1975) and prescriptive (Barsalou 1985) (Lynch, Coley and Medin 2000) considerations. As with gradable adjectives, Bear & Knobe argue that this can be understood in terms of normality. The thought is that when we are considering whether something is a prototypical example of a category, we are really considering whether it is a *normal* example of that category. So, Bear & Knobe's aim with this study is to show that our judgements of prototypicality are influenced by our beliefs about average and ideal examples of a category.

To test this, Bear & Knobe (2017, 30-31) took 542 participants and split them into three roughly equally sized groups. All the participants were randomly given one out of six exemplars from each of the eight categories that were included in the study (categories like teacher, grandmother, vacation, car, etc.). They were then asked to indicate on a seven-point scale to what extent they agreed/disagreed with a claim that varied depending on the group.

The first group were asked if they thought the exemplar was an average member of the category. The second group were asked if they thought it was an ideal member of the category. Finally, the third group were randomly presented with one of three questions: one concerning whether they thought the exemplar was a *good* example, another concerning whether they thought it was a *paradigm* example, and a third concerning whether it was a *prototypical* example.

To illustrate this, one exemplar for the category 'grandmother' was as follows:

An 80-year-old woman who is constantly grumpy and mean to her grandchildren. Detests spending time with other people, but always demands that her children do favors for her. Talks in a loud and shrill voice. (Bear and Knobe 2017, 34)

Members of the first group who were presented with this were then asked to what extent they thought it described an average grandmother. Members of the second group were asked if they agreed that it described an ideal grandmother. While members of the third group were asked if they agreed that it described a *good/paradigm/prototypical* example of a grandmother. Bear & Knobe assume that these terms prompted the participants to consider which exemplars represented a normal instance of each category. The judgements that the participants made can therefore be understood as judgements of normality.

Bear & Knobe then found the mean for the average, ideal, good example, paradigm example, and prototypical example ratings for each of the exemplars. As with the first two studies, the values that related to normality fell in between the other two values at a rate that was far higher than would be expected by chance, providing further support for Bear & Knobe's conclusion. Both this and the previous study rely on the claim that the concept of normality

is invoked in our judgements of prototypicality and standards, respectively. I am in agreement with this claim, and thus take all three studies as providing evidence for Bear & Knobe's conclusion.

Discussion

There may be some worries that Bear & Knobe's conclusion is a little too strong here. They are certainly correct, that what the subjects took 'normal' to mean was distinct from what they understood as 'average' or 'ideal', as the normal values consistently fell between the average and ideal values. Yet it might seem to be too soon to declare that this captures what 'normal' means in most contexts. After all, it might very well be the case that the sense identified by Bear & Knobe is one that only gets used within contexts that involve being the subject in an experiment, or that otherwise share some other feature in common with such a context. That is, are Bear & Knobe really justified in the move from their results to a general conclusion about the meaning of 'normal'? Personally, I think that they are, because it is difficult to see why their results would simply be confined to contexts of the same kind as being a subject in an x-phi study. For this to be the case, it would have to be true that there is a highly specific sense of 'normal' that can only be properly used in a fairly limited pool of cases. Therefore, I think we are fairly safe in accepting Bear & Knobe's generalised conclusion.

Another worry might be that the judgments made by the participants are not the result of the semantic meaning of 'normal'. On this understanding, they were not making their judgements using an as yet unidentified general sense of normal that is shaped by both descriptive and prescriptive considerations. Rather, their judgements were an expression of the statistical sense of 'normal' that were then adjusted post hoc towards the ideal for pragmatic reasons.

In an everyday context it can be easy to see why we might do this for some domains. As an example, we might not want to be seen as justifying international conflicts by saying that it is 'normal' for there to be 19 per decade. However, there are a few problems with this explanation. Firstly, it does not work for every domain, why would we consistently choose to inflate the number of books it would be normal to read in a year? If anything, it seems like we might want the number to be lower if we feel we haven't read enough. Secondly, it seems difficult to cash out this explanation without tacitly admitting that the semantics of 'normal' include some kind of evaluative component – otherwise what would be the problem with saying that 19 international conflicts per decade was normal? We could simply insist that we were using the statistical sense of 'normal' in the same way that Brad tried to appeal to such a sense in the example from the previous section. Finally, it is difficult to see how such pragmatic considerations could carry so much sway in the context of an anonymised survey.

From this, it is reasonable to say that, in the absence of a better explanation of the data or a plausible criticism of their methodology, Bear & Knobe's conclusion is warranted. This conclusion, that there is a general sense of 'normal' which is being employed by the participants when they are asked to make normality judgements, is one that the orthodox view will struggle to elegantly account for. They obviously cannot take this general sense to be the only sense of 'normal', so there are really only two options available to them: either the general sense is its own sense, in addition to the ones already identified, or Bear & Knobe are wrong to draw the conclusion that they do.

The option of postulating an additional, general sense of 'normal' seems to be the less appealing option of the two. The reason for this is that it seems to invite some difficult

questions regarding why we would have this extra sense of 'normal' and how we are able to express a more specific sense over the general one. One question asks why we would need a general sense of 'normal', if the more precise senses are already available to us. The proponent of the orthodox view might claim that the general sense is for our everyday, imprecise talk of 'normal', while the more precise senses are for more formal contexts. This is not completely implausible, but if it were true, 'normal' would be unique amongst ambiguous terms for having two or more specific senses for formal contexts and a general sense for everyday contexts that is shaped by all the other senses. This also leads us to the question of how we are even able to detect whether a speaker is using the general sense or a more specific sense. Again, the proponent of the orthodox view could argue that it just comes down to whether the term is being used in a formal or everyday context. Yet even if such an account were forthcoming, we would still have reason to prefer a univocal account here just by appealing to Grice's razor.

This means that the proponent of the orthodox view will likely be more successful by simply rejecting Bear & Knobe's conclusion. To do this, they will have to demonstrate that Bear & Knobe's subjects were not, in fact, using the general sense of normal that Bear & Knobe identify, but were instead using a specific sense of normal. This route seems more promising, but it would require some kind of explanation for why the normal values so consistently fell in between the average and ideal values. In the absence of this kind of explanation, or some other explanation that presents an alternative to Bear & Knobe's conclusion, or the identification of some kind of methodological failing on their part, we have no reason to discount their conclusion, and thus no reason to endorse the orthodox view.

The alternative dual-axis view, which I develop in the next chapter, seems to fare far better in this regard. Such a view allows for the possibility that our judgements of normality can be shaped by both descriptive and prescriptive considerations. In the following chapter I will explain exactly how that is to be done.

It should be clear at this point that, although the possible responses to Bear & Knobe's results by the proponent of the orthodox view are not completely implausible, they seem far less appealing than Bear & Knobe's hypothesis that there is one singular sense of 'normal', namely the sense deployed by the participants in study. This then provides us with an additional reason to prefer a univocal account to the orthodox view.

6. Moving Past the Orthodox View

It should be clear by now that, beyond the intuition that statements such as (1)-(4) at the start of this chapter deploy different senses of 'normal', there is no reason to prefer the Orthodox view to a univocal account, and there *are* good reasons to prefer a univocal account to the orthodox view.

Moreover, there is an additional reason for preferring a univocal account when the ultimate goal is to provide a normative account of causation. The reason is that a univocal account will help us to avoid the unclarities objection, as it would make it impossible for the meaning of 'normal' to be underspecified with respect to which sense is being used, which is a distinct possibility on the orthodox view. Therefore, even those who remain unconvinced by my argument against the orthodox view should be able to see why we ought to prefer a univocal account if we want to endorse a normative account of causation.

In the next chapter I will therefore provide my own univocal account of normality, which I will then use in a normative account of causation.

Chapter 6 – A Normative Account of Causation

In this chapter I shall provide my own normative account of causation. So far I have established that many of the existing attempts to develop such an account have failed as a direct result of not including a complete analysis of the normative concepts that they employ. To address this, I will begin by offering my own analysis of the concept of normality, and I will then deploy that analysis in a new account of causation. This analysis of normality will, for the reasons provided in the previous chapter, will be univocal. In that chapter I also raised the possibility of a dual-axis view of normality, which I will expand upon and develop into a complete account.

It is important to note that my aim is not to provide a full conceptual analysis of causation, as doing so would require further discussions of issues that fall outside the scope of this thesis (e.g., the relata of causation, the preemption problem, whether or not causation is transitive, etc.). Rather, my goal will be to explain what the constraints on contextually relevant counterfactual scenarios are, if we accept that the relevant constraint is whether or not those scenarios are normal. I will remain neutral on what exactly a complete analysis of causation will look like, aside from continuing to assume a broadly counterfactual view. My hope is that the account I offer here can be invoked by other theorists to address the two key questions that I posed all the way back in Chapter 1:

- I. Is causation normative?
- II. Is causation contextual?

My answer to both will be a resounding ‘yes’, and in what follows I will explain how and why this is the case.

This chapter will be divided into three main sections. The first will focus on presenting and arguing for the dual-axis view of normality. In section two I will then make use of this in a new normative account of causation before making use of this account to solve the puzzle cases presented in Chapter 2. Finally, in the third section I will explain my answers to the two main questions raised at the beginning of this thesis.

1. The Dual-Axis View of Normality

In the previous chapter I hinted at an account of normality according to which what counts as normal is defined with reference to two axes: one representing descriptive considerations and the other representing prescriptive considerations. This account is partly informed by the experimental research from Bear & Knobe (2017) that I discussed in the previous chapter, although the account itself is wholly original. Bear & Knobe used their experimental results to argue that the folk have a generalised concept of normality that they understand as being typically expressed by the term ‘normal’. The fact that the participant’s judgements of what is ‘normal’ tended to fall in between their judgements of what is ‘average’ and what is ‘ideal’ is also taken to show that this general kind normality is shaped by both descriptive averages and prescriptive ideals. Unfortunately, these results do not provide are not fine-grained enough to allow us to truly understand how normality is shaped. As such, the account I offer will presented as a somewhat speculative hypothesis, that could hopefully be verified through further testing.

My account represents normality as a field whose boundaries are drawn on two different axes: one for the average and the other for the ideal. The fact that our judgements of normality seem to vary across different contexts can thus be explained by the idea that these axes carry different degrees of significance in different contexts, which results in the reshaping of the boundaries of what counts as normal.

To spell out the precise details of this account, I will use Bear & Knobe's data as a starting point. I will then use this data to derive three specific claims that I wish to make about our judgments of normality:

- 1) Judgements about what's normal do not typically fall on the exact mid-point between the average and the ideal. Rather, these judgements are almost always significantly closer to one side than the other.
- 2) The sense of 'ideal' that is used in arriving at judgements of normality is not a best-of-all-possible-worlds sense of ideal, but rather a 'what is the best that we can reasonably hope for' sense of ideal.
- 3) Normality judgements are context-sensitive.

I believe that each of these points suggest an important truth about normality. I will discuss each point in turn and explain its impact on my analysis.

(1) Degrees of Similarity

To discuss point (1) it will be useful to examine the degree to which the normal values given in Bear & Knobe's results are aligned with the average values or with the ideal values. To measure this, I developed a measure of the degree of similarity that each normal value had to either its corresponding average or ideal value. To calculate the degree of similarity to the average (sa), we can use the following equation:

$$sa = 1 - \frac{|n - a|}{|a - i|}$$

Where 'n' stands for the normal value, 'a' stands for the average value, and 'i' stands for the ideal value. Calculating the degree of similarity to the ideal (*si*) is very similar, with only one minor change:

$$si = 1 - \frac{|n - i|}{|a - i|}$$

For clarity, I then converted the results of both formulas into percentages. One problem with this method is that in cases where the normal value falls outside of the range between average and ideal you end up with one of the numbers being negative. In such instances, I adjusted any negative numbers up to zero, as having negative degrees of similarity doesn't make much sense.

To illustrate this, let's apply the methodology to one of the domains that Bear & Knobe tested. I will use the values given for 'servings of vegetables per month'. There the average value was 34.81, the ideal value was 67.67, and the normal value was 51.97. So, the calculation for the degree of similarity to the average in this domain is as follows:

$$1 - \frac{|51.97 - 34.81|}{|34.81 - 67.67|}$$

Which comes to 0.4778, or 47.78%. This is the degree of similarity that the normal value has to the average value in the domain of 'servings of vegetables per month'.

Applying this to all the data gathered from the first study generates the following results:

Domain	Degree of Similarity to Average	Degree of Similarity to Ideal
Books read/yr	98.94%	0.00%
Percent middle school students bullied	98.69%	1.31%
Mins doctor is late/appointment	95.00%	0.00%
Money cheated on taxes	93.16%	0.00%
Romantic partners/lifetime	88.65%	0.00%
Percent high school dropouts	82.88%	17.12%
Drinks of frat brother/weekend	77.11%	22.89%
Calories consumed/day	76.10%	23.90%
Mins waiting for customer service	75.05%	24.95%
Times checking phone/day	74.67%	25.33%
Sugary drinks/wk	61.46%	38.54%
Servings of vegetables/mnth	47.78%	52.22%
Hrs TV watched/day	41.57%	58.43%
Percent students cheat on exam	40.04%	59.96%
Times cleaning home/mnth	27.97%	0.00%
Hrs exercising/wk	27.84%	72.16%
Lies told/wk	26.42%	73.58%
Computer crashes/mnth	25.70%	74.30%
International conflicts/decade	18.24%	81.76%
Times calling parents/mnth	0.00%	0.00%

Figure 1

One immediate observation we can make about this data is that the normal values tend to be closer to the average values than they are to the ideal values. This becomes more apparent when we plot the data onto a graph:

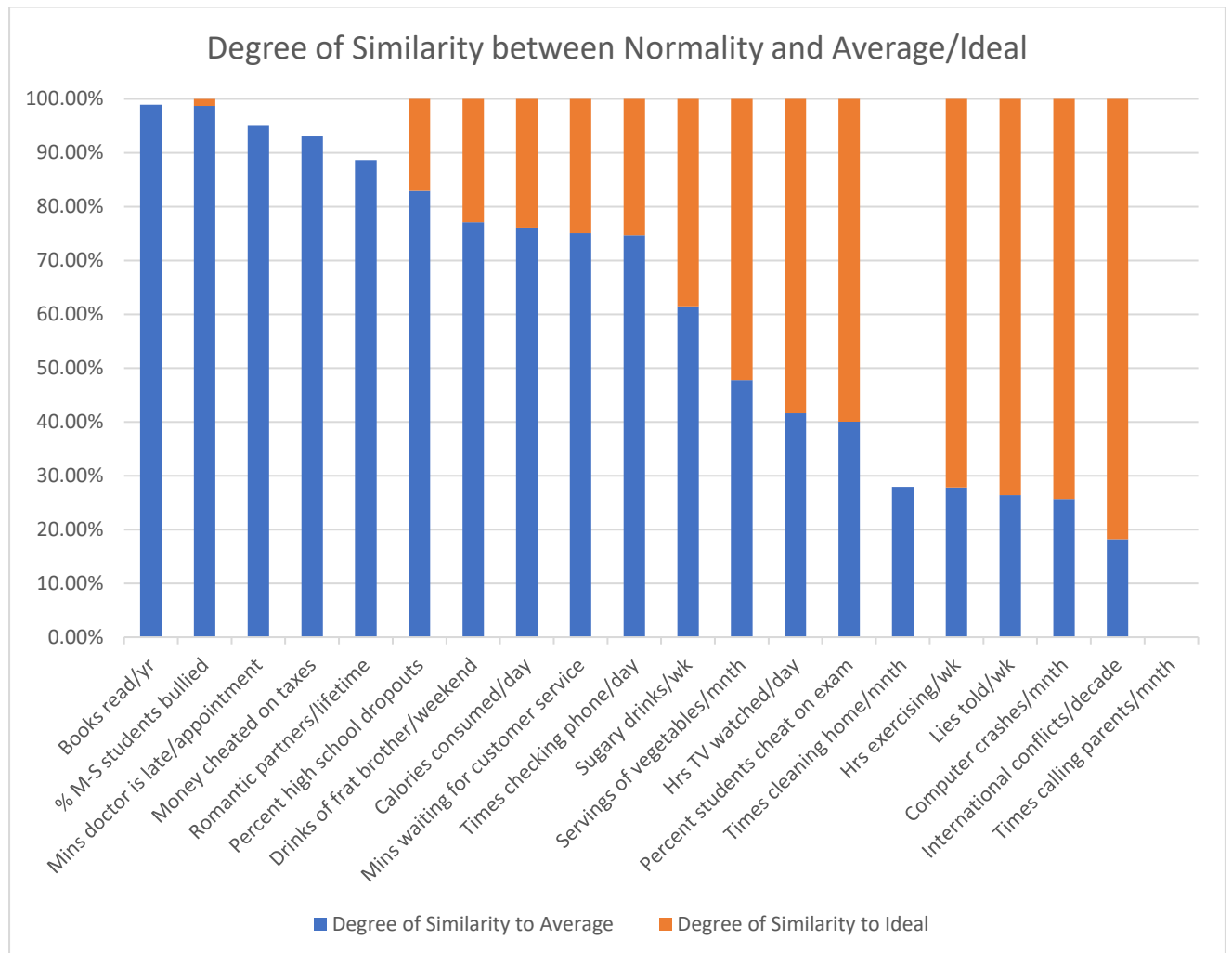


Figure 2

To make this more precise, we can categorise the domains based on the degrees of similarity that the normal values have to both the average and the ideal values. There are 6 distinct sets that the domains can be divided up into based on whether their normal values tend towards either the average or the ideal. These sets are as follows:

- Average Tracking – This set contains all the domains that have a 90-100% degree of similarity to the average. These are the domains where the normal value picked out by Bear & Knobe’s subjects almost exactly follows the average value.
- Average Leaning – The domains in this set all have a 60-90% similarity to the average. So here, the normal values are closer to the average values, but not to the extent found in the average tracking domains.
- True Intermediate – This set contains the domains that have a roughly equal degree of similarity with both the average and the ideal. Consequently, the domains in this set will have a 40-60% similarity to both the average *and* the ideal.
- Ideal Leaning – This is the mirror of the average leaning set. This set contains all the domains that have a 60-90% similarity with the ideal.
- Ideal Tracking – This is the mirror of the average tracking set. Here we find all the domains that have a 90-100% degree of similarity to the ideal.
- Non-Intermediate – The final set is for all the domains that don’t meet the conditions for membership in any of the other sets. The domains in this set will all have normal values that fall outside the average-ideal range by a significant amount. As such, they do not have the required degrees of similarity to either the average or the ideal for membership in one of the other sets.

We can now categorise each domain according to the rules set out above. If we do this then we get the following:

- **Average Tracking – 4 domains** (books read, students bullied, mins doctor is late, money cheated on taxes).
- **Average Leaning – 7 domains** (romantic partners, school dropouts, frat drinks, calories consumed, mins waiting for customer service, checking phone, sugary drinks).
- **True Intermediate – 3 domains** (servings of vegetables, watching TV, students cheating).
- **Ideal Leaning – 4 domains** (hours exercising, lies told, computer crashes, international conflicts).
- **Ideal Tracking – 0 domains.**

- **Non-intermediate – 2 domains** (cleaning home, calling parents).

The first thing to notice about this is that significantly more domains fall into the average leaning and tracking categories (11) than the ideal leaning and tracking categories (4). Furthermore, none of the domains meet the conditions for membership in the ideal tracking category. From this, we can conclude that overall, the normal values identified by Bear & Knobe's subjects tend to be much closer to the corresponding average values than the corresponding ideal values. This suggests that, in the context of the respondents for the first study, the average played a much more significant role in determining what was judged to be normal than the role played by the ideal. However, I don't think that this will be the case for all contexts, as we saw in the previous chapter with the homophobia example, there are some contexts that can plausibly be understood as being more normatively charged, so the ideal will have a much stronger role to play.

It may be tempting to conclude from this that normality is merely a statistical concept that can be understood along more or less the same lines as the average. As discussed further down, this view is already advocated for by some (Strössner 2015), so there already a degree of independent plausibility to it. However, the domains that are ideal leaning pose a clear counterexample to this, but they can be dismissed as merely being the result of some kind of bias. Specifically, it could be argued that these domains only generate normality judgements that lean more towards the ideal because they are more normatively loaded. That is, domains like 'Hours exercising per week', 'Lies told per week', 'International conflicts/decade', and 'Computer crashes per month' generate very strong feelings about what ought to be the case. Culturally we place a lot of emphasis on health, honesty, and peace, which explains why the first three domains were ideal leaning. In the case of computer crashes, the fact that this

domain is ideal leaning can be explained by the importance that computers have for our careers and the corresponding need to have them function correctly. What this means is that the true value for the normal number of computer crashes should be closer to the average than what the subjects judged it to be. That they in fact judged it to be closer than the ideal is a result of the subjects having strong feelings about what ought to be the case. A very similar story can also be told for every other domain that leans more towards the ideal than the average.

Unfortunately, it does not seem that this effect is systematic, as many domains that should also elicit strong normative feelings are not ideal-leaning at all. The domains, 'Percent middle school students bullied' leans strongly towards the average, despite concerning a morally charged topic. This may be explained by the fact that middle-school bullying is not likely to be a significant concern for adult test subjects. The same cannot be said of 'Money cheated on taxes', which also skews heavily towards the average. In this case, I do not think the answer is particularly informative without data on what the subjects actually think about people who cheat on their taxes. If they do not see it as a strong moral bad and think of it as something they would do if they had the opportunity, then that may explain why their answers lean more towards the average than the ideal. The root of the issue here is that these results represent the subject's immediate reactions to some fairly broad questions, which will invariably result in answers that are influenced by a wide-range of factors. The only remedy to this is to perform more studies with the aim of generating increasingly fine-grained data that will help us to identify the specific factors that go in to determining our judgements of normal values.

Further difficulties arise when we apply my methodology to the results of the second and third studies. For study two, which focused on testing the subjects' intuitions about *standard* amounts, we get the following degrees of similarity:

Domain	Degree of Similarity Average	of to Degree of Similarity to Ideal
Mins doctor is late/appointment	99.93%	0.00%
Books read/yr	99.56%	0.00%
Romantic partners/lifetime	99.47%	0.53%
Servings of vegetables/mnth	97.78%	2.22%
Percent high school dropouts	97.05%	2.95%
Money cheated on taxes	95.67%	0.00%
International conflicts/decade	95.09%	4.91%
Lies told/wk	94.33%	5.67%
Mins waiting for customer service	94.28%	5.72%
Hrs exercising/wk	93.81%	6.19%
Sugary drinks/wk	93.33%	0.00%
Percent students cheat on exam	93.00%	7.00%
Times cleaning home/mnth	89.83%	0.00%
Hrs TV watched/day	89.76%	10.24%
Times checking phone/day	86.40%	13.60%
Percent middle school students bullied	84.61%	15.39%
Computer crashes/mnth	77.57%	22.43%
Drinks of frat brother/weekend	72.43%	27.57%

Calories consumed/day	62.21%	37.79%
Times calling parents/mnth	0.00%	0.00%

Figure 3

And plotted on a graph:

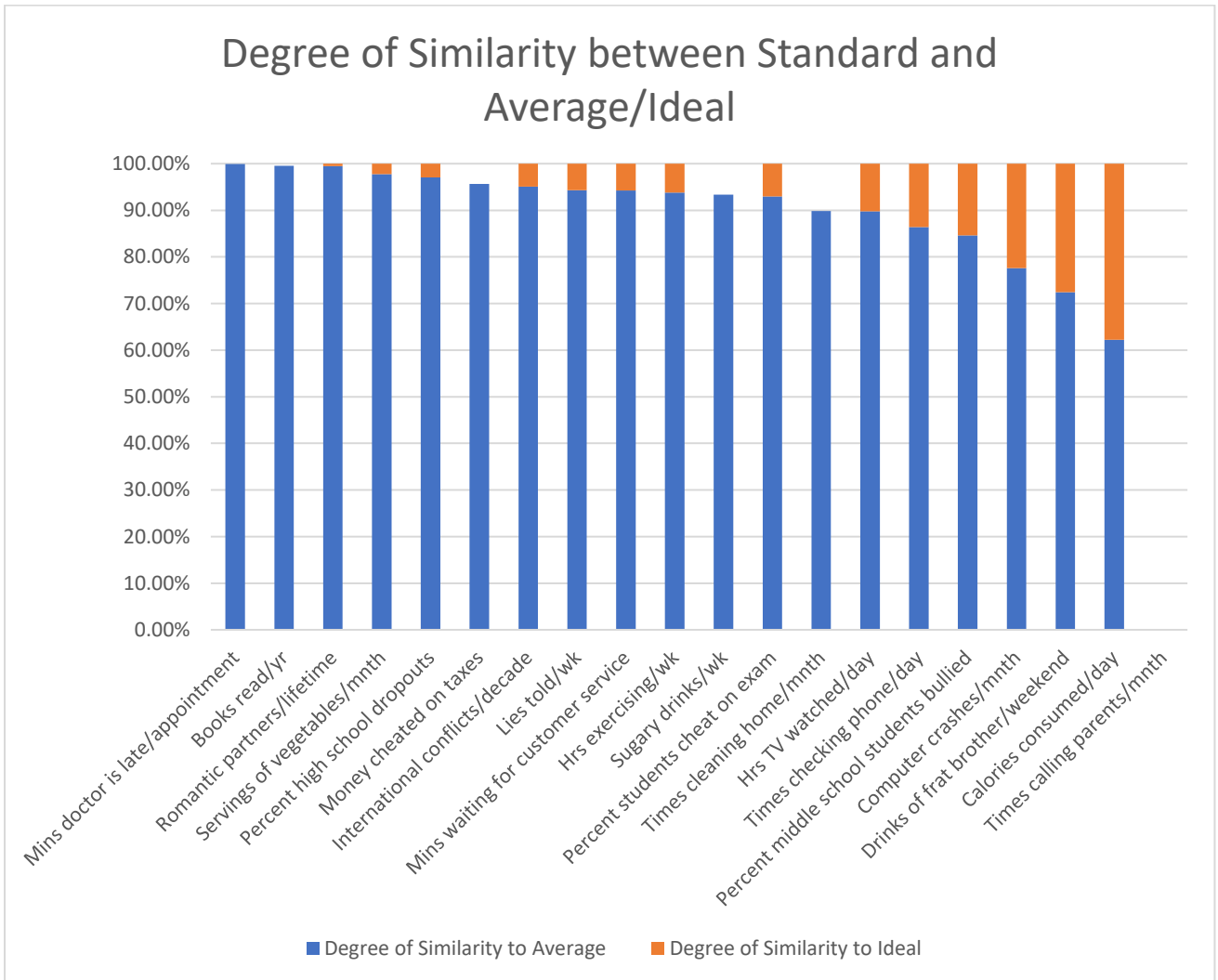


Figure 4

When comparing these results to those of the first study, it is immediately apparent that the domains had a much higher degree of similarity to the average than they did before. Here twelve domains are average tracking, and a further seven are average leaning. This is plausibly

a result of the fact that the question Bear & Knobe presented their subjects with referred to ‘large’ and ‘small’ amounts, which prompted the subjects to think more mathematically and less about values, resulting in them being pushed more towards the average.

Now compare this with the results from the third study. Recall that this test involved subjects rating the extent to which they agreed that something was a good/paradigm/prototypical example of a given category. For the sake of readability, I will present the results for just three out of the eight categories. Firstly, we can examine the results from those who were asked if what they read was a ‘good’ example:

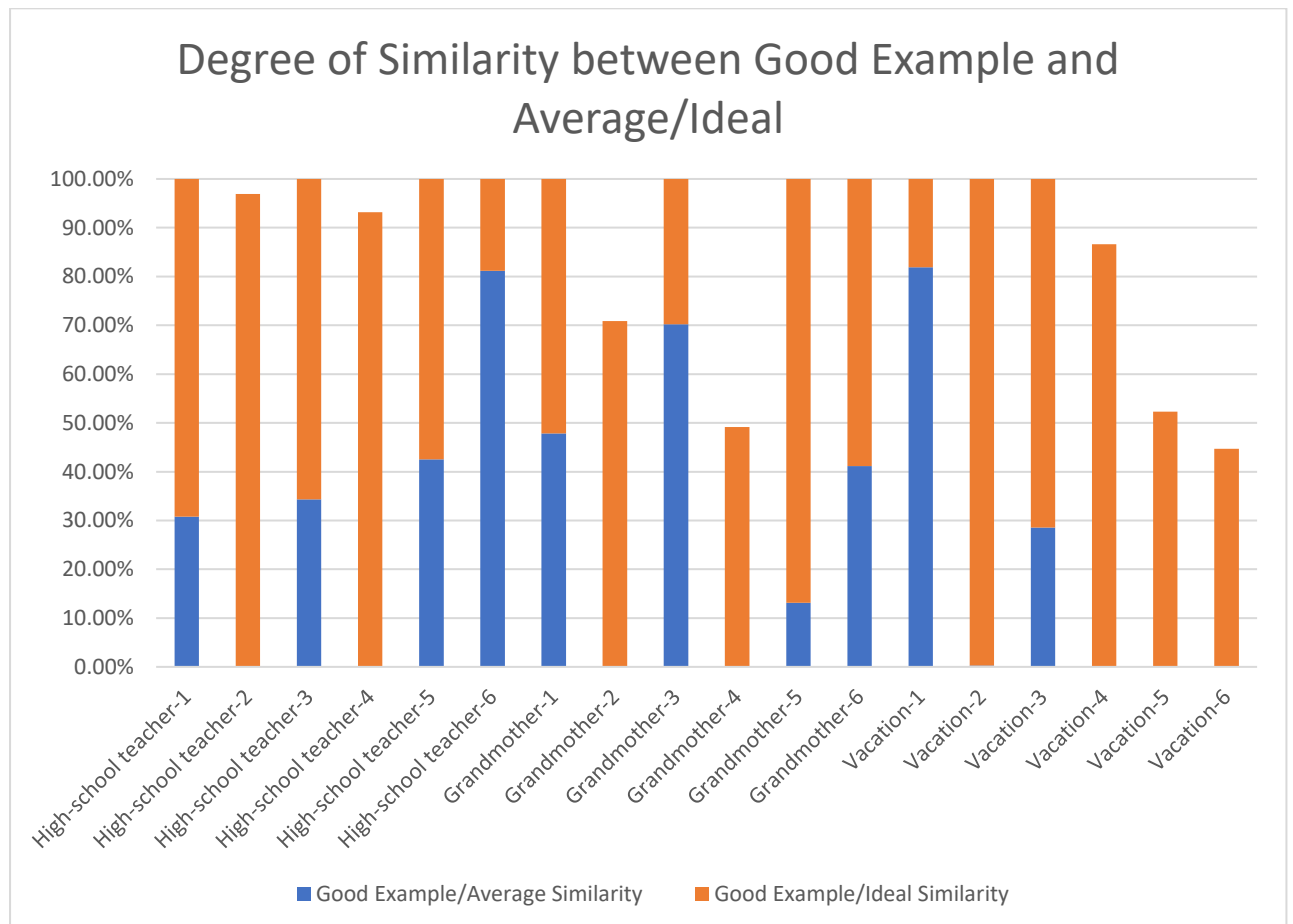


Figure 5

Just from looking at this we can see that, unlike the results from the first two studies, most of the domains lean more towards the ideal. Out of the eighteen results shown here, three of them are ideal tracking and six of them are ideal leaning. In contrast, only three of these results are average leaning and none are average tracking. So clearly the ideal is exerting a very high degree of pull on these results. This fact becomes even more significant when we examine the results from the subjects who were asked about 'paradigm' examples (Figure 6) and 'prototypical' examples (Figure 7):

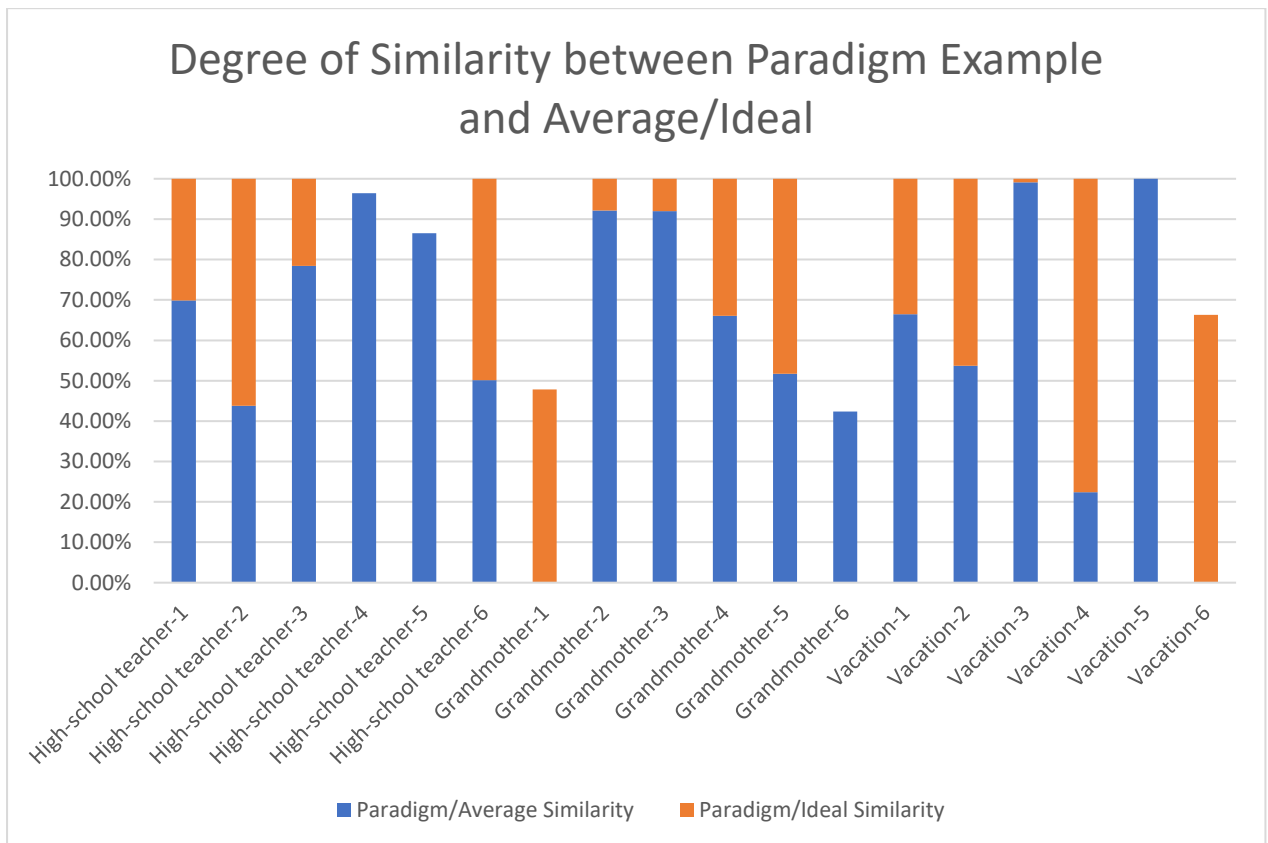


Figure 6

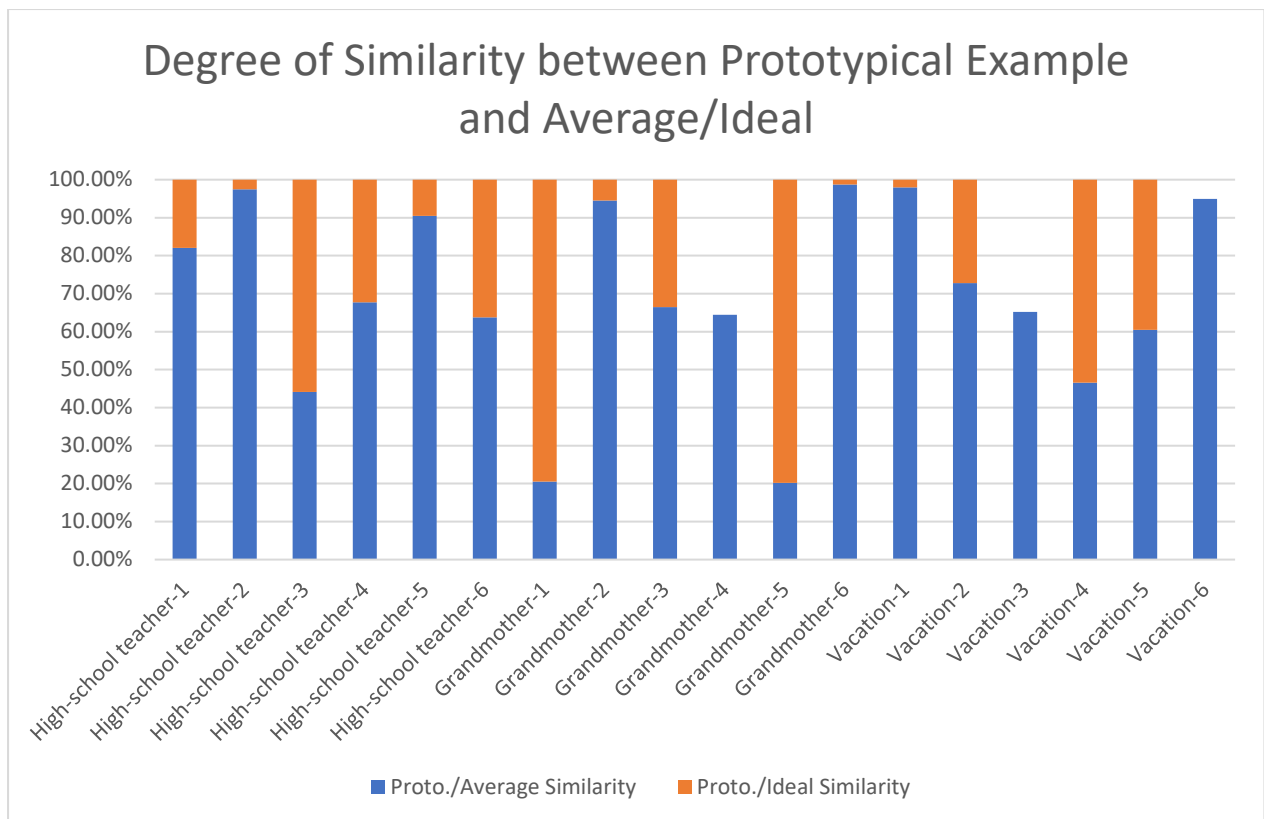


Figure 7

With these two sets of results, we have a return to the pattern of most of them being either average leaning or average tracking (although they both still include some judgements that are ideal leaning). This indicates that using the word ‘good’ instead of either ‘standard’, ‘paradigm’ or ‘prototypical’ leads to normality judgements that lean more towards the ideal than the average. More generally, it seems that the way that we frame the question can have quite a significant effect on whether or not the resulting normality judgement leans more towards the average or towards the ideal.

Let’s return to the possible view I raised earlier that attempts to explain normality in terms of averages and dismisses any ideal leaning judgement as being the result of a bias that is triggered in normatively charged domains. There seems to be two ways for such a view to

incorporate these findings about the degrees of similarity in studies two and three. One is to extend the bias explanation for the ideal leaning domains from the first study to cover the ideal leaning domains in all three studies. The argument here would be that the use of the word 'good' in the question triggered the bias due to being a normatively charged term. 'Good' is a term that is closely connected to morality and desirability. Therefore, its use in one of the questions for the third study prompted the subjects to treat their judgement as being more normatively charged than it would have been otherwise. As a result, their judgements leaned towards the more normative ideal and away from the non-normative average.

An alternative explanation would be to challenge Bear & Knobe's methodology by arguing that we cannot infer anything about the nature of normality from the answers people give to the question of whether or not something is a 'good' example of a category: the concepts of goodness and normality are not that closely related. Asking if a thing counts as a 'good example' of something will necessarily introduce normative considerations into the process of deciding on an answer. Since the question of whether or not normative considerations factor into competent judgements of normality is precisely what is at issue, it would be question-begging to claim that the answers to the 'good' question are judgements of normality. As a result, the judgements made by the subjects who were presented with the 'good' question should not be considered judgements of normality and thus don't need to be accommodated by an account of normality.

Both these possible explanations have their advantages. The first provides a unified explanation for all the instances where the ideal appears to strongly influence our normality judgements, while the second raises an important objection to Bear & Knobe's methodology.

What they both have in common though is that they suggest an account of normality that appeals solely to the average: where normality is reduced to purely statistical facts. On such an account, any normality judgement that is even slightly influenced by the ideal will be mistaken and possibly the result of some bias.

This kind of account does provide an explanation for the data presented by Bear & Knobe. However, we may want to question why this phenomenon of people's normality judgements leaning more towards the ideal – in domains/contexts that are normatively charged – should be considered a bias, rather than an instance of the concept of normality being competently deployed. Why should we assume that our normality judgments are only ever free from bias when they happen to line up with the average? The only advantages such an account would seem to possess is that it would be far simpler than an account that incorporates the ideal into the truth-conditions of normality judgements, and it would also allow us to easily explain why people tend to appeal to statistical facts to justify their claims about what is normal. Yet these advantages disappear when we think about some of the cases where 'normal' is used to make a normative claim that does not obviously seem to be the result of a bias. For example, in the previous chapter we had the statement 'it's not normal to lie to your partner'. This statement seems to be doing more than just stating a statistical fact, as it would still seem true even if it was the case that most people lied to their partners (and even if they were not just white lies). This normative feature of some normality statements will prove difficult to explain on a purely statistical account of normality.

This is not to say that it is impossible to provide such an account, but what it does mean is that there would need to be some story given for why many normality statements seem to

have this normative feature. An example of an account where the truth-conditions for normality statements are reduced to mere statistical facts is the one offered by Strössner in her 'Normality and Majority: Towards a Statistical Understanding of Normality Statements' (2015). There she addresses the apparent normativity of normality statements in a footnote, where she states that 'the view that normality is essentially normative is a misunderstanding that is caused by the appearance of the words' (2015, 795). She recognises that 'normal' can be used in normative contexts but argues that this is also true of many words that we don't generally understand as being normative. This is further supported by the fact that many normality statements appear to not be normative. Strössner concludes that we should try to analyse what 'normal' means outside of these contexts prior to any considerations about any normative uses of the term.

Strössner does not elaborate on this claim that the view that apparent normativity of normality is just a misunderstanding, so it is difficult to assess how plausible this claim is. However, what I think should be clear by now is that there is no reason to suppose that it will be straightforward to provide a statistical account of normality that captures the varying degrees of similarity to both the average and the ideal found in Bear & Knobe's results. This suggests that we may want to instead consider an account of normality that appeals to both the average and the ideal in explaining why something is or is not normal.

As indicated in the previous chapter, an analysis of the term 'large' provides a useful analogy for how we might want to go about providing such an account. Intuitively, we may want to say that a thing's being large is directly tied to its volume, since the more space something occupies, the more appropriate it is to describe that thing as 'large'. Yet as we saw with the

tunnel case, there are contexts where the rightness of calling something large is not determined purely by that thing's volume. In that case, the train had a greater volume than the bus, but it is more appropriate to say that the bus is 'too large' to go into the tunnel because it is taller than the train. The way to understand this is to hold that a thing's largeness is determined across three dimensions: height, width, and length. In most contexts, these three dimensions are all equally relevant when considering whether or not something is large. This is why volume and largeness tend to go together. However, as in the tunnel case, there are some contexts where certain dimensions are given more weight than others. The dimensions of height and width are more relevant to the assessment of whether or not something is too large to fit in the tunnel. As a result, the train, whose volume mostly comes from its length, is not judged to be large in this context. Meanwhile, the fact that the bus is taller than the train ends up being more relevant to its being appropriately described as large than its volume, which is far less than that of the train, due to it not being as long. So, with 'large' we have a term where the appropriateness of its use is determined across three dimensions, where those dimensions are given different weightings depending on certain features of the context within which the term is used.

How might this be applied to 'normal'? Well, from what we have said so far we can hold that the truth-values of normality statements are determined across two dimensions: proximity to the average and proximity to the ideal. This will best be captured by mapping these two dimensions onto two different axes, allowing us to construct a graphical representation of normality. This graphical representation should, when properly developed, capture what it means for something to be normal. A good place to start is to try to pin down exactly what we mean when we talk of the 'average' and the 'ideal', as this will help us to understand

precisely what the two axes will be representing. To this end, we should now turn to the second of my three key points.

(2) The Ideal is Not Ideal

Returning to the results of Bear & Knobe's first study, the values given for one particular domain stand out: 'Percent of middle school students bullied'. Intuitively we may want to say that the ideal percentage would be 0%. After all bullying is something that is almost universally considered to be wrong, and so the best possible state of affairs would be one where there was no bullying. Yet the mean response to this question about the ideal percentage of students being bullied was 2.31%. This suggests that several of the participants gave a higher percentage than 0 as their answer, contrary to what we might have expected to happen. We see a similar phenomenon occurring in the domains of 'Minutes the doctor is late per appointment', 'Money cheated on taxes', 'Percent of high school dropouts', 'Minutes waiting for customer service', 'Percent of students to cheat on their exams', and 'International conflicts per decade'. It seems to be intuitively plausible to say of all (or at least some) of these domains that their ideal values would be 0. That is, in an ideal world we would never have to wait for a doctor to see us or for customer service to respond. No one would ever cheat on their taxes or in their exams. No one would ever be bullied at school, and no one would ever drop out. And, most importantly, there would be no conflicts between nations. Yet the subjects consistently judged that the 'ideal' value for all of these domains was higher than 0. Bear & Knobe do not attempt to explain this fact; however, since my aim is to analyse 'normal' in a way that pitches 'normal' somewhere between 'average' and 'ideal', we need to understand what an 'ideal' quantity or percentage or value is.

My explanation for why the subjects so often gave answers to the 'ideal' questions that did not correspond to what would be the case in the best of all possible worlds is that this was not the world that the subjects were thinking about when the question was presented to them. That is, they did not understand the question to be asking them about what would be the case in an 'ideal' or perfect world. Rather, they understood it as asking them what is the best that we could reasonably hope for within these domains. In the case of bullying, they may have thought something like: 'we may not be able to get rid of all bullying, but what is the best that we can reasonably expect given what we know about middle-school aged children?'.

In thinking like this they are holding certain assumptions about the actual world fixed and considering what might be possible given those limitations. The source of these limiting assumptions can be somewhat varied, from the subject's own beliefs and experiences to the prevailing discourses around a particular topic. In the case of bullying, it may be that the assumptions come from their experiences being a middle-school child, from raising a child of their own, from depictions of bullying in media, or from somewhere else. All these sources can (individually or jointly) create the impression that bullying is so prevalent at school that it cannot be completely eradicated and the best we can hope for is that it be minimised. As a result, they can only respond to the question about the 'ideal' percentage of bullying with the lowest value that they think bullying levels could be reduced to, whilst still holding that it cannot, in practice, be stopped entirely. More generally, this process of coming to a judgement about the 'ideal' percentage of bullying provides us with a way of capturing the sense of 'ideal' that is relevant for normality. That is, the ideal for a specific domain is the best that we can hope for within that domain given certain limiting assumptions.

It may be that some philosophers would object to this sense of 'ideal' on the basis that it is a mistake to allow these limiting assumptions to affect our judgements about the ideal and, ultimately, the normal. The proponent of such a view would likely try to argue that the best-of-all-possible-worlds sense of 'ideal' is the only true sense of 'ideal', and the judgements made by Bear & Knobe's subjects are simply the result of a failure to challenge their own assumptions. In the case of bullying, the subjects have incorrectly assumed that bullying is an essential part of school life, and that assumption has led to them making a false judgement about the ideal. If the subjects were all epistemically virtuous agents, they would have challenged that assumption and eventually discarded it. This would then have led to them making a true judgement. Indeed, we might even go further and say that it is potentially harmful to not challenge those assumptions, as it can lead to us being satisfied with imperfect solutions to problems, where better solutions are available. With bullying, this might be seen in teachers being unwilling to work at completely eliminating bullying in situations where it has already been substantially reduced. It should be easy to see how this might be generalised to individuals failing to fully eliminate as many wrongs as they possibly could, or to provide as much good as they possibly could, and on a big enough scale this could allow huge systemic harms to occur whilst being largely ignored³⁹.

³⁹ To give a topical example, a common argument made against lockdowns as a measure for reducing the rate of COVID-19 infections is that the disease is only lethal in less than 1% of cases, and so implementing a lockdown simply isn't worth the cost. In the terms that we have been using so far we can understand this as their judging that the 'ideal' fatality rate of a highly contagious disease is around 1% of those it infects (given the limiting assumption that the disease will result in death for some patients regardless of what we do), and that any disease that either meets or exceeds (is lower than) this ideal rate does not provide a sufficient reason to endure the costs of a lockdown. So, as long as the fatality rate stays at or below 1%, a lockdown is unwarranted as a means of preventing the spread. After all, if you catch it, it's very unlikely to kill you. The problem with this line of thinking is, of course, that even 1% of all the people who will likely be infected by COVID is an atrociously large

This argument raises an important point: we should be careful about the source and veracity of the limiting assumptions that are involved in our construction of the ideal. However, I think that this issue is distinct from the question of whether or not there can be a proper sense of ‘ideal’ that includes limiting assumptions. It can indeed be extremely harmful to falsely believe that things are as good as they can get, but the problem seems to stem from the content of the assumptions, rather than the mere fact that they are there. It does not seem obvious to me that the epistemically virtuous agent should challenge and discard all their limiting assumptions. For a start, there may well be cases where this is not even possible. Yet even if we concede that this is something we *can* do, there may be reasons to think that it is not something that we *should* do.

The reason why we might want to keep our limiting assumptions lies with the function of the more constrained sense of ‘ideal’ that I have been describing here. This function can be captured by the aphorism that ‘the perfect is the enemy of the good’. The point of this aphorism is to highlight the fact that fixating on having things be the best they can feel overwhelming and demanding to the point where we are prevented from implementing small-scale changes that will be of considerable benefit to people. If our only ideals are completely unconstrained, they will be extremely difficult to achieve, and attempting to do so will likely be both extremely impractical and extremely demoralising. Having an ideal that is constrained by certain (unproblematic) limiting assumptions will allow us to develop goals that are more achievable and less likely to make us burn out trying to reach them.

number of people to sentence to a preventable death (not to mention that there may be other long-term harms that come from being infected). This remains true even if the fatality rate ends up being much lower, say 0.01%.

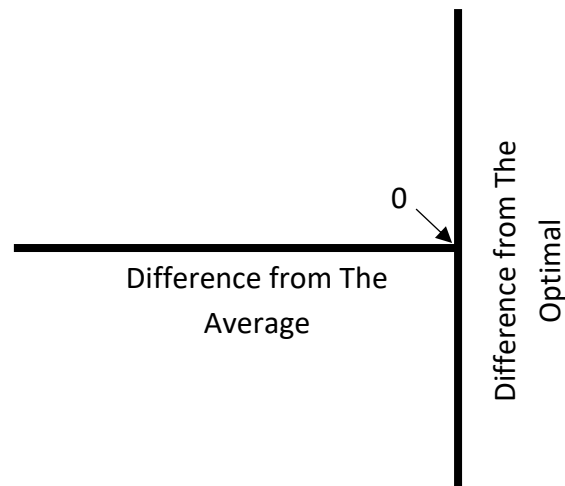
What this suggests is that there are advantages to both the constrained and unconstrained senses of 'ideal'. The unconstrained ideal allows us to clearly identify a direction for our efforts and ensures that we are not overlooking any possible courses of action that would be even more beneficial than what we are already doing. Meanwhile, the constrained ideal provides us with achievable goals that can be met on the way to our unconstrained ideals. Therefore, we have reason to maintain both senses of the term.

For my purposes, it will help to be more precise about which sense of 'ideal' I will be using. As such, I will henceforth refer to the constrained ideal as the 'optimal', as it has a similar meaning to 'ideal', but also suggests a recognition of the limiting factors. I will reserve the use of 'ideal' to refer to the unconstrained ideal. So, to state it more formally, something is ideal if it is the case in the best of all possible worlds, while something is optimal if it is the case in the best of all worlds where the limiting assumptions are held fixed. The ideal is thus the best that is possible, while the optimal is the best that we can reasonably hope for.

To bring this back to my account of 'normal', we can now say that one of the axes in the graphical representation will stand for the similarity/proximity to the optimal. An effect of this is that this allows for the possibility that something may be more ideal than the optimal. For example, if we are thinking about the domain of philanthropy, we might think that the optimal amount of your income to give to charity is 50%. So, someone who gives more than that has exceeded that optimal. To represent this, we can understand the optimal as being at the zero-point on the y-axis, where something that exceeds the optimal will be placed above zero and something that fails to meet the optimal will be below zero. The further something

is from the optimal in a positive direction, the higher it will be. The further something is from the optimal in a negative direction, the lower it will be.

Likewise, the x-axis will represent the distance that something is from the average. Yet here it does not make sense to say that something is more average than the average, since a thing is either average or it is not. Of course, if we are discussing amounts, something can exceed the average, but that does not mean it is *more* than the average. To see what I mean, consider the fact that the average number of cups of coffee for one person to drink in a day is 3. Now imagine that I drink 1 cup per day, and you drink 5. If we remove any talk of what is ideal, then it is difficult to draw any meaningful distinctions between how my coffee habits and your coffee habit, differ from the average, since they both do so by exactly 2 points. There is value in exceeding the optimal that is not present for exceeding the average, as with the average it does not matter which side of it you are on. To represent this graphically, the x-axis will stand for the points of difference from the ideal in both a positive and negative direction, with the average itself sitting on the zero-point. Putting both axes together will yield the following:



Here something that is both the average and the ideal will be placed at 0. Anything that differs from the average will fall in the negative part of the x-axis, and the further it is from the average, the further down the axis it will go. It should be noted here that I don't have a specific kind of 'average' in mind when it comes to the mode/median/mean distinction. Rather I am using 'average' to mean a more a general sense of something that is representative of a whole class, or that could be appropriately described as typical. With the optimal, anything that instantiates the values present in the optimal more than the optimal itself will be higher on the y-axis, while anything that falls short of the optimal will fall lower down on the y-axis.

To demonstrate this, suppose that 3 cups of coffee per day is not only the average, but it is also the optimal with drinking more cups being desirable. If we were to then plot every (whole) number between 0 and 10 of cups of coffee drunk in a day on the graph, it would be as follows:

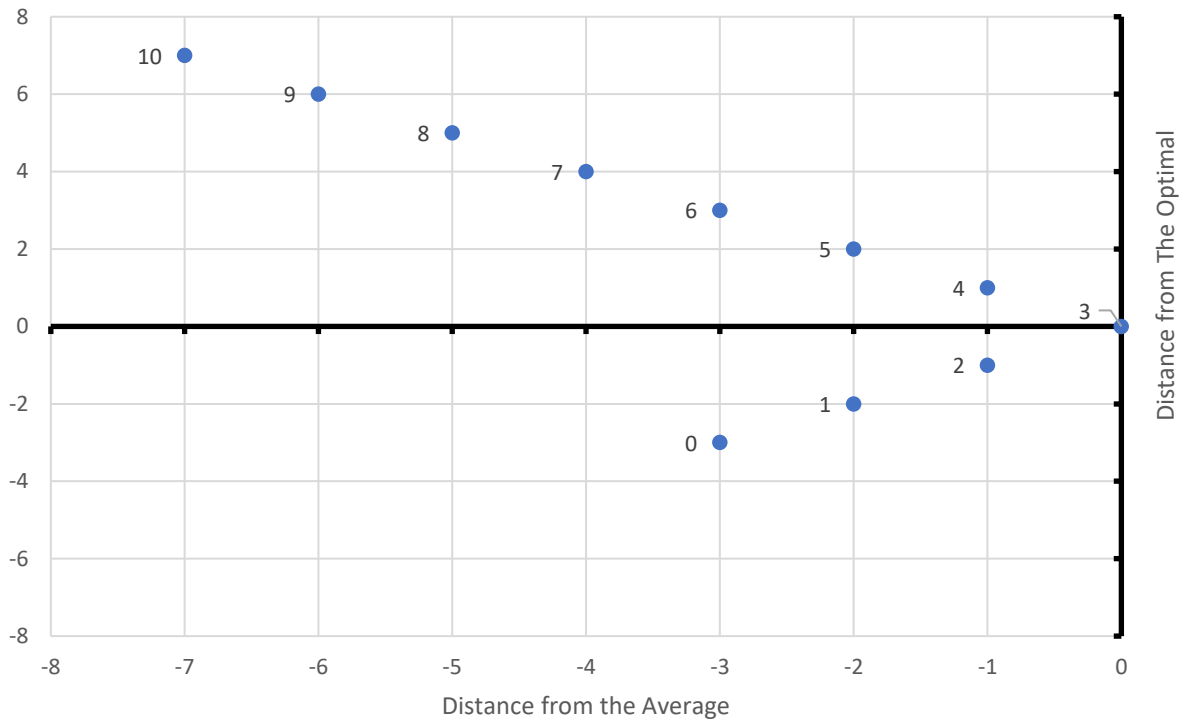


Figure 8

Now if we modify it so that 1 is the optimal and drinking fewer cups of coffee is desirable, we get the following:

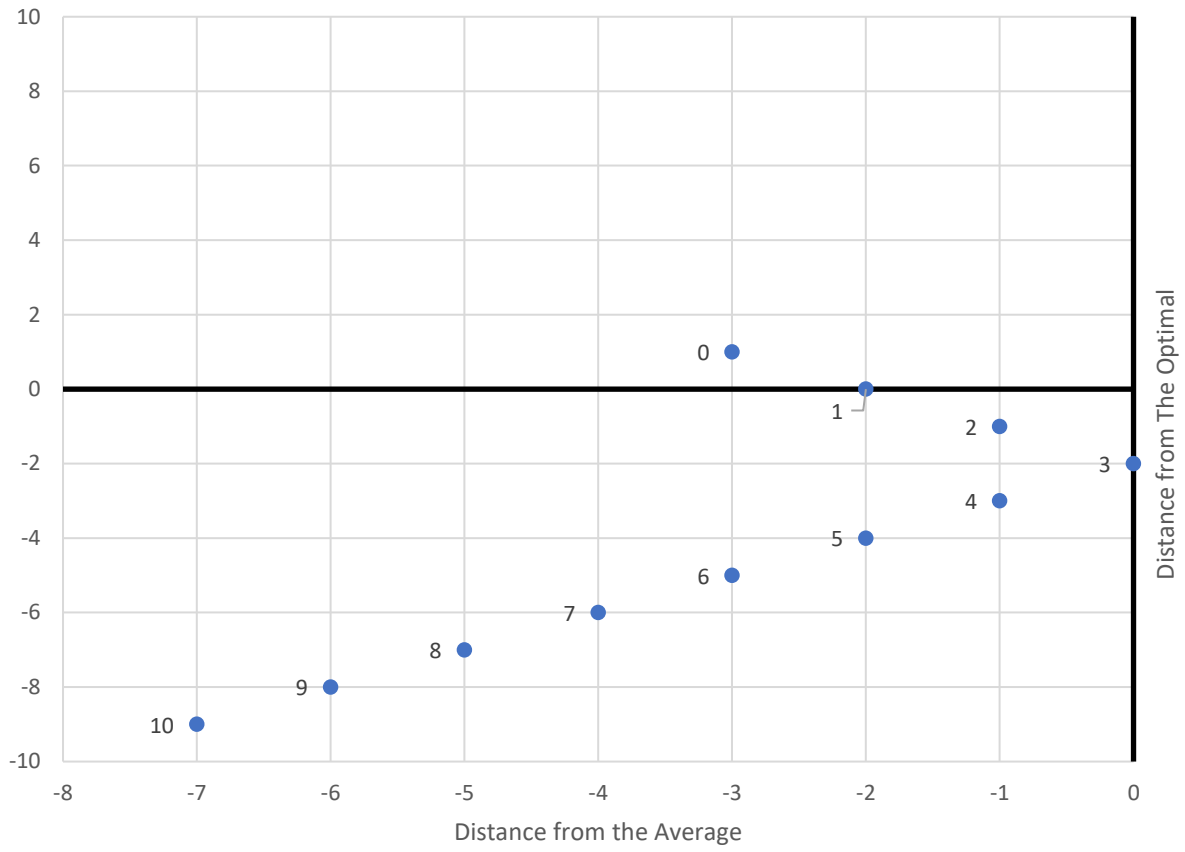


Figure 9

Finally, let's keep the ideal at 1, but change the average to 7:

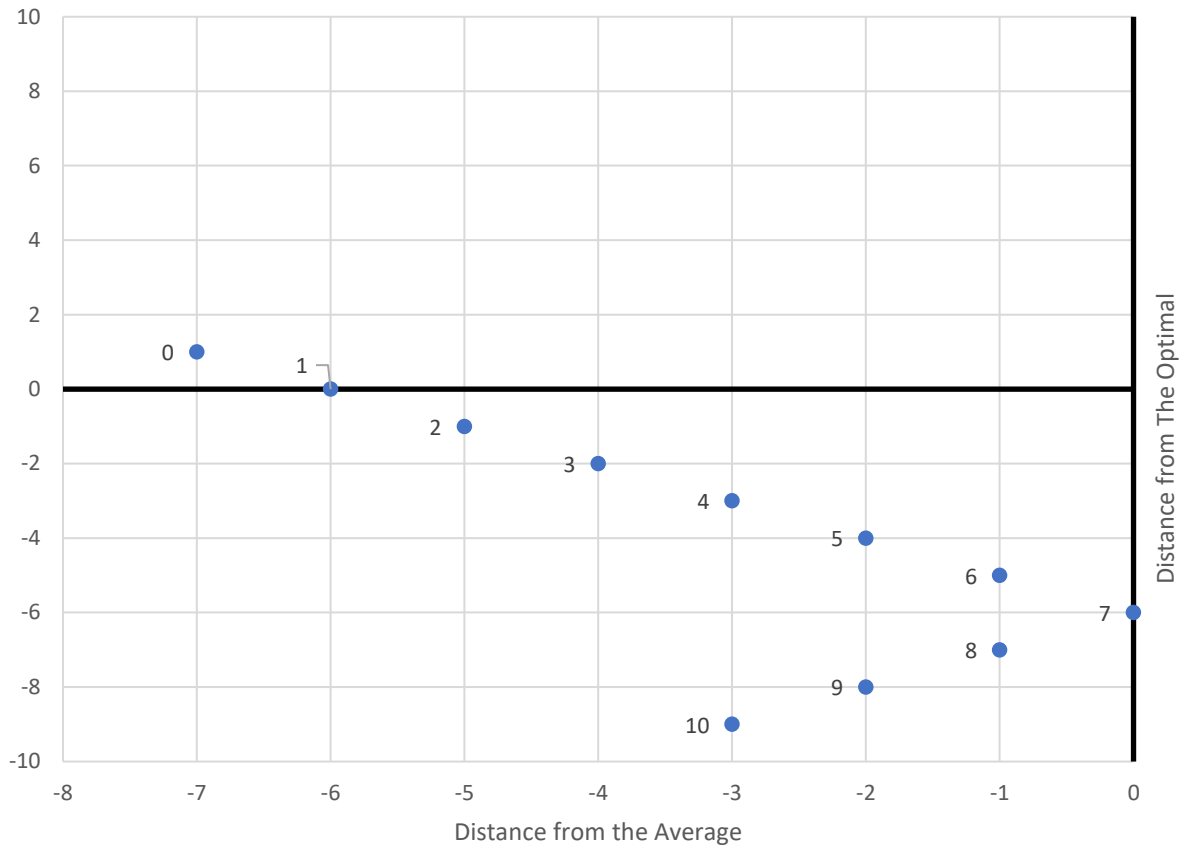


Figure 10

Of course, all this really provides us is a set of points on a graph. The question we need to now be asking is which points get to count as normal. I argue that the points within a certain region will count as normal, and the bounds of that region are determined by the context that the normality judgement is made in, which brings me onto my third key point.

(3) The Context-Sensitivity of Normality Judgements

The final point that I wish to make about our normality judgements is that, if we accept the points I have offered so far, these judgements will likely vary across different contexts. On the framework I have provided, there are three different contextual factors that could and most

likely would lead to variations in our judgements of normality across different contexts. These three factors are:

- i) Differences in which reference class is salient.
- ii) Differences in the limiting assumptions used for constructing the optimal.
- iii) Whether or not the domain is normatively charged.

In the case of (i), context determines which reference class we are appealing to when our normality judgements are made. This not only affects what is normal, but also what is average and optimal. So, something that may be average/optimal for, say, a teacher, may not be average/optimal for people in general. In cases where a normality judgement is being made, the same reference class will be salient for the average, the optimal, and the normal.

With (ii), the limiting assumptions that are involved in the construction of the optimal may change across contexts. As an example, the limiting assumptions held by a group of anti-bullying campaigners may be different from the assumptions held by a teacher at a school where bullying is rife. As a result, the two are likely to reach very different conclusions about what the optimal level of bullying is, and consequently they will also have very different views about what a normal level of bullying is.

Finally, (iii) determines what the region is for whether or not something gets to count as normal. Now as we have seen with the results from Bear & Knobe's studies, most neutral contexts seem to produce judgements that are average leaning. That is, the average in neutral contexts, the average carries more weight for determining whether or not something is normal. We can represent this graphically by highlighting the region within which something

will count as normal. For illustrative purposes, let's reuse the first coffee drinking graph (Figure 8):

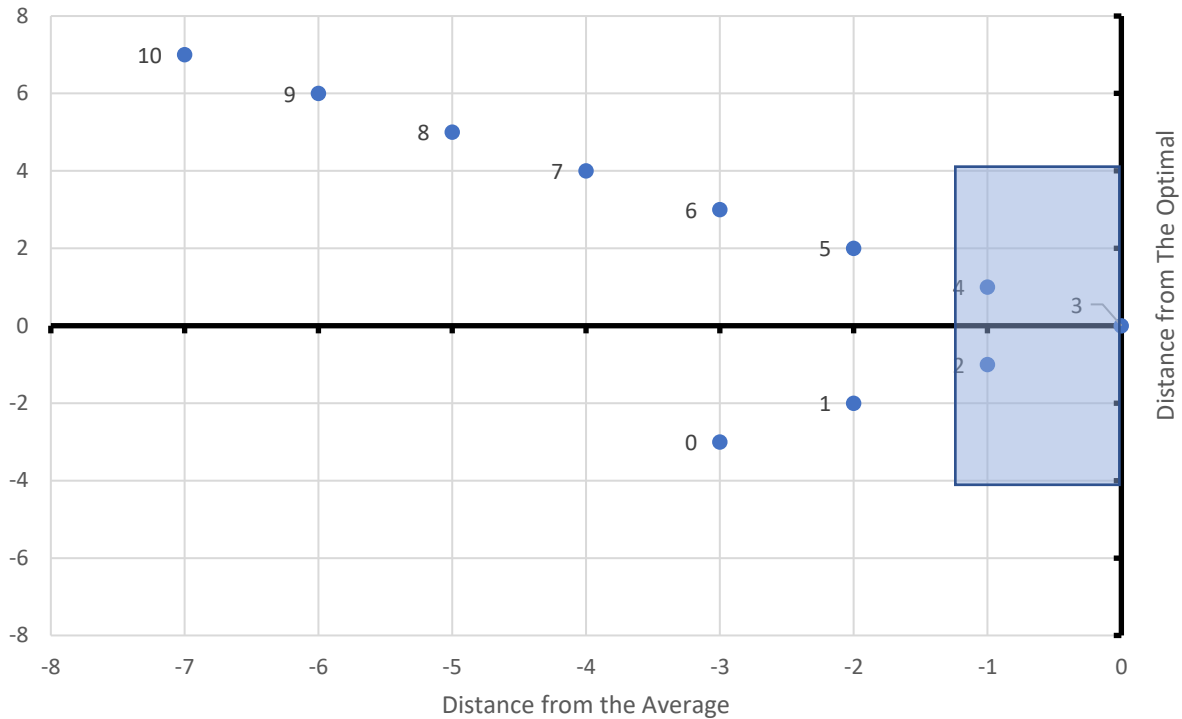


Figure 11

The fact that the average carries more weight is indicated by the fact that the width of the normal region is considerably smaller than its height, meaning that in order to be normal it is much more important to be closer to the average than it is to be close to the optimal. As a result, it is normal to drink 2, 3, or 4 cups of coffee a day, but anything more or less than this is considered abnormal.

However, in a more normatively charged context, the region shifts to provide more weight to the optimal. To see this, suppose that it is discovered that drinking multiple cups of coffee per day is extremely unhealthy, leading to the situation represented by Figure 9, where it is

optimal to only drink 1 cup of coffee per day (with the limiting assumption being that it is very difficult to quit completely). This research about the health impacts of coffee becomes widely circulated, and people who drink a lot of coffee become stigmatised for it. This results in the domain of coffee drinking becoming normatively charged. In this case, the graph would look like the following:

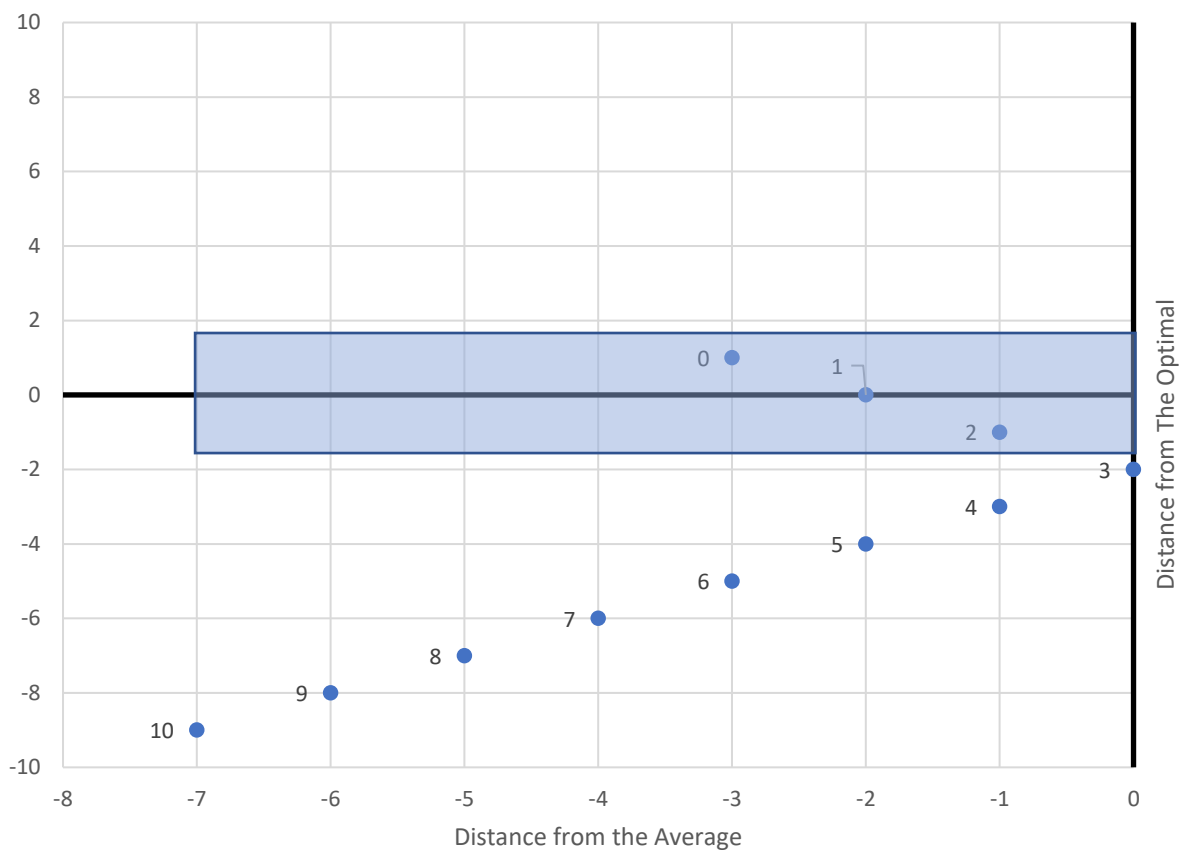


Figure 12

Since the domain is normatively charged, the optimal is given more weight. This is represented by the normal region having its width be greater than its height, so that the distance from the optimal is more likely to make the difference between something's being or not being normal. In this case, normal means drinking 0, 1, or 2 cups of coffee a day.

It is in this way that our judgements of normality can respond to differences in the contexts where those judgements are made. It is important to reiterate that I do not just take this account to be an explanation of how we arrive at our judgements of normality, but also an account of the truth conditions of those judgements. Unlike Strössner I do not take this contextual variance to occur at the level of pragmatics, nor do I think that this is all the result of some cognitive bias. I fully intend the preceding analysis to be an account of what it means to describe something as 'normal'.

Having now provided an account of normality, it is time to make use of it to help us in developing a normative account of causation.

2. Returning to Causation

Now that I have provided my account of normality, we are in a position to see how it can be applied to causation and then be used to solve the puzzle cases presented in Chapter 2. In Chapter 3 I demonstrated how the problem that these puzzle cases raise cannot be simply dismissed by appealing to pragmatics and biases, and that we needed to take seriously the notion that our intuitions about these puzzle cases were the result of a competent application of the concept of causation. However, as I showed in Chapter 4, many of the existing attempts to deliver on this are hampered by the fact that they do not offer any sort of analysis on what it means for something to be 'normal'. Having now provided such an analysis, I am now in a position to address this gap in the literature by examining how my analysis applies to causation.

As a starting point, I shall draw upon Menzies'/Hart & Honoré's idea that a cause is something that makes the difference between the normal course of events for a causal system and the

actual course of events for that system. That is, when we reason causally, we are trying to identify the moment where the actual course of events diverged from a hypothetical normal course of events. The challenge I raised against this account in Chapter 4 was that it does not specify what the normal course of events for a given system is. With my account of normality in place, we are now able to do so.

Recall that a key notion for Menzies' account was the idea that there is a 'normal' state of a causal system. To understand what this means, given my account of normality, we first need to understand what the average state of a causal system is, and what the optimal state is. The average state of a system is what happens most of the time in causal systems of the contextually salient reference class. The optimal state of a system is the best version of the system that we can hope for, given certain limiting assumptions. The normal state will be located in a region whose borders are determined by these other two states. In most ordinary contexts, the average state will be given more weight when it comes to shaping these borders, but in normatively charged contexts more weight will be placed on the optimal state. Note that my account will remain neutral on whether there can be more than one normal state of a system.

Now that we have specified what the normal course of events for a system is, we can specify a set of conditions for integrating this into Menzies' account of causation:

The statement 'c caused e' is true only if:

- 1) c and e both actually occur.

- II) e is an event that does not occur as part of any normal state of the causal system under consideration.⁴⁰
- III) If c had not occurred, then the system would have been in a normal state (so c makes the difference between events following a normal course and the actual course).

It is again worth mentioning that these are not intended to be a complete set of necessary and sufficient conditions for causation. Rather, they are meant to provide plausible constraints on what counterfactual scenarios are relevant when making and justifying our causal judgements.

It should also be noted that by embedding these conditions into an account of causation, the resulting account will be contextualist. This is because the conditions refer to the notion of a normal state of the causal system, and whether or not a particular state of the system is a normal state will depend on the context within which we are making our causal judgement. So, the same causal statement can be evaluated against two different states of the appropriate causal system – that is, two different counterfactual scenarios – in two different contexts of evaluation.

I will now use the analysis presented here to solve the puzzle cases presented in Chapter 2. This will serve to demonstrate how the account works in practice.

⁴⁰ It may be objected that this condition makes it impossible for an event that is part of the normal course to ever be caused. I will respond to this objection in the next chapter.

The Venusians and the Forest Fire

Let's start with the forest fire case. Recall that the puzzle concerned the occurrence of a lightning strike which resulted in a forest fire. We were then asked to consider the following two statements:

- 1) The lightning strike caused the forest fire.
- 2) The presence of oxygen caused the forest fire.

Statement (1) was acceptable in the context of a group of forest rangers but not in the context of the space-faring Venusian scientists. Meanwhile the opposite was the case for statement (2): it was unacceptable in the context of the forest rangers and acceptable in the context of the Venusians.

To see how my account can solve this puzzle, first consider the context of the forest rangers. The reason that they judge (1) to be acceptable is that it is true. It is true because it meets conditions (I)-(III).

Condition (I) is clearly satisfied because both the lightning strike and the forest fire actually occur. Condition (II) is satisfied because the forest fire does not occur in any normal state of this causal system, and condition (III) is met because the lightning strike is what makes the difference to whether or not the forest fire occurs. To see how, we can provide a graphical representation for the normality of the different states of the causal system in question. There are four possible states for the system to be in:

- S1) The lightning strike occurs and there is oxygen present. Consequently, the forest fire occurs. (This is the actual state of the causal system).

S2) The lightning strike does not occur and there is oxygen present. Consequently, the forest fire does not occur.

S3) The lightning strike occurs but there is no oxygen present. Consequently, the forest fire does not occur.

S4) The lightning strike does not occur and there is no oxygen present. Consequently, the forest fire does not occur.

S1 is the actual state of the system, so when the forest rangers are reasoning counterfactually about this case it will be one of the other states of the system that they end up appealing to. Here, the relevant counterfactual state will be S2 because that is the normal state of the system in this context. This then raises the question of *why* S2 is the normal state of the system. My account of normality can help us to answer this question.

To show why S2 is the normal state of the system in the context of the forest rangers, we first need a means of assessing the degree to which each of these worlds are average and optimal. The causal system itself is fairly simple, with only three binary variables that each represent one of the following:

- a) The occurrence/non-occurrence of the lightning strike.
- b) The presence/absence of oxygen.
- c) The occurrence/non-occurrence of the forest fire.

It will therefore be fairly easy to precisely model the degree to which each state of the causal system is both average and optimal. To do this, we can identify a value for each of these variables that is the average value for that variable, alongside a value that is the optimal value for that variable. These two values can be the same, but they may also be different. Which values we identify as both average and optimal will at least be partly determined by the context, which in this case is that of the forest rangers.

For (a), the non-occurrence of the lightning strike is the average value since lightning strikes are less common than the absence of lightning strikes. Given that we are here considering the context of the forest rangers, it also seems reasonable to claim that the optimal value of (a) is also the non-occurrence of the lightning strike, since lightning tends to be harmful and the non-occurrence of the lightning strike (at least at the location where it actually occurred) is something that we could reasonably hope for.

With variable (b), the average value would be the presence of oxygen, since the salient reference class for the forest rangers is Earth's surface, where oxygen is almost always present. This is also the optimal value for this variable since the presence of oxygen is also superior to the absence of oxygen.

Finally, the average value for variable (c) will be the non-occurrence of the forest fire since, as with the lightning strike, the non-occurrence of forest fires is more common than the occurrence of forest fires. Likewise, the non-occurrence of forest first seems preferable to the occurrence of forest fires, so that will be the optimal value for (c).

We can now assign each state of the system a rating out of three based on the degree to which they conform to the average state of the system (R_a), or the optimal state of the system (R_o), where 2 represents the greatest deviation from the average/optimal and 0 represents the least deviation:

$$S1) R_a = 2; R_o = 2$$

$$S2) R_a = 0; R_o = 0$$

$$S3) R_a = 2; R_o = 2$$

$$S4) R_a = 1; R_o = 1$$

These values can then be mapped to a normality graph:

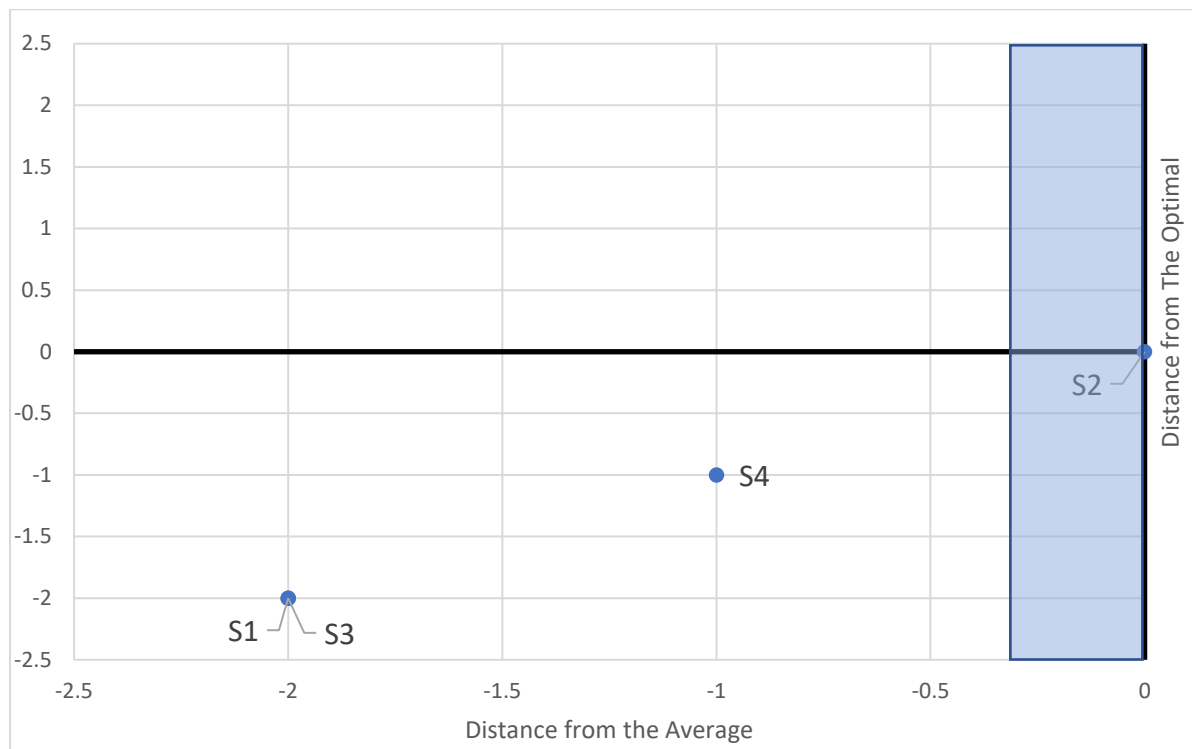


Figure 13

There's no reason to think that this domain is one which is particularly normatively charged, so I have drawn the bounds of the normal region to represent the fact that the average is given more weight here in determining what is normal. Here, there is clearly only one normal world, S2. In this world, there is no forest fire, and so condition (II) of my account is met. Condition (III) is also met because the lightning strike is the thing that makes the difference between the non-occurrence of the forest fire in S2 and its occurrence in the actual state, S1. Thus, the lightning strike meets all three conditions for causation within the context of the forest rangers. The presence of oxygen, however, does not meet Condition (III) because it does not make the difference between S2 – the normal state – and S1, the actual state, since in both S1 and S2 oxygen is present. Therefore, statement (1) is true, and statement (2) is

false within the context of the forest rangers. This explains our intuitions about the acceptability of these statements.

The fact that these intuitions differ in the context of the Venusians is a result of there being a difference in which counterfactual scenario is relevant for their causal judgements. Specifically, due to a difference in the values of (a) and (b) that are the average values for those variables, the normal state of the system in this context is S3. This is because, in *this* context, the salient reference class is one that includes the goings-on on Venus, where lightning strikes are incredibly common, and there is no oxygen. As such, the average value for (a) is the occurrence of the lightning strike, while the average value for (b) is the absence of oxygen. As a result, the possible states of the system will all have a different R_a (this is even assuming that the R_o s also remain the same). The new ratings will be:

S1) $R_a = 2$; $R_o = 2$

S2) $R_a = 2$; $R_o = 0$

S3) $R_a = 0$; $R_o = 2$

S4) $R_a = 1$; $R_o = 1$

This changes how the systems are modelled on the normality graph:

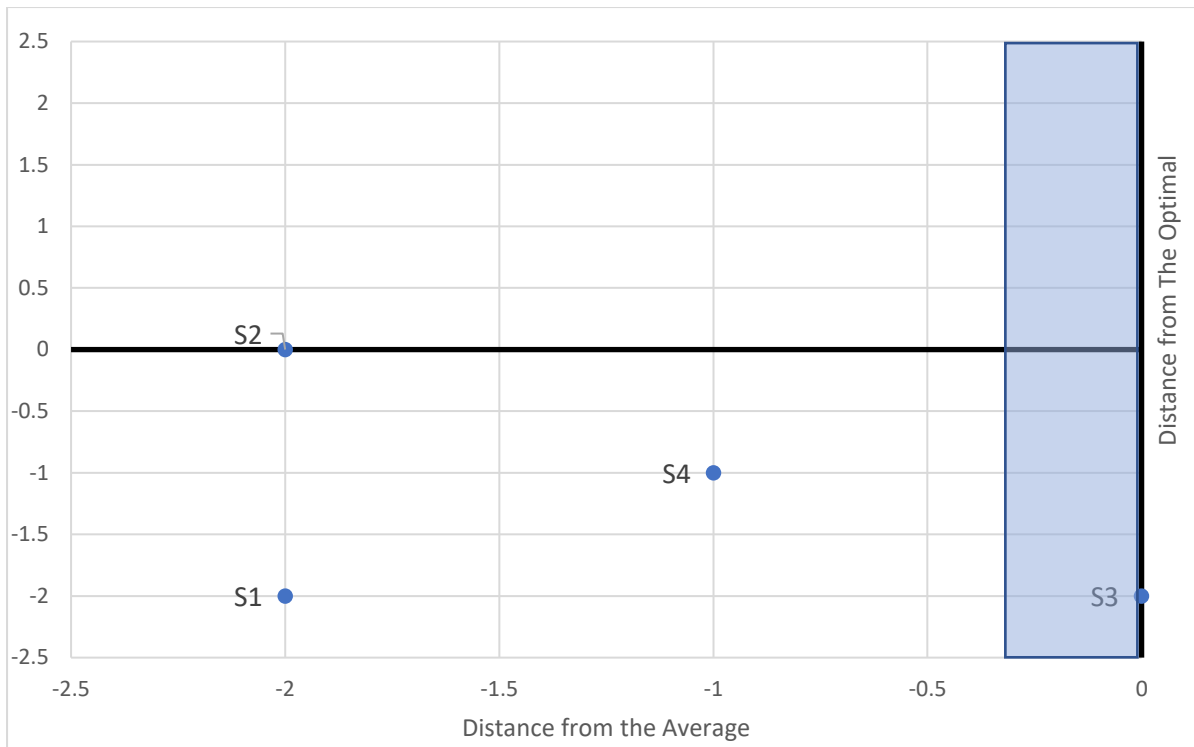


Figure 14

As we can see, S3 is the only normal state of the system in the context of the Venusians. So, in this context, statement (1) does not meet condition (III) because the lightning strike occurs in both the normal state (S3) and the actual state (S1), and so cannot make the difference between the two. Yet the presence of oxygen *does* make the difference between the two states, which means that statement (2) meets all three conditions for causation. Therefore, in the context of the Venusians, statement (2) is true, while statement (1) is false. This explains why our intuitions about the acceptability of these statements differs from that of the forest rangers, thereby solving the puzzle.

The Gardener and the Queen

Having now presented my account and demonstrated how it can be applied to solve a specific puzzle case, it should now be fairly straightforward to see how it applies to the other puzzle

cases. With this in mind, I will run through the remaining puzzle cases more quickly, slowing down only to discuss certain important issues as they arise.

Now let's turn to the second puzzle case: The Gardener and the Queen. Recall that the case involved a gardener neglecting to water some flowers, which resulted in them wilting. These flowers had no connection to the Queen. The puzzle comes when we assess the following two causal statements:

- 3) The gardener's not watering the flowers caused them to wilt.
- 4) The Queen's not watering the flowers caused them to wilt.

Intuitively, (3) seems acceptable, while (4) seems unacceptable.

The fact that (3) and (4) collectively refer to only three distinct events would suggest that, as with the Venusians case, we can model this causal system using just three variables. One can represent the gardener's watering/not watering the flowers, another can represent the Queen's watering/not watering the flowers, and the third can represent the flowers' wilting/not wilting. When modelled this way, the case is structurally isomorphic to the forest fire case, with the values of two independent variables jointly determining the value of a third variable.

The fact that this case is structurally isomorphic to the forest fire case means that the puzzle it raises can be solved in a similar fashion. Considering all the possible states of the system, the only one that is normal is the one where the gardener *does* water the flowers, the Queen *does not* water the flowers, and the flowers remain alive as a consequence of being watered. This state of the system is closest to the average because gardeners typically water flowers that are under their care, whilst queens typically do not water flowers that they have no

connection to. This state is also closest to the optimal state because gardeners should water plants that are under their care, and it would be too demanding to expect people to water flowers that they are not responsible for. As a result, the gardener's not watering the flowers is a difference maker, while the Queen's not watering the flowers is not. Therefore statement (3) is true, as it meets all three conditions, and statement (4) is false in virtue the fact that it does not meet condition (III). This accords with our intuitions about the case, thereby solving the puzzle.

Knobe Effects (The Pen Vignette)

As stated in Chapter 2, our intuitions about the previous two puzzle cases are confirmed by the results of many different experimental studies into causation. There I discussed one such study by Knobe & Fraser, where participants were presented with the pen vignette and then asked to assess two causal statements based on the information they were provided in the vignette. The story they were told took place in a philosophy department where the administrative assistants are allowed to take pens from the receptionist's desk, but the faculty are not. One day, both Professor Smith and an unnamed administrative assistant each take one of the last two pens from the receptionist's desk, resulting in a problem when the receptionist is unable to take a message. The subjects were asked to what extent they agreed with the following two statements:

- 5) Professor Smith caused the problem.
- 6) The administrative assistant caused the problem.

The subjects who were presented this case were significantly more likely to agree with (5) than they were to agree with (6).

This case is again structurally isomorphic to the forest fire case. We can include variables to represent both Professor Smith's and the administrative assistant's taking/not taking a pen, as well as the occurrence/non-occurrence of the problem that followed. We have little information to go on regarding the frequency with which either person takes a pen. In such cases, it seems reasonable to suppose that people typically follow the rules that apply to them, so Professor Smith's not taking a pen and the administrative assistant's taking a pen would be the average values of their respective variables. With the third variable, we can reasonably assume that the receptionist's problem of not having any pens is not a common occurrence, thus the non-occurrence of the problem is the average value.

For the optimal values, it seems obvious that Professor Smith's not taking a pen and there being no problem are both preferable. The status of the administrative assistant variable is less clear, intuitively it seems that the best possible situation is one where the administrative assistant already has a pen, and so does not need to take one of the receptionist's pens. This suggests that the optimal value is the administrative assistant's not taking a pen. However, it also seems that the administrative assistant's taking a pen is not as bad as Professor Smith's taking a pen, which is a fact that should be captured by our model. To do this, we can stipulate that a given state of this causal system will lose 1 point from its R_o if Professor Smith takes a pen, and if the administrative assistant takes a pen it will only lose 0.5 from its R_o .

The possible states of this causal system along with their R_a s/ R_o s are therefore:

- S1) Both Professor Smith and the administrative assistant take a pen, so the problem occurs. ($R_a = 2$; $R_o = 1.5$)
- S2) Professor Smith *does not* take a pen, but the administrative assistant *does*. The problem does not occur. ($R_a = 0$; $R_o = 0.5$)

S3) Professor Smith *does* take a pen, but the administrative assistant *does not*. The problem does not occur. ($R_a = 2$; $R_o = 1$)

S4) Neither Professor Smith nor the administrative assistant take a pen. The problem does not occur. ($R_a = 1$; $R_o = 0$)

Represented graphically:

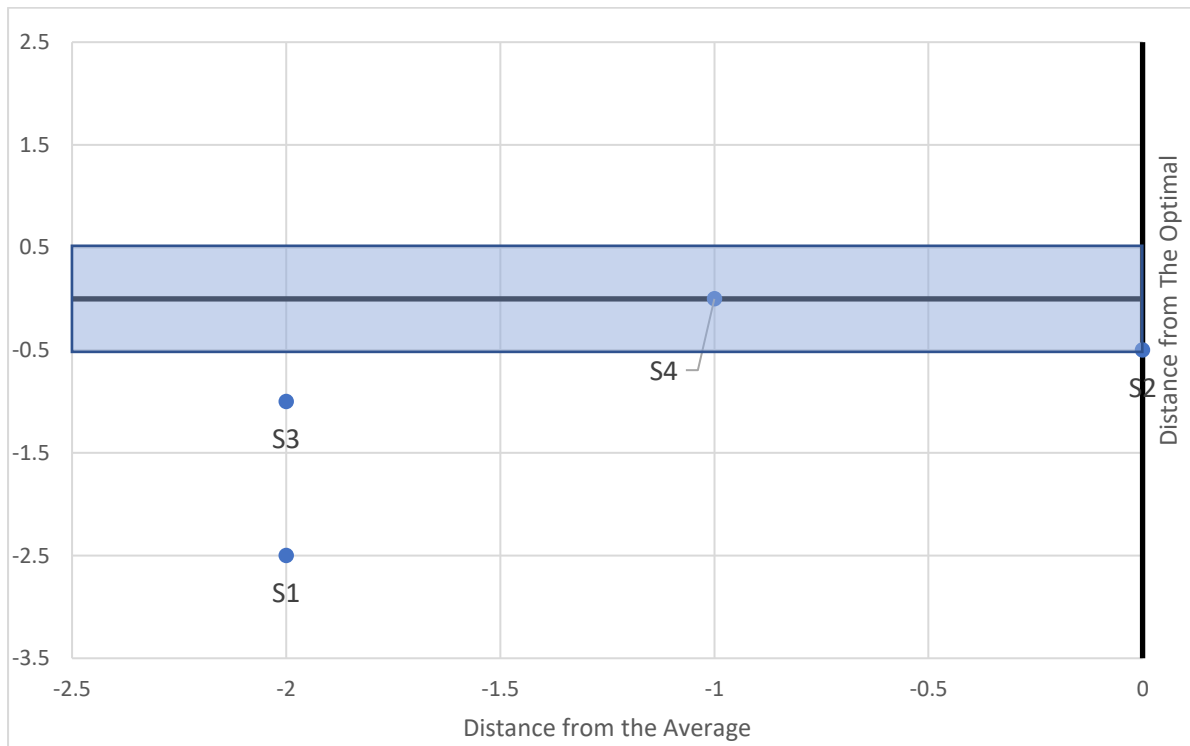


Figure 15

This case could plausibly be understood as normatively charged because of the emphasis on rules, which is represented by where the normality region is drawn. As such, the normal states of the system are S4 and S2. This means that statement (5) comes out as true by virtue of the fact that Professor Smith's taking a pen makes the difference between the actual state of the system (S1) and a normal state of the system S2. The same is not true of statement (6), which comes out as false in virtue of the fact that the administrative assistant taking a pen does not make the different between the actual state of the system (S1) and any normal state of the

system. Though note that it is not far off, as if we were to count S3 as a normal state, the administrative assistant's taking a pen *would* be a difference maker, and consequently (6) would be true. For this to happen, the normal region would only need to be slightly larger than it currently is, which may be achieved by only slightly altering the context within which the causal statements are evaluated. I suspect this is why the subjects were less willing to disagree with (6) than they were willing to agree with (5).

This way of explaining the results of the study can be straightforwardly applied to any other experimental results that confirm the existence of Knobe effects in causation. However, there is one important set of results that still needs to be addressed. The results in question are those presented by Sytsma et al. (2012), which I previously mentioned in Chapter 4. These results are crucial because the existing Competence Strategy views are not in a position to explain them, in virtue of the fact that they do not tell us how we are supposed to decide what 'normal' amounts to.

Recall that in this study, subjects were presented with a version of the case that also included information about statistical regularities. These regularities were on both a population-level (how often faculty members and assistants took pens) and an agent-level (how often Professor Smith and the administrative assistant, who Sytsma et al. name John, took pens. The results that come from examining both these levels are interesting, although only the agent-level results present a significant challenge.

With population-level statistical regularities, Sytsma et al. conclude that they 'have no discernible effect on ordinary causal attributions for the Pen Case' (2012, 817). In other

words, the fact that it is common or uncommon for the faculty/assistants to take pens has no impact on our overall causal judgements.

Unfortunately, the results from the studies involving agent-level statistical regularities raise deeper concerns. The inclusion of these kinds of regularities in the vignette did have an effect on the judgements, however it was the opposite of what we might have expected (Sytsma, Livengood and Rose 2012, 817-818). Specifically, when the subjects were presented with a version of the case where Professor Smith was said to have rarely taken a pen and John was said to have frequently taken a pen, the subjects were less likely to agree with (6) and more likely to agree with (5). That is, when it was typical for John to take the pens, the subjects were more inclined to agree that he caused the problem, even though it was normal in both the sense that he often did it and it was not against the rules. Likewise, in this instance the subjects were less willing to agree that Professor Smith caused the problem, even though her taking pens was extremely abnormal, both because she had rarely done it before, and it was against the rules.

The conclusion suggested by these results is that we don't always identify abnormal events as causes, contrary to the kind of normative account that I have provided. It must therefore be the case that something else is going on. Sytsma et al. suggest that the subjects of this and all the other Knobe Effect studies are merely confused, and the judgements they make are tracking judgements of responsibility instead of causation. The reason that they are more likely to grant causal status to an action that is agent-level typical is that we are more likely to attribute responsibility when the agent has repeatedly performed the offending action. John is more blameworthy if we know that he has frequently taken pens in the past, because

every time he took a pen he made it more likely that something bad would happen. Thus, we hold him morally responsible for what happened, as the blame that we ascribe to him is far greater than the minor amount of blame we ascribe to Professor Smith for breaking the rules. As a result, the subjects' confusion between judgements of causation and responsibility lead them to pick out John as having caused the problem instead of Professor Smith, even though, strictly speaking, they are both equally causal.

I believe that Sytsma et al. are too hasty in concluding that the subjects have simply confused the notion of causation and moral responsibility. To see why, we should consider how my account might handle the information given about the agent-level typicality of both Professor Smith and John's taking a pen. To start with, it will certainly affect the R_a s for all the possible states of the system. If we assume that Professor Smith's taking the pen is typical, then any state where she takes a pen will be closer to 0 on the x-axis (Distance from the Average) than one where she does not take a pen. The same will be true of John if we assume it is typical for him to take a pen, and the inverse will be true of either person if we assume that it is atypical for them to take a pen. Of course, this will not address the problem that Sytsma et al. raise because it will only affect the movement of the possible states of the system along the x-axis, meaning that there will be no change regarding which of those states are normal.

Yet my account allows for the possibility that the values can also shift across a different dimension: the optimal. Sytsma et al. have already acknowledged that it seems intuitively worse for someone who frequently takes pens from a public supply to take another pen, than it does for someone who rarely takes a pen to take one. We assume that the person who always takes pens acts selfishly, due to the lack of regard they show to the other people who

may need to borrow pens from that supply. Such a person should probably take more care not to lose the pens that they have already borrowed, rather than continually getting new ones. On the other hand, the person who rarely borrows a pen can easily be seen as more sensible and considerate of others. They can be seen as someone who only borrows a pen when they absolutely have to, and never without good reason. Now when it comes to determining how optimal each possible state of the system is, these judgements about the characters of Professor Smith and John are going to carry a lot more weight than the departmental rules about who is allowed to take a pen. As a result, the states where John takes yet another pen will be less optimal than the states where Professor Smith borrows a pen for the first time. This means that the states of the system where John takes a pen will be further from the optimal than the states where Professor Smith takes a pen.

Earlier I noted that, because this case involves the violation of a rule, the taking of someone else's property, and a harm being done (albeit a fairly minor one), it is plausible to view this case as normatively charged. Accepting that this case is normatively charged, then the change in the $R_{\circ}s$ for each state of the system that results from the inclusion of information about agent-level typicality will have a much greater effect on the overall normality of those states than whatever change that information brings about in the $R_{\circ}s$. This explains why the Sytsma et al.'s results seem to suggest that the information about typicality had the opposite effect on our causal judgements than what we would expect from a normative account.

At this point, Sytsma et al. could object that if the subjects really did make these judgements about the moral characters of both Professor Smith and John, and then allowed those judgements to influence their evaluations of (5) and (6), then that would be a mistake. I do

not think that this is right, but for now I will pause any discussion of this until the next chapter. For now, I feel that I have successfully demonstrated how my account can be used to explain the results from the studies that I have looked at so far. I will now move on to the final set of puzzle cases.

Causal Superseding

Recall that causal superseding was where the norm violations of one agent affects the causality attributed to another agent. The example we looked at came from a study by Kominsky et al. (2015). They presented two versions of a case to two different groups of participants, before asking them to evaluate two causal statements. Both versions of the case involved a couple, Bill and Sue, trying to acquire a pair of antique bookends. In both versions of the case, Sue buys one of the bookends. The two versions of the case then differed regarding whether or not Bill bought or stole the other bookend. The two statements that the subjects were asked to assess were:

- 7) Sue caused them to possess the paired set of bookends.
- 8) Bill caused them to possess the paired set of bookends.

Two groups of subjects were presented with two different versions of the case: one where Bill stole one of the bookends, and the other where he buys it. The subjects who were presented with the theft version of the case were less willing to agree with (7) than the group who were presented with the other version. The puzzle is to explain this difference in the acceptability of (7).

Here we have two different versions of the case where both causal systems are made up of three distinct variables: one where Sue buys/doesn't buy her bookend, one where they both

do/don't end up with both bookends, and one where, depending on the version of the case, either Bill buys/doesn't buy his bookend or Bill steals/doesn't steal his bookend⁴¹. In terms of the average values of these variables, the only thing that strikes me as atypical is Bill's stealing the bookend. When it comes to the optimal values, all values seem to be plausibly optimal except for Bill's stealing the bookend (because it is morally wrong) and their not being in possession of both bookends. The possible states of the system and corresponding R_a/R_o s are as follows:

Buying Case:

- S1) Sue buys one bookend and Bill buys the other. They have both bookends. ($R_a = 0$; $R_o = 0$)
- S2) Sue buys one bookend, but Bill does not buy the other. They don't have both bookends. ($R_a = 0$; $R_o = 1$)
- S3) Sue does not buy her bookend, but Bill buys his bookend. They don't have both bookends. ($R_a = 0$; $R_o = 1$)
- S4) Neither Sue nor Bill buys a bookend. They don't have both bookends. ($R_a = 0$; $R_o = 1$)

Theft Case:

- S1) Sue buys one bookend and Bill steals the other. They have both bookends. ($R_a = 1$; $R_o = 1$)
- S2) Sue buys one bookend; Bill does not steal the other. They don't have both bookends. ($R_a = 0$; $R_o = 1$)
- S3) Sue does not buy her bookend; Bill steals his bookend. They don't have both bookends. ($R_a = 1$; $R_o = 2$)

⁴¹ Notably, in the theft version of the case, buying it is not an option.

S4) Sue does not buy her bookend and Bill does not steal his bookend. They don't have both bookends. ($R_a = 0$; $R_o = 1$)

In the buying case, all the states of the system have the same R_a , and all the non-actual R_o s are also the same. So, when evaluating this case, all the states matter when evaluating statements (7) and (8). As a result, both statements come out as true because both Sue's purchasing her bookend and Bill's purchasing his bookend make a difference between the actual state of the system and at least one normal state (S2 for Bill and S3 for Sue).

Now compare this to the theft version. This version is plausibly normatively charged, as the case involves a clear moral wrong. As such, the R_a is likely to have little effect on our judgements of what the normal state of the system is in this case. Instead, we ought to look to the R_o , and specifically to the fact that S3 has a lower R_o than the other possible states. This is important because this is the state that, if it were normal, would make Sue's buying the book a difference maker. The fact that this state is less likely to be considered normal provides a good explanation for why the subjects were less likely to agree with (7) in this version of the case. This therefore explains the results of Kominsky et al.'s study, and thus solves the puzzle.

3. Answering the Big Questions

In Chapter 1 I identified two key questions that I hoped to answer in this thesis. Those questions were:

- I. Is causation normative?
- II. Is causation contextual?

It should now be clear that my answer to the first question is 'yes'. In this chapter I have provided an account of the normative constraint on causation. I have explained in detail what

it means for something to be normal and have provided numerous examples of how normality, thus conceived, determines which counterfactual scenarios are relevant to the truth or falsity of a given causal claim. I have also demonstrated that the kind of normativity involved in causation is not a mere statistical, value-neutral kind of normativity, but one that is prescriptive and value-laden. If my account is correct, then all causal judgements will involve some sort of consideration of the optimal, even if its effects may be minimised by the context.

This then leads me on to the answer to the second question: causation *is* contextual, because facts about what causes what are determined, in part, by facts about what is normal – and facts about what is normal are themselves sensitive to many different contextual factors. So, if the truth-values of causal statements are sensitive to what is normal, and what is normal can vary across contexts, then the truth-values of causal statements will also vary across contexts. Put simply, causation inherits its contextuality from normality.

Having now provided the answers to the two main questions of this thesis, I will in the final chapter set about defending those answers from some potential objections that may be raised by those who oppose normative accounts of causation.

Chapter 7 – Objections and Replies

In the previous chapter I developed an analysis of normality and applied it as part of a new, normative constraint on causation. In this seventh and final chapter I will consider and respond to a series of possible objections that may be raised against my account. This will be done over the course of four sections, with each section setting out the objection, followed by my response. These objections will be presented in order of the challenge that I believe they pose to the account, with the weakest first and the strongest last. The objections that I will consider are as follows:

1. My account does not respect the intuition that causation is a natural relation.
2. My account (along with many other existing normative accounts) does not allow for the possibility that normal events can both cause and be caused.
3. A reiteration of Blanchard & Schaffer's point that the use of normative notions in this way constitutes a 'free parameter' that renders my account unfalsifiable.
4. My account opens the door to more indeterminacy than we ought to accept in an account of causation.

Each of these objections will be explained and responded to in turn.

1. The Natural Relation Objection

All the way back in Chapter 1 I argued that we ought to understand the contextualist/invariantist/contrastivist debate within the realm of causation (and beyond) as resulting from two competing intuitions. The contextualist supports their view by appealing to the intuitions that we have about certain puzzle cases, such as the forest fire case. The invariantist disagrees because they want to uphold the intuition that causation is a natural

relation. That is, they see causation as some kind of objective feature of the world that we uncover through empirical investigation. To maintain *this* intuition, they are forced to deny that the intuitions we have about the cases that the contextualist appeals to tell us anything about causation itself and are merely a result of pragmatic considerations or cognitive biases. This means that the choice between the contextualist and invariantist positions ultimately boils down to the question of which intuition (or set of intuitions) we would rather give up: the natural relation intuition or the contextual intuitions. The contrastivist suggests a third option, according to which we must abandon the intuition that causation is a two-place relation.

Following on from this, in Chapter 2 I argued that a similar dilemma can be presented regarding the question about whether or not causation is normative. Those who wish to defend a normative account of causation will often support their position by appealing to cases where our intuitions seem to suggest that our causal judgements are determined to some extent by judgements about normality. This conclusion is then rejected by those who seek to uphold the natural relation intuition, since normativity seems to be a non-natural phenomenon. Once again we have a conflict between the intuitions we have about certain cases on the one hand, and the natural relation intuition on the other.

As noted in both Chapters 1 and 2, there is a close affinity between contextualism and the claim that causation is normative. Those who argue that causation is normative tend to also assume that the normative factors that affect our causal judgements are contextual. If normality is the relevant contextual factor here, then this means assuming that something can be properly described as normal in one context, and not normal in another context. This

view of normality has a lot of intuitive plausibility, since it seems that statements of normality are, at the very least, sensitive to whatever reference class happens to be salient in a given context. Now, if we accept this about normality and also hold that causal statements are sensitive to our judgements about what is normal in a given context, then the conclusion that causation is contextual will immediately follow. Most theorists who argue for a normative account of causation *do* believe that normality (or whatever other concept they appeal to) is contextual, and so they tend to also be committed to the view that causation is contextual (or they are a contrastivist, which as I noted in Chapter 1, is often treated as a mere form of contextualism). This is why the normative and contextual questions are so frequently treated as the same question, and hence why those who answer in the affirmative to either question tend to be grouped together.

I maintain that the two positions can come apart. We could have a contextualist account of causation where causal statements are sensitive to some non-normative feature of the contexts that they are used in. Likewise, we could have a normative account, where the relevant normative notion is not understood as being contextual. The claim that we should endorse the natural relation intuition constitutes an objection to both of these accounts. This is why there tends to also be a lot of overlap between invariantists and those who argue that causation is not normative. Responding to the kinds of arguments that are offered by these defenders of the natural relation intuition will be the focus of this first section.

So far, what we have seen is that I agree with most contextualists that normality is contextual and consequently so is causation. As a result, my account is incompatible with the natural relation intuition on two counts, and so those who argue that causation is neither normative

nor contextual will have a strong motive for rejecting it. Their claim would be that by abandoning the natural relation intuition, I have willingly discarded an extremely appealing intuition for the sake of being able to accommodate our intuitions about a set of obscure puzzle cases. That is, by abandoning the natural relation intuition we lose something of immense value, and this is hardly outweighed by the comparative gains we get from being able to explain our intuitions about the puzzle cases without appealing to pragmatics and biases. Someone who pursues this line of thinking could even acknowledge that the existing attempts to provide an account on the pragmatics-and-biases strategy suffer from some rather serious problems, including those that I identified in Chapter 3. Yet despite this they may still go on to argue that the correct way to respond to those problems is to try and provide a better pragmatics and biases account – one that ‘keeps the ... causal calculus pure’, as Blanchard & Schaffer put it (2017, 211). We should not respond by abandoning the natural relation intuition, as the results of that will be completely unpalatable.

It is at this point that we ought to raise an important question: what is the value in holding that causation is a natural relation? The answer would seem to be that it allows us to more easily argue that causal statements make substantive claims about the structure of reality. That is, it ensures that our causal statements are making claims about something *real*. This is important because of the role that causation plays, which is to allow us to make sense of reality and the relations that events have to one another. This is true for all ways of understanding causation, even those that fall outside of the realm of counterfactual theories. The question we then need to ask is whether understanding causation as a natural relation will help us to achieve this end of making sense of reality.

A defender of the natural relation intuition will argue that understanding causation as a natural relation absolutely helps us to make sense of reality. After all, if we choose to uphold the natural relation intuition, then our causal statements are making claims about relations that really are out there in the world. These statements will therefore inform us about how the universe is structured, in a manner that is independent from human centred norms and conversational contexts. Contextualism cannot have this because the truth or falsity of those statements will always be determined to at least some degree by the conversational context within which they are uttered. Such statements are fundamentally limited by contextual parameters in what they can tell us about reality. They can refer to real, context-independent relations; but this process will be in some way muddied by the context within which the statement is used. To see what this means, consider how a contextualist account handles the forest fire case. There the forest rangers cannot truthfully assert that the presence of oxygen caused the fire, which obscures the fact that the fire is equally counterfactually dependent on both the lightning strike and the presence of oxygen. By insisting that causation is a natural relation, we are able to do justice to this fact about reality. We get a complete story of what needed to be the case in order for the fire to occur. We have a full understanding of the reality of the situation, with nothing being obscured by the context.

Now it seems fairly evident that such a concept of causation will be helpful if our account of causation is non-reductionist. This includes views such as Armstrong's account of causation as a universal (1999) and Dowe's account of causation as transfer of energy (2000). The supporters of such theories are realists in the strongest sense of the word. They take the existence of causal relations to be an ontological matter, independent of any considerations of normality or context. Causal relations are either a part of reality, or they are not. Our own

concept of causation should therefore do justice to the natural relation intuition because there really are relations of causation out there in the world, and our causal statements will need to refer to those relations. A contextualist account like mine has the potential to completely overlook certain causal relations, and so will be overlooking a part of reality. Therefore, on a non-reductive view of causation, understanding causation as a natural relation allows the concept to perform its role of helping us to make sense of reality by allowing us to capture a part of that reality. This role that causation plays really would be undermined if we understood causation to be contextual, since it could no longer perform the function of fully capturing an objective feature of reality.

This idea that we should uphold the natural relation intuition to ensure that our concept of causation can help to give us a complete picture of reality can also be found amongst reductive accounts of causation, including the counterfactual theories that are my target for this thesis. Many of these theorists have clearly felt it necessary to uphold the natural relation intuition, as they believe that by doing so, we can have a concept of causation that will aid our understanding of reality by being able to describe all relations of counterfactual dependence between distinct events⁴² as causal. It is for precisely this reason that Lewis chooses to focus his efforts on providing an analysis of a 'broad and nondiscriminatory concept of causation' (1973, 559) in his original treatment of the topic. A contextualist (or otherwise discriminatory) concept of causation is seen as unfit for this purpose.

⁴² See Lewis (1986b, 259).

I challenge this on the basis that we don't *need* a non-discriminatory concept of causation to make sense of the web of counterfactual dependence relations that exist in reality. The reason for this is that we already have something that can fulfil this role: the concept of counterfactual dependence itself⁴³. This can be seen by the fact that we can describe the structure of the forest fire case without ever using the term 'caused'. Of course, most people will not be familiar with the technical term 'counterfactual dependence', but the basic idea of 'if this hadn't happened then this other thing also wouldn't have happened' is something that should be fairly easy for people to wrap their heads around. With this in mind, there does not seem to be any work for causation to do when it comes to understanding the relations of dependence that hold between distinct events.

One way that we might respond to this is to simply make the move that causation identical with counterfactual dependence. That is, saying that 'e counterfactually depends on c' is just the same as saying that 'c caused e'. Such a position has historically been unpopular due to the existence of various pre-emption cases that seem to demonstrate that counterfactual dependence is not necessary for causation. However, this position has been recently defended by Northcott (2018), who observes that there are many preemption cases where our intuitions count in favour of such a simple counterfactual view and argues that any intuitions we have that count against such a view are mistaken. Of course, such a view will

⁴³ Hitchcock and Knobe seem to make a similar point to this when they state that 'If the concept of actual

causation were entirely egalitarian...the truly important information would already be contained in the causal structure, and it seems that people would be better off ignoring questions of actual causation altogether and simply talking about patterns of counterfactual dependence' (2009, 593). Yet they also hold that 'egalitarianism is entirely appropriate at the level of...causal structure' (*ibid.*). They seem to take this causal structure to be the object of a 'scientific concept' of causation (Hitchcock 2007), but it is not made clear what this concept adds to our understanding of reality that cannot be achieved beyond mere relations of counterfactual dependence.

also have to discard any intuitions we have that motivate the problem of selection, including those about the puzzle cases from Chapter 2. It is for this reason that I argue for a different way of addressing this issue, which Northcott characterises as the view that ‘Causation is some mixture of simple counterfactual dependence and other factors’ (2018, 17).

For such a move to be justified, we need to provide a reason as to why these additional factors are a part of our concept of causation. That is, how do they ensure that the concept of causation is able to perform its function of making sense of reality? I believe the answer to this becomes very apparent on the kind of normative, contextualist account that I am defending. Specifically, a normative and contextualist concept of causation will provide us with a kind of curated causal knowledge that could not be gained through the use of a non-discriminatory concept, where all we will get is a long list of counterfactual dependence relations with no obvious criteria for distinguishing them. The normative account provides good grounds for selecting some of those relations and elevating them to the status of causal relations, whilst also allowing us to filter out those relations that are not selected for. Therefore, the reason why normative factors form a part of our concept of causation is that they allow us to acquire curated causal knowledge: knowledge of the kinds of causal relations that we are selecting for. This knowledge is distinct from the broad and indiscriminate causal knowledge that we get from a non-discriminatory concept.

The question we now need to ask is why we ought to value this curated causal knowledge. Another way to put this is to ask why we would want to select certain relations of counterfactual dependence and raise them up to being causal relations. One possible response to this comes from causal interventionism, which understands causation as enabling

us to identify the ways in which we can intervene on the world to make things happen (Ramsey 1978) (Menzies and Price, Causation as a Secondary Quality 1993) (Woodward, Making Things Happen: A Theory of Causal Explanation 2004). Hitchcock also offers a more general view where he states that this kind of causation plays a role in our 'goal-directed reasoning' (Hitchcock, Actual Causation: What's the Use? 2017, 124), by showing us how our actions can lead to our achieving our desired outcomes. On this kind of view, we select for the counterfactual dependence relations that will help us to bring about these outcomes. This means that the curated causal knowledge we acquire will be instrumentally valuable in allowing us to perform certain actions with the confidence that they will help us to achieve our goals. The fact that it is *curated* causal knowledge allows us to filter out those factors which we cannot intervene on and hence are irrelevant to figuring out what we need to do to meet these goals.

The goal of the forest rangers is to prevent future forest fires, and the knowledge that the earlier forest fire was caused by a part of the forest being struck by lightning will help them to achieve this. They now know to set up a lightning rod to prevent further fires. A belief that the fire was caused by the presence of oxygen would be entirely unhelpful in this context. The fact that the presence of oxygen is common enough on Earth to be normal makes it all but impossible to remove and doing so would be disastrous to the forest's ecosystem. The beauty of this account is that it also works in the context of the Venusians, where the goal may be to prevent such fires on their own planet. Here the knowledge that the presence of oxygen caused the forest fire will be valuable because it instructs them to ensure that oxygen is never present for one of Venus's many lightning strikes. So, we have an explanation for the value of curated causal knowledge, and how that value might vary across contexts.

My opinion on this kind of interventionist account is mixed. On the one hand, I can see that it provides a plausible explanation for the value of curated causal knowledge, but on the other hand I also think that there may be some curated causal knowledge that is still valuable *without* informing us of how best to intervene or contributing anything to our goal-directed reasoning. Such knowledge may just be intrinsically valuable, in the same way that any significant knowledge about the world seems to be valuable. I will not defend this position here, as such an epistemological question falls outside the scope of this thesis. My conclusion here is therefore that the point of the concept of causation is to provide us with a kind of curated knowledge that we cannot acquire by merely discovering relations of counterfactual dependence. I remain neutral on where the value of this knowledge comes from, though I have demonstrated that it is plausible to believe that such knowledge has value, as the interventionists clearly do. So, contrary to what the defender of the natural relation intuition claims, we have not lost out by abandoning that intuition.

On top of this, there is also a question of why we ought to value the raw, impartial, non-discriminatory kind of causal knowledge that we get if we uphold the natural relation intuition over the curated causal knowledge that I defend. As I have already indicated, the role that causation plays in our making sense of reality has to be something that cannot be achieved using counterfactual dependence alone. For there to be a distinct notion of a 'causal relation' that notion must have a role. On a normative and contextualist account, the role of such a notion is to enable the formation of curated causal knowledge. An account that upholds the natural relation intuition can make no such claim, since on such an account causal knowledge will be completely impartial.

My opponent may respond to this by insisting that a natural relation account of causation can still provide us with curated causal knowledge, and crucially, this knowledge will be free from the muddying effects of normative and contextual considerations. After all, very few theorists argue that causation and counterfactual dependence are identical, and anyone who attempts to identify one with the other will inevitably run afoul of counterexamples involving either preemption, overdetermination, or transitivity. So, they already accept that the role causation plays in our understanding of reality is different to that of counterfactual dependence, and that causation involves the selection of particular relations of counterfactual dependence to deliver us curated causal knowledge. Crucially, this selection process is not something as hard to pin down as normality, instead it will involve some more precisely defined notion that explains why some counterfactual dependence relations are selected to be causal relations while others are not.

This kind of curated causal knowledge is typically taken to be what is delivered to us by science. The idea here is that the scientific method ensures that conclusions are reached in a manner that is impartial, context-independent, and value-free. This process is not free of some kind of selection; the knowledge it provides us with is still curated. However, it is not a process that involves claims about what is and is not 'normal'. When an astronomer lays out the entire causal story for why Mercury has a more eccentric orbit than any other planet in the solar system, we can feel assured that the relations they identify have not been selected based on their adherence or non-adherence to some norm, or based on the conversational context that we happen to find ourselves in. The story will not include reference to every event that Mercury's orbit counterfactually depends on, but rather will provide a curated selection of causal facts for us to learn about, thereby increasing our understanding of reality.

Rather than relying on normality for the selection process, they instead use a much more precise notion of scientific relevance when selecting relations of counterfactual dependence to be granted the status of being causal relations.

This, of course, paints a very rosy and likely inaccurate picture of scientific practice, which has often been challenged as not being impartial (Kuhn 1962), context-independent (Longino 1990), or value-free (Harding 1992). Even if these points turn out to be ill-founded, it does not seem like the everyday practice of science deploys the kind of non-discriminatory concept that Lewis offers. When experiments are conducted, the testers tend to be focused on the effect that one variable has on another, with all other variables being controlled for. Causal selection is built into the way such experiments are conducted, and this is not merely laziness on the part of scientists: it is understood as how such experiments ought to be done. Notably, the choice of which variables to focus on will also be determined by the contextual factor of what the scientists in question happen to be investigating. Even our humble astronomer will only select events to highlight as causes based on their own interests and the audience they are providing the story to. So it is in fact not at all clear that a non-discriminatory concept of causation is valuable when it comes to scientific enquiry.

Is there an area where a non-discriminatory concept *can* plausibly be said to have value? Perhaps, but it is far from obvious where that would be if it is not scientific enquiry. Indeed, one of the reasons that someone may try to uphold the natural relation intuition is to preserve a degree of purity in scientific enquiry, by characterising it as something that is unmuddied by normative or contextual considerations. Yet if this is the motive behind the intuition, it is misguided. In what I have said so far, it is clear that a non-discriminatory concept does not

provide our scientific enquiries with anything that cannot already be provided by the concept of counterfactual dependence or a normative and contextual concept of causation.

To summarise, a normative and contextual concept of causation provides us with a particular kind of curated causal knowledge that is plausibly of value to our making sense of reality. Meanwhile, a non-discriminatory concept of causation does not seem to contribute anything towards this end. We can therefore conclude that the natural relation objection is mistaken; we are not losing anything of significant value by giving up the natural relation intuition in favour of the contextual intuition. As a result, the kind of analysis that we ought to prefer is one that takes causation to be partial and discriminatory.

As a final point, I shall return to the thread that I began in the previous chapter regarding Sytsma et al.'s experimental results. Those results seemed to suggest that agent-level statistical typicality affected the subjects' causal judgements in the opposite way to what my analysis predicts. That is, the subjects were more inclined to agree that John's or Professor Smith's taking a pen caused the problem if the act of taking the pen was agent-level typical. Initially this seemed to provide a counterexample to my position, as the subjects were selecting a more normal cause over one that was clearly abnormal.

Sytsma et al.'s explanation for this was that the subjects were confused about what the study was asking of them and making judgements of moral responsibility rather than causation. In the case where it was typical John the administrative assistant to take the pen, the subjects viewed him as more blameworthy, as his constant taking of the pens made it more likely that none would be left on the receptionist's desk for when one was needed. By agreeing to the statement that John caused the problem and disagreeing with the statement that Professor

Smith caused the problem, the subjects were, according to Sytsma et al., asserting that John was to *blame* for what happened, rather than asserting that John was the only one to have *caused* the problem. Thus, the Knobe Effect in causation is merely the result of subjects making judgements of moral responsibility when they should be making judgements of causation.

My response to this acknowledged that Sytsma et al. were right to explain their results as stemming from certain moral assessments made by the subjects. The explanation I gave for this was that the subjects' judgments were informed by implicit character assessments they had made regarding the two agents described in the vignette. In the case where John's taking a pen is agent-level typical, we cannot help but see him as someone who is both selfish and inconsiderate, so his taking another pen will likely be unjustified. By comparison, Professor Smith is seen as far more sensible and respectful, so when she takes a pen it is seen as morally justified, even though it is against the department's rules. As a result, the normal state of the system will be one where Professor Smith takes a pen, but John does not. With this as the normal state, John's taking a pen will be a difference maker, while Professor Smith's taking a pen will not; hence the subjects being more likely to agree that the former caused the problem.

Crucially, the difference between my explanation of these results and Sytsma et al.'s explanation, is that my explanation does not hold that the subjects were confused when they made their judgements. The causal judgements were reached through a competent application of the concept of causation and the moral considerations were involved in their reaching that judgement in just the way that such considerations ought to be factored in. I

take it to be an advantage of my account that we don't have to deny people's intuitions about the truth-values of causal statements. However, as I noted in the previous chapter, Sytsma et al. could respond by arguing that by allowing these moral considerations to interfere with their judgements at all, the subjects were allowing their own evaluative judgements of the persons involved in the case to influence their causal judgements – thus falling prey to the same confusion between causation and moral responsibility.

This possible response to my explanation fails because it relies on the assumption that the role of causation within our cognition is to provide impartial, non-discriminatory judgements on the situations we are presented with. As I have argued in this section, such an assumption is mistaken, since such a concept of causation does not obviously hold any value when it comes to helping us make sense of the world. Sytsma et al.'s subjects got it right with their judgements, using a discriminatory concept. The value in this can be seen if we imagine that someone is tasked with reprimanding one of the agents to ensure the problem of there no longer being any pens does not happen again. Here it would make more sense to target John the administrative assistant, as his apparent selfishness makes him more likely to take a pen in future than the more considerate Professor Smith. This approach can be justified by appealing to the kind of curated causal knowledge that we get from a contextualist and normative account. The explanation therefore holds up and continuing to treat it like a bias would only beg the question against a contextualist and normative account.

So, Sytsma et al. are wrong to insist that the subjects' judgements were the result of a mistake on their part. This will apply equally to any other theorist who tries to uphold the natural relation intuition by trying to explain away contrary intuitions as being the result a mistake or

a bias. Meanwhile, it is a virtue of my account that we are not forced to insist that people's ordinary causal intuitions are mistaken.

2. Normal Events as Causes and Effects

The second objection I will consider concerns how my account handles cases where normal events are either the cause, the effect, or both. The universe is full of events that lead to other events as part of the normal state of things. As an example, consider the following causal statement:

- 1) The sun emitting light at noon caused the plants to photosynthesise shortly after.

This causal statement describes an entirely unremarkable process: the sun does what it normally does, and the plants then do what they normally do as a result. The causal system here can be appropriately modelled using just two variables: one to represent the sun's emitting and not emitting light, and the other to represent the plants photosynthesising and not photosynthesising. Here the normal state of the system seems to also be the actual state, where the sun emits light and the plants photosynthesise. Yet if this is the case, then on my view the causal statement will be false, because the sun's emitting light cannot make the difference between the normal state of the system and the actual state of the system, because the two states are numerically identical and so there is no difference to be made. Also, for the statement to be true the effect also must not be present in any normal state of the system. So, contrary to our intuitions, my analysis gives the result that (1) is false because it does not meet conditions (II) and (III).

My response to this is to highlight the effects that context can have on which state of a causal system gets to count as normal. In most contexts, events like those referred to in (1) – events that seem to be a part of the ‘normal’ everyday running of the universe – are rarely the subject of any kind of causal inquiry, and so they will often be implicitly relegated to the background. Yet when we acknowledge these events and directly pay attention to them, that can also come with a change in context that in turn affects what the normal state of the system is. Any context where we are seriously evaluating (1) is going to be a context where we are in some way engaging with question of how plants photosynthesise, even if that process of engagement consists in recalling a fact that we already knew. By making photosynthesis the object of our study in this way, we make relevant all of the things that do not photosynthesise. Thus, the salient reference class here will include all matter (or maybe just all matter with chlorophyll) that does not photosynthesise. Since the average plant is not constantly under sunlight, what is normal relative to this reference class will be for the plant to not be photosynthesising. As a result, the plants’ photosynthesising will not be in any normal state of the causal system, so condition (II) is met.

Similar considerations apply to the sun’s emitting light. In a context where plant photosynthesis is under a microscope, all the events that it counterfactually depends upon become relevant, along with their possible absences. Thus, the salient reference class here will then include all the possible places and times where the sun’s light is not visible. With this as a reference class, the sun’s shining light on the plants starts to seem more atypical when compared to how things are in the rest of the universe. So, the sun’s emitting light is not average, and so the average state for the system will be one where the sun’s omitting light at noon does not occur. Plausibly, we can assume that this domain is average leaning since the

case does not involve any agents. This means that the average state of the system will have a much greater impact on what the normal state of the system will be in this case. So, the normal state of the system is one where the sun's emitting light does not occur and consequently the plants do not photosynthesise. Here condition (III) will be met because the alleged effect does not occur in any normal state of the system, and the emitting of light from the sun makes the difference between the normal and actual states of the system. This means that (1) will be true in this context.

There may well be other examples where it is less obvious how my analysis can explain how an event is a cause or effect. Yet I think that generally, we can account for these cases by thinking about how context affects what the normal state of the system is.

One worry might be that this allows for the possibility that we can make causal statements true simply by uttering them. An example would be one of the forest rangers claiming that the presence of oxygen caused the forest fire. It could be argued here that, by analogy with the explanation of (1) just given, by even raising this claim the context is shifted such that the salient reference class includes all the possibilities where oxygen was not present, and the end result of this would be that the forest ranger's claim would come out as true. More broadly, this would also seem to undermine the whole contextualist project, which argues that the intuitions we have that certain statements are unacceptable to say in certain contexts is a result of those statements being false in that context. This move could no longer be made if we were to concede that the utterance of any causal statement could change the conversational context so as to render that statement true.

This is a valid concern, but it is one that can be addressed by highlighting the ways in which the context of the forest rangers renders such a claim false in the first place. Recall that the forest rangers are attempting to identify the cause of the fire so that they can understand why the fire happened in the way that it did. To reach this understanding they need more than just a simple list of all the events that the fire counterfactually depends on. They need to know why the fire started in the time and place that it did, rather than at any other time or place. Claiming that the presence of oxygen caused the forest fire does nothing to help them achieve this goal, and so it is an instance of the concept of causation being inappropriately applied. Now the erroneous forest ranger could continue to insist on their claim that the presence of oxygen caused the fire, perhaps by referring to scientific facts regarding the occurrence of fires, and with enough force this may end up changing the conversational context enough that the other forest rangers concede that the fire *was* caused by the presence of oxygen. However, this would represent a significant shift in the goal of the conversation. Rather than being about understanding this fire and what led it happening in the way that it did, the conversation is now about the physical processes involved in the occurrence of fires, which represents a significant enough shift in context that the truth-values of the relevant causal statements will also be affected. Such a shift can only be brought about with the full cooperation of everyone involved in the conversation.

At this point, it may be objected that this seems like something of a cheat. That all I've done is just explain away the problem by merely stipulating that the context leads us to construct the case in a way that guarantees we will get the result that I need. This is hardly an unreasonable criticism at this point, and it leads us nicely to the third objection.

3. Is this all Completely Unfalsifiable?

For this third and final objection, let's return to Blanchard & Schaffer's original statement of the unclarities objection. There they make the following remark:

[D]efault-relativity often seems to us to come close to a free parameter in an otherwise so precise and objectively constrained formalism, which basically gives the theorist leeway to hand-write the result she wants. (Blanchard and Schaffer 2017, 192)

While my analysis answers the unclarities objection by providing a clear set of truth-conditions for normality statements, the point made in this quotation still has some sting. Put simply, the concern here is that my analysis may turn out to be unfalsifiable in the same manner that Popper used when talking about homeopathy and psychoanalysis (Popper 1963). That is, any time a potential counterexample can be raised against the account, I can simply use the machinery provided by my analysis to simply 'hand-write' whichever answer that will confirm that same analysis. This can be most clearly seen in my talk of the normality region. It could simply be argued that the size and shape of this region is entirely arbitrary and can be used by anyone to justify just about any judgement of normality and, consequently, causation. In other words, all my analysis has achieved is to provide additional tools for a bad faith advocate of my analysis to use as a means of handwaving any counterexamples (including the mechanism of appealing to reference classes that I used in the previous section). If this were the case, then the normative constraint on causation I have offered would not be much of a constraint, since it could always be applied by just fiddling around with what counts as normal within the context in question.

To respond to this thoroughly, I will need to consider all of the parts of my account where an objection could be raised along these lines, and then show why those concerns are ill-founded. However, I shall first make a general remark regarding how my analysis ought to be taken. Over the course of developing the analysis in Chapter 6, I represented the normal states of causal systems using causal models, numerical values, and graphs. These representations should not be taken as an accurate depiction of how we undertake our causal reasoning process, but instead should be treated as a means of representing how a variety of different factors interact to shape the objects of our causal judgements. As such, there is likely to be some degree of vagueness involved, particularly in cases where very different states of a causal system can plausibly be considered normal. This is not to say that a more precise and constrained framework is not possible. Indeed, I think that providing such a framework would be an excellent way to build on the ideas presented in this thesis. Yet such a framework would also require a considerable amount of work to develop and is not needed for my present purpose of understanding everyday talk of causation.

Now to address the specific parts of my analysis where Blanchard & Schaffer's point is likely to carry the most weight. I suspect that this objection will be most likely raised against the parts where context plays a big role. The reason for this is that contexts contain a wealth of different factors that could all potentially be appealed to when trying to show how the analysis is able to deliver the right answer. In the previous chapter, I identified three different ways in which contextual factors can affect what is normal, which will consequently impact on our causal judgements. These three factors were:

- i) The reference classes that are salient.
- ii) The limiting assumptions used for constructing the optimal.

iii) Whether or not the domain is normatively charged.

It is first worth noting that (i) seems like it would be a feature of most accounts of normality, as it seems intuitively plausible to claim that the truth-values of many normality statements are sensitive to whatever reference class is contextually salient. So, while it could be argued that the bad faith advocate can use this sensitivity to hand-write whatever results they need in response to a given counterexample, this criticism will apply equally to any account of causation that involves considerations of normality. In response to this, we should aim to develop an independently plausible set of principles that can help us to constrain the kinds of reference classes that can be appealed to in a specific context, meaning that the bad faith advocate cannot just appeal ad hoc to whatever class they want. Such a project will be a fairly broad undertaking in the philosophy of language, and so falls outside the scope of my inquiry. Right now, it is sufficient to simply reiterate the point made in the previous section, which is that a conversational context is determined by its participants. If the participants all agree that a conversation is about one thing, there is no room for the theorist to argue that it is really about something else. As a result, any counterexample where the salient reference classes are clearly stipulated will be immune to the bad faith advocate's tricks. That is, in such an example, the context must be defined in such a way that it is immediately apparent what reference class is being appealed to.

More concerning is the use of the unfalsifiability objection against (ii) and (iii), which are both unique to my account. (ii) concerns the assumptions that impose limits on our construction of the optimal, which then feeds into our judgements of normality. Recall that I defined the optimal as being the best that we could hope for given the truth of certain assumptions. What those assumptions are will vary across contexts, which means that the optimal and therefore

the normal will also vary across contexts. For example, when it comes to middle school bullying, the members of an anti-bullying organisation are likely to have very different limiting assumptions to a jaded schoolteacher, resulting in them making very different judgements about what is optimal. This difference in judgements about what is optimal might then lead to them making very different causal judgements. With (iii) the effect instead results from how the normal region is shaped by the context. The main factor that determines this is the extent to which the domain we are investigating is normatively charged. If it is, then the region will be shaped such that the optimal will have a much greater impact on whether or not something is normal. If the domain is not normatively charged, then the shape of the region will reflect the fact that the average carries much more weight.

Both (ii) and (iii) are potential targets of the unfalsifiability objection, as they both present some clear opportunities for the bad faith advocate to exploit a lack of clarity in order to hand-write answers in response to legitimate counterexamples. With (ii), they can invent some previously unmentioned limiting assumption, which just so happens to produce the result that they want. Likewise, with (iii) they can pick out some feature of the context which allows them to plausibly argue that the relevant domain either is or is not normatively charged, depending on what allows them to achieve their desired result.

Again though, we can respond to this concern by highlighting the fact that the content of a conversational context is shaped by its participants and this is not something that the theorist can override. This means that when a critic attempts to present a counterexample to my account, they will need to be careful to ensure that the relevant limiting assumptions are identified, and that the domain is plausibly set out as being either normatively charged or not

normatively charged. This certainly makes the process of setting up a counterexample far more demanding, but it will ensure that the bad faith advocate will have no recourse to invent or stipulate some feature of the context that can let them hand write a successful response. So, put simply, a critic of my account must ensure that they are precise in their descriptions of the conversational context surrounding any alleged counterexample that they provide. This will ensure that they will experience no trouble with the bad faith advocate.

Despite this, there may well be some lingering worries here. Specifically, there may be some concern that (ii) and (iii) usher in an unacceptable level of indeterminacy into my account. This leads me onto the fourth and final objection.

4. The Indeterminacy Objection

In the previous section I considered how a bad faith advocate could use the mechanisms of my account to essentially render the view unfalsifiable. My response was that this issue could be avoided as long as any potential counterexamples to the account were stated with the right level of precision. However, there may still be a concern that my account admits far too much indeterminacy, as we cannot always expect a case to be presented to us with the level of precision offered by a philosopher who is attempting to refute a theory. We will encounter cases where vagueness in the limiting assumptions or shape of the normality region will make it difficult to come to any definitive causal judgements. To return to the quotation from Blanchard & Schaffer, this kind of indeterminacy runs counter to the ‘precise and objectively constrained formalism’ (2017, 192) that is so desirable in an account of causation.

The main sources of this objection will be features (ii) and (iii) from the previous section. Starting with (ii), critics like Blanchard & Schaffer may observe that it is often unclear what

assumptions are being used to limit the optimal in any given context. For instance, suppose we are discussing what the optimal number of times to clean your house is per month. Obviously, the amount of time someone is assumed to have will place the most significant limitations on this, but how much time exactly are we supposed to assume is available to a given person? Someone who holds down multiple jobs whilst also raising two children will have far less time available to them than an unemployed childless teenager who is financially supported by their wealthy parents, so which one are we to use as our assumed limit on the optimal number of times to clean one's home? One answer might be that we simply assume that the time available is whatever a normal person would have, but this clearly results in an infinite regress that, while not vicious, is unhelpful in this instance.

To see how this might be a problem, consider a case where I ask a friend to house-sit for me over the course of a single month, which they promise to clean a reasonable number of times, with the assurance that if they don't, they will pay for a cleaner. When I get back, I'm not happy with how often my friend has cleaned the house. They claim to have cleaned it twice and argue that it is unreasonable of me to expect them to have done it more often. However, I insist that they have not done enough and get them to pay for a cleaner.

It seems that there are two causal statements to consider here:

- 2) My friend's not cleaning the house more than twice caused them to have to pay for a cleaner.
- 3) My attitude towards cleaning caused my friend to have to pay for a cleaner.

How do we determine whether or not these statements are true? It seems to hinge on whether or not the normal state of the system is one where my friend cleans the house more

than twice. Since we are dealing with a case that involves a promise, it is plausible to hold that the domain here is normatively loaded. So, the question now is whether or not my friend cleans the house more than twice in the optimal state of the system. This is where it becomes important what the limiting assumptions are. If we are in a context where the limiting assumptions entail some very heavy constraints on how often my friend can clean the house, then we might plausibly hold that the optimal state of the system will be one where the house is only cleaned twice. However, if we are in a context without those assumptions, the optimal state of the system may be one where the house is cleaned more than twice. If we were in the former context, then (2) would be false and we would be more likely to assert (3). However, if we were in the latter context, then (2) would be true. This demonstrates how a difference in limiting assumptions across contexts can affect the truth-values of causal statements. So, it seems here that there is a lot of potential indeterminacy regarding the truth-values of both (2) and (3), which will be off-putting to anyone looking for a precise formalism.

In response to this I would first draw attention to the fact that cases like this – where a lack of clarity regarding limiting assumptions lead to indeterminacy in our causal judgements – will be fairly rare. The reason for this is that our causal judgements are reached by implicitly determining which possible states of a causal system counts as normal, and then appealing to those in our counterfactual reasoning. A change in a limiting assumption will only have an effect on the normality of a possible state of the system if the state is one where it is unclear how close that state is to the optimal. Additionally, the effects of this change will also be extremely limited if the domain is one that is not normatively charged.

Of course, such cases do still exist, as evidenced by the fact that I was able to provide one myself. That said, even in these cases some of the potential for uncertainty can be dispelled by identifying what limiting assumptions are in play. With the cleaning case, we just need to question what assumptions the different parties are working with that limit the optimal number of times to clean the house. By making these assumptions clear, we can potentially work towards a kind of consensus regarding whether or not my friend's only cleaning the house twice was optimal/normal. From here we can then give determinate answers regarding the truth-values of (2) and (3). Of course, this may not work for every case, but it will dissolve many of them, and those that are left will be few and far between.

When it comes to (iii), the problem may not be quite as easy to address. The degree to which a given domain is normatively charged will always determine the size and shape of the normality region and there will consequently be a lot more cases that will result in the kind of indeterminacy we are worried about. To see what I mean, consider the following graph depicting the relative normality to the two possible states of a particular causal system:

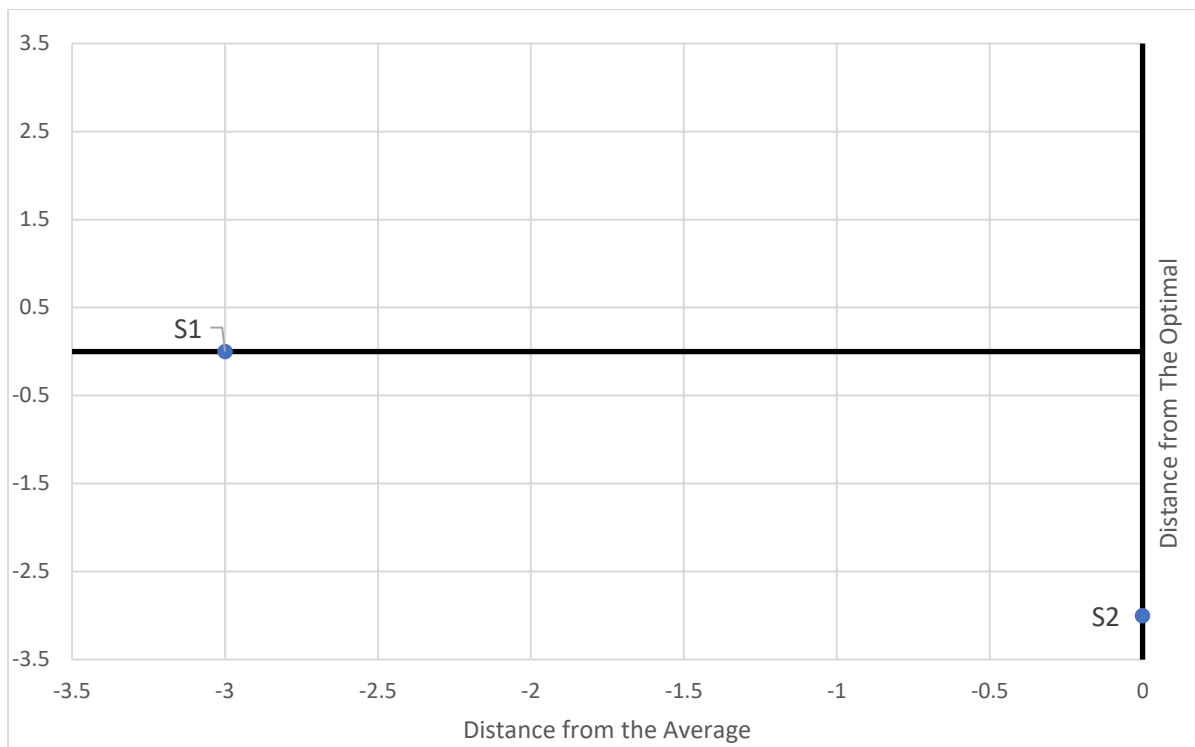


Figure 16

Here, S1 is close to the optimal but far from the average, while S2 is close to the average but far from the optimal. In cases like this it really matters what shape the normality region takes. If a causal claim will be true if only S1 is normal, but false if only S2 is normal, then the truth or falsity of the claim will depend entirely on whether or not the domain is normatively charged. Yet there will likely be cases where this is not apparent, meaning that those cases come with an unacceptable amount of indeterminacy.

An ideal way to respond to this concern would be to provide a precise schema for determining the size and shape of the normality region but given the number of different contextual factors that could plausibly have an impact on this, I will not be able to provide such a schema here. However, I can offer a general rule of thumb for determining the amount of influence the optimal exerts on the normality region. That is, a domain can be considered normatively

charged within a given context if the conversational participants within that context see the domain as important and an appropriate arena for attitudes of praising, commending, blaming, or condemning. This is perhaps still broad enough that it would still allow for some indeterminacy, but there will inevitably be some fuzziness when dealing with everyday folk judgements of causation.

At this point, my opponent could stamp their feet and insist that any account which has failed to provide a precise formalism has failed at providing any sort of account of causation. The thing is, this is now just a more sophisticated version of the natural relation objection that we began with. My aim throughout this thesis has been to provide an analysis of folk judgments of causation, and border cases are to be expected when dealing with judgements that are made in an everyday, imprecise context. These cases will also generally be the exception and can be largely eliminated through more precise specifications of conversational contexts, which should be possible in more formal settings. To insist that an account of causation should be able to offer us the kind of precise conditions for causation that are being asked of here is to insist that our account be able to carve nature at the joints, which already presupposes that causation is a natural relation. So, as long as we are willing to abandon the natural relation intuition (and I have already provided ample reason for doing this), then a small amount of indeterminacy will be completely unproblematic. Indeed, this may be an advantage, since a lot of our causal talk is indeterminate and vague, and we might want our accounts to reflect that fact.

Conclusion

Over the course of the prior seven chapters I have answered two fundamental questions about the nature of causation:

- I. Is causation normative?
- II. Is causation contextual?

My answer to both questions is a resounding ‘yes’. Over the course of providing these answers I have presented and defended an original account of normality, where ‘normal’ is understood as a univocal term whose usage is shaped by two distinct dimensions: the average and the optimal. I have then applied this account to the topic of causation and used it to address a series of persistent problem cases. In this chapter I have also considered and responded to four possible objections that might be raised against my analysis.

I have demonstrated that my analysis provides a more compelling solution to the puzzle introduced in Chapter 2 than any other option that is currently available. This puzzle concerned the fact that our intuitions about certain cases seem to suggest that our causal judgements are sensitive to our judgements of normality. As shown in Chapter 3, our intuitions about these puzzle cases cannot simply be dismissed by appealing to pragmatics-and-biases; we instead have to hold that they are the result of people competently applying the concept of causation. However, as shown in Chapter 4, the existing attempts to provide such an account on the competence strategy have failed as a result of not providing a clear or complete analysis of the concept of normality. My analysis succeeds because it is able to plug that gap in the literature.

What I would hope to see in future is that my analysis is incorporated into many of the existing counterfactual accounts of causation, as by doing so they will be able to solve a persistent problem faced by those accounts. This will, in essence, be a new normal for accounts of causation.

Bibliography

- Alicke, Mark. 2008. "Blaming Badly." *Journal of Cognition and Culture* 8 (1): 179-186.
- Alicke, Mark, David Rose, and Dori Bloom. 2011. "Causation, Norm Violation, and Culpable Control." *The Journal of Philosophy* 108 (12): 670-696.
- Apresjan, JU D. 1974. "Regular Polysemy." *Linguistics: An Interdisciplinary Journal of the Language Sciences* 12 (142): 5-32.
- Armstrong, David. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Armstrong, David. 1999. "The open door: Counterfactual versus singularist theories of causation." In *Causation and Laws of Nature*, by Howard Sankey, 175-185. Dordrecht: Springer Science & Business Media.
- Barner, D, and J Snedeker. 2008. "Compositionality and statistics in adjective acquisition: 4-Year-Olds interpret tall and short based on the size distributions of novel noun referents." *Child Development* 79 (3): 594-608.
- Barsalou, L W. 1985. "Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11 (4): 629.
- Bear, Adam, and Joshua Knobe. 2017. "Normality: Part descriptive, part prescriptive." *Cognition* 167: 25-37.

- Beebe, Helen. 2004. "Causing and Nothingness." In *Causation and Counterfactuals*, by L A Paul, Ned Hall and John Collins, 291-308. London: The MIT Press.
- Bennett, Jonathan. 1988. *Events and their Names*. Indianapolis: Hackett Publishing.
- Bensing, Steven, Adam Bear, and Joshua Knobe. 2016. "Normality judgments and explicit sampling." *Unpublished raw data*.
- Blaauw, Martijn. 2013. "Contrastive Belief." In *Contrastivism in Philosophy*, by Martijn Blaauw, 88-100. London: Routledge.
- Blanchard, Thomas, and Jonathan Schaffer. 2017. "Cause without Default." In *Making a Difference: Essays on the Philosophy of Causation*, by Helen Beebe, Christopher Hitchcock and Huw Price, 175-214. Oxford: Oxford University Press.
- Boorse, Christopher. 2014. "A Second Rebuttal On Health." *Journal of Medicine and Philosophy* 39: 683-724.
- Campbell, Keith. 1990. *Abstract Particulars*. Oxford: Blackwell.
- Chadwick, Ruth. 2017. "Normality as Convention and as Scientific Fact." In *Handbook of the Philosophy of Medicine*, by Thomas Schramme and Steven Edwards, 1-12. London: Springer.
- Cheng, Patricia W, and Laura R Novick. 1991. "Causes versus enabling conditions." *Cognition* 83-120.

- Clarke, Randolph, Joshua Shepherd, John Stigall, Robyn Repko Waller, and Chris Zarpentine. 2015. "Causation, norms, and omissions: A study of causal judgements." *Philosophical Psychology* 28 (2): 279-293.
- Cohen, Stewart. 1988. "How to be a Fallibilist." *Philosophical Perspectives* 2: 91-123.
- Cohen, Stewart. 1998. "Contextualist solutions to epistemological problems: Scepticism, Gettier, and the lottery." *Australasian Journal of Philosophy* 76 (2): 289-306.
- Cohen, Stewart. 1999. "Contextualism, Skepticism, and the Structure of Reasons." *Philosophical Perspectives* 13: 57-89.
- Davidson, Donald. 1967. "Causal Relations." *The Journal of Philosophy* 64 (21): 691-703.
- Davis, Wayne. 2014. *Implicature*. Edited by Edward N. Zalta. 21 September. <https://plato.stanford.edu/archives/fall2014/entries/implicature/>.
- DeRose, Keith. 1992. "Contextualism and Knowledge Attributions." *Philosophy and Phenomenological Research* 52 (4): 913-929.
- DeRose, Keith. 1995. "Solving the Skeptical Problem." *The Philosophical Review* 104 (1): 1-52.
- DeRose, Keith. 1999. "Contextualism: An Explanation and Defense." In *The Blackwell Guide to Epistemology*, by John Greco and Ernest Sosa, 187--205. London: Blackwell.
- Dowe, Phil. 2000. *Physical Causation*. Cambridge: Cambridge University Press.
- Dretske, Fred. 1977. "Referring to Events." *Midwest Studies in Philosophy* 2 (1): 90-99.
- Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge MA: MIT Press.

- Driver, Julia. 2015. "Luck and Fortune in Moral Evaluation." In *Contrastivism in Philosophy: New Perspectives*, by Blaauw Martijn, 154-172. London: Routledge.
- Dupré, John. 1998. "Normal People." *Social Research* 65 (2): 221-248.
- Eddington, Chelsea M, and Natasha Tokowicz. 2015. "How meaning similarity influences ambiguous word processing: the current state of the literature." *Psychonomic Bulletin & Review* 22: 13-37.
- Egré, Paul, and Florian Cova. 2015. "Moral asymmetries and the semantics of many." *Semantics and Pragmatics* 8: 1-45.
- Garfinkel, Alan. 1981. *Forms of Explanation: Rethinking the Questions in Social Theory*. London: Yale University Press.
- Grice, H P. 1975. *Logic and Conversation*. Vol. 3, in *Syntax and Semantics*, by Peter Cole and Jerry L Morgan, 41-58. New York: Academic Press.
- Grice, H P. 1989. "Further Notes on Logic and Conversation." In *Studies in the Way of Words*, by H P Grice, 41-57. Harvard University Press: Cambridge.
- Halpern, Joseph Y, and Christopher Hitchcock. 2010. "Actual Causation and the Art of Modeling." In *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, by Rina Dechter, Hector Geffner and Joseph Y Halpern, 383-406. London: College Publications.
- Halpern, Joseph Y, and Christopher Hitchcock. 2015. "Graded Causation and Defaults." *British Journal for the Philosophy of Science* 66 (2): 413-457.

- Harding, Sandra. 1992. "After the Neutrality Ideal: Science, Politics, and "Strong Objectivity"." *Science and Politics* 59 (3): 567-587.
- Hart, H.L.A., and Tony Honoré. 1985. *Causation in the Law*. Oxford: Oxford University Press.
- Haslanger, Sally. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them To Be?" *Nous* 34 (1): 31-55.
- Heller, Mark. 1999. "The Proper Role for Contextualism in an Anti-Luck Epistemology." *Philosophical Perspectives* 13: 115-129.
- Hilton, Denis J, and Ben R Slugoski. 1986. "Knowledge-Based Causal Attribution: The Abnormal Conditions Focus Model." *Psychological Review* 93 (1): 75-88.
- Hitchcock, Christopher. 1996a. "Farewell to Binary Causation." *Canadian Journal of Philosophy* 26: 267-282.
- Hitchcock, Christopher. 1996b. "The Role of Contrast in Causal and Explanatory Claims." *Synthese* 107 (3): 395-419.
- Hitchcock, Christopher. 2003. "Of Humean Bondage." *The British Journal for the Philosophy of Science* 54 (1): 1-25.
- Hitchcock, Christopher. 2007. "Three Concepts of Causation." *Philosophy Compass* 2 (3): 508-516.
- Hitchcock, Christopher. 2012. "Portable Causal Dependence." *Philosophy of Science* 79 (5): 942-951.

- Hitchcock, Christopher. 2017. "Actual Causation: What's the Use?" In *Making a Difference: Essays on the Philosophy of Causation*, by Helen Beebe, Christopher Hitchcock and Huw Price, 116-131. Oxford: Oxford University Press.
- Hitchcock, Christopher, and Joshua Knobe. 2009. "Cause and Norm." *The Journal of Philosophy* 106 (11): 587-612.
- Hornsby, Jennifer, and Jason Stanley. 2005. "Semantic Knowledge and Practical Knowledge." *Proceedings of the Aristotelian Society* 107-145.
- Icard, Thomas F, Jonathan F Kominsky, and Joshua Knobe. 2017. "Normality and actual causal strength." *Cognition* 161: 80-93.
- Kahneman, Daniel, and Amos Tversky. 1972. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3: 430-454.
- Kahneman, Daniel, and Amos Tversky. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5: 207-232.
- Kahneman, Daniel, and Amos Tversky. 1974. "Judgment under uncertainty: Heuristics and Biases." In *Judgment under uncertainty: Heuristics and Biases*, by Daniel Kahneman, Paul Slovic and Amos Tversky, 3-20. Cambridge: Cambridge University Press.
- Kahneman, Daniel, and Amos Tversky. 1982. "The simulation heuristic." In *Judgement under uncertainty: Heuristics and biases*, by Daniel Kahneman, Paul Slovic and Amos Tversky, 201-208. Cambridge: Cambridge University Press.

- Kahneman, Daniel, and Dale T Miller. 1986. "Norm Theory: Comparing Reality to Its Alternatives." *Psychological Review* 93 (2): 136-153.
- Kamp, Hans. 1981. "The Paradox of the Heap." *Journal of Symbolic Logic* 147: 225-277.
- Kennedy, Christopher. 1999. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. London: Routledge.
- Kennedy, Christopher. 2007. "Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives." *Linguistics and Philosophy* 1-45.
- Knobe, Joshua. 2009. "Folk judgments of causation." *Studies in History and Philosophy of Science* 40: 238–242.
- Knobe, Joshua. 2010. "Person as Scientist, Person as Moralist." *Behavioral and Brain Sciences* 33 (4): 315-365.
- Knobe, Joshua, and Ben Fraser. 2008. "Causal Judgment and Moral Judgment: Two Experiments." In *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, by Walter Sinnott-Armstrong, 441-448. Cambridge MA: MIT Press.
- Kominsky, Jonathan F, Jonathan Phillips, Tobias Gerstenberg, David Lagnado, and Joshua Knobe. 2015. "Causal Superseding." *Cognition* 137: 196-209.
- Kornblith, Hilary. 2002. *Knowledge and its place in nature*. Oxford: Oxford University Press.
- Kratzer, Angela. 1977. "What 'Must' and 'Can' Must and Can Mean." *Linguistics and Philosophy* 1: 337-355.

- Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kumar, Victor. 2014. "'Knowledge' as a natural kind term." *Synthese* 191: 439-457.
- Lewis, David. 1973. "Causation." *The Journal of Philosophy* 70 (17): 556-567.
- Lewis, David. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13 (4): 455-476.
- Lewis, David. 1986a. "Causation." In *Philosophical Papers Volume II*, by David Lewis, 159-213. Oxford: Oxford University Press.
- Lewis, David. 1986b. "Events." In *Philosophical Papers Volume II*, by David Lewis, 241-270. Oxford: Oxford University Press.
- Lewis, David. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74 (4): 549-567.
- Lewis, David. 2000. "Causation as Influence." *The Journal of Philosophy* 97 (4): 182-197.
- Lewis, David. 2004. "Void and Object." In *Causation and Counterfactuals*, by John Collins, Ned Hall and L A Paul, 277-290. Cambridge, Massachusetts: The MIT Press.
- Lipton, Peter. 1990. "Contrastive Explanation." *Royal Institute of Philosophy Supplements* 27: 247-266.
- Livengood, Jonathan, and Edouard Machery. 2007. "The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence." *Midwest Studies in Philosophy* 31: 107-127.

- Lombrozo, Tania. 2010. "Causal-explanatory pluralism: how intentions, functions, and mechanisms influence causal ascriptions." *Cognitive Psychology* 61 (4): 303-332.
- Longino, Helen. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- Lynch, Elizabeth B, John D Coley, and Douglas L Medin. 2000. "Tall is typical: central tendency, ideal dimensions, and graded category structure among tree experts and novices." *Memory & Cognition* 28 (1): 41-50.
- MacFarlane, John. 2007. "Relativism and Disagreement." *Philosophical Studies* 132: 17-31.
- MacFarlane, John. 2009. "Nonindexical Contextualism." *Synthese* 166 (2): 231-250.
- Maslen, Cei. 2004. "Causes, contrasts, and the nontransitivity of causation." In *Causation and Counterfactuals*, by Ned Hall, L A Paul and John Collins, 341-357. Cambridge: MIT Press.
- McDermott, Michael. 1995. "Redundant Causation." *The British Journal for the Philosophy of Science* 46 (4): 523-544.
- McGrath, Sarah. 2005. "Causation By Omission: A Dilemma." *Philosophical Studies* 123 (1/2): 125-148.
- Mellor, David Hugh. 1995. *The Facts of Causation*. London: Routledge.
- Menzies, Peter. 2004a. "Difference-making in Context." In *Causation and Counterfactuals*, by John Collins, Ned Hall and L A Paul, 138-180. London: The MIT Press.

- Menzies, Peter. 2004b. "Causal Models, Token Causation, and Processes." *Philosophy of Science* 71 (5): 820-832.
- Menzies, Peter. 2007. "Causation in Context." In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, by Huw Price and Richard Corry, 191-223. Oxford: Clarendon Press.
- Menzies, Peter. 2009. "Platitudes and Counterexamples." In *The Oxford Handbook of Causation*, by Helen Beebe, Christopher Hitchcock and Peter Menzies, 341-367. Oxford: Oxford University Press.
- Menzies, Peter. 2010. "Norms, causes, and alternative possibilities." *Behavioral and Brain Sciences* 33 (4): 346-347.
- Menzies, Peter. 2011. "The Role of Counterfactual Dependence in Causal Judgements." In *Understanding Counterfactuals, Understanding Causation: Issues in Philosophy and Psychology*, by Christoph Hoerl, Teresa McCormack and Sarah R Beck, 186-207. Oxford: Oxford University Press.
- Menzies, Peter, and Helen Beebe. 2020. *Counterfactual Theories of Causation*. Edited by Edward N Zalta. 21 December. Accessed January 1, 2021. <https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/>.
- Menzies, Peter, and Huw Price. 1993. "Causation as a Secondary Quality." *The British Journal for the Philosophy of Science* 44 (2): 187-203.

Millikan, Ruth Garrett. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge MA: MIT Press.

Moldovan, Andrei. 2019. "Descriptions and Tests for Polysemy." *Axiomathes* 1-21.

Montminy, Martin, and Andrew Russo. 2016. "A Defense of Causal Invariantism." *Analytic Philosophy* 57 (1): 49-75.

Neta, Ram. 2003a. "Contextualism and the Problem of the External World." *Philosophy and Phenomenological Research* 66 (1): 1-31.

Neta, Ram. 2003b. "Skepticism, Contextualism, and Semantic Self-Knowledge." *Philosophy and Phenomenological Research* 67 (2): 396-411.

Northcott, Robert. 2008. "Causation and Contrast Classes." *Philosophical Studies: An International Journal for Philosophy* 139 (1): 111-123.

Northcott, Robert. 2018. "Pre-emption cases may support, not undermine, the counterfactual theory of causation." *Synthese*.

OED Online. 2020a. *bank*, n.3. September. Accessed 10 01, 2020.
<https://oed.com/view/Entry/15237?rskey=zoWosN&result=3>.

OED Online. 2020b. *bank*, n.1. September. Accessed October 1, 2020.
<https://oed.com/view/Entry/15235?rskey=zoWosN&result=1>.

Paul, L A. 2000. "Aspect Causation." *Journal of Philosophy* 97 (4): 223-234.

- Phillips, Jonathan, Jamie Luguri, and Joshua Knobe. 2015. "Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities." *Cognition* 145: 30-42.
- Popper, Karl. 1963. *Conjectures and Refutations*. London: Routledge.
- Putnam, Hilary. 1982. "Why There Isn't a Ready-Made World." *Synthese* 51 (2): 141-167.
- Rabagliati, Hugh, and Jesse Snedeker. 2013. "The Truth About Chickens and Bats: Ambiguity Avoidance Distinguishes Types of Polysemy." *Psychological Science* 24 (7): 1354–1360.
- Ramsey, Frank. 1978. "General Propositions and Causality." In *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, by Frank Ramsey, 133-151. London: Routledge.
- Reichenbach, Hans. 1956. *The direction of time*. Berkeley: University of California Press.
- Rosch, Eleanor, and Carolyn B Mervis. 1975. "Family resemblances: Studies in the internal structure of categories." *Cognitive Psychology* 7 (4): 573-605.
- Rose, David. 2017. "Folk intuitions of actual causation: a two-pronged debunking explanation." *Philosophical Studies* 174 (5): 1323-1361.
- Samland, Jana, and Michael R Waldmann. 2016. "How prescriptive norms influence causal inferences." *Cognition* 156: 164-176.
- Sartorio, Carolina. 2010. "The Prince of Whales Problem for Counterfactual Theories of Causation." In *New Waves in Metaphysics*, by Allan Hazlett, 259-276. Basingstoke: Palgrave Macmillan.

- Saul, J M. 2017. "Racial Figleaves, the Shifting Boundaries of the Permissible, and the Rise of Donald Trump." *Philosophical Topics* 45 (2): 97-116.
- Schaffer, Jonathan. 2000. "Causation by Disconnection." *Philosophy of Science* 67 (2): 285-300.
- Schaffer, Jonathan. 2004. "From Contextualism to Contrastivism." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 119 (1/2): 73-103.
- Schaffer, Jonathan. 2005. "Contrastive Causation." *The Philosophical Review* 114 (3): 297-328.
- Schaffer, Jonathan. 2008. "The Contrast-sensitivity of Knowledge Ascriptions." *Social Epistemology* 22 (3): 235-245.
- Schaffer, Jonathan. 2012. "Causal Contextualism." In *Contrastivism in Philosophy: New Perspectives*, by Martijn Blaauw, 35-63. London: Routledge.
- Schaffer, Jonathan. 2016. "Grounding in the image of causation." *Philosophical Studies* 173 (1): 49-100.
- Schaffer, Jonathan, and Joshua Knobe. 2012. "Contrastive Knowledge Surveyed." *Noûs* 46 (4): 675–708.
- Schurz, Gerhard. 2001. "What Is 'Normal'? An Evolution-Theoretic Foundation for Normic Laws and Their Relation to Statistical Normality." *Philosophy of Science* 68 (4): 476-497.
- Sedgwick, Peter. 1982. *Psychopolitics*. New York: Harper and Row.

- Sinnott-Armstrong, Walter. 2004. "Classy Pyrrhonism." In *Pyrrhonian Skepticism*, by Walter Sinnott-Armstrong, 188-205. Oxford: Oxford University Press.
- Sinnott-Armstrong, Walter. 2006. *Moral Skepticisms*. Oxford: Oxford University Press.
- Sinnott-Armstrong, Walter. 2008. "A Contrastivist Manifesto." *Social Epistemology* 22: 257-270.
- Smith, Martin. 2016. *Between Probability and Certainty: What Justifies Belief*. Oxford: Oxford University Press.
- Snedegar, Justin. 2017. *Contrastive Reasons*. Oxford: Oxford University Press.
- Spellman, Barbara A. 1997. "Crediting causality." *Journal of Experimental Psychology: General* 126 (4): 323-348.
- Stalnaker, Robert. 1999. "Pragmatic Presuppositions." In *Context and Content: Essays on Intentionality in Speech and Thought*, by Robert Stalnaker, 47-62. Oxford: Oxford University Press.
- Stanley, Jason, and Zoltán Gendler Szabó. 2000. "On Quantifier Domain Restriction." *Mind and Language* 15: 219-261.
- Stapleton, Jane. 1994. *Product Liability*. London: Butterworth & Co.
- Steglich-Petersen, Asbjørn. 2012. "Against the Contrastive Account of Singular Causation." *British Journal for the Philosophy of Science* 63 (1): 115-143.

- Strawson, P. F. 1992. "Causation and Explanation." In *Analysis and Metaphysics: An Introduction to Philosophy*, by P. F. Strawson, 109-131. Oxford: Oxford University Press.
- Strössner, Corina. 2015. "Normality and Majority: Towards a Statistical Understanding of Normality Statements." *Erkenn* 80: 793-809.
- Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Pub. Co.
- Sytsma, Justin, Jonathan Livengood, and David Rose. 2012. "Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions." *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 814-820.
- Vácha, Jiri. 1978. "Biology and the Problem of Normality." *Scientia* 113 (9-12): 823-846.
- van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Vicente, Agustín, and Ingrid L. Falkum. 2017. *Polysemy*. 27 July. <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-325>.
- Viebahn, Emanuel. 2018. "Ambiguity and Zeugma." *Pacific Philosophical Quarterly* 749-762.
- Wachbroit, Robert. 1994. "Normality as a Biological Concept." *Philosophy of Science* 61 (4): 579-591.
- Westerståhl, Dag. 1989. "Aristotelian Syllogisms and the Generalized Quantifiers." *Studia Logica: An International Journal for Symbolic Logic* 48: 577-585.

Williams, Michael. 1991. *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*. London: Blackwell.

Woodward, James. 2004. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, James. 2006. "Sensitive and insensitive causation." *Philosophical Review* 115 (1): 1-50.

Zwicky, Arnold M, and Jerrold M Sadock. 1975. *Ambiguity Tests and How to Fail Them*. Vol. 4, in *Syntax and Semantics*, by John P Kimball, 1-36. New York: Seminar Press.