An Investigation into the Nature of Mathematical Objects

Samuel Dickson

PhD

University of York
Philosophy

March 2023

Abstract:

My thesis primarily concerns metaphysics, the philosophy of mathematics and the philosophy of causation. A common assumption is that mathematical objects, if they were to exist, would be non-spatiotemporal acausal abstracta. But this allows for some classic problems. The Benacerraf challenge concerns how it is that we could know about mathematical objects, given their acausal nature. The "makes no difference" objection says that, given the acausal nature of mathematical objects, the world would be the same as it is whether or not they existed, so we need not believe in them. Indispensability arguments may respond by saying that we have to believe in mathematical objects, but they do not respond to the core of these issues. I aim to solve these problems by suggesting that mathematical objects are a kind of 'in-between' object that is neither abstract nor concrete, I call such objects "exotic objects". I present a view that mathematical objects might be exotic by being non-spatiotemporal but causal. Mathematical objects are causal in virtue of constraining the physical world, i.e. the reason that 23 strawberries are indivisible between 3 people is *because* 23 is indivisible by 3. I argue that this constraint relation can be viewed as a kind of causation because it exhibits the relevant kind of counterfactual dependence, e.g. "Had 23 been divisible by 3 then…". But such statements are counterpossible rather than merely counterfactual so I also offer an account of why we should treat counterpossibles as non-trivial. I argue that the theory I propose maintains the advantages of platonist theories of mathematics whilst avoiding the classic problems mentioned above, but also avoiding committing to nominalism.

Contents:

Declaration:

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Large parts of chapter 4 have been published as a paper titled ‘Mathematical Causation’ in the journal *Analiza* [Dickson, S. (2021). Mathematical Causation. Analiza. (1). p. 85-102.] which developed from drafts of that chapter. Likewise, a version of the ideas presented throughout chapter 5 has been published as a paper titled ‘Against Vacuism in the journal Studia Semitoyczne (Semiotic Studies) [Dickson, S. (2022). Against Vacuism. Studia Semitoyczne (Semiotic Studies). 35(2). p. 7-33.].

Acknowledgements:

I am grateful to the Department of Philosophy at York for supporting my research by awarding me a Graduate Teaching Scholarship. Teaching was a fun experience, and I like to think it helped me to be more productive. In particular, I would like to thank David Worsley for facilitating this and making it as enjoyable as possible.

I would like to offer special thanks for all the general support along the way provided by the department. Julie Kay and Janet Eldred have always been available to answer any questions I had and offer help. I am grateful for this.

Thanks need to be offered to the philosophical community at York. I am grateful to have been able to help revitalise the postgraduate community in recent years with both Eleanor Byrne and Robin Pawlett-Howell by organising the social side of the PG community. We have facilitated more ridiculous but philosophical chats than I can count. It was helpful to remember that philosophy is not just something we happen to be doing, it is something we *enjoy* doing. The postgraduate community at York demonstrates that incredibly well.

I need to offer thanks to my supervisors, Mary Leng and Paul Noordhof and my TAP member, Rob Trueman. All have been great sources of clarification and discussion throughout my PhD and I would not have been able to develop my views in the way I have without them.

I would like to thank my partner Gaenor for tolerating my philosophical ramblings and being willing to engage and offer support when I needed it. Finally, I would like to give some credit to my mum. I talk about causation quite a lot in this thesis, so it's probably appropriate to note the counterfactual dependence that had she not encouraged me to read so much as a child, I wouldn't have done this, thanks

0 - Introduction

0.1 Scope of the thesis

Standard mathematical platonism holds that there are abstract mathematical objects that exist independently of our talk and practice concerning them. The oft-used illustration of this is that Pythagoras did not invent his theorem, but instead *discovered* that it was true. How platonists standardly characterise the abstract nature of mathematical objects is that they are non-spatiotemporal and acausal. Mathematical objects then are taken to be inert, they are idle entities. It is therefore unusual that they seem to have an indispensable role in our theories (see, Colyvan, 2001, 11). It is unusual because we tend to have strong intuitions that, in order for us to believe in the existence of x, x should *do* something. I share this intuition. Mathematical objects should *do* something. But it isn't clear that, as abstract objects, they *can* do anything. But nor do mathematical objects seem to be concrete. The focus of this thesis is to try to see if problems for platonistic conceptions of mathematics can be solved by appealing to a third kind of object in between the traditional concrete and abstract categories. I call this theory - according to which mathematical objects stand between the traditional categories of 'abstract' and 'concrete' - *Exotic Realism*. My thesis makes a contribution by carving out this category of 'in-between' objects that I call "exotic objects". A concrete object is something which exists in space and time and which has causal powers, by contrast, an abstract object is something which does not exist in space and time and does not have causal powers. Exotic objects, therefore, are *some but not all* of these things. For example, an object which is non-spatiotemporal but causal would be an exotic object. I will argue that this framework of exotic objects will be fruitful for metaphysics generally. We need a way to meaningfully talk about the different kinds of objects that theories posit, and the standard abstract/concrete dichotomy is unable to do this efficiently.

I make contributions by beginning to categorise the exotic objects framework, but also by offering up *Exotic Realism* as an option and arguing for it. I argue that mathematical objects are best conceived of as exotic objects. I hold that mathematical objects impact the world via mathematical constraint (as initially characterised by Lange, 2017). I draw an analogy between constraint and causation, via defining the key characteristics of constraint and emphasising its similarity to counterfactual accounts of causation. This analogy could be used to support either a strong or a weak claim, which I will respectively call **Con=G&C** and **weakCon=G&C**. **WeakCon=G&C** claims that constraint is not a kind of causation, but shares many similarities to causation so is part of the same wider class of dependence relations[1] and this licences us to appeal to constraint as an existence-indicating relation, much as we do with causation[2]. **Con=G&C** claims that constraint is a kind of grounding and a kind of causation, it represents a crossover case. There are, therefore, two different kinds of exotic object that one can view mathematical objects as. On **Con=G&C**, mathematical objects are exotic objects that are non-spatiotemporal but causal. On **weakCon=G&C**, mathematical objects are a different kind of exotic object, one that is non-spatiotemporal and acausal but involved in another important dependence relation, constraint. For the purposes of this thesis, I will aim to defend **Con=G&C,** but even if I am not successful in establishing constraint is a form of causation, the arguments made will nevertheless make **weakCon=G&C** plausible, and in fact only the weaker claim is essential for my account of the role of mathematical objects in our theories[3].

The constraint/causation analogy will ultimately be put in terms of counterfactuals. However, this analogy rests on the usage of counterpossibles. Given that the orthodox view of counterpossibles says that they are vacuously true, I make further contributions by providing a systematic explanation of why it is that we should treat counterpossibles as non-vacuous. Overall, my account is a “top-down” approach rather than “bottom-up”. I do not make assumptions about the nature of mathematical objects and then try to explain their usage in science etc., I look at their usage in science and the roles they are supposed to play and use this to make arguments about what their nature must be like.

---

[1]Karen Bennett’s (2017) ‘building relations’ are a good illustration of this. Whether constraint would form a non-causal subset of grounding as a building relation (2017, 12-13), or would be a distinct kind of non-causal building relation is not fixed by **weakCon=G&C** but the claim would be adaptable to either route.
[2]For example, see Armstrong’s (1978) discussion of the eleatic criterion in this way.
[3]Although see 7.4.1 of the conclusion for some discussion of future research opportunities that could help to decide the case between **Con=G&C** and **weakCon=G&C**.

0.2 Context

It will be helpful to briefly consider the context of the debate around mathematical objects. There are many theories of mathematical objects but I will focus on two. Firstly, mathematical platonism is arguably the 'orthodox' view of mathematics. Accordingly, it is worth highlighting some problems facing platonism in order to justify the move to *Exotic Realism*. Perhaps the fiercest competitor to platonism is nominalism, in its various forms. Nominalism claims to solve the problems of platonism by denying the existence of abstract mathematical objects. Accordingly, after considering the problems with platonism, I will likewise consider some problems with nominalism. Again, doing this will allow me to justify *Exotic Realism*.

0.2.1 Platonism

Standard mathematical platonism alleges that the existence of mathematical objects seems to follow from highly plausible features of mathematics. We refer to mathematical things with singular terms, a practice usually reserved for objects. We take mathematical statements to be informative and true, for example, it is an interesting and non-trivial fact that there are infinitely many prime numbers. This seems to be something we discover about mathematics rather than something we invented. Platonists claim that this can only be the case if the objects we are talking about actually exist. However, they cannot be the sort of object we are normally used to; they are unlike trees and tables, for example. Questions such as 'where is the number seven?' do not make sense, whereas such questions about more familiar objects are commonplace. Unless we are willing to accept substantial revisions to our standard assumptions about the nature of mathematical objects - for example about their lack of spatiotemporal location - mathematical objects could not be concrete. Given that the distinction between the concrete and abstract is taken to be exhaustive, it is claimed that mathematical objects are abstract. This is the platonist account of mathematical objects. However, this account faces tough challenges. For example, if numbers are intended to be abstract objects then it is mysterious how we could ever come to know about them. This problem is known as the epistemological challenge, inspired by Paul Benacerraf (1973). David Liggins formulates Field's (1989) version of the epistemological challenge rigorously in the following way:

Epistemological challenge:

1 - Mathematicians are reliable, in the sense that for almost every mathematical sentence **S**, if mathematicians accept **S**, then **S** is true.

2 - For belief in mathematics to be justified, it must at least in principle be possible to explain the reliability described in 1.

3 - If mathematical platonism is true, then this reliability cannot be explained even in principle.

Conclusion: If mathematical platonism is true then mathematical beliefs cannot be justified.

(Liggins, 2010, 74)

The main point of discussion has often been with 3. Penelope Maddy (1990, 37-67) formulates a defence of premise three based on the causal theory of knowledge. If we have knowledge of something then that thing must be involved in a causal chain that generates our belief. Given that mathematical objects are supposed to be abstract objects, they are causally inert. Therefore, we cannot have knowledge of them. This means that we cannot explain the reliability alluded to in 1 and so we should reject mathematical platonism. The platonist account is in a tricky position, if mathematical objects are abstract, we cannot know about them. But it seems we do know some mathematical truths.

Another challenge the platonist must address is the makes no difference argument (**MND**). As Balaguer says:

> "if mathematical objects are causally inert, then whether or not there really *exist* any mathematical objects has no bearing on the physical world…"
> Balaguer (1998, 157)

Baker (2003) formalises this challenge in the following way:

Makes-no-difference:

(1) If there were no mathematical objects then (according to platonism) this would make no difference to the concrete, physical world.

(2) Hence (on the platonist picture) we have no reason to believe in the existence of mathematical objects.

(Baker, 2003, 247).

This is a good formalisation but it builds in that it is an objection to Platonism. However, this argument can be levelled at many different realist conceptions of mathematics. For example, it seems that one could make a parallel challenge to platonist-structuralist conceptions of mathematics. These arguments rely on the alleged acausality of mathematical objects. Whilst platonists posit this as a key feature of mathematical objects, one would not have to be a platonist to describe mathematical objects as acausal. The core of the argument seems to be: Mathematical objects play no causal role in the world so the world would be the same as it is if they did not exist, so we do not need to believe in their existence. Therefore, we should formalise a general version of this argument (**MND**) in the following way:

MND:
i. If something plays no causal role in the world, we do not need to believe it exists.
ii. Mathematical objects play no causal role in the world
C. So, we do not need to believe in the existence of mathematical objects.

In response to these issues, platonists have often turned to indispensability arguments. Indispensability arguments take many forms but here is a simplified form of the standard Quine/Putnam version:

IA:
a - We ought to have ontological commitment to all and only those entities apparent reference to which is indispensable to our best scientific theories.
b - Apparent reference to mathematical entities is indispensable to our best scientific theories.
C - Therefore, we ought to have ontological commitment to mathematical entities.
(Colyvan, 2001, 11)

By appealing to **IA**, platonists hope that we can make an abductive leap to mathematical objects and sidestep the concerns of Benacerraf’s epistemological challenge as well as **MND**. Benacerraf challenges us that we have no reason to believe in the existence of mathematical objects, based on their acausality. The **IA** aims to challenge premise **2** of the epistemological challenge. We can be justified in mathematical beliefs based on indispensability rather than on explaining our belief-forming processes' reliability. In the case of **MND**, the **IA** aims to challenge **i**. Sometimes, belief in entities is indispensable even if they are acausal. An issue remains, however. Indispensability arguments tell us that we ought to believe in mathematical objects but tell us nothing about what mathematical objects are like, and they tell us nothing about their role in the world. In a way, this actually makes the epistemological challenge and **MND** even more severe because it amplifies the mysterious nature of mathematical objects. If mathematical objects are acausal, it is surely bizarre that they are indispensable for science. It also does little to address the core of the epistemological challenge. It is strange that we are capable of having reliable beliefs about acausal entities.

Indispensability arguments are silent about the nature of mathematical objects. They might tell us that we have to believe in mathematical objects, but they do not say anything about what mathematical objects *are*. Given the assumption that the abstract/concrete distinction is exhaustive, the theories that rely on **IA** tend to stipulate that mathematical objects are abstract. But this means that mathematical objects do not *do* anything for us. They are idle indispensable entities. But it seems that things are indispensable, or should only be indispensable, by virtue of *playing some role*.

0.2.2 Nominalism

There are many different kinds of nominalism and they all face interesting issues. I will focus on fictionalism as a candidate for being a strong nominalistic alternative to platonism. Fictionalism itself comes in many forms but a fair summary of fictionalist views is that, whilst mathematics, taken at face value, is a body of claims about the things that platonists like to talk about, abstract objects and the like, mathematical theories are literally false because there are no such abstract objects.

It is claimed that indispensability arguments pose a serious problem for nominalism. The fictionalist might want to say that we should talk *as if* mathematical objects existed, whilst resisting commitment to them, but this flies in the face of indispensability. Indispensability arguments tell us that we *have* to believe in mathematical objects. A linked problem here is to point out that it would be strange if belief in something non-existent in the fictionalist sense was indispensable. The fictionalist would have a difficult task in explaining why it is that talking about things which do not exist is indispensable. Of course, some fictionalists try just this. One way to respond is to say that the indispensable role that mathematics plays is representational rather than ontologically indicative, e.g. see Saatsi, 2011.

However, even if mathematics merely plays a representational role, it is undeniably an incredibly successful tool. The success of this tool is easy to make sense of under *Exotic Realism*. Under nominalism, this success is a fact in need of explanation but with no easy explanation in sight[4]. If mathematics is just a useful fiction, what might explain its success is a mystery. Fictionalists do indeed have responses to such issues, however, one may not wish to endorse such responses and these risk losing some of the intuitive force of fictionalism. Instead, it is worth noting that *Exotic Realism* is simply not affected by these issues. One key motivation of nominalism is that it removes our commitment to 'spooky' metaphysical entities like mathematical objects. *Exotic Realism* aims to show that exotic objects are not spooky. Indeed, I will argue that we are already committed to the existence of some exotic objects. In particular, in 1.2.2 I argue that the objects of physics are exotic. Modern physics has started to blur the lines about whether it investigates concrete objects or not[5]. Phenomena in quantum and particle physics make us question our idea of what it is for something to exist in space. I believe quantum fields and subatomic particles like electrons and quarks are good candidates for exotic objects, and recognising this will hopefully make us more open to the idea that mathematical objects could occupy a similar ontological space. By making mathematical objects less spooky, I remove a key motivation to adopt nominalism. By also responding to the problems that platonism faces, I position *Exotic Realism* as an attractive alternative theory of mathematical objects. These two sections have not presented knockdown arguments to platonism and nominalism, but they highlight issues that those accounts face that *Exotic Realism* does not. This in itself is an advantage.

0.3 Advantages of *Exotic Realism*

Platonists and nominalists assume that, if they were to exist, mathematical objects would be abstract, and most importantly acausal. But this assumption is rarely argued for. With this acausality assumption, the success and applicability of mathematics might seem mysterious. This is why my thesis challenges this assumption. I argue that mathematical objects might

[4]One might also think that it is mysterious under platonism as well. If mathematical objects play no active role in the world, why we need to appeal to them and why they matter so much is mysterious.
[5]See 6.3.2 for a discussion of how the terms 'concrete' and 'physical' come apart.

play a causal role in the world via constraint. Because they are causal, we can easily respond to Benacerraf and **MND**. We do have a causal path to knowledge about mathematics and mathematical objects *do* make a difference to the world. This account, therefore, supports the **IA**. It could provide an answer as to *why* mathematical objects are indispensable. But it also strengthens the **IA**. When a platonist appeals to the **IA**, it is mysterious how acausal mathematical objects could play this indispensable role. For *Exotic Realism*, how they could play this indispensable role is no longer mysterious, because they are causal. By conceiving of mathematical objects as exotic objects with causal powers my account can avoid the traditional problems with platonism in a straightforward way. I can also avoid the complexities and difficulties of nominalism without appealing to mysterious entities.

0.4 Summary of argument

As stated, I want to form a theory of mathematical objects as exotic objects. The thesis will build towards this theory and then respond to problems that arise with it. For the reasons outlined above, and as will be emphasised again later, *Exotic Realism* is an attempt at providing a better account of what mathematical objects would have to be in order to play the role they are supposed to play in our theories. This is the principal aim of my thesis, but there is a secondary project that should be highlighted. This thesis serves as a form of case study for the exotic objects framework. This thesis is intended to show the kind of work that can be done, and the kind of theories developed, within the framework of exotic objects. The thesis is limited by scope to mathematical objects, but other theories in metaphysics could benefit by changing to exotic objects rather than abstract ones. As noted, by conceiving of mathematical objects as exotic objects, we can solve some of the most important problems afflicting platonism and other realist theories. Furthermore, the key arguments *for* realism are actually stronger when used to support exotic objects.

0.4.1 Chapter 1 - Exotic objects

I start by discussing the notion of an exotic object. The aim here is to expand the theoretical space concerning the different kinds of objects to allow room for new theories concerning purported abstract objects. This theoretical room is described and the beginnings of a framework for categorising this space are offered. This opens avenues for future research in categorising the different kinds of exotic objects and what examples of them we might already be appealing to/discussing. Some preliminary issues with this account are highlighted and responded to. I will be describing mathematical objects as the kind of exotic object that is causal but non-spatiotemporal. To say this, I need to be clear about what notion of causation I am using. I argue that mathematical objects are causal because they constrain the way the world is, and further that constraint is a form of both grounding and causation. To make this analogy I first need to discuss grounding and its parallels with causation.

0.4.2 Chapter 2 - Grounding

This chapter discusses the notion of grounding and its parallels to causation. There is supposed to be a great deal of crossover between the hallmark features of grounding and causation. Problem cases can be produced for each of these hallmark features. Importantly, not only are the problem cases structurally similar for causation and grounding, the proposed solutions are as well. This points to a deeper similarity between causation and grounding than people often credit. This might be enough to make the case that causation and grounding are the same thing, but one does not have to make this commitment. One can settle for the weaker thesis that they are both examples of a broader class of dependence relations. This does not harm my overall account of mathematics because I want to restrict my claims to a specific subset of grounding and causation, rather than make a global claim about both relationships. What this chapter does, though, is make a compelling case that grounding and causation are very close, and as such should be treated analogously.

0.4.3 Chapter 3 - Causation

To make the analogy between grounding and causation, and indeed constraint and causation, I need to specify the notion of causation I am appealing to. This chapter proposes some theories of causation that would be compatible with the claims I want to make about mathematics. These theories generally lie in the counterfactual tradition, but probabilistic versions are also available. In any case, counterfactual theories of causation are successful and can solve a lot of the problems that afflict process theories. For simplicity, it will be easier to talk about my account if I settle on one account of causation in particular. I settle on the interventionist approach because, given the tricky notion of an intervention, it might seem to be difficult to apply to mathematical causation. I show how my view of mathematical constraint can be seen as causal under this interventionist picture. By doing this with a difficult theory, I show the flexibility of my account. Given that I can work with some difficult theories of causation, my account of mathematics will equally work with simpler theories. After this chapter, therefore, I adopt an interventionist interpretation of causation.

0.4 4 Chapter 4 - Mathematical constraint

With the groundwork of the previous chapters, it becomes possible to talk about mathematical constraint more specifically. In this chapter, I distinguish between **Con=G&C** and **weakCon=G&C** in terms of the causation/constraint analogy. I discuss the constraint relationship, as defined by Lange 2017, including mathematical constraint. There are constraint relationships that we should judge to be straightforwardly causal relationships, I call these ‘physical constraint’ cases. These physical constraint cases have a specific structure and behave in a specific way under interventions/counterfactual scenarios. I argue that mathematical constraint cases are structurally identical. Moreover, they behave in the same way under interventions/counterfactual scenarios. In the physical constraint case, this is sufficient to describe the relationship as causal, so it should be enough in the mathematical case. We should therefore endorse **Con=G&C**. At this stage, the full picture of *Exotic Realism* is clear, what remains is to address the various problems that have arisen.

0.4.5 Chapter 5 - Counterpossibles

In giving a counterfactual analysis of grounding, and in particular, in dealing with counterfactuals about mathematics, we end up dealing with a specific class of counterfactuals, counterpossibles. These are counterfactuals whose antecedent (and/or consequent) is an impossibility. Orthodoxy states that counterpossibles are vacuously true. This is a problem for *Exotic Realism* because I require non-vacuously true/false counterpossibles to back up the kinds of causal claims I want to make. This chapter deals with this problem. I offer an account of why we should treat counterpossibles as non-vacuous. With this account in place, we can make the causal claims we need to about mathematical objects. I also make substantial contributions to the literature on counterpossibles; combining work done in multiple areas showing that the orthodox view is wrong as well as providing the wider framework that is needed to explain *why* orthodoxy is wrong.

0.4.6 Chapter 6 - General problems

Throughout Chapter 2, multiple resistant problems emerge that suggest important differences between causation and grounding. This chapter shows how these problems either do not affect the comparison of constraint with causation or how they can be responded to. Problems specifically concerning the constraint/causation analogy from Chapter 4 are also dealt with here. Some of these problems include notions of determinacy, synchronicity and the causal closure of the physical. This chapter deals with all of these problems. In doing this, I aim to show that *Exotic Realism* is stable and can respond to a slew of difficulties.

0.5 Conclusion

Having argued for **Con=G&C**, I will argue that we can view mathematical objects as being causal via mathematical constraint, and so as exotic rather than abstract. *Exotic Realism* is a better account of mathematical objects than alternatives because it better accounts for the role they are intended to play. Moreover, I argue that some of the objects of physics are exotic. A key strength of nominalism lies in platonic mathematical objects being a unique kind of object that we can reject. In claiming that we are already committed to exotic objects, I remove this option from the nominalist and put them on the back foot. We cannot simply reject mathematical objects on the basis that they are exotic, because we would end up rejecting vast swathes of science. *Exotic Realism*, then, is stronger than platonism. But I also discuss **weakCon=G&C** and in particular how some problems that will require future research might tell in favour of it. Either way, this reconception of mathematical objects within the exotic objects framework avoids some key problems of platonism by committing to mathematical objects playing an active role in the world. We can spell out this role by appealing to counterpossibles. To support this usage, I developed an account that says why mathematical counterpossibles are non-vacuous that can be extended to counterpossibles more generally. My thesis, therefore, has applications beyond the philosophy of mathematics. By developing the exotic objects framework, I contribute to the wider metaphysical debate. This account of mathematical objects can be seen as a case study of what the exotic objects framework can do for us, and why we should adopt it. This framework is available to other metaphysical theories that want to posit non-concrete objects. But it also makes it easier for people who want to be deflationary about ontology to explain exactly what it is they are deflating. We need to talk about objects, and we need to do so in a more systematised way than we currently do, that is what the exotic objects framework can provide.

Chapter 1 - Exotic Objects

1.1 The basic idea

A common thought amongst philosophers is that in order for us to justifiably class something as existing, it must impact the world in some way, it must make a difference. I will refer to this affecting of the world as impingement[6]. If there were to be entities that failed to impinge in any way then, as Armstrong says "...if these other sorts of entity do not act upon this world of particulars, there seems to be no methodological reason for postulating them" (Armstrong, 1978, 132). Three common kinds of impingement are the following: existing in space; existing in time; and being causal[7]. Typically, a concrete object is defined as something which does all of these things. It is also typically thought that these characteristics are tied together somehow, that either an object exhibits all of these things, or it exhibits none of them. Therefore, opposed to concrete objects, there are abstract objects, things which fail to impinge in any of these ways. This characterisation of abstract objects in terms of something they fail to do (in this case, failing to impinge) is an example of Lewis' (1986c, 83-84) 'negative way'. Given the typical view that these impingement characteristics are tied together, the distinction between the abstract and the concrete is viewed as exhaustive. Nominalists tend to use the claim that abstract objects fail to impinge in one of the usual three ways as evidence that they do not exist[8]. However, the claim that those characteristics mentioned above are tied together is rarely argued for independently. Moreover, there are many examples in the literature which seem to assume some independence between them.

I want to propose that between concrete and abstract objects is a third category, that I will term the category of 'exotic' objects. Exotic objects impinge in at least one, but not all, of those three ways, meaning that there are more kinds of objects than we previously thought. I use the word "object" in the broadest possible sense, synonymous with "thing" or "entity". There are lots of different sorts of things that exist, and these are all different sorts of objects.

---

[6]This term will be explored and defined later.
[7]I will refer to these as impingement characteristics.
[8]There may well be other characteristics, or ways of impinging, that are important for ontology, but for the purpose of discussing mathematical objects and this thesis, I will merely be focussing on these three. These three seem to be particularly important impingement characteristics, and I will discuss a distinction between these and other kinds of characteristics that objects might have in 1.3.

They are alike in that they all impinge on the world, even if not in terms of how[9]. For example, we could consider objects which lack a spatiotemporal location but which enter into causal or causal-like relations and as such still impinge on the world. It is this kind of object which I will suggest mathematical objects are. The difficulties facing theories of mathematical objects as abstract objects can be avoided if we ensure that mathematical objects affect reality in some way. I contend that the best formulation of this is to suggest they enter into causal relationships, and this will be explored elsewhere in the thesis; but if mathematical objects are causal, then they are not abstract, and are therefore exotic objects.

The suggestion of in-between objects, as we shall see below, is not unique. Most often objects will be posited that simply do not fit as either abstract or concrete (as posited). On occasion, people will use an explicitly in-between term, e.g. Parsons (2009, 37) refers to quasi-concrete objects in this way and Robinson (2014) refers to quasi-abstract objects similarly. Even less frequent in the literature is any effort to systematise these in-between objects. Davies (2019) can be seen as such an attempt, although a different method to mine. Davies attempts to create a hierarchy of abstract objects, differentiated by their identity conditions. For Davies, the hierarchy is differentiated because some abstract objects include spatiotemporality/causality in their equivalence relations (2019, 822), whereas others do not. Davies does not explicitly talk in terms of in-between objects but merely different kinds of abstract objects. This point will come up again in 1.6.3 and as I shall argue, this is functionally equivalent to my proposal. So, although Davies categorises mathematical objects differently from how I ultimately plan to, the spirit of his idea is consistent with my own.

One of the key contributions of this thesis lies in an effort to systematise these in-between objects. We need a framework to talk about what proponents of them are actually talking about. Lowe (1998) discusses his own categorisation of the different kinds of things. Lowe brings up an important point when discussing his proposed categories, he says that they "should be admitted as genuine categories, at least in the sense of being categories of entity which could intelligibly be supposed to be exemplified, whether or not exemplars of all of them do in fact exist." (1998, 83). This is what the exotic objects framework can do for us. One does not *have* to believe in any exotic objects, but in order to have fruitful metaphysical

[9]In this sense, I take "object" to be a family resemblance term in the sense that "game" is a family resemblance term. Two games may share no mutual properties but are both nonetheless examples of games, in the same way, we may have different kinds of objects that share no relevant characteristics but are still nonetheless objects.

debate, one should be able to describe and understand what rival metaphysicians are talking about, and what sorts of things would exist if they were right. The project of starting to construct this framework is one of the key aims and contributions of this thesis and will also be a fruitful point for future research (see 1.3 for more discussion of this).

Many theories posit in-between objects, implicitly or explicitly. These theories may go wrong, but they do not go wrong merely by positing in-between objects. Labelling all non-concrete objects as abstract does not do justice to the differences in the *kind* of things that these objects are. As such, the table below is aimed at clarifying the different kinds of non-concrete objects that there might be. We need a system to describe these in-between objects and develop this new characterisation of objects. With this in mind, it is time to develop the idea of an exotic object. Exotic objects may impinge in some ways, but not all, as such there are many more different types of objects than usually thought. The table below will show each potential[10] type of object, I will then discuss some potential examples in the literature of things that already seem to be treated as exotic objects, even if not by that name.

1.2 Exotic Objects

[10]In the sense of Lowe, 1998, 83 above.

Object	Examples	Spatial	Temporal	Causal
Concrete objects	Tables, chairs etc.	1	1	1
i.		1	1	0
ii.		1	0	1
iii.		1	0	0
iv.		0	1	1
v.		0	1	0
vi.		0	0	1
Abstract objects	Universals, propositions	0	0	0

I take it that criteria for existence are rules about how the values can be distributed in the table, and consequently what sort of things we can allow. For example, the Eleatic criterion can be interpreted to state that it is not possible to set the value of the causal variable to 0. Likewise, one can envision spatial and temporal criteria, stating that their respective variables must also be 1. Taken together we get the thesis that only concrete objects exist. But we will now consider some examples of objects which do not fit neatly as either abstract or concrete. These objects are, in my terms, all different kinds of exotic objects, and they can be described in the table above. A clarification is necessary at this stage that I am not committing to the existence of objects to fill every row on this table, indeed I think that some rows will remain unfilled. This could be due to metaphysical laws, e.g. perhaps it is not possible to be spatial without also being temporal (but not vice versa) for example. Again, I refer to the Lowe quote from 1.1, we need a framework to describe what metaphysicians are discussing, and the exotic objects framework can do that for us. Accordingly, I will not give an example object for each row and merely focus on some examples that are already being appealed to in philosophy, as well as some justifiable reclassifications of objects. In particular, I will also discuss where mathematical objects can fit as exotic objects.

1.2.1 Row i: Spatiotemporal but acausal

Certain interpretations of epiphenomenalism describe mental phenomena in ways which would suggest they fit into this category. Clearly mental phenomena have a finite temporal existence, thoughts are often fleeting and pass away in moments. One might also describe them as spatial, if we cannot narrow them down to the brain we can at least narrow them down to the specific body that the mind inhabits. However, epiphenomenalism asserts that whilst being caused by physical things, these mental phenomena cannot themselves have a causal influence on the physical world. There are strong and weak versions of epiphenomenalism, the weak version says that mental phenomena can be caused by other mental phenomena. The strong version says that only physical phenomena can cause mental phenomena, in other words, mental phenomena have no causal powers whatsoever. It seems then, that what the strong epiphenomenalist is describing is something that does not neatly fit into the concrete object category, but prima facie also seems non-abstract. Given this, it seems that mental phenomena would be best described as exotic objects. Another object which would fit into this category is the Epiphenomenalons described by Forrest (1982). These are so-called ‘useless particles’ that occupy positions in space and time but never interact with anything or cause anything to happen. If such objects existed they would be neither concrete nor abstract strictly speaking, but rather exotic.

Colyvan (1998) argues that when we call something causal, what we really mean is ‘causal with respect to us’. Colyvan goes on to discuss things which fail to be causal with us and so should be classed as acausal, e.g very distant stars. One might respond that we can still describe such objects as causal because they might causally interact with other things or themselves, even if they do not causally interact with us. We can still call such objects causal because we are familiar with the *type* of causation they engage in. For abstract objects, not only are they not causal with us, we cannot even understand what it would mean for them to be causal with anything, so we say they are acausal. I think this is a good point. However, there are objects which this might not cover. For example, possible worlds that have different physical laws to our own. In such worlds, we simply are not familiar with the kinds of causal relations that take place. One might think that we might simply be able to say that the entities in such worlds are causal but in unfamiliar ways. However, such a response could also be used in the case of abstract objects. Perhaps abstract objects causally interact with one another in some unfamiliar way. If we can postulate that in the case of objects in possible worlds, we should be able to in the case of abstract objects. In order to avoid this conclusion, we would have to say that such things are simply not causal. But this means that such things are exotic rather than abstract/concrete.

1.2.2 Row iv: Temporal and causal but non-spatial

Robinson (2014) proposes that nation-states are objects of this type. He proposes that nation-states do not have definite boundaries, borders shift all the time and when territories get brought into the mix, they become non-contiguous objects. Rather than attempt to explain the spatial properties of such an object, Robinson denies that states exist in space. However, they clearly have an existence in time. At a certain point in history, legislation was signed and passed that brought a state into existence and at a point in the future, this state may well dissolve. Additionally, Robinson wants to say that these states have causal powers in virtue of the representatives who act on their behalf. Notice how we say that the UK declared war on Iraq, not that the UK delegation declared war on the delegation from Iraq, much like the fact that Iraq was not required to rejoin the UN after the overthrowing of Saddam Hussein's regime, because the state was the entity that was a UN member, not the particular government (Robinson, 2014, 21). It seems plausible to say that states exist in time and have causal powers, whilst failing to exist in space, given the above difficulties with specifying an exact spatial location for them. But this means that states cannot be concrete objects in the way they are usually conceived. Nor do we want to say that states are fully abstract objects. Instead, Robinson dubs them 'quasi-abstract' objects, but these are easily accounted for as exotic objects within my framework.

Subatomic particles and the objects of quantum physics are arguably also examples of this kind. It seems pretty clear that electrons, for example, are causal, we can detect them on instruments. It also seems pretty clear that they are temporal, they came to exist at a certain time and could cease to exist if they collided with an anti-electron. However, the status of existing in space is trickier[11], wave-particle duality and phenomena like superposition throw doubt on whether or not, prior to measurement, it even makes sense to talk about electrons having a position in space at all. For example, consider the following quotes:

[11]This will be discussed in more detail in 6.3.2 when I will argue that the notion of spatial is ill-defined and underpins the definition of physical/concrete, so it is questionable whether such examples are in fact spatial/concrete.

> If we ask, for instance, whether the position of the electron remains the same, we must say "no"; if we ask whether the electron's position changes with time, we must say "no"; if we ask whether the electron is at rest, we must say "no"; if we ask whether it is in motion, we must say "no." (Oppenheimer, 2014, 31)

> At every instant a grain of sand has a definite position and velocity. This is not the case with an electron. (Born, 1954)

> It is not possible to give “a description of the motion of the corpuscles in our familiar space-time concepts” (Heisenberg, 1926, 412) [in (Camilleri, 2007, 181)].

These quotes are of course by no means conclusive yet they should lead us to doubt whether our intuitive concept of something existing in space is even applicable to the objects of fundamental physics. This is once again nicely illustrated in the following quote:

> According to Democritus, atoms had lost the qualities like colour, taste, etc., they only occupied space… but in modern physics, atoms lose this last property, they possess geometrical qualities in no higher degree than colour, taste, etc…
> (Heisenberg, 1952).

There is much more research to be done in this area, the question is by no means settled. But the above should at least motivate us to consider the idea that the objects of physics have radically different features to, or at least lack some of the features we think are inherent to, concrete objects. As Matthias Egg says:

> the “particles” of fundamental physics in fact have so little in common with ordinary objects that it becomes doubtful to even place them in the same category
> (Egg, 2021, 10)

This is why I want to suggest that the objects of fundamental physics are in fact exotic objects. Hopefully the above has made this a less radical idea. There is also an advantage for me in doing this. We are already committed to the existence of the objects of fundamental physics (or at least most of us are). If these turn out to be exotic, then we are already committed to exotic objects and so it becomes much harder for the nominalist to dismiss other examples of proposed exotic objects merely because they fail to be concrete. This makes construing mathematical objects as exotic objects much more plausible.

1.2.3 Row vi: Non-spatiotemporal but causal

The folk conception of causation seems to require that it involves spatiotemporal interaction, so one might wonder if this category presents a problem. However, we might simply reject this intuition that causation requires spatiotemporal interaction. Counterfactual accounts of causation make no inherent mention of spatial interaction, instead, causation is a counterfactual dependence relation, in whichever way it is realised. One way we might use this counterfactual notion of causation is to highlight the parallels with the grounding relation (see chapter 2). If we say that grounding is a kind of causation, then in fact lots of purportedly abstract objects which are intended to do some grounding work will turn out to be this kind of exotic object. A weaker version of this view would be to consider grounding and causation as both being kinds of a broader class of dependence relations.

This is the category I want to assert mathematical objects belong to. Mathematical objects ground certain things in the world by constraining them. This constraint is a kind of causation, even though the objects themselves do not exist in space and time. This conception could appeal to metaphysicians who normally postulate abstract objects, but are concerned with makes-no-difference or epistemological worries about them. On my view, mathematical objects may be non-spatiotemporal but they operate on the world via constraint, which is a causal relation. To use a classic example, one cannot divide 23 strawberries (without cutting them) equally between 3 people *because* 23 is prime. If this relation is a kind of causation, then mathematical objects cannot be abstract and must be exotic. One objection to this will be discussed later but is worth mentioning here. The issue is the mysteriousness/impossibility of something non-concrete or non-physical affecting the physical world. This problem is discussed widely in the literature on the philosophy of mind as the problem of the causal closure of the physical/completeness of physics (e.g. Gibb, 2010 & Lowe, 2000). This problem will be discussed in detail in chapter 6.3 but to preview responses now I do not think the notion of ‘physical’ used in these arguments is well defined and I contend that this fails to threaten the view of mathematical objects being causal.

To describe the examples so far discussed is not to say that all the theories which appeal to them are correct. It is not my focus to defend these theories, I merely want to pull on intuitions to show that the notion of exotic objects is not as mysterious as it might seem. It is something that people have already begun to appeal to, implicitly or explicitly. Recognising this fact opens the door and allows us to formalise the kinds of exotic objects and begin to talk about them. For my purposes I wish to investigate whether mathematical objects might best be described as exotic, to do this I will restrict myself to row **vi** as I think this is the most promising route to help describe mathematical objects. Having considered some potential examples of exotic objects we can now see the table more filled in with those examples, before moving on to critically discuss the framework.

Object	Examples	Spatial	Temporal	Causal
Concrete objects	**Tables, chairs etc.**	**1**	**1**	**1**
i.	**Mental Epiphenomena**	**1**	**1**	**0**
ii.	**?**	**1**	**0**	**1**
iii.	**?**	**1**	**0**	**0**
iv.	**Nation states, subatomic particles**	**0**	**1**	**1**
v.	**?**	**0**	**1**	**0**
vi.	**Mathematical objects**	**0**	**0**	**1**
Abstract objects	**Universals, propositions**	**0**	**0**	**0**

1.3 Is this table complete?

One point to address is whether or not the table above is complete, i.e. whether there are types of exotic objects beyond those listed above. I think it is very likely that there are more kinds of exotic objects for several reasons. Firstly, it might seem that rather than differentiate a lot of supposed abstract objects, the above table just forces lots of different examples into one row and that this characterisation is inaccurate. This is worth exploring in detail.

Take the example of subatomic particles as discussed above, I argue that these can be characterised as exotic because they are straightforwardly temporal and causal, but fail to be spatial in the right sense. However, take the example of so-called fictional entities, things like Sherlock Holmes and Narnia. These are also things that fail to be spatial because they do not *actually* exist, they do not have a concrete physical presence at the actual world, however, they seem to be temporal (they were created, and will be forgotten), and it is arguable that they are causal (they have influenced a lot of people). This would mean that these were also a kind of exotic object, but problematically, exactly the same *kind* of exotic object as subatomic particles. This seems mistaken. Whatever the subject matter of physics is, it is a fundamentally different *kind* of thing to the subject matter of metaphysics, they simply do not study the same kinds of objects. We should maintain the commitment that these are fundamentally different *types* of things, and the exotic objects framework gives us a way to do that. One way that we might say that fictional objects are a different kind of thing to subatomic particles is that fictional objects are mind-dependent, whereas quarks would exist whether there were minds or not. This is not something covered in the table above and so shows that something is missing in that above characterisation. This does not need to be a fatal problem for the idea of exotic objects, just for the way they have been characterised above. There is a difference between the ways that things can impinge on the world, and between the characteristics that impinging objects have. An illustration of this will be helpful.

The key point is this; there are many ways for an object to be, but only a few ways for an object to impinge on the world. This can be illustrated by a simple example, let us take a man-made artwork such as the Mona Lisa, and a natural object, e.g. the white cliffs of Dover. The Mona Lisa is, in a relevant (but very weak) sense, mind-dependent, not in the sense that its existence is only sustained by being entertained by minds, but in the sense that it would not have come into existence were it not for minds. On the other hand, the white cliffs of Dover are mind-independent, their existence is neither sustained by minds nor were they created by minds, they exist independently of minds. This suffices to make them different kinds of objects, but note that they are not different in terms of how they impinge on the world. Both are spatiotemporal and causal, they both impinge on the world in exactly the same way. These are different kinds of objects, not in terms of how they impinge, but in terms of their other characteristics. Their impingement is, in a sense, *how* they exist, how they take up ‘ontological space’, and how we come to know about them. Their mind-dependence/independence is not how they impact the world but is instead a feature of that impact. This illustration should suffice to show the difference between impingement and other characteristics. For the purposes of my thesis, I am interested in impingement, to truly categorise the different kinds of objects, exotic and otherwise, we would need to add many more columns to the table above to differentiate them all. Mind-independence is likely to be one of these features, perhaps fundamentality is another, in the sense that a quark (if it is a fundamental particle) is a different *kind* of thing to a neutron. Identifying these characteristics and categorising the objects they suggest into distinct kinds would be an interesting task but is beyond the scope of this thesis.

The table above is definitely not complete concerning non-impingement characteristics of objects, but it also seems it might not be complete concerning impingement characteristics. This can be illustrated by an appeal to contingently non-concrete objects (CnCs). These are concrete objects which are not actual, but might have been (Williamson, 2013, 8), e.g. a sibling that was not born or a continent that was not formed. Williamson plausibly claims that such objects are not abstract, because abstract objects are necessarily abstract, and that CnCs are concrete at some possible worlds. However, some theorists are willing to admit that CnCs do exist at our world, just not as concrete objects, e.g. Linsky & Zalta (1996, 284). Clearly, such things would have to be some kind of exotic object, but it is not clear which type. CnCs are clearly not causal, they are not there to do any causing. But CnCs are also non-spatiotemporal, it makes no sense to ask where or when the continent that did not form exists. In the table above, such things would be categorised as abstract objects, yet intuitively they are not, because such things *could* have been concrete. CnCs represent an object type that does not yet have a row on the table. Given that an object must impinge in some way to be counted as existing, what this suggests is that there are in fact more ways to impinge than we have covered. Perhaps grounding or truthmaking (or something else) are ways to impinge, and Williamson’s CnCs impinge by doing one of these things. Once again, this presents an interesting project for metaphysics but this falls beyond the scope of this thesis. I do not need to provide a complete table of exotic objects, I need only provide the characterisation above because ultimately I want to say that mathematical objects exhibit causal powers whilst being non-spatiotemporal. The only part of the exotic objects table that I need to consider is the part shown above.

1.4 Why think there are exotic objects at all?

Part of the motivation for the exotic objects framework is that, as should be clear, we are already tacitly relying on something like this. Different philosophers have different conceptions of what exactly abstract objects are. A lot of philosophical theories already seem to rely on objects which are neither concrete nor abstract as traditionally defined, but they are more often than not reluctant to place them in a middle category and so describe them as concrete or abstract, or change the definition in an ad hoc way. This leads to problems. Nominalists are often keen to reject any appeal to abstract objects, but if we can show that they are already committed to objects which are not traditionally concrete, then these views may soften.

Using the term ‘abstract object’ to refer to all these things is unhelpful because it suggests that they are all tokens of the same type and that they can therefore be subjected to the same objections. Explicitly using ‘exotic objects’ as an umbrella term for the different kinds of non-concrete objects would clear up some of this confusion. It would become easier to see why theories that postulate non-spatiotemporal, causal objects cannot be criticised in the same way as those which postulate acausal, nontemporal but spatial objects, for example.

Terminological issues aside, this seems an accurate way of categorising the kinds of things there are. Quantum physics seems to describe real entities, and they might not be concrete as traditionally defined. I also think mathematical objects exist, again as non-concrete objects. But I do not think these are the same specific type of object. Aluminium, magnesium and gold are all elements, and they are all metals, but that does not mean they are all *the same element*. Similarly, the above-mentioned examples are all kinds of objects, and they are all non-concrete, but they are not the same *kind* of non-concrete object. I want to label the non-concrete as exotic because it makes the dialectic easier by distancing itself from traditional worries and confusions with abstract objects.

1.5 Why think mathematical objects are exotic?

One might wonder why this re-categorisation of mathematical objects as exotic objects helps, and so it will be worth making that explicit. The main (and possibly the strongest) argument for mathematical objects comes in the form of the **IA** as discussed in the introduction. Typically, this is offered as an argument for mathematical platonism, but the argument is, at its core, neutral about the specific kind of mathematical object we must believe in. Part of the motivation behind this argument is the mysterious nature of mathematical objects. Mathematical objects are *so* unusual that we have to provide independent motivation for believing in them. Hence we say that even if mathematical objects are disconnected abstracta, we *have* to believe in their existence thanks to the practices of science. This goes some way to defeating the "makes no difference" (MND) arguments that get levelled at mathematical objects. It is worth restating the **MND** from the introduction:

MND:
i. If something plays no causal role in the world, we do not need to believe it exists.
ii. Mathematical objects play no causal role in the world
C. So, we do not need to believe in the existence of mathematical objects.

The **IA** targets premise **i**. It provides an independent reason that we need to believe in mathematical objects, even if they are non-causal. The purpose of my account is to nullify the force of **MND** as well as Benacerraf-style objections against mathematical objects. This is done via a denial of **ii**. Platonists and nominalists often assume **ii**, but rarely argue for it. Perhaps one reason for this is because it is assumed that mathematical objects are non-spatiotemporal and it is difficult to imagine how causal interaction could take place between non-spatiotemporal objects and ordinary spatiotemporal objects. Balaguer seems to note this assumption:

"For since platonists maintain that mathematical objects exist outside of spacetime, they endorse what we might call the principle of causal isolation (PCI), which says that there are no causal interactions between mathematical and physical objects." (Balaguer, 1998, 110)

Chapter 4 builds a case to argue that **PCI** is false, this interaction might be difficult to imagine but it is coherent. The advantages of denying **ii**/**PCI** should now be clear, we straightforwardly sidestep **MND** and Benacerraf challenges in their tracks, mathematical objects *are* causal and so how we can come to know about them etc. can be accounted for because we can show that they are causal with respect to us.

This answers one form of the question, “what is a good motivation for thinking mathematical objects are exotic?”. We can solve some key problems in the philosophy of mathematics by doing so. But one might also want to ask why the conception of mathematical objects as exotic is a good fit for what mathematical objects might be. This question would target how the role of mathematical objects would be filled by their being exotic over abstract. This is where my view will have significant strength over nominalistic and platonistic theories. *If* mathematical objects exist, it is quite clear they must play *some* role in the world. They have to *do* something, to be responsible for *something*. This “have” is not necessarily a prerequisite for existence, perhaps there could be idle entities, we simply would not know about them, but as **MND** says, we should not believe in such things. Instead this “have” means that mathematical objects must do something for *us to count them as existing*. Platonism has a lot of positive features, it captures the apparent meaning and truth value of our mathematical language straightforwardly. However, it holds that mathematical objects are non-spatiotemporal and acausal, they do not *do* anything. This is an issue, there is a strong intuitive pull that things that exist have an impact on the world, the world is different because they exist (and not merely by the fact of their existence), platonic mathematical objects do not have this impact. Nominalism dispenses with this worry in any number of ways, mathematical objects do not exist as abstract objects, so these worries do not arise. But it has its own problems, our intuitive views have to be substantially changed, we also might have to be radically revisionary about mathematical/scientific practice. If we dispense with mathematical objects completely (like in Field, 2016) then our scientific language can become much more complex, and the ideas we wish to express become harder to express and harder to understand. If we are left with a choice between this complexity or with the mystery of platonism, neither an entirely attractive option, we can see the appeal of *Exotic Realism* as it aims to sidestep these issues.

1.6 Objections

1.6.1 Mathematical objects are paradigm abstracta

One might object to my account by saying that surely if anything is abstract, mathematical objects are, they are the archetypal abstract object. Reclassifying mathematical objects as non-abstract would be a category mistake. There is certainly something to this objection, abstract objects are often defined by way of example. More often than not, abstract objects are defined as things like mathematical objects. If we say that mathematical objects are no longer abstract then we might lose the sense of the terms we are talking about. This is part of the motivation to construct the exotic objects framework. The orthodox view of mathematical objects is that, if they were to exist, they would be abstract. Given this, an analysis of mathematical objects as non-abstract (but not concrete) might seem illegitimate. It seems that if we fail to account for that significant datum, we miss something important from the account. Philosophers seem to think the project should be "Given that mathematical objects would be abstract, let's provide an account of them". We have changed the subject if we say they are not abstract. This objection would be misguided because it is perfectly legitimate to ask the question "Are we sure mathematical objects would be abstract?".

I would like to borrow some work from Dennett (1993) to explain this.

> "Suppose anthropologists were to discover a tribe that believed in a hitherto-unheard-of god of the forest, called Feenoman. Upon learning of Feenoman, the anthropologists are faced with a fundamental choice: they may convert to the native religion and believe wholeheartedly in the real existence and good works of Feenoman, or they can *study the cult* with an agnostic attitude." (Dennett, 1993, 82).

Let us suppose there are various disagreements over some features of Feenoman. Some claim his eyes are blue, others brown. But eventually anthropologists manage to establish a definitive list of the features of Feenoman as described by this religion (Dennett, 1993, 82). The Feenomanists obviously believe in the existence of their deity, but they accept they may have got some details wrong. They could be corrected about these, perhaps by interaction with Feenoman itself. Now we imagine what would happen if the anthropologists partially confirmed these beliefs:

> "...consider what would happen if an anthropologist confirmed that there really was a blue-eyed fellow named Feenoman, who healed the sick and swung through the trees like Tarzan. Not a god, and not capable of flying or being in two places at once, but still undoubtedly the real source of most of the sightings, legends and beliefs of the Feenomanists".
> (Dennett, 1993, 84).

Some believers might then revise their beliefs to fit in line with these new findings, because the anthropologists could point and say "what you thought was a god, was actually this fellow over here". Some may resist this, "no, that isn't the *real* Feenoman, that's a pale imitation". But that seems illegitimate. We have accounted for enough of the facts that were described by the sightings and legends to say that this is what was being talked about, it was just being misdescribed.[12] To circle back to mathematics, we should engage in a similar process. Orthodoxy told us that mathematical objects are abstract. We have encountered conflicting data to this, arguments like Benacerraf's challenge and **MND** objection seem to suggest that abstract mathematical objects would not exist (or we should not believe they do), but we might still want to believe in mathematical objects. To account for these problems we will have to shift our perspective. Maybe we can still believe that mathematical objects are non-spatiotemporal, but we might have to reject this acausal assumption. To account for this conflicting data, mathematical objects need to *do* something. Of course, people may shake their heads and say "no, you're not talking about *real* mathematical objects, *real* mathematical objects are abstract", but this seems illegitimate. Of course we are no longer talking about mathematical objects that are abstract, but that is because mathematical objects are not abstract. Mathematical objects as exotic objects fits enough of our prior intuitions of them to be considered as what we were *actually* talking about in these descriptions of platonic entities, it is just that we have been misdescribing them. Sticking to believing in mathematical objects being abstract is comparable to still insisting that Feenoman exists in light of the explanation of its actions having been carried out by a person. It might be right, but it seems that in light of the evidence we ought to alter our beliefs accordingly. We should characterise mathematical objects as exotic.

[12]A third option is to simply dispense with the belief all together "there is no Feenoman". In light of how many features of the god Feenoman have been captured by the flesh and blood Feenoman, this does not seem the correct strategy, but one could take it.

1.6.2 Are there any traditional abstract objects?

One point which could be made about my account is whether, under it, there are any abstract objects as traditionally conceived, i.e. non-spatiotemporal and acausal. It seems that all the things we might normally think of as abstract, e.g. mathematical objects and universals, instead find a space under exotic objects, so there is no need for the category of abstract objects.

I think this is a good point, and that it is likely that all the things that different theories currently conceive of as abstract are actually exotic. I am happy to say that abstract objects probably do not exist, as in 1.6.1 I think that for any abstract object, we can better account for its role if they turn out to be exotic. However, I do not want to rule out abstract objects by definition, perhaps there is something that cannot be satisfactorily accounted for as exotic/concrete. As in 1.3, it is also probable that there are many more ways for something to impinge on reality than merely being spatial, temporal or causal. There may be many more kinds of exotic objects than those that are displayed in the tables. Perhaps one of these object kinds turns out to be a good stand-in for abstract objects, but for my purposes, I do not need to concern myself with other kinds of exotic objects, or with whether there are abstract objects. The main point I am arguing for is that mathematical objects would be best described as a specific kind of exotic object.

1.6.3 Are there any concrete objects?

Likewise, one might wonder whether my account commits me to saying that there are no concrete objects. It seems quite clear that spatiotemporal and causal things exist, so if my account denies this is the case, that is a problem. The reason my account might lead to this is that I want to say that objects studied by quantum physics, e.g. quarks, are not concrete because their existence in space is not as clear-cut as tables and chairs and any spatial properties we ascribe to them are very different. Given that such things build up the ordinary objects we see around us, we might have to say ordinary objects are not concrete. I can respond to this in two ways: I can argue that this is simply not a feature of my account, or I can accept this is what my account says, but explain why it is not a problem.

Firstly, I could simply say that whilst quarks and other subatomic particles are exotic, that does not mean that macroscopic objects like tables and chairs are exotic, they are of course concrete. There might be a difficulty accounting for this though. The hard problem of consciousness challenges materialists to explain how consciousness can arise from purely unconscious material things. I might be faced with a similar issue accounting for how the concrete can emerge from the non-concrete; how spatial things like trees can emerge from non-spatial quarks. This initially might seem a difficult problem to solve but this is not the case. We can look to physics for some precedent in at least saying that such emergence is not mysterious. Special relativity tells us that matter and energy are the same thing, or more specifically that matter is in some sense condensed energy. Energy does not have the physical *presence* that we think of matter as having, it does not have a straightforward existence in space and its causal powers can be questioned (energy is *how* things have causal potency, but it is not immediately clear that energy is itself causal). But scientists are happy to say that matter emerges from energy in some way. During a process of emergence, new properties can develop. We already accept this so it should not be considered mysterious to say that the concrete could emerge from the non-concrete. *That* something concrete emerges from the non-concrete should not be mysterious, *how* something might do this is mysterious but that is up to science to explain and plausible accounts of how things are built up already exist.

Alternatively, I could accept that there are no concrete objects. One (incorrect) reading of this would be that my account says that those objects which are described as concrete do not exist. But this is not the case, my account would at most say that such things do not exist *as* concrete objects. Tables and chairs exist, but they are non-concrete. This is a much more plausible view, I believe in the same sorts of things as the proponent of concrete objects, merely categorising them differently. This categorisation is not ad hoc, if the objects of quantum physics are not concrete, and the concrete cannot emerge from the non-concrete, then tables and chairs are non-concrete. But that conditional is something we should all accept, that just explains how our description of the content of the world could be wrong. That conditional does not change the actual content of the world. Moreover, there is some plausibility to this view. We tend to think the kinds of things there are in the world is a discrete range, but things inevitably turn out to be more of a spectrum than we initially think. This could easily also be the case for objecthood. We might have initially thought there were only abstract or concrete objects, but in fact, it turns out there are neither of these kinds of objects and everything is somewhere in the middle. For the purposes of my thesis, I am indifferent to which of these responses to take, all that matters is that they are available to *Exotic Realism* should they be required.

One might wonder whether, given the above, we should simply redefine concrete objects rather than postulate exotic objects. On a straightforward view of space, electrons might not exist in space[13], but clearly electrons do exist in space, so we need to be clearer in what we mean by spatial. Perhaps a probabilistic location is enough, or perhaps even just the potential of having a spatial location is enough to say that something is spatial, and therefore concrete. The problem with this is that it may commit us to saying some obviously non-spatial things are spatial. For example, we might want to say that electrons are spatial because they have a chance of being located somewhere. But compare this with Williamson's CnCs discussed earlier. They are objects which do not exist in space, because they do not exist actually, but they might have done. In a sense then, CnCs have a chance of being located, so they would turn out to be spatial. But this seems false. If CnCs were spatial, space would be filled with innumerable quantities of them. It just seems obvious that CnCs are non-spatial. There will be other 'metaphysically weird' entities that will similarly fall under this definition of spatial. Some of these may already be considered causal and temporal so it will turn out that such entities would be the same *kind* of thing as the objects of fundamental physics, and as noted by Egg (2021, 10), this seems wrong.

One way to respond to this would be to bite the bullet and say that by fulfilling this definition of spatial, things like CnCs simply might be concrete. But this is not a problem because there are different *kinds* of concrete objects. To do that, they would have to show that such different types of concrete objects differ with respect to some non-impingement characteristics. At this stage, the parallels with my earlier suggestions should be clear. Fundamentally, this picture is not too different from my own. We both would agree that there were many different types of objects, but where I would say there are different types of exotic objects, they would say there are different types of concrete objects. At this point, if these different kinds of concrete object capture what I mean by exotic objects, then this fails to be a significant metaphysical disagreement. As such, I will simply put this disagreement to one side.

[13] Again, this will be discussed more in 6.3.2.

1.7 Section Summary

I want to maintain the intuitions about mathematical objects that platonism captures (along with the explanatory power they bring), but also avoid the mysteriousness that nominalistic theories criticise. The intuition that mathematical objects must have some impact if they exist is a compelling one, and I agree with it, but maintain that the complexities introduced by nominalistic theories can be avoided. As in the table in 1.2, I think there are three main ways in which something can 'impact' or 'impinge' on the world: having spatial existence; having temporal existence; or being causal. I do not think mathematical objects are spatial, one cannot go out and discover an object that is the number "2" in space. Nor does it seem that numbers are identical with all their token instantiations. For one, it is not clear what it would mean for something to instantiate "twoness"[14]. Nor do I think we can identify mathematical objects with things like physical sets, for example, we would run into difficulties determining if a set containing a bat and a ball, and a set containing a table and a chair are both in some way *the* number two[15]. This would seem to take us straight back to a non-spatial universal-like object. Further, I do not think a trope-theoretic account of numbers akin to an account of properties would work. I do not think the sets mentioned above would both be number-two-tropes, but distinct *things* (in that sense of tropes). There are already an infinite amount of natural numbers, and the infinite number of sets of two things that can be produced would lead to an infinite number of tropes (of the infinite series of natural numbers), this seems to be too many things in existence to commit to. Finally, if mathematical objects were purely concrete objects, it brings the question of whether or not they could be destroyed. If all the concrete mathematical objects were destroyed (if that were possible) it would seem to mean that mathematical truths no longer held true, and this seems mistaken.

Mathematical objects do not seem to be temporal objects. A strong intuition for many people is that before Pythagoras developed his theorem, it was already the case that within a right-angled triangle, the area of the square whose side is the hypotenuse is equal to the sum of the areas of the squares on the other two sides. It was not the case that Pythagoras invented

[14]An aristotelian property view of mathematical objects is discussed and argued against along these lines in 4.8.2.

[15]Another alternative would be to identify the natural numbers with classes. One could define the number 2 as the class which contains all the classes of individuals which contain two elements (Coquand, 2022). One issue with this theory is how this extends into the very high numbers and indeed into infinities. A googol is larger than the (estimated) number of atoms in the universe and so it seems difficult to define that as a class containing all the classes of googol-many-individuals. This difficulty magnifies with even higher numbers. Perhaps such difficulties are fixable, but at first glance this seems an unattractive theory.

or *produced* this truth, he merely discovered it. The mathematical truths were already true before any minds started thinking about them, and given that they were already true, we have no reason to think that they will cease to be the case when minds cease to exist; mathematical truths are always true. The objects that mathematics describes did not come to exist at a particular time, and they will not cease to exist at another, they exist outside of time.

This leaves the one remaining place that mathematical objects might impact or impinge on the world as causality. The orthodox view is that non-spatiotemporal entities cannot engage causally with spatiotemporal entities. So, if mathematical objects are non-spatiotemporal, then it seems they should not be able to causally interact with the ordinary world. This was Balaguer's **PCI** and a large motivation for the assumption of **PCI** has surely been the inability to imagine or describe how this causal process might take place, this was the problem levelled at Descartes' substance dualist account of the mind (Robinson, 2020). This objection arose when our conception of causation was very much of the billiard ball variety, one object bumps into another. Many different theories of causation do not place so much pressure on physical interaction in this way, for example, counterfactual accounts and more specifically, interventionism. It is under such accounts that I can propose how we can understand causation between mathematical objects and the straightforward physical world. One way to start such a project is to look at the grounding relation, as it seems that mathematical facts ground certain physical facts. However, grounding is typically taken to be a relation distinct from causation. Chapter 2 will investigate the parallels between the grounding relation and causation. Ultimately, these might well be distinct relations but they are not as far apart as one might initially think, indeed they are so close together that I hold there is a slight overlap between a kind of grounding and a kind of causation, constraint. This relation will be investigated more thoroughly in Chapter 4. It is via this constraint relation that mathematical objects impinge on and impact the world, it is the way in which they *do* something, and so the way in which we can say that they exist. In developing this account of *Exotic Realism*, I hope to account for the platonist intuitions and sidestep the "mysteriousness" challenge that nominalists may make.

Chapter 2 - Grounding

2.1 Grounding vs Causation

This chapter aims to bring grounding and causation as close together as possible. The arguments developed here will aim to show that grounding is simply a kind of causation, however, there is substantial pressure against this in the literature. It may turn out to be impossible to generally identify grounding as a type of causation, but this need not be detrimental to my overall thesis. In the end, all that I require is that a specific kind of grounding relationship, constraint, is a kind of causation. In arguing for the unity of grounding and causation, I hope to make the idea of a small crossover between them much more plausible. Having the unity of grounding and causation throughout this chapter, I will move on to discuss constraint more specifically in chapter 4. There I will argue that, under an interventionist treatment, we can show that constraint is causation. Regardless of whether one agrees that grounding and causation are identical, what the following does is bring them closer together. Even if distinct, grounding and causation are close enough that we should treat them similarly. For example, the Eleatic criterion is treated as a hallmark for existence, given that grounding is arguably close to causation, we might account for this with a 'directed dependence relation' criterion instead. In order to be posited to exist, something must enter into either a causal or grounding relationship. With this understanding of the chapter's scope we will start off with preliminary discussion of causation and grounding before moving onto the analogy between them.

Grounding is a technical notion capturing the function of natural language statements like 'because' and 'in virtue of'. It is a privileged dependence relation like causation as it backs explanations. An 'in virtue of' claim is a case of grounding when the conditional holds of some sort of necessity. For example, the ball is red and round in virtue of the fact that the ball is red and the ball is round; in this case, the necessity involved is logical. Different kinds of necessity will produce grounding statements of particular interest to different areas (Fine, 2012). In the natural necessity case, the in-virtue-of relation will be of use to science, such as in the sentence "the fact that the particle is accelerating obtains *in virtue of the fact* it is being acted on by some net positive force". Similarly, "The fact that action *X* is the good thing to do obtains *in virtue of* being the action which produces the most net happiness." will be of particular interest to ethicists. Because the in-virtue-of relation is different in each case, the explanatory task of the discipline involved will be slightly different because the explanatory relationship is different. Fine (2012) asserts that grounding is of particular importance to philosophy, indeed, grounding is a central notion to metaphysical inquiry. Specifically to what he calls realist metaphysics. This is the kind of metaphysics associated with questions of what is real. Ultimately, questions in realist metaphysics are questions of what grounds what, and so plainly without grounding these questions would disappear. To explain phenomena in science we look to what causes their occurrence, in philosophy we should look to what grounds their occurrence.

In contrast to realist metaphysics, Fine (2012, 40) articulates the idea of naive metaphysics. Whilst realist metaphysics is concerned with questions of what is real, e.g. the discussion of whether or not numbers are real, or instead merely a useful fiction; naive metaphysics is concerned with the nature of things, irrespective of whether they exist or not, for example, the debate around whether or not fictional characters are created by their authors. Naive metaphysics can obviously involve the notion of ground, but need not. In realist metaphysics, however, Fine (2012, 41) asserts that the notion of ground is essential and that without it the discipline would disappear. This is because, at the base level, questions about whether or not numbers exist are questions about what grounds the facts about numbers and the statements we make involving them. Someone committed to the existence of numbers would say that such statements are ultimately grounded in mathematical objects, whereas a nominalist would resist this claim. Many philosophers may consider themselves as not concerned with the nature of grounding. For example, fictionalists are not concerned with what grounds the truth of mathematical claims because they believe mathematical objects do not exist and all mathematical claims are literally false. Fine could argue that such views would be translatable into the grounding context. Perhaps he could interpret fictionalists as saying that mathematical claims are false because they do not have grounds. What is clear is that grounding discussions are present in many distinct areas of metaphysics, making it a very important notion in philosophical enquiry and its adoption and usage would illuminate many different enquiries (Fine, 2012, 41).

Causation is a relation supposed to capture the meaning of similar statements, such as "domino X fell because domino Y hit it". Different causal statements will be linked with different kinds of necessity. Both "lightning causes forest fires" and "electrons repel each other" appear to be causal, but appear to involve different sorts of necessity/contingency. Because of this, the explanation of why each holds is likely to be different. It seems that causation and grounding are similar, and one might go so far as to assert that they are one and the same relation. Wilson (2018a, 723) seems to make this strong claim (which he labels **G=MC**) when he says that grounding simply is metaphysical causation. But one might think that only a weaker claim is being made here, that grounding and causation are simply of the same broader class of dependence relations (e.g. see Bennett 2017, 71), call this **weakG=MC**. Ultimately, all my account requires is **weakG=MC**. Nonetheless I think **G=MC** is plausible. This chapter will survey the case for **G=MC**. To some, the issues with **G=MC** that are discussed will suffice for its falsity. But hopefully this chapter will bring grounding and causation close enough together to show that **weakG=MC** is plausibly true. In chapter 4 I will make a similar distinction between strong and weak claims, and all that is required for either of those is **weakG=MC**, so this chapter serves to lay some convincing groundwork towards those claims.

2.2 What is grounding?

Having considered a general comparison of the two relations, it is worth discussing grounding in more detail so that we can better grasp the comparison being made. Grounding is most often described by way of paradigm cases. Grasping these cases helps to form a familial notion of grounding so that we can identify it in further cases. Take some paradigm cases:

Singleton: The existence of Socrates grounds the existence of singleton Socrates.
Double-negation: The truth of P grounds the truth of ¬¬P.
Truthmaking: The existence of Socrates grounds the truth of 'Socrates exists'.
Mind/body: My being in brain state B grounds my being in mental state M.
Part/whole: The existence of my head grounds my existence.
(Wilson, 2018a, 731-732)

In all these examples we are supposed to intuitively grasp a common relation, grounding. But, a critic might allege that rather than one relation in all the cases, we have different relations mistakenly grouped under the same heading of 'grounding'. Conflating these relations is what confuses the concept of grounding. We could respond by highlighting how causation seems similarly diffuse. Take these paradigm cases:

Entanglement: The measurement of the spin of electron x fixes the spin of electron y.
Pool balls: The first ball impacting the second ball caused the second ball to travel into the pocket.
Planets: The gravitational forces between them cause both planets to be exactly where they are.
Inspiration: The inspirational talk I listened to caused me to go to university.

Arguably, in each example we have a straightforward case of causation, however, the examples are very distinct. An initial option to unite these causal examples and distinguish them from grounding is energy transfer. We can tell a story about energy passing from the cause to the effect, which is not the case between grounder and groundee. However, **Entanglement** does not seem to involve energy transfer, given that it happens instantly and at distances which would make it *impossible* for energy to be transferred. But some scientists seem to treat **Entanglement** as a causal case, we should not dispute this[16]. Indeed, if we denied it was causal despite the fact it exhibits other hallmark features of causal relations, e.g. we can repeat this process in a laboratory, then we must either deem it to be a new kind of dependence relation or a kind of grounding relation. Introducing a new dependence relation just to account for **Entanglement** seems to be unnecessary, because it seems that causation can do the job. One thing that does seem to be common in all these cases is counterfactual dependence. Without the cause, the effect would not have happened. It could be suggested that this dependence is the identifying mark of the causal relation. However, such counterfactual dependence is also present in the grounding examples above, so we have failed to distinguish these causal relationships from the grounding examples. It seems tricky to identify a common thread between all cases which marks them as causal, which will not also be present in the grounding case. But this should not lead us to say that causation is an

[16]Friedell (2020, 136) raises the methodological worry that we should not let our views about causation influence interpretations of quantum mechanics. Whilst there are these interpretations that see phenomena like entanglement as causal, we should instead aim to account for that in our causal theories.

incoherent notion, or that there are different relations at play. Instead, some may allege that it seems reasonable and pragmatic to say that there is one relation at play here and the differences only appear because of the differences in relata (Wilson, 2018a, 730).

Before beginning the analogy with grounding and causation, I will first briefly explain what is meant by constraint, as proposed by Lange (2017). Lange develops the idea of what he calls explanation by constraint by way of example. A mathematical explanation by constraint would be that the reason Jane cannot evenly divide her 23 strawberries equally between her 3 children is *because* 23 is not divisible by 3; the mathematics involved here constrains the actions she can perform. I want to assert that mathematical constraint is a kind of causation. Chapter 4 argues for this but, for now, this brief overview will suffice as and when constraint comes up during this discussion. Now that we understand grounding, we can begin developing the analogy proper with causation.

2.3 G=MC

Wilson (2018a) proposes the idea that grounding is metaphysical causation, a thesis he calls **G=MC**. Grounding is simply a case of causation, and hence grounding explanations are causal explanations. Wilson alleges that grounding and causation have a lot of “persistent parallels” that are indicative of their identity. Furthermore, **G=MC** is theoretically attractive from parsimony concerns because we no longer need separate theories of grounding and causation (2018a, 724). Many of the persistent parallels emerge from alleged *problems* with analyses of grounding and causation. The fact that the problems are structurally analogous is an indication of a deep underlying structural similarity between the relations. This similarity is even further supported by the fact that the *responses* to these difficulties are parallel. Wilson argues that **G=MC** is the best explanation of these facts. To build this case, and draw grounding as close together as possible, we will consider these alleged disanalogies in turn and slowly build the case for the deep structural similarity that Wilson identifies. We will first consider transitivity, asymmetry and irreflexivity. These are alleged to be hallmark features of causation and grounding, but suffer similar problem cases. After that, we will consider some criticisms of **G=MC** considered by Wilson (2018a) and others. In each case, we will aim to show that the problems are parallel and can be responded to in parallel ways, or at least that the disanalogy is not genuine. As a fallback, remember that I am not endorsing **G=MC**, although I am sympathetic. All I need to do is draw grounding and causation as close together as possible to make the case that constraint is a kind of causation. I will note along the way those disanalogies that seem most significant for **G=MC** and we will turn back to those in chapter 6 where I will argue that they do not present a problem for my claim that constraint is a kind of causation.

2.3.1 Transitivity

If one billiard ball hits a second, which goes on to hit a third, it seems reasonable to say that the first billiard ball caused the third to move; causation seems transitive. Likewise, grounding seems transitive, as Schaffer (2012) says:

"if the physical system grounds the chemical arrangement, and the chemical arrangement grounds the biological organism, then it is natural to thereby infer that the physical system must ground the biological organism."
(Schaffer, 2012, 122)

Intuition tells us that both grounding and causation are transitive, that is to say that for both '$\forall x \forall y \forall z((Rxy \wedge Ryz) \rightarrow Rxz)$' is true. However, there seem to be counterexamples. Hall (2004) develops the example of a hiker who sees a boulder rolling towards them and ducks; the boulder then passes overhead and they survive. Plausibly, the falling rock caused their ducking, and their ducking caused their survival, but we want to resist saying that the falling rock caused their survival. It seems in this case that we have true causal relations in operation but that transitivity has failed. For the transitivity of grounding, Schaffer (2012, 126-127) discusses a case involving a dented sphere, **O**, arguing that the dent in **O** grounds **O** having determinate shape **S***, and that **O** having **S*** grounds **O** being near-spherical, but that the dent does not ground **O** being near-spherical. Again, it looks like we have genuine cases of grounding but that transitivity has failed. It is worth noting that the similarity in these problem cases suggests a similarity between the relations of grounding and causation, which can be seen as a positive in favour of **G=MC**. However, we should consider how we might respond to the above. Three options present themselves when considering how we might handle these cases: we can explain the problem away; we can deny that transitivity fails, or; we can accept the transitivity failure. We will consider these in turn.

Skow (2018, 3-6) discusses a distinction between stative and non-stative verbs. Skow claims that stative verbs commit us to the existence of states, whereas non-stative verbs commit us to the existence of events. The distinction can be seen with a simple example. The phrase "Jones was paddling the boat" is a non-stative verb as it commits us to an event, namely Jones' paddling of the boat. Conversely, the phrase "Jones is tall" is stative, and does not commit us to an event (Skow, 2018, 3). Skow is committed to the idea that only events can be causes, and not states, and so it seems that stative phrases would not be causal. Something similar might be happening in the Hall and Schaffer cases above. It seems that 'near-sphericality' and 'survival' are states rather than events. I suggest that some implicit understanding of this stative/non-stative distinction is what causes people to reject that these are both causal claims. Perhaps people are not truly denying transitivity in these cases, but instead denying that states are involved in causal relationships. When Hall says that falling rocks are not the sort of things that cause survival, that is because 'survival' is not the sort of thing that is caused. Similarly, near-sphericality is not the sort of thing that is grounded by dents, because it is not the sort of thing that is grounded.

Alternatively, one could bite the bullet by saying that transitivity is not violated. The dent does ground near-sphericality and the falling rock does cause survival. We can motivate this a bit more by redescribing the boulder case in a way that does not change its meaning. Instead of saying that the boulder causes the hiker to survive, we say that the boulder causes a near-miss. It seems that falling boulders are exactly the sorts of things that cause near-misses, so I assert that we should just accept transitivity in this case. Similarly, I do not feel the intuitive pull that the dent in the ball does not ground it being near-spherical. Schaffer claims (2012, 127) that the reason the dent does not ground the dented sphere having determinate shape **S*** is that the sphere would have had that shape regardless of the dent. I do not feel the force of Schaffer's intuition here. Dents in balls are the sorts of things that ground near-sphericality because they are the sorts of things which prevent balls from being fully spherical. To use a causal analogy, we can accept that in getting four out of five numbers, you nearly win the lottery. If the reason you got four out of five is because you were unable to pick a fifth number, your inability did cause you to nearly win the lottery. It did so because inability to pick a number is exactly the sort of thing that causes you to not pick all of the correct numbers, and this just is what it is to nearly win the lottery. In both cases above, we could similarly maintain transitivity. It seems very strange to say that the actual shape that the sphere has, dent included, has absolutely no effect on the determinate shape that the sphere has, but this is what Schaffer's view seems to require.

Finally, we could maintain that transitivity does indeed fail in both cases, but because the cases are structurally similar then an explanation of, or different response to, this will also be structurally similar. Once again, this would suggest a structural similarity between the relations. As it happens, this is the route that we will ultimately take when discussing constraint[17]. It is more helpful to view causation and constraint as contrastive relations which dodges any transitivity problems like those above.

2.3.2 Irreflexivity & asymmetry

Both causation and grounding are traditionally considered to be irreflexive. Irreflexivity is a property of relations, specifically it is the property that holds of a relation *R* when, '$\forall x \neg Rxx$' is true. In the causal case, this means that for any **X**, **X** cannot cause itself. In the grounding case, **X** cannot ground itself. In the causation case, this has been questioned by appeal to time travel cases, e.g. consider the case of someone who constructed a time machine and then travels back in order to give their past self the designs, or consider (allegedly) self causing things, e.g. the universe or a deity. Reflexive causation as above has been described extensively in the literature so it seems plausible to allow it in principle (Wilson, 2018a). Likewise, grounding is traditionally thought of as irreflexive but it is possible to construct cases where it is a reflexive relation. For example, some people think that identity is an explanatory notion and therefore a potential for a reflexive grounding relation. There may be such a relationship present between mental states and brain states under certain formulations of mind-brain identity theory (e.g. Jenkins, 2011 as cited in Wilson, 2018a, 728). The irreflexivity of grounding and causation can be challenged in very similar ways again pointing to a shared underlying structure. Another hallmark feature of traditional accounts of causation and grounding is that they are asymmetric. This is the feature that is intended to grant the inherent order to causal relationships. A relation is asymmetric when '$\forall x \forall y Rxy \rightarrow \forall x \forall y \neg Ryx$' is true. Because irreflexivity and asymmetry concern the directionality of the relationship, we can reuse the problem cases. The time travel case would be a symmetric causal relationship. Having the time machine plans would cause the traveller to go back and

[17]See 6.4 for a discussion of how a move to contrastivity can help us avoid transitivity problems.

give themselves the plans, **X** causes **Y**. But giving themselves the plans causes them to have the plans and use them to build a machine and travel back, **Y** causes **X**. In the grounding case, if identity is a grounding relation then we can see again how the relationship might be symmetric. If irreflexivity fails then the asymmetry of causation and grounding will also fail, and vice versa. This again hints at a deeper structural similarity[18].

2.3.3 Schaffer's criticisms

2.3.3.1 Causation is indeterministic, grounding is deterministic

Schaffer (2016) considers grounding and causation to be distinct notions, nonetheless sharing many similarities. However, because of three important disanalogies, he maintains they are separate. I will deal with these in turn and show how the apparent differences are merely a mistake. Schaffer's first allegation is that causation can be non-deterministic, whereas grounding cannot. It seems reasonable to say that causation is probabilistic rather than deterministic. Not doing so would leave us with the conclusion that smoking does not cause lung cancer because in some cases people smoke without developing it. Whereas it seems that grounding has to be deterministic; if parts ground wholes then we cannot make sense of a situation in which we would have all the parts of an object (compresent) and not have the whole object (Schaffer, 2016).

One might try to point to a distinction between full and partial grounds/causes to respond here but such a response would not be successful. The issues of the partial cases aside, it seems that full grounds are deterministic and guarantee that their groundee obtains, but the same cannot be said for full causes. Take any case of full grounding, e.g. **Singleton** above. **Singleton** seems to be deterministic, if Socrates exists, the singleton set containing Socrates will exist. On the other hand, take the case of an electron travelling towards a device which will measure its spin. It is entirely probabilistic which effect the electron will have on the detector, whether it is measured spin up or spin down for example. Arguably this is a case of full causation but one that appears to be entirely probabilistic[19] based on what we know of quantum mechanics. This seems problematic for **G=MC** as the disanalogy between causation and grounding has resurfaced, seemingly stronger. Luckily though, one may still yet defend

[18]Again, issues like this will be discussed more in-depth in 6.4 when we discuss contrastivity.
[19]Although see 6.6 for a discussion that this might nonetheless be deterministic.

G=MC by denying the probabilistic nature of causation. It is true that many interpretations of quantum mechanics are stochastic and say that the world is non-deterministic at bedrock. Given that it seems grounding is deterministic we have a significant disanalogy with causation. However, stochastic approaches are not the only ones available in quantum mechanics. Many so-called 'hidden variable' theories say quantum mechanics is non-stochastic and deterministic. These interpretations seem to possess as many theoretical virtues as stochastic interpretations so there is no immediate reason to favour one over the other. Indeed, given that we treat the macro world as deterministic it seems we would want to pick a quantum theory that is likewise deterministic for theoretical unity. Regardless of a preferred interpretation of quantum mechanics, we can agree that it is not a settled matter as to whether or not causation is deterministic; the answer could turn out to fall either way. For now, this discussion will suffice, but as noted we will discuss the deterministic nature of causation/constraint in 6.6 and in fact will conclude that constraint is always deterministic. But as constraint was only ever supposed to be identical to a specific subcategory of causation, this does not threaten my account but the issues of the deterministic nature of grounding could cause problems for **G=MC**.

2.3.3.2 Distinct Portions Of Reality

Schaffer (2016, 76) claims that causation connects distinct portions of reality whereas grounding does not. For example, a grounding statement such as "parts ground wholes" does not involve distinct portions of reality. Causal statements such as "lightning causes forest fires", however, do involve distinct portions of reality. However, this rests on assumptions which, whilst intuitive, are unfounded. It is at least logically possible for things to be self-causing, via causal loops. Consider the time travel case from 2.3.2. This seems to be a case of causation connecting arguably indistinct portions of reality. Or consider quantum tunnelling. Let us imagine a particle is in a state **x**, **x** is a stable state but there is a state, **y**, which is equally stable but lower energy; state **y** is, therefore, 'preferable' to the particle. However, to get to state **y**, the particle would have to overcome a massive energy threshold, consider a graph with a bell curve, state **x** is on the left of the bell curve and state **y** is on the right. Because state **y** is a lower energy state, the particle can, probabilistically, 'tunnel' through the bell curve and arrive at state **y**. It does this by 'borrowing' energy from its surroundings to surpass this energy threshold. It seems that this borrowing of energy is a causal process connecting distinct portions of reality, however, it is not so clear that the reason the borrowing happens is concerned with distinct portions of reality. The particle borrows this energy *because* state **y** is lower energy than state **x**. It seems that by merely being a possible state, the particle is caused to initiate this borrowing process and this seems to be a causal notion, but one that is not between distinct portions of reality.

Similarly, grounding may operate between distinct and non-distinct portions of reality, particular facts may ground other particular facts, therefore grounding events. Facts about treaties signed/revoked will ground the fact that nations are in a state of war/peace, this would be an example of grounding between distinct portions of reality. Additionally, it may be the case that something can ground itself in a roundabout way if we accept grounding loops. We can consider the case of three facts which are each grounded in the other two and nothing else, for example, volume, mass and density of an object[20]. Any value can be explained and grounded in terms of the other two but it would seem strange to privilege one with being the most fundamental, instead, it seems reasonable to say they all ground each other (Thompson, 2016, 47). The point is that grounding and causation, whether statistical broad claims or specific individual cases, equally seem to be able to connect distinct/indistinct portions of reality, this once again strengthens **G=MC**.

2.3.3.3 Well-foundedness

Schaffer's final disanalogy is that whilst causation can be non-well founded, grounding must be well-founded (Schaffer, 2016, 95). The assumption that causation does not have to be well-founded is controversial, people may resist this claim which potentially admits infinite chains of causation. People might want to appeal to a self-causing event as the start of the causal chain such as a creator or some natural process, but it is by no means a common consensus that the causal chain goes back infinitely. This cannot present the challenge to the causation/grounding analogy that it was intended to as we can hold both relations to be ones that must ultimately terminate, rather than be of infinite length.

Alternatively, we could argue that grounding can be non-well-founded. Schaffer believes grounding must be well-founded because grounding involves a 'transference of reality'. The more fundamental things ground and provide reality to the less fundamental. As such the

[20]This also serves as another potential counterexample to the irreflexivity and asymmetry of grounding as discussed above.

relation must ultimately terminate somewhere so that the reality actually comes from somewhere. This 'transference of reality' argument is limited, for example Bennett (2017, 120-122) challenges this in a number of ways. One way she does so is to point out that whilst the intuition that reality must 'bottom-out' is a strong one, it is in direct conflict with the intuition that everything exists for a reason, there are no things which just simply come into being. An ultimate ground for reality might seem to be a prime example of something that 'just exists' without further explanation and this would seem to contradict Schaffer's intuitions. These intuitions seem equally strong, so it is not clear which one we should reject in favour of the other. Alternatively, if one is not convinced by Bennett's argument, we could construct a case of infinite grounding. As Schaffer calls it, a 'gunky' universe could provide such a case in which each level of reality is grounded in the level below it, but the levels do not terminate, provided one does not accept his priority monism (Schaffer, 2010). If this is the case it seems we could have non-well-founded grounding and maintain the analogy with causation.

2.3.4 Fundamentality

Consideration of this point brings us neatly on to another potential disanalogy, that grounding is intrinsically linked to fundamentality in a way that causation simply is not (Wilson, 2018a, 730-731). Groundees are grounded in something more fundamental, this is part of what it is to be a grounding relation. Causation does not have this link to fundamentality, causes and effects are more often than not on the same level of fundamentality. This seems to be a structural disanalogy between the relations and so threatens **G=MC**. However, we can respond. It is perfectly possible for grounding to operate on the same level of fundamentality; consider the cases from 2.3.3.2. None of the three values of volume, mass and density can be privileged as the most fundamental, instead, they all seem to ground each other. In this way, grounding can operate between different or the same levels of fundamentality, just like causation can.

However, such examples are undoubtedly unusual and uncommon, the point remains that nearly all grounding relations operate between different levels of fundamentality and nearly all causal relations do not. People may allege that this is still in need of explanation because this fact is peculiar for **G=MC**. Here I can agree that this is an unusual fact in need of explanation, however, it is not a knockdown objection. Furthermore, this only seems to affect the general **G=MC** thesis rather than my restricted account concerning constraint. In constraint cases, the grounds/causes *are* in fact more fundamental, this is actually what brings about the constraining relationship. This, again, is something that will be discussed more fully later on, in 6.5.

2.3.5 Specific Spatial Location

Causation seems to have a specific spatiotemporal location whereas grounding does not. This disanalogy is closely linked to 2.3.3.2. If causation is supposed to connect distinct portions of reality, it will involve specific locations. If grounding does not connect distinct portions of reality, it might not even need to talk of a spatial location at all. Take an example, "the lightning hitting the tree caused the forest fire", this relation is clearly tied to a specific spatiotemporal location, when and where it occurred. Now take a grounding example, "parts ground wholes", it seems that this contains no information tying it to a spatiotemporal location. This is not a disanalogy however as it rests on a simple confusion between type and token causation/grounding.

Token instances of causation, e.g. the forest fire above, are tied to a specific spatiotemporal location. Grounding relations such as "parts ground wholes" are not tied to a specific spatiotemporal location, but this is because this is not a token instance of grounding, it is a type. We should not be surprised at this, because this is merely a difference between type and token. If we have a token instance of a grounding relation, this will be tied to a specific spatiotemporal location. An example of a token grounding relation would be that the existence of the particular bulb, shade, wiring etc. ground the existence of the particular lamp. Clearly, this token grounding relation does have a spatiotemporal location, that of the lamp in question. We can also take a type instance of a causal relationship "smoking causes lung cancer"; this likewise has no specific spatiotemporal location. To head off a potential objection I will clarify that, clearly, this has some sort of spatiotemporal location, i.e. wherever there are lungs and whenever there is smoking together; but this is a general spatiotemporal location, not a specific one. Moreover, grounding exhibits this same general location, "parts ground wholes" is presumably tied to wherever and whenever there are parts and wholes. The only difference is that the spatiotemporal location is much more general in the grounding case, but this is merely a difference in degree not in kind. It would be tricky to find a threshold above which this generality means the relation is grounding and below which it is causation without being arbitrary and so we should not conclude from this that the relations are different. It is also worth noting that counterfactual accounts of causation simply make no mention of spatial location in the counterfactuals. A dependence relation is highlighted but there is no requirement that the dependence relation relies on a specific spatial location. To demand that we include a specific spatial location in any case of causation begs the question against counterfactual accounts. In summary, this objection fails because it either rests on a misunderstanding or simply assumes that counterfactual dependence is not causation.

2.3.6 Energy Transfer

Energy transfer presents multiple difficulties to the comparison between grounding and causation. These issues broadly come in two kinds. The first kind says that causation simply is energy transfer, and so grounding cannot be causation because it does not transfer energy. The second kind of issue concerns the causal closure of the physical/the completeness of physics; it might seem that giving mathematical objects causal power would violate these closure principles in problematic ways. The second issue requires discussion in more depth and so that will be delayed until 6.3. It is worth saying more about the first issue now though.

One failed response to this would be to say that grounding does transfer a mark, it transfers existence or being (Bernstein, 2016, 26). This is a mistake though because existence is not a conserved quantity. Energy and any other mark that causation allegedly transfers can come in degrees. However, existence/being are absolute concepts, they are on or off. Saying that the grounder transferred 0.5 being to the groundee makes no sense, whereas saying that the cause transferred 3.6 joules to the effect arguably does. The fact that 23 is indivisible by 3 does not transfer a mark to the fact that Jane is unable to divide her 23 strawberries between her 3 children. The allegation is that this signals a substantial difference between causation and grounding that needs to be addressed. I agree that some causal relationships seem to transfer conserved quantities, and I am happy to concede that it seems that no grounding relations do this[21].

However, it is not at all clear that *all* causal relationships straightforwardly transfer a mark. In 2.2 we discussed the fact that some scientists treat entanglement as a causal relationship. But given that entanglement seems to happen across vast distances instantaneously, it does not seem possible that any mark is actually being transferred. The causal relationships I will ultimately be interested in identifying with grounding relationships do not transfer marks, so this objection will simply not affect my account. A broader point can be raised though that energy transfer objections seem to beg the question against counterfactual accounts of causation. Counterfactual accounts make no mention of energy transfer in determining which relationships are genuinely causal. Moreover, intuitively it seems that there are instances of absence causation/causation by omission which cannot be accounted for by energy transfer accounts but can be accommodated by counterfactual accounts. Given that I will be endorsing a counterfactual account of causation later on, I think it is fine to sidestep the first kind of issue with energy transfer. Even if this is a genuine difference between some causal and grounding relationships, it is not a feature present in the kinds of causal relationships I am interested in.

2.3.7 Temporal Separation

It seems that we can observe causes and effects and individuate them by their temporal separation (Wilson, 2018a, 731), we can see causal processes happening but the same cannot be said for grounding. This objection comes in two parts. The first part is that the causal process (and the individual causes and effects) can be observed. The second part is that causes and effects are distinguished by their temporal separation, which grounds and groundees lack. The second part of this objection is addressed over 2.3.8 and 2.3.9, so I will put that to one side for now.

The first assertion can be responded to by denying that we can observe the causal process. Humean thinking cautions us that we actually never observe causation (Hume, 1978, 93), all

[21]Although see 3.2 where I discuss energy transfer accounts in more detail. The more abstract the conserved quantity becomes, the more likely that grounding will arguably transfer it. If this is the case, then in fact energy transfer problems simply present no issue for **G=MC** in this way.

we ever see is regularities and constant conjunction. This is also what we see in grounding cases, so, on a Humean picture, grounding and causation might appear very similar. For those not persuaded by Humean thinking, there is an alternative response. We can observe some causes at the macro level but not all. For example, if there are true causal relationships at the quantum level, such as those suggested by hidden variable theories, we cannot observe them. This could be a failure in principle and so there would be unobservable causal relationships. We cannot, therefore, distinguish grounding and causation by observability of the 'process'. Overall then, the first part of the temporal separation problem does not seem to unduly threaten the **G=MC** thesis. As a last (minor) point the entire idea of the interventionist treatment of grounding, as we shall see in 3.5, is to show that we can determine grounding relations and so 'observe' them. To deny this is to beg the question against a grounding interventionist like Wilson (2018a).

2.3.8 Synchronicity

Grounding is typically taken to be a synchronous relation, relating things at time ***t*** to other things at time ***t***. Causation, however, is typically taken to be diachronic, relating the cause at time ***t*** to the effect at time ***t+n***. This seems a significant disanalogy and creates issues for **G=MC** (Wilson, 2018a, 729-730). This disanalogy is overstated however. Some grounding relations are diachronic. For example, someone's property of being human is grounded in past causal history, which seems straightforwardly diachronic. Alternatively, there are potential cases of synchronic causation[22] such as quantum entanglement. If this is a case of genuine causation, then the measurement of an electron's spin at time ***t*** fixes the measurement of its entangled partner at the same time. However, there still seems to be a disanalogy between causation and grounding; because the majority of grounding cases do not seem to be diachronic, unlike the majority of causal cases.

Wilson (2018a) responds to this by denying that this is a significant difference. Perhaps the reason that grounding does not seem diachronic is that it is not a temporal relation like causation after all; as such it is a category mistake to describe grounding as synchronic or diachronic. This might seem detrimental to **G=MC** but Wilson (2018a, 730) explains this difference in terms of what laws the relata are governed by. The relata of causation are

[22]In 6.1 I will argue that mathematical constraint is synchronic. Importantly, this is also the case for those constraint relations that seem straightforwardly causal.

governed by the laws of nature so we should expect the relation to be temporal. Grounding relata are governed by metaphysical laws and so we should not be surprised that they generally turn out to be synchronic. We could also appeal to this to respond to Schaffer's concerns from 2.3.3 as well. However, I do not want to rely on this response too much as I would prefer to bring the relations of grounding and causation even closer together first. We should be willing to accept cases of synchronous causation such as entanglement as genuine. Importantly, in the constraint cases I will discuss, constraint is in fact synchronic[23]. So, whilst these issues might pose a problem for **G=MC**, they need not pose a problem for my claim.

2.3.9 Bernstein's Criticisms

Bernstein (2016) objects to the analogy between grounding and causation for a number of reasons. Although she concedes the two relations share similarities, these are mainly superficial and there remain significant structural and logical differences between the relations. These differences suffice to show, not only that the relations are distinct, but that any analogy between them is unhelpful. One of the key features she picks up on is that causation seems to be diachronic, whereas grounding is synchronic. As we have already discussed, it seems eminently plausible that this distinction is only apparent. The other option is to accept the difference and say that it is not significant because it can be explained away by appeal to law governance. However, Bernstein believes that this will not work because the disanalogy is more significant than supporters of **G=MC** typically assume. Bernstein claims the diachronic/synchronic disanalogy leads to three significant structural differences:

- Questions about the relationship between time and causation are much richer and more common than those about grounding and time;
- Supporters of G=MC will have a tricky time distinguishing grounding from synchronic causation;
- The diachronicity of causation allows for hasteners and delayers in causal relationships (Bernstein, 2016, 24-25).

I will address each of these and respond. These fail to be as significant as Bernstein intends.

[23] Again, see 6.1.

2.3.9.1 Metaphysical Debate

It is true that there is a difference because there is an open question about whether effects can precede causes; but there is no corresponding question about whether groundees can precede grounders. This is not a major issue. Dialectics and how we use the concepts might be responsible for this difference. It may also turn out that effects cannot precede causes. Backwards causation is a controversial matter and whether or not it is even possible has yet to be determined. The idea itself could potentially lead to contradiction, for example, if an effect occurs before the cause, it should be possible, in principle, to intervene and stop the cause from occurring, but this would lead to paradox. Backwards causation is not a settled matter and until such a time as it is it cannot be used as an argument for a distinction between causation and grounding[24].

Additionally, there is a subtle shift Bernstein makes, based on a slight misconception, which generates the issue. For Bernstein, “It is a substantive metaphysical question whether causes always precede their effects...” (Bernstein, 2016, 24). This is something I agree with, there is a question about whether or not causes *temporally* precede their effects. Bernstein goes on to say “But it is not a substantive question whether grounders are metaphysically prior to what they ground: they must be...” (Bernstein, 2016, 24). The mistake here is that metaphysical priority is not the analogue of temporal priority.

It seems that grounding often relates the more fundamental to the less fundamental[25], but temporal priority is not the analogue of this metaphysical priority. The actual causal analogue of metaphysical priority is causal priority. Causes must *cause* their effects, they must be causally prior. There is no substantive debate on whether causes cause their effects, because that is simply what it is for something to be a cause, it must be causally prior to its effect. There may be a lack of questions about whether grounds are metaphysically prior to groundees, and there may be a rich literature on whether causes are temporally prior to effects. But this does not signal a difference between the relations because these are not analogous priorities. Bernstein’s mistake is in thinking that temporal priority and metaphysical priority are counterparts.

---

[24]We earlier discussed potential time travel cases which could cause problems for the reflexivity of causation. If backwards causation is a genuine open possibility, then we are fine to use it to challenge traditional features such as irreflexivity. If backwards causation is not a live option, then both problems dissolve.
[25]Although recall the discussion from 2.3.4 that this might not need to be the case.

Moreover, it is in fact the case that there are parallel questions about the temporal priority of grounding. In cases of synchronic grounding it obviously is not a substantive metaphysical question whether grounds temporally precede what they ground, because the relation is synchronic, just as with synchronic causation. But this is a feature not a bug. We can have substantive metaphysical debate about whether grounds temporally precede their groundees. We can ask questions such as "what grounds the fact that there will be a sea battle tomorrow?" or "what grounded the fact that Louis Phillipe I would be the last king of France?". In both cases it seems a candidate for the ground is an object/event in the future, meaning that the groundee may temporally precede what grounds it. It could turn out that in both these cases the relevant facts are grounded in things which temporally precede them, just as it may turn out that in cases of purported backwards causation the causes do actually temporally precede the effects[26]. What is clear at this stage, however, is that we can have substantive metaphysical debate about whether causes or grounds temporally precede their effects/groundees.

It also seems that we can have substantive debate about 'backwards' grounding. On Schaffer's Priority Monism account, what is fundamental is the cosmos as a whole, the ultimate ground of existence is the totality (Schaffer, 2010). But the cosmos is built out of (potentially infinite) parts, so it seems in a sense "backwards" if what they build up is actually the thing that grounds them. Bernstein's disanalogy does not, then, present a problem to **G=MC**, or to my claim.

2.3.9.2 Hasteners & Delayers

Bernstein classifies a hastener as something in a causal relationship that would bring forward the occurrence of the effect, for example, an assassin who targets someone already dying of a disease hastens the occurrence of their death (Bernstein, 2016, 24). Conversely, a delayer is something present in a causal relationship that pushes back or prevents the occurrence of the effect. For example, the doctor treating the same sick person, thus prolonging their life, is a delayer in the causal relationship. In cases of diachronic causation, these are readily present and a key phenomenon of the causal relationship. Bernstein claims that because grounding is

[26]This is not a view I need to endorse for the purposes of the thesis, instead this discussion is just to illustrate how this sort of debate could in fact be present in both the grounding and causation cases.

synchronic there can be no analogous hasteners and delayers, representing a significant disanalogy.

Bernstein's point is a little unclear. She seems to treat hasteners/delayers as exclusively 'intermediary' phenomena. Bernstein characterises hasteners/delayers akin to catalysts, they are things which 'fit in' between the cause and effect and so make the effect happen sooner (or later). As Bernstein claims that grounding is unilaterally synchronic there is no 'between' for a hastener/delayer to fit in, thus there is a disanalogy. There are two problems with this argument. Firstly, it is not clear that 'intermediary' hasteners/delayers are the only kind of hasteners/delayers. The obtaining/occurring of P would ground $P \vee Q$, if Q occurs/obtains earlier than P, then it has hastened the grounding of $P \vee Q$. Or consider a delaying case, if I bet that a tossed coin will come up heads, the coin landing heads grounds my victory. If the coin lands tails up five times preceding the heads outcome, those five tosses delayed the grounding of my eventual victory. Quite clearly, grounding relationships can be hastened/delayed.

Bernstein may object that the lack of 'intermediary' hasteners/delayers is nonetheless significant in some way. *If* this lack of intermediary hasteners/delayers is genuine, and if it is, as Bernstein seems to think, due to the synchronicity of the relationship, then it is unclear why Bernstein is surprised by this development. If intermediary hasteners/delayers are features of a diachronic relationship, then of course synchronic grounding relationships would not have them. This would be a feature of the account, not a bug. Similarly, we should expect no such cases in examples of synchronic causation. But all this rests on the assumption that grounding is always synchronic and as we discussed in 2.3.8, it is not clear that this is true. Diachronic grounding relationships are liable to be subject to hastening/delaying in much the same way that diachronic causation is. Hasteners/delayers are no threat to **G=MC**. In section 6.1 I will turn to discussions of hasteners/delayers in constraint relationships and argue that the case is much the same there.

2.3.9.3 Synchronic causation

Proponents of **G=MC** aim to unite grounding and causation. In doing so, they may be unable to distinguish cases of synchronic causation from synchronic grounding. For example, an iron ball depressing a cushion seems to be a case of synchronic causation rather than grounding

(Bernstein, 2016, 24). But the **G=MC** theorist has no way of making this distinction. Depending on your interpretation of **G=MC**, this might not be a problem. If one endorses **weakG=MC**, one could say that Bernstein's example is simply of the same broader kind as grounding. If one wants a stronger **G=MC** thesis, then one could simply bite the bullet and say that there is no causal/grounding distinction in this case.

Alternatively, a defender of **G=MC** might see Bernstein's counterexample as significant. How that debate will go is up to the defenders of various positions, but it is worth noting that my ultimate claim concerns merely a subset of grounding and causal claims. However, since the relations I am interested in will generally be synchronic, it is worth saying something. As mentioned in 0.1 of the introduction, I am arguing for **Con=G&C** but intending that that argument provides background support **weakCon=G&C**. If **Con=G&C** is true, I am happy to say that there is no difference between synchronic grounding and synchronic causation, this is just a feature of the small overlap. If **weakCon=G&C** is true, and if the above represents a difference between grounding/constraint/causation, this is no problem; on that view it was only ever required that grounding/constraint/causation be *similar*. Either way, Bernstein's problem simply dissolves.

2.3.10 Problem of iterated ground

There is substantive metaphysical debate about whether there is anything that grounds grounding relations (Litland, 2017). Grounding is intended to be a relation between truths. So Take the statement "parts ground wholes", this is itself a truth so we can ask what it is that grounds it, what grounds the fact that parts ground wholes. This is the problem of iterated ground. Moreover, in causation, there is no parallel question about what causes causal relationships to hold. This could be a sociological coincidence or it could be a signal of a deeper difference between the relations.

My intuition tells me that this is simply a sociological coincidence and a bit of a philosophical red herring. For starters, there are in fact these further questions about causal relationships. Indeed, this forms the basis of the scientific endeavour. When confronted with the causal relationship "smoking causes lung cancer", we naturally want to know why this is the case, what causes this relationship to hold, so we further investigate and discover the carcinogenic properties of the substances involved. So such iterated questions are not unique to grounding. As Skow (2016) puts it, we in fact often have these kinds of questions about causation. For example, one can imagine the following interaction:

"Why is the museum so much more crowded today than last time?"
"Because today is Friday."
"What does that have to do with it?"
(Skow, 2016, 75).

Here, an explanation was asked for, and a causal explanation was given, A further question was then asked about why it is that this causal relationship holds. This is an iterated question exactly like those which plague grounding. Furthermore, these are perfectly natural questions to ask, indeed such questions are a large part of the project of science. Kovacs argues that questions of iterated causation (QIC) "can be asked just as intelligibly as QIG [questions of iterated ground]" (2022, 464). Kovacs bases this on the existence of causal events, i.e. the event of c's causing e. Indeed, Kovacs thinks that this problem is at least as serious as the question of iterated ground. If this is a problem for grounding, it is also a problem for causation. Far from suggesting a significant disanalogy, this, then, brings the relations even closer together as they are subject to a parallel problem.

For Sider (2020), the existence of questions of iterated ground is not in itself problematic, because even if such questions form an infinite series, this need not be problematic (2020, 749-751). What creates the issue is that iterated grounding claims seem to be about the fundamental, but seem to involve appeal to some non-fundamental terms. For example, cities are not thought to be a fundamental constituent of reality. In explaining what grounds the existence of cities, we might (ultimately) point to various quantum facts (Sider, 2020, 748). In a specific case, this grounding fact, that a particular quantum state grounds the fact that New York City is a city, is itself a fact that is subject to grounding. We might then ask "What grounds the fact that quantum states ground the fact that New York City is a city?". If we stipulate that this is merely a brute fact, then part of the fundamental description of the universe will include a statement about cities. For Sider, this violates a key principle he endorses called **Purity**.

Purity:

No ungrounded facts can involve a non-fundamental concept.
(Sider, 2020, 748).

Ultimately, Sider thinks we can maintain **Purity** by responding to individual cases of iterated ground with a selection of tools (2020, 756). Kovacs (2022, 469) is happy with Sider's solution in the grounding case and thinks that such a tactic is extendable to the case of iterated causation.

Although there may be questions of iterated grounding, they need not be problematic. Not only do we have responses available as Sider (2020) argues, but there are also iterated questions about causation. Kovacs (2022) argues that the grounding solution can be mapped onto the causal case. But regardless of possible solutions to these problems. The problem of iterated grounding was supposed to indicate a significant disanalogy between grounding and causation. The fact that both relations are subject to the same problem in the same way is not indicative of this disanalogy, in fact it speaks in favour of the opposite, grounding and causation have yet another similarity[27].

2.5 Section Summary

The aim of this chapter was to highlight the deep similarities between causation and grounding. I have surveyed some key alleged differences in the literature and hopefully deflated their significance. But some may remain, bringing Wilson's strong **G=MC** claim into doubt. Fortunately, I do not need to commit to **G=MC**, instead all I need is **weakG=MC**. Ultimately, I just need grounding and causation to be close enough for a small crossover in the constraint cases. One might worry that the features which posed a problem for **G=MC** might also pose a problem for my strong claim about constraint (**Con=G&C**). I will delay discussion of these issues until Chapter 6 so that we have a full picture of constraint in mind before addressing them. For now, whilst we have considered an analogy between causation and grounding, we have not considered *what kind* of causation. Accordingly, before we discuss constraint in Chapter 4, we will now move on to discuss the kinds of causation that are compatible with **G=MC** or **Con=G&C**.

[27]Once we have established a picture of constraint, I will discuss an analogous problem of iterated constraint in 6.7.

Chapter 3 - Causation

3.1 The grounding analogy

In chapter 2 I compared grounding to causation by pointing to their key characteristics. In some cases, problems emerged for these characteristics of causation (e.g. transitivity), but there are parallel problems for grounding. Through the treatment of these problems, I aimed to establish **weakG=MC**, i.e. that grounding and causation are both examples of a broader class of dependence relations such as building relations as proposed by Bennett (2017, 71-83), but not identical. This might have shown us that grounding is similar to causation, but we have not yet specified what *kind* of causation it is similar to. To support this analogy, it would be helpful to be clear on what theories of causation this analogy works with. This chapter will assess different conceptions of causation that are compatible with the analogy with grounding, i.e.: simple counterfactual accounts; probabilistic accounts; and the interventionist conception of difference-making. I will briefly consider these in turn before focusing on interventionism as a kind of case study for the comparison. Firstly, it will be worth talking explicitly about which accounts of causation seem to be *incompatible* with the analogy, and so would be unsuitable for my account.

3.2 Incompatible accounts

Any kind of mechanistic/process/mark-transfer account of causation seems incompatible with a grounding analogy and therefore as a backdrop for mathematical causation. These accounts all differ subtly, but the reasons that they are unsuitable for my project are similar and so for my purposes, I shall simply treat them as one type of view about causation. I will refer to these as process accounts. There are three broad reasons why process accounts are not suitable for an analogy with grounding/constraint.

Firstly, grounding does not involve energy transfer, or indeed any kind of mark being transferred. We have already discussed this in 2.3.6 but it is worth noting explicitly that process accounts all rely on some sort of mark transfer to be indicative of causation. It is therefore unlikely that one of these theories could be modified to allow for grounding to be causal. Secondly, process accounts struggle with apparent cases of causation by omission, often simply denying this as a kind of causation. However, the existence of grounding relations that involve omission seems likely, as we shall discuss later (3.6.2.1). Finally, process accounts seem to view causation as an inherently diachronic phenomenon. We saw in 2.3.8 and 2.3.9 that grounding seems to be a synchronic relation and that causation seems diachronic (although this was challenged, and will be challenged further in 6.1). A view that builds diachronicity into a notion of causation is obviously incompatible with views that take grounding and causation to be comparable in that they can both be synchronic/diachronic.

The reasons that process accounts are incompatible with my account of mathematics are also, I argue, broad reasons why they should be rejected, regardless of my view on mathematics. Firstly, there are compelling examples of counterfactual causation which do not seem easily explicable in terms of energy transfer. Given that grounding does not seem to involve energy transfer,[28] this would suggest a deep incompatibility with my account. Examples which illustrate this are often ones which illustrate a point from the previous paragraph, i.e. that process accounts struggle with alleged cases of causation by omission. If process accounts of causation cannot account for this datum then this is a mark in favour of accounts which can deal with them. As we will discuss in 3.6.2.1, examples of grounding by omission seem plausible. If process accounts cannot deal with them, this again signals a deep incompatibility. Finally, whether or not we think causation is as a matter of fact a diachronic relation, we should not inherently build diachronicity into an account of causation. Examples of synchronic causation are interesting points of discussion and should not be dismissed by mere definition; we should allow theoretical room for synchronic causation in order to assess such cases properly. It is not clear that process accounts could satisfactorily deal with synchronic cases. It would be very unclear how energy or any other mark could be transferred synchronically. As discussed in 2.3.8 and 2.3.9, grounding seems to be synchronic more often than not. It therefore seems that process accounts might simply judge grounding/constraint to be non-causal by default.

One might caution that more sophisticated process accounts are possible that can respond to the problem cases surveyed so far. Perhaps one could make the mark that is transferred more

[28]Although see the next paragraph for more discussion of this.

abstract. However, imagine that the more abstract mark was 'mass' or 'stuff'. It seems that we might be able to say that such marks are transferred or at least maintained in grounding cases. The existence of the flat surface and the legs ground the existence of the chair. These parts have certain mass and energy levels. In this grounding relationship, there is a conserved quantity because the chair itself maintains the same level of mass and energy as its constituent parts. The point here is that the more abstract the conserved quantity becomes, the more likely that such accounts can actually allow grounding to be a kind of causation, or at least very similar. Given this, I will instead focus on some accounts that are clearly compatible with the analogy to constraint. If process accounts are modified and become suitable for the analogy, this will just turn out to be a strength of my theory. As such, I will in turn consider: basic counterfactual accounts of causation; probabilistic accounts; and finally the interventionist approach to causation.

3.3 Basic counterfactual account

3.3.1 Counterfactual accounts

Due to the difficulties with process theories, many people have favoured counterfactual analyses of causation. There are many different forms of counterfactual account with subtle differences, but Noordhof (2020, 9) presents an intuitive and simplified form of counterfactual account as follows:

For distinct events e_1 and e_2, e_1 causes e_2 if
(1) If e_1 were to occur, then e_2 would occur.
(2) If e_1 were not to occur, then e_2 would not occur.

If we see **A** followed by **B**, and we know that if ¬**A** had been the case, then ¬**B** would be the case, then **A** caused **B**. Typically, we go on to assess the truths of these counterfactuals in terms of possible world semantics as follows:

> "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, 1986a, 41)

It seems correct to say that the first domino falling caused the second to fall because in the closest world where the first domino did not fall, nor did the second. Such simple accounts are compatible with my account of grounding because there seems to be an aspect of counterfactual dependence to grounding. Consider the following case.

Socrates:
i. "Were Socrates not to have existed, then the singleton set containing Socrates (written as {Socrates}) would not have existed either." i. seems to be true, whereas:

ii. "Were Socrates not to have existed, then {Socrates} would have existed" seems false.

ii. is false because {Socrates} is counterfactually dependent upon Socrates. Had Socrates not existed, then the singleton set containing him could not have existed.

However, it is worth discussing some of the issues that counterfactual accounts face and how they might respond.

3.3.2 Wrong-trackers

It is worth considering the problem of wrong-trackers here because this problem will affect all theories that take a counterfactual approach to grounding. It is, therefore, worth responding now to put those concerns to one side. As has been mentioned, for counterfactual accounts, if it is true to say "if **¬A** had been the case, then **¬B** would be the case", then it is the case that **A** caused **B**. However, there are situations in which it also seems to be true to say that "if **¬B** had been the case, then **¬A** would be the case". Consider the case of barometers. Atmospheric pressure causes the reading on barometers. Sometimes, this atmospheric pressure is indicative of an approaching storm. It is true to say that the atmospheric pressure caused the barometer reading (or that the approaching storm caused the atmospheric pressure). Had the atmospheric pressure been different/had the storm not been approaching then the barometer reading would have been different. We want to commit to that being a true causal claim.

But notice that the same holds in reverse. Had the barometer reading been different, the atmospheric pressure would have been different. But we do not want to say that the barometer reading is a cause of the atmospheric pressure/the approaching storm. Something has gone wrong here, there seems to be counterfactual dependence between the barometer reading and the atmospheric pressure but this is not causation. Such cases are known as wrong-trackers because the relevant counterfactuals track the wrong relationship.

It is also worth pointing out that such problems are present with counterfactual analyses of grounding. For example, consider **Socrates** above. The relevant counterfactual that we want to associate with the counterfactual dependence (i) is true. However, there is another counterfactual that seems true:

iii. "if {Socrates} did not exist then Socrates would not have existed"

But we want to say that whilst there is a dependence relation at play in i. there is not one in iii. This is, as Wilson (2018a, 736) says, a case of "wrongtracking". Issues with wrongtracking often result in sophisticated developments of the counterfactual view, e.g. interventionism as we will see in 3.6.2.2. Of course, responses showing why backtracking counterfactuals are false are prevalent in the literature, so this need not be a fatal problem for theories. But every counterfactual theory of causation *must* deal with them. Ultimately this chapter argues that my account of mathematical causation is compatible with multiple accounts of causation. To do this, I choose interventionism as a particularly tricky theory of causation that does not immediately seem applicable to the mathematical case. By showing that it *is*, I aim to demonstrate the flexibility of my account. So, rather than showing how each an every account responds to wrong-tracking issues I merely flag the issues here, and will show how interventionism specifically responds to it later on. Before moving to that interventionist picture though, it is worth discussing probabilistic causation.

3.4 Probabilistic causation

Probabilistic accounts of causation are a popular alternative in the causal debate. Fortunately, certain forms of this could work with my account of mathematical objects. Menzies (1989) discusses one such account. An early formulation of a probabilistic theory of causation is developed in Lewis (1986b, Postscript B). Lewis says that c causes e “just in case if c were to occur, the chance of e's occurring would be x; and if c were not to occur, the chance of e's occurring would be y, where x is much greater than y.”. There are of course problems with probabilistic accounts of causation, and this is merely a first effort at a definition. But all I need to show is that this core idea is compatible with my account of mathematical objects.

To see the compatibility we will consider a simplified case. Constraint, as we shall see in 4.3.2, is a matter of limiting ranges of possibility. Constraint is the making impossible/ruling out of some options. Imagine a scenario with ten possible outcomes. Without any constrainers (imagine this as a lawless world), ‘anything goes’, anything could happen. The probability of any event is equal (0.1). If we introduce a constrainer that rules out five of the options, the probability of the remaining options will increase to 0.2. When one of these happens, it seems that, on a probabilistic account of causation, we should judge the constrainer as a contributing cause, because it raises the probability of the effect. This is easier to see in stronger cases. Imagine the constrainer ruled out not five, but nine of the ten options. The probability of the remaining option will rise to 1, i.e. it is guaranteed. The constrainer raised the probability of the effect happening and so it seems we should judge it as having truly been a probabilistic cause[29]. One issue with the above is that it suggests that constraint is a deterministic relationship, whereas it is traditional thought that causation is (or at least can be) an indeterministic relation. This problem will be more thoroughly addressed in 6.6 but what I will say here is that one can simply opt to bite this bullet but point out that there are cases of constraint that we want to say are causal (4.4) which are likewise deterministic, so this does not actually threaten the compatibility of probabilistic causation with my account of mathematical objects. It is also worth noting that whilst some probabilistic accounts are framed in counterfactual terms, not all are. It should be seen as a positive feature of my account that it is compatible with more than merely counterfactual accounts.

[29]In 3.6 onwards we will discuss the notion of a structural equation model. It is worth noting here that a probabilistic account also fits with the use of structural equation models, instead of 1s and 0s being the only values we give to variables, we simply give them a value within that range, where the value of downstream variables is still affected by, and dictated by, the upstream variables.

3.5 Interventionism

So far we have surveyed some different accounts of causation and considered some problem cases for process theories. Interventionism is a particularly difficult kind of counterfactual account for my theory. Ultimately, I will discuss claims concerning mathematics having been different, even though it is *impossible* for mathematics to have been different. One might worry that my account will be incompatible with the interventionist perspective as it takes these "what if things had been different?" claims as being interventions that *could*[30] be made on the system to test if the relationship is causal. However, I aim to show that *Exotic Realism* is compatible with this difficult counterfactual account. In doing this, I hope to show the general strength and adaptability of *Exotic Realism*.

Counterfactual accounts of causation naturally lend themselves to an account of grounding/constraint, consider how Hume characterises causation when he says "Or, in other words, where, if the first object had not been, the second never had existed." (Hume, 1995, 54). We tend to think that had the grounder not existed, the groundee would not exist. A notion of causation that involves us treating the effect as counterfactually dependent on the cause in this way parallels this nicely. Interventionism is a particularly good illustration of this idea. As we will see shortly, it considers cases where the cause is absent to determine whether or not the effect happens. This matches the spirit of Hume's idea above. Accordingly, Wilson (2018a) appeals to interventionism when arguing that grounding is a kind of metaphysical causation.

Interventionism specifies whether a relationship between **X** and **Y** is causal based upon what would happen if an intervention on **X** with respect to **Y** took place. An intervention is a technical notion that is generally defined by four criteria:
I is an intervention on **X** iff

- **I** causes **X**
- **I** isolates **X** from previous causes of **X** so that the value of **X** is fixed by **I** alone. This is to screen off **X** from background conditions and other factors so that its causal relationship or lack thereof with **Y** can be investigated in isolation.
- Any path from **I** to some effect **Y** goes through **X**. This is to prevent **I** from itself causing **Y** via some unexpected route. For example consider when one takes a placebo

[30]See 4.7.3.2 and chapter 5 for this discussion.

but gets better anyway, the administration of the drug has caused recovery, not the drug itself.

- **I** is independent of any variable **Z** which causes **Y** and is on a path which does not go through **X**.

(Woodward 2003, 105)

If these criteria are met, an action can appropriately be described as an intervention and the causal relationship can be tested. If an intervention on **X** with respect to **Y** causes **X** and proceeds to cause **Y** we can confirm that **X** causes **Y**.

It is worth clarifying here that interventionism should be thought of as a way to test when it is appropriate to describe a relationship as causal. As Woodward states “We can explain what is for a relation between **X** and **Y** to be causal by appealing to facts about other causal relations involving **I**, **X** and **Y** and counterfactual claims involving the behaviour of **Y** under interventions on **X**” (Woodward, 2003, 105). Interventionism is not a reductive account of causation because the aim is not to give an account of causation in non-causal terms. The aim is to test when it is that we can call a relationship causal.

Having got this basic understanding of interventionism in place, we can move on to discuss structural equation models (SEMs) and how they can be used to model causal relationships. With that in place, we can start to see how they fit together. Having done this, I will then take an interventionist approach for the rest of the thesis but it is important to remember that *Exotic Realism* is compatible with the other accounts of causation flagged above. I merely attach myself to interventionism because sticking to one theory of causation will make the discussion of constraint easier. As mentioned earlier, this also shows that *Exotic Realism* is a flexible theory that can fit with some tricky, sophisticated accounts of causation. The idea that mathematical objects are causal is counterintuitive but interventionism does a good job of illustrating the idea and pulling intuitions in the other way to provide a compelling case that mathematical constraint is a causal relationship.

3.6 Structural equation models

3.6.1 Basic models

Wilson (2018a, 740-741) provides a good definition of SEMs. A model involves a set of variables representing the desired features of the world we aim to assess, coupled with a set of structural equations that link the values of the variables to the underlying structure of reality as well as a function that specifies which values the variables in fact take. For the models we will look at, the values will be either a 1 or a 0, but as noted in 3.4, one can easily build SEMs with values that represent probabilistic cases. As stated, the structural equations themselves represent the relevant dependence relations that we want to assess as being causal or not. The '=' in the structural equations represents this counterfactual dependence. As Wilson (2018a, 740) says "Thus, each causal model encodes a set of counterfactual dependencies: if B, C, D were set to specific values by an intervention, A would take a specific value.". With this understanding in place, we can begin considering the models themselves.

Simple: Window

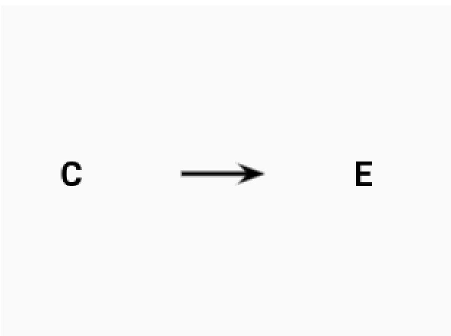
Variables

- C: Suzy throws the rock
- E: Window smashes

Structural Equations E=C

Assignment C=1; E=1

Graphical Representation



Wilson, 2018a, 24

One might wish to see whether Suzy's throwing of the rock caused the window to break, so appropriate interventions might be made. One might start by recreating the scenario to see if the same things happen. At that stage, one might start manipulating the **C** variable in various ways. One might straightforwardly prevent the throw and see what happens to the window, one might intercept the throw and again see what happens. Interventionism, then, would show this relationship to be causal. We of course already know that **C** caused **E** in this case, but one can see the procedure that should be undertaken with less clear cases to determine causal dependence. Interventionists accounts can respond to wrongtracking worries with simple tests like these as well. We can illustrate this with the classic example of a barometer and a storm. Take a perfectly working barometer. When there is a drop in the barometer, a storm soon follows. Indeed, all drops in the barometer are followed by storms, and when there is no drop, there is no storm. One might then invoke the counterfactual "had the barometer not dropped, the storm would not have happened", which is true, and claim that the barometer causes the storm. This, of course, is a case of wrong-tracking. Performing simple tests like isolating the barometer and artificially causing a drop in it will reveal this. When we see that no storm follows, we might wish to repeat this a few times but will eventually conclude that our initial hypothesis was mistaken.

The simple model above also maps straightforward cases of grounding. **Socrates** above is one such example. We know that the existence of Socrates grounds the existence of {Socrates}. Let us take the variables to be as follows:

C: Socrates exists
E: {Socrates} exists
(Wilson, 2018a, 24).

Wilson proceeds to argue that the interventionist account of causation will determine this to be a causal relationship. We can imagine a metaphysical intervention that brought about the non-existence of Socrates and it seems that in such circumstances it would be correct to say that {Socrates} also ceased to exist. It is worth flagging up the issue of wrong-trackers with regards to grounding once again. Just as we can imagine a metaphysical intervention that would remove Socrates, and therefore {Socrates}, it seems we can also imagine the reverse.

We can imagine a metaphysical intervention where the value of the {Socrates} variable is changed, i.e. {Socrates} does not exist. It seems that in such a world, it would also be the case that Socrates did not exist. But this does not track the grounding relationship as we see it. This is another wrong-tracker. There are two responses available.

Firstly, we might look back at the criteria that define an intervention above. One of the criteria for **I** to be an intervention to test the relationship between **X** and **Y** is that "Any path from **I** to some effect **Y** goes through **X**. This is to prevent **I** from itself causing **Y** via some unexpected route." (Woodward 2003, 105). We can see how this might work in the barometer case above. Short of breaking it open and it no longer being a properly working barometer, we might find that to artificially cause a drop in the barometer is to manipulate the pressure around it. But doing so just is causing storm-like conditions to bring about the drop in the barometer. Once again, the appropriate path stipulated by the definition of an intervention simply is not available. Similar reasoning applies in the grounding case. There would be no method of producing the non-existence of {Socrates} that did not go via Socrates, i.e. there is no path from **I** to some effect **Y** that goes through **X**, the only path goes through **Y**.

Alternatively, we also have another way to distinguish the right-tracker from the wrong-tracker. Lowe (1998, 145) points out that it is easy to spot wrong-tracking identity relationships. We know that the identity of {Socrates} depends upon the identity of Socrates, and not vice versa. In this way, we know that Socrates must ground {Socrates} and not vice versa (Lowe, 1998, 145-150). It is worth noting that although Lowe's test seems to work here, it may not work in all cases. Let us assume that there is a multiply realisable mental property, **M**, that is grounded in various physical brain states $\mathbf{P^1}$-$\mathbf{P^n}$. The grounding relation clearly goes from the operant **P** to **M**, however, it seems that the identity relation might go the other way. The identity of the brain states might be thought to be dependent upon the mental property they give rise to. If this is the case, then although Lowe's test helped us in **Socrates** it might not help us elsewhere. But Lowe's test does not need to be a cure-all. It is important to recall that interventionism is a test to determine when we should call a relationship causal and when we should not. In the above case, no intervention can be performed and so we cannot test and say that the relationship is causal, however, we have Lowe's test to help out. If one is concerned about cases where it seems that this sort of response might not be possible it is worth remembering that interventionism is not reductive. If we sometimes cannot perform a test to see if a relationship is causal, that is not a mark against interventionism, it merely means it is not possible to determine if that particular relationship is causal.

3.6.2 A case for interventionism

As stated, for the purposes of this thesis I will attach myself to the interventionist picture. It is worth justifying interventionism over process accounts and some of the other accounts that we discussed earlier. With this understanding of SEMs and interventionism, we are in a position to do that. Process accounts often struggle to account for causation by omission, and standard counterfactual accounts seem to struggle with cases of overdetermination. However, by coupling the interventionist approach with SEMs, we can deal with such cases. Using these same tools, we can then move on to show that constraint is a causal relationship.

3.6.2.1 Omission

We discussed earlier the fact that productive accounts of causation often struggle with cases of causation by omission. However, counterfactual/probabilistic accounts can deal with these readily. This is a point in favour of these accounts. As Hall & Paul (2013, 186) say:

> "there are clear-cut cases where omissions cause things. If Billy fails to take his medicine and falls ill, his omission causes his illness. To deny causal efficacy in such cases is to give up any hope of a reasonable model of causation".

Schaffer (2000, 286-289 [cited in Hall & Paul, 2013, 186]) also makes the case that we ought to accept causation by omission because to not do so would require us to reject some paradigm cases of causality as non-causal. Such intuitive cases are significant, we ought to prefer accounts that can accommodate causation by omission. The interventionist approach can easily describe cases of causation by omission via SEMs. A classic example of omission is that the plant dies because it is not watered.

Omission: Plant

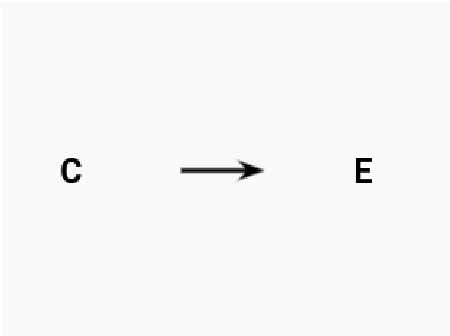
Variables

- C: Plant is watered
- E: Plant dies

Structural Equations E=1-C

Assignment C=0; E=1

Graphical Representation



(Wilson, 2018a, 742)

There are also examples of grounding that can be mapped with the same model. For example, the set of unicorns is empty because unicorns do not exist. We can imagine interventions that would show the plants example to be causal, perhaps you arrive back from holiday to find all your plants have died. You suspect this might be due to them not being watered but want to test if this was indeed the cause. You might start by putting plants in a laboratory setting and isolating them. You give them adequate space, heat and light but water only half of them. If you monitor this over time, you will see that the watered plants survive and the unwatered plants die. You can then conclude that a causal relationship was at play between not watering the plant and the plant dying. Similarly in the omissive grounding case, we can see a dependence relation at play, just as in **Socrates**, if we imagine that unicorns had not existed at a world, then we will see that this leads to the set of unicorns (at that world) being empty.

In 2.3.1 I noted that a possible move to deal with transitivity problems is to accept a contrastive account of causation[31]. A similar move might help in dealing with omission cases. For those unwilling to accept causation by omission, one might think that it is nonetheless true to say that "had the plant been watered rather than not watered, it would have survived rather than died". A move to contrastivity is a way of avoiding committing to absence

[31]This is discussed in more depth in 6.4.

causation, and if people wish to take that route, counterfactual theories are particularly effective at explaining contrastivity.

3.6.2.2 Overdetermination

Overdetermination has classically been a tricky kind of causal situation for counterfactual accounts to deal with. However, the interventionist approach does a good job of capturing them. Consider the following case:

Firing Squad:
A prisoner (**E**) is ready to be executed by a firing squad of two guards, **A** and **B**. The execution is set up such that, if **A** does not fire, then **B** will. Let us assume **A** indeed fires and as such kills **E**. In this case, **A** caused the death of the target (abbreviated as **E**), but the corresponding counterfactual 'If **A** had not fired, then **E** would not have died' (i.e. if ¬**A** then ¬**E**) is false because **B** would have fired instead and killed **E**.

The interventionist account, paired with SEMs, can deal with cases of overdetermination and deliver the correct results. What seems to go wrong here is that, whilst we know that **A** caused **E**, the relevant counterfactual is false. It seems false because a relevant fact has been missed out, the presence of **C**. A SEM with a specified degree of appropriateness will include variables such as the second assassin and allow the causal relationship to be traced so that correct conclusions about it can be drawn. If we want to capture the relation in **Firing Squad**, we can construct a model like so:

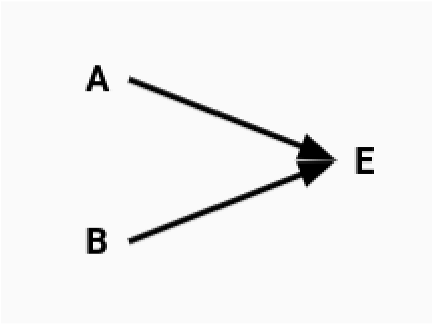
Variables

- A: Guard A fires
- B: Guard B fires
- E: Prisoner dies

Structural Equations
E=max(A, B)

Assignment A=1; B=1; E=1

Graphical Representation



(Wilson, 2018a, 742-743)

By including the crucial overdetermining variable, we can perform more complex interventions (e.g. had **A** not fired and **B** was restrained, then…) that allow us to determine the causal relationships. Moving from a naive counterfactual view allows us to account for cases like this. Interventionism in particular lends itself well to this because we can make sense of combining interventionism, i.e. as above "if **A** had been stopped from firing and **B** had been restrained then…". Interventionism makes it particularly easy to see the causal relationships in overdetermination cases like this where a naive view would fail.

It is worth noting that we can also similarly map grounding cases. For example, if we plug in the following variables then we have a case of grounding overdetermination on the above model.

Variables

- A: Potion contains 1 gram of arsenic
- B: Potion contains 1 gram of strychnine
- E: Potion is poisonous

(Wilson, 2018a, 743)

The potion contains two poisons, arsenic and strychnine, so the overall potion is poisonous. Arsenic or strychnine alone would ground that property independently, but as they are both present it is overdetermined.

The construction of appropriate models is how interventionism deals with these wrongtracking counterfactuals. As Wilson states:

> "The distinction between right-tracking and wrong-tracking counterfactuals is then derived in the interventionist framework from a distinction between appropriate and inappropriate causal models. Right-tracking counterfactuals are those with antecedents specifying some combination of interventions on model variables in some appropriate model, and with consequents specifying some values for other model variables in that model."
> (Wilson 2018a, 738).

A model that does not include both of the guards in **Firing Squad** is not appropriate because it does not allow for testing of the relevant variable in isolation. We need to factor **B** into our models so that we appropriately screen off its influence and test to see the relationship between **A** and **E**. Just as we would need to screen off **A** if we wanted to see the relationship between **B** and **E**.

3.7 Issues for interventionism

As stated, I am considering mathematical causation from the interventionist perspective precisely because it is a tricky account of causation to mesh with my view. Doing so will show the flexibility of *Exotic Realism*. One of the main reasons that an interventionist account of grounding/constraint will be difficult is that the interventions in question will be *impossible*. It is therefore considering what straightforward causal interventionism says about impossible interventions and what a grounding interventionist can say.

3.7.1 Causal interventionism & impossible interventions

If no intervention is possible on **X** with respect to **Y** then there would be no way to test if the relationship is causal. This would not be a problem but for the fact that scientists readily describe such relationships as causal, so cannot be appealing to interventionism. The interventionist framework can accommodate this practice, however. For relationships in which an intervention is impossible, we can construct an analogous model and test interventions upon it. How analogous the model will be will depend on the strength of the results we hope to attain. An illustration will be helpful; consider the case of earthquakes, which we currently have no means of producing. What we can do is construct a physical model of an area of land complete with tectonic plate-like objects that we can control. With this model, we could create earthquake-like events and see what effects they have. We can then 'scale up' our results to actual cases of earthquakes and determine the actual causal relations. For cases in which this would not be similar enough, such as gases in the centre of stars, we can instead use a computer model. Indeed, this is often how scientists work within cosmology as it is obviously in principle impossible to intervene on things such as the early conditions after the big bang and the density of dark matter said to affect star formation. Given this practice of using such models to generate causal conclusions, we have an easy response to the problem.

3.7.2 Grounding interventionism

Before discussing how grounding interventionism can respond to these issues as well, it is worth considering a more neutral notion of what an intervention is. As Woodward defined them, interventions are an inherently causal notion. But we do not necessarily want that causation built into testing grounding relationships. We should make the language around interventions neutral. One way to intuitively grasp what this neutral intervention is would be to imagine it as if a god clicked their fingers and 'made it so', this would not necessarily have to be a causal 'making', but could be. Therefore we define an intervention in the following way. **I** is an intervention iff:

- **I** fixes the value of **X**
- **I** isolates **X** from previous value fixers of **X** so that the value of **X** is fixed by **I** alone.
- Any path from **I** to some variable **Y** goes through **X**.
- **I** is independent of any variable **Z** which fixes the value of **Y** and is on a path that does not go through **X**.

(Adapted from Woodward, 2003, 105)

If we want to see if **X** has a causal/grounding relationship to **Y**, our intervention should be the thing that fixes the value of **X**. If **X** is previously caused/grounded by something else, we should ensure that, instead, only **I** fixes its value now. For **I** to constitute a good test, we do not want it to directly cause/ground **Y**, it should only do so via **X**. And finally, in cases where there are multiple causes/grounds but we only want to test one, we should aim not to affect any other value fixer, **Z**. Anything which fulfils these criteria is an appropriate intervention and can be used to test the causal/grounding relationship. If a relationship passes this test (and, as we shall in chapter 4, constraint seems to) then we should not resist calling it causal, just as any relation that passes Woodward's original test is deemed to be causal.

3.7.2.1 Impossible interventions (again)

The problem of impossible interventions also arises for grounding interventionism, but is even more significant. Far from dealing with contingent physical objects and events, we are now dealing with metaphysically necessary features of the world. What it would mean to perform an intervention on such things is unclear, and the notion of things having been different is unclear[32]. It seems mysterious as to how we could even begin to test grounding relations with interventions. This is partially misguided, and it is possible to construct a case in which one might perform an intervention to test a grounding relation in a similar way to the model response for causal interventionism. Arguably, this is exactly what happens in thought experiments. For example, our scientific reasoning might lead us to the conclusion that mental states simply are brain states/are ultimately grounded in brain states. To test this, we try to imagine a being with the same brain states as us, but lacking any mental states (a zombie). Chalmers (1996, 96) does this to show that, because we can do this imaginative test, brain states cannot ground/be identical with mental states. A more concrete way we might test this is by building machine minds that lack mental states but do all the same 'physical stuff' as human brains. Alternatively, we might try things the other way around. We might try to conceive of a scenario in which a being has the same mental states as us but a completely different set of physical states, i.e. whether mental states are multiply realisable. Obviously, controversy surrounds the results of any of these thought experiments, and responses can be made to explain the 'data' we get from them but the point is that a model can be used to test a grounding relationship. The fact that there is not a unified take on what the model shows is not a problem, as even in the sciences, the results of models can be questioned in this way.

[32]Such interventions would naturally involve counterpossible conditionals, this is what makes them tricky to understand. Discussion of how we should treat counterpossibles will be delayed until chapter 5 when they can be given full attention.

3.7.2.2 Are models close enough?

A further objection can be pressed, however, that the models in the grounding cases, as above, are not close enough to the actual case being modelled. In the earthquake example discussed, we can see the model is simply a scaled-down version of the actual event and this is what allows us to manipulate it and draw conclusions from the interventions made. However, things seem different in the case of zombies. The worry is that the differences between the model and the actual case invalidate any results of interventions. Again though, this response is mistaken, such objections do not seem to land against imaginative tests like the case of philosophical zombies. Because, aside from being mere thought experiments, these cases are *exactly* like what they are supposed to model.

We can also lend ourselves to some help from causal models. Because we cannot perform interventions on the inside of stars, and we cannot build a physical model, we use mathematical models. But these mathematical models are radically different from the actual situation being described, the model is not even made up of concrete objects. Instead, we are intervening on mathematical structures and equations. I maintain that the reason scientists are happy to use such models is that they are as close as it is possible to achieve, and so have to suffice. A similar line of reasoning can help us in the grounding case, until (and unless) we can get much closer, thought experiments will simply have to suffice. It seems better to have some models and postulated relations, even if not totally accurate, rather than reject the models and have nothing. Therefore, we can defend thought experiments as sufficient for our purposes and a way of making a preliminary test on the relevant hypothesis.

A sceptic about interventionist treatments of grounding may be willing to accept the general reliability of thought experiments, but deny that we will be able to do this for all types of grounding. A response to this would be to show how we can perform interventions on other grounding relations. For example, take the truth-making example of grounding, the truth of a statement is typically taken to be grounded in the state of affairs the statement refers to. Take the example statement "**s** believes **r** about **Q**" this is grounded in the fact/state of affairs that **s** believes **r** about **Q**. As the statement is about **s**'*s* belief, it is grounded in **s**'*s* belief. An intervention we could perform would be to tell **s** that **r** is lying. This would remove **s**'*s* belief in **Q** and would render the statement false overall. This would seem to show that the truth of the statement is indeed grounded in the facts the statement is about.

The problem is still present, however, for certain cases where a φ is grounded in a ψ that holds of metaphysical necessity. For example, let us assume that the statement "1+1=2" is grounded in the fact that 1+1=2. No intervention can be performed to change this fact and test this relationship. The allegation would be that we cannot perform thought experiments on such matters. A response is available, however. Because we are dealing with metaphysical necessity and possible worlds we quickly run into the use of counterpossibles, e.g. "Had it been the case that parts did not ground wholes…". The orthodox view is that regardless of the content of the consequent, counterpossibles are (trivially) true. A full discussion of this point is the topic of chapter 5 but for now, I will simply say that I deny that all counterpossibles are true; some are false. We can properly perform these thought experiments because it is possible to get a false result and so highlight the dependence relation at play via 'intervention'. We can think about impossible matters and reason what would be the case if they held.

3.8 Section summary

In this chapter, we saw some different options for theories of causation which are compatible with the causation/grounding analogy. Where process theories face challenges, the interventionist picture does not. However, the issues with the interventionist picture, particularly those surrounding impossible interventions, are magnified in the case of mathematical constraint. With the groundwork laid for the analysis of grounding under the interventionist picture in this chapter and the previous, we are in a position to move on. Given that interventionism potentially represents a challenging causal theory for my account, in the next chapter, I will show that it is in fact compatible with the notion of constraint I will be working with. I take it to be a strength of my theory that it is adaptable to this tricky case, but it is important to keep in mind that constraint does not *have* to be tied to the interventionist picture. Constraint can also be viewed under a straightforward counterfactual view or a case of probabilistic causation. With a specific causal picture in mind, we can now go on to make the case that constraint seems to be a crossover case between grounding and causation. Given that constraint seems to be a causal relationship, and given that mathematical objects seem to be involved in constraint relationships, we can make the claim from chapter 1 that mathematical objects are exotic in virtue of being causal.

Chapter 4 - Mathematical Constraint[33]

4.1 What is constraint?

4.1.1 The intuitive idea

Amongst the types of explanation, there seems to be a distinctively constraining sort. We have an intuitive conception of things constraining the world and our potential actions in it. Consider the following example of legal constraint, **Driving**. In the UK one cannot (legally) drive until the age of 17 *because* it is illegal to drive under that age; the law in place (the constrainer) constrains the kind of actions we can (lawfully) take. Part of the explanation of why a 16 year old cannot (legally) drive will involve appealing to this law. This is a very weak illustration of constraint but a more familiar example would be the idea that the law of gravity will be involved in an explanation of why I do not float off into space. These two types of explanation by constraint should be familiar and intuitive enough for us. I want to argue that, in some cases, constraint relationships exist. In some straightforward physical cases, this constraint relationship is a causal one. I will then go further and argue that we can treat mathematical constraint as being causal for exactly the same reasons.

In this chapter, I will argue for a strong claim (**Con=G&C**), as mentioned in the introduction. I aim to argue that constraint is simply a form of both causation and grounding. In arguing for **Con=G&C**, I hopefully make the fallback weaker position more plausible. The weaker position (**weakCon=G&C**) states that constraint is simply in a broader class of dependence relations along with grounding and causation. Again, this would mirror a Bennett-style (Bennett, 2017) view of building relations. Constraint would simply be one of the building relations.

[33]Large parts of this chapter draw on a paper published as "Mathematical Causation" (Dickson, 2021).

4.1.2 Mathematical constraint

All of this leads us to an account of mathematical explanation proposed by Marc Lange (2017). Lange identifies constraint explanations as playing a key role in science but he maintains that these explanations are in fact not causal. Lange alleges that mathematical facts are similarly involved in a constraining relationship with the world, restricting the sorts of events that can happen in the world. Consider the following:

Strawberries:
The reason that Jane cannot divide 23 strawberries equally between her 3 children (without cutting) is *because* 23 is not divisible by 3.

Facts about mathematical objects constrain the way the world can be. If 23 had been divisible by 3, then it would have been possible to divide the strawberries between the children, although this is a metaphysical impossibility. Lange (2017, 6) describes such explanations as distinctively mathematical scientific explanations. He also claims that there are similar constraint explanations in the sciences which are not distinctively mathematical. An example of such a non-mathematical constrainer is the principle of energy conservation which, in basic terms, states that there is a set quantity of energy in the world which merely changes from one form to another. Energy conservation could either be a constraint on the way the world must be, or a coincidence. Given that scientists tend to treat this as more than a mere coincidence, it seems reasonable to therefore say that energy conservation is indeed a constraint. Laws concerning forces such as gravitation and electromagnetism are alike in conserving energy (Lange, 2017, 49). Given that energy conservation is a constraint, the force laws are the way they are *because* energy conservation holds. The reason that the law of gravitation conserves energy is *because* energy conservation holds. Lange wants to assert that constraint explanations are non-causal. In 4.8, I dispute this. Moreover, the constraint relation itself is a causal relation, as I argue in 4.2 and throughout the chapter. Constraint is causal because it shares a structure with physical constraint relations that we would intuitively want to claim are causal relations. Given that these physical cases are causal, we should conclude that the constraint relation is causal, and that therefore mathematical objects are causal. Given that constraint explanations describe a causal relationship, they are also causal.

4.1.3 Mathematical constraint as grounding

I also argue that it is plausible to describe constraint as a grounding relation. Let us recall that grounding is intended to be a technical notion which captures the meaning of "in virtue of" relationships. A classic example of grounding is "the ball is red and round in virtue of being red and being round". This seems parallel to the constraint notions discussed above, *p* cannot legally drive a car in the UK *in virtue of the fact that p* is under the age of 17. Jane cannot divide her 23 strawberries equally between her 3 children *in virtue of the fact that* 23 is indivisible by 3. The failure in each case is grounded in the fact that does the constraining. It is *because* of this fact that the failure occurs. We have made the *in-virtue-of* explicit in each case but it would normally be implicit, perhaps substituted by a *because*. There seems to be a clear thread running through all these uses of *because,* and I contend that what holds them together is that they are all examples of the grounding relation. Causation and grounding are privileged in that they both back explanation. In the cases above, offering the constrainer as the explanans seems genuinely informative, providing a genuine explanation. It seems that the constraint relation also backs explanation. One might think that constraint is a third kind of privileged relation that backs explanation, but I resist this. Constraint and grounding share many similarities that grant us justification to claim constraint simply is a form of grounding. For example, both seem to be synchronic relations which do not admit of hasteners and delayers[34], both seem to benefit from a contrastive treatment[35], and both intuitively have some connection with fundamentality[36].

Moreover, if one wishes to be resistant in saying that constraint is grounding, then (assuming that one also wishes to resist classifying it as causation), one will have to postulate at least three fundamental and distinct relations in their ontology, causation, grounding and constraint. This seems unattractive, we should try to reduce the fundamental terms in our ontologies as much as possible and unifying constraint under either or both of grounding and causation seems a good way to do that. One may alternatively wish to simply reject that constraint is a genuine relation. The problem with this is that there seem to be genuine cases of constraint, e.g. **Driving** and more to be discussed in this chapter. *Some* relation exists in all of these constraint cases, and it seems that what all these cases have in common is that it is the same relation at play in each and every one. Given this, unifying constraint as a form of

[34]Although see 6.1 for a discussion of this.
[35]See 4.7.1 and 6.4.
[36]See 6.5.

grounding is well founded. That is to say, given that constraint is a form of both grounding and causation, I argue that **Con=G&C** is true. From this point on, that constraint is a form of grounding will be assumed and the remainder of the chapter will be devoted to arguing that constraint is also a form of causation.

4.2 What does it take for something to be causal?

It is worth clarifying the aims of this chapter. It does not seem as though something is causal if it cannot in principle form a part of a causal explanation. I want to distinguish this from the claim that part of what it is for something to be causal is to be a part of a possible causal explanation[37]. But if something is causal, then we should, in principle, be able to describe the causal 'processes' it engages in. Describing this process from start to finish will suffice to be a causal explanation. Consider that if we describe the causal processes occurring when a domino falls and follow this process through until we have reached the fact that it hits another domino, we have explained why the second domino falls. The same principle should apply in the case of constraint.

An issue is that there is 'less space' for a causal explanation. What I mean by this is that we can give many explanations for many explananda, some of which will be causal. In constraint cases, the constrainer is 'remote' in an important sense. It seems that the only explanations that constrainers could possibly be a part of are constraint explanations as Lange describes them. If these explanations are not causal, then it would seem strange to describe the constrainers as causal. There is a bit of a complication here because Lange sometimes describes the mathematics at play in constraint cases as causes, e.g. Lange (2017, 19). Because of intuitions like those above, I find it difficult to reconcile this with the explanations in question being non-causal. As such I will distinguish between two important claims about constraint cases:

1 - Constraint is a causal relation.

[37]By 'possible' I merely mean 'could be articulated'. I remain neutral on whether or not causal explanations exist regardless of us having an articulation of them or not. The point is we need not have actually articulated a causal explanation for something to *be* causal.

This will be the main focus of this chapter and I plan to establish **1** through the use of interventionism and structural equation models (SEMs). Because I take Lange's thesis that constraint explanations are non-causal as being in tension with this point, I will spend 4.8 trying to show that he is mistaken in this respect. This brings up another claim:

2 - Constraint explanations are causal explanations.

The argument above should hopefully pull on intuitions that if constraint is a causal relationship, then constraint explanation is causal explanation. So, given **1**, we are licensed to believe **2**. For now, I will focus on **1** but in 4.8 and in particular 4.8.1 I argue that constraint explanations are causal explanations. Independent arguments for **2** would be important but fall beyond the scope of this thesis as interesting avenues for future research.

Ultimately, the account I will develop is going to say that our folk conception of causation does not capture all of the causal entities. This may be counterintuitive to some, so before moving on it is worth offering some support to the intuitions behind my account.

The constraint relation may indeed not look like paradigm cases of causation, but this is because we ordinarily think causation is a productive force, e.g. billiard balls hitting one another. As we discussed in 3.6.2.1, inactions and omissions are sometimes considered causes, but the relationship they have to their effects is very unlike this productive causation. Moreover, there are alleged causal relations in science unlike either of these. For example, some argue that it is causal when an entangled electron is measured for its spin, fixing the spin measure of its entangled partner (Einstein's 'spooky action at a distance'). When a fire is put out, we would ordinarily consider the water being poured on it as a cause of its being extinguished and not the person praying nearby. But modern physics tells us that both should be counted among the causes of the extinguishing. This is because everything within an event's past light cone has influenced that event. Modern physicists accept this despite it being a radical re-understanding of our folk conception. I use this example here to highlight that the folk intuition about causation is not an infallible guide and we might find one reason or another to reject those intuitions, as some interpretations of modern physics require us to do. If it is acceptable in this case to distance oneself from the folk conception, then it is also acceptable for my account to distance itself from the folk conception.

For most people, causation is thought of in the productive sense. But this conception does not easily match with common examples of causation. Again consider causation by omission which need not be an example of energy transfer. Considerations like this are generally seen as a failure of energy transfer accounts of causation and require a re-understanding. Given the failure of these folk conceptions, and the developments of physics. We are licensed to consider including mathematical constraint as causal as more reasonable. We have previously amended our view of causation in seemingly radical ways, there need not be as much resistance to doing so again. Our intuitions about causation are not reliable, and so given that this new understanding can be explained with interventionism, it is reasonable. There is precedent and justification to make this move even if it does go against our intuition. Ultimately, as we will see in 4.4 and 4.5, when we look at constraint abstractly via SEMs we can see that the mathematical constraint and physical constraint share the same structure. It is this similarity in structure that should lead us to conclude that both relations are causal.

4.3 How constraint works

4.3.1 Pigeonholing

One notion we might use to intuitively gain a grasp of constraint is that of pigeonholing in mathematics (Lipton, 2009, 46-47). Given a number of objects, ***n***, and a number of containers, ***m*** into which the objects are distributed, where ***m<n***, it is inevitable that one container must contain more than one item. The restrictions in place on the system, the constrainers, constrain the possible positions that objects can take. There are similar examples of constraining relationships which seem intuitively causal, take the following:

3-doors:
An agent, ***s***, *must* pass through one of three doors, ***a***, ***b*** or ***c***; however, an agent, ***t***, has partially obstructed ***s*** in this by locking two out of the three doors, ***b*** and ***c***. Given that ***s*** *must* pass through one of the doors, it seems straightforward to say that ***t*** *caused* ***s*** to pass through door ***a***.

This might seem radically different to explanations by constraint discussed earlier because it is agential. However, we can consider **Automated-3-doors**, whereby ***b*** and ***c*** are locked randomly by a computer program. Plausibly still, this computer program would *cause* the agent to travel through ***a***. One can consider a similar 2-door system in which the lock operates by detecting the spin of one of a pair of entangled electrons. If the electron is detected spin-up, the door will lock, if spin-down, it will remain open. Given that the detection of one electron will fix the spin of the other, this will ensure the system has one unlocked and one locked door. Again, it would seem plausible to say that the detection of the electron as spin-up, thus locking the door, would cause the agent to go through the unlocked door. Finally, we can remove all mention of agents at all and instead talk about pipes with flowing water, or rivers. A blockage in two of three pipes, or two of three distributaries, will cause the water to flow down the other. In each of these cases, a constrainer on conditions results in a further condition holding.

4.3.2 Is this constraint?

Some may be sceptical that these are constraint relationships. The examples discussed above seem wildly different to, for example, Lange's example of the principle of energy conservation (2017, 49). However, I argue that these are the same class of explanations. Lange's examples and other examples of constraint all possess a specific structure. Although these cases may seem different, they also possess this structure and so are examples of constraint. Mapped into structural equation models (SEMs), these cases will all display the same features.

What is relevant in describing whether or not a relationship is causal is the structure of the case. As I will assert in 4.4 and 4.5, mathematical constraint shares a structure with physical constraint cases that we would ordinarily describe as causal, so we should say that mathematical constraint is a causal relationship. Considering the cases discussed in 4.3.1 above, the essence of what happens in a constraint case is that a potential path is being cut off, resulting in another path obtaining. This 'screening off' relation is constraint, in some cases it may be necessary to distinguish physical constraint (e.g. **3-doors**) from mathematical constraint (e.g. **Strawberries**). This is not a distinction in kind, it is just two examples of the same relation. I merely rely on a terminological distinction to show that in the physical case the relation possesses certain features, i.e. a certain structure and a certain pattern of counterfactual dependence, and so is uncontroversially causal. Given that the mathematical cases possess the same structure and pattern, we should conclude that they are causal.

4.4 Physical constraint

4.4.1 Intuitive example

To build the case for constraint being causal, it will be helpful to consider a specific case of physical constraint in-depth, exploring the features it possesses and how an interventionist account would assess the relation. I will then show that the same structure applies to cases of mathematical constraint. The example we will use is as follows:

River:
A river flows down a hill and comes to two potential distributaries (**A** and **B**) the river could then follow to continue descending. As it happens, the river only flows down one of these, **B** (perhaps because it is closer or lower). One day, a tree falls and blocks the path of distributary **B**, resulting in the river having to carve a new path through a different distributary, **A**.

It should be intuitive to say that the fallen tree has caused the river to flow down distributary **A**. What would also be correct to say is that the fallen tree has made it impossible for water to flow down distributary **B**. This latter feature implies this is a constraint relationship. The tree has restricted the range of possible actions the water can take. In **Strawberries**, the mathematical fact results in a certain distribution of the strawberries being impossible. The mathematical fact is a constrainer, it constrains the physical world. The cases are parallel and both seem to be equally constraining.

It is worth noting that **River** is causal and constraining in virtue of the same elements. The cause is the fact (or event) that the tree has fallen and the effect is the fact (or event) that the water flows the way it does. The constrainer is the fact the tree has fallen, this results in the fact the water flows the way it does. Now of course one could say that these relationships are not identical and we instead have two at play at once, but I see no strong arguments for this. As I will argue, these relationships behave in the same way under similar interventions and the patterns of dependence remain the same. Given the fact that they are between the same elements, motivations from parsimony should compel us to conclude there is merely one relation at play here.

4.4.2 Interventionist recap

It will be worth briefly recapping interventionism and SEMs that we discussed in 3.5 and 3.6 before going into the constraint model. Recall that interventionism is best described as a test for when it is apt to describe a relationship as causal. **X** causes **Y** if it is the case that an intervention on **X** with respect to **Y** causes **X** and proceeds to cause **Y**. Essentially, causation is this distinctive pattern of counterfactual dependence that we test by interventions. SEMs are a set of variables, which can be considered as questions, e.g. “Is the rock thrown?” or “Does the window break?”. The assignment of values to these variables can be considered the answers to these questions, i.e. a 1 would be a yes, and a 0 would be a no (Wilson, 2018a, 719). Along with this, we construct a graphical representation which displays the dependence relations at play. If the model is accurate, we will be able to perform interventions and demonstrate the dependence relations, i.e. if we intervene on the rock being thrown we will be able to affect whether or not the window breaks, but not vice versa. With this overview in mind, we can move on to consider the constraint relationship as a SEM and how it behaves under an interventionist treatment.

4.4.3 Structure of constraint

River:

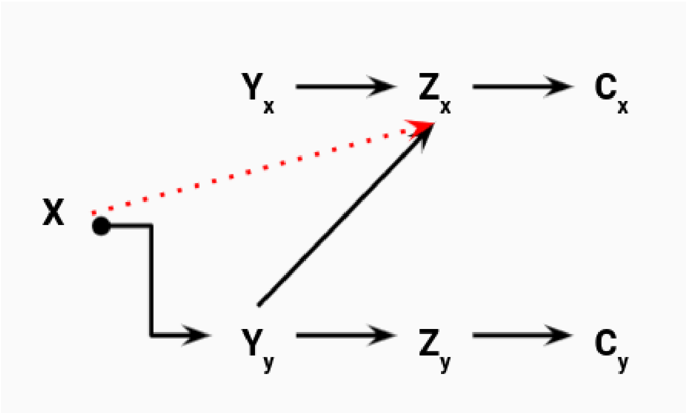
Variables:

- **X**: Tree falls
- $\mathbf{Y_x}$: Distributary **A** open
- $\mathbf{Z_x}$: Water flows through distributary **A**
- $\mathbf{C_x}$: Consequences
- $\mathbf{Y_y}$: Distributary **B** open
- $\mathbf{Z_y}$: Water flows through distributary **B**
- $\mathbf{C_y}$: Consequences

Structural equations: $\mathbf{Y_x{=}Y_y{+}1, Z_x{=}Y_x, C_x{=}Z_x, Y_y{=}X{-}1, Z_y{=}Y_y, C_y{=}Z_y}$.

Assignment: $\mathbf{X{=}1; Y_x{=}1; Z_x{=}1; C_x{=}1; Y_y{=}0; Z_y{=}0; C_y{=}0}$.

Graphical representation



The dashed red line from **X** to $\mathbf{Z_x}$ represents the transitive (or contrastive as we will see later) dependence of $\mathbf{Z_x}$ upon **X** because **X** ultimately rules out certain other possibilities ($\mathbf{Y_{y,}} \ldots \mathbf{Y_L}$); it is the ruling out of these other possibilities that causes $\mathbf{Z_x}$ (combined with the 'active' possibility of $\mathbf{Y_x}$). We can see how this abstract structure applies to the river case by assigning variables as above.

This is a causal case because interventions on the value of **X** will alter the value of the **Y** variables and this effect will remain stable across counterfactual situations. But also the value of the $\mathbf{Y_y}$ variable will affect the value of the $\mathbf{Z_x}$ variable. In this way, **X** is a transitive cause of $\mathbf{Z_x}$. We can see this is the correct result by the fact we would readily say that the blocked distributary was the cause of the water flowing down distributary **A**. Interventions such as preventing the tree from falling, dropping trees in similar locations and seeing where the water flows etc. will allow us to test the causal nature of this relationship and establish these dependence patterns.

4.5 Mathematical constraint

However, we can plug in variables relating to mathematical constraint and the structure will still apply, moreover, interventions on the same variables will result in the same effects, like so:

Strawberries:

Mathematical Constraint

- **X**: 23 is indivisible by 3.
- $\mathbf{Y_x}$: 23 objects divisible between 3 people only non-evenly
- $\mathbf{Z_x}$: The strawberries divisible in a particular way
- $\mathbf{C_x}$: Further consequences
- $\mathbf{Y_y}$: 23 objects divisible evenly into 3 groups
- $\mathbf{Z_y}$: 23 strawberries divisible evenly between 3 people
- $\mathbf{C_y}$: Further consequences

The mathematical fact that 23 is indivisible by 3 constrains it to be the case that Jane cannot divide her strawberries equally between 3 people. Parallel to **River** it does so transitively. The mathematical fact constrains it to be the case that 23 objects cannot be divided evenly into 3 groups, which in turn constrains the strawberries.

4.5.1 An interventionist treatment

In **River**, let us suppose we perform an intervention on **X**, changing its value to 0, i.e. stop the tree from falling. We will see that the value of the $\mathbf{Z_y}$ variable will change to a 1, i.e. water will flow down *B*. In this case it is conceivable that we could build a scale model of **River** and perform interventions on it, allowing us to scale-up those conclusions to the 'real' case. Equally with **Strawberries**, let us suppose that 23 had been divisible by 3, well then it would have been possible to divide 23 objects evenly into 3 groups, and furthermore Jane would have divided her 23 strawberries evenly between her 3 children[38]. The $\mathbf{Z_y}$ variable would have changed in exactly the same way. So, on an interventionist treatment, 23 being indivisible by

[38]There may be some concerns that we cannot simply suppose such a thing, I recognise this issue and it is discussed initially in 4.7.3.1 but more fully in chapter 5.

3, caused Jane to fail to divide her 23 strawberries between her 3 children. Distributary **B** being blocked is the cause of the water flowing down distributary **A** in the same way that 23 objects being indivisible into 3 groups is the cause of Jane distributing the strawberries in the way she ultimately does. In these cases, there is a distinctive pattern of counterfactual dependence that is maintained across the interventions. Recall that this pattern is exactly what interventionism tells us to look for in determining whether a relationship is causal. We can determine that **River** is causal with certain interventions and this seems to be the correct result. But interventions on equivalent variables will show the same pattern of counterfactual dependence in **Strawberries**. It seems then, that we should conclude that **Strawberries** is a case of causation.

4.5.1.1 Impossible interventions

In 3.7, I discussed how one might use interventions on models to form causal judgements about relations upon which it is impossible to perform interventions. This seems to be a reasonable response in the case of straightforward interventionism, but one might naturally wonder if we can extend this success to mathematical constraint. The worry is that in the cases of laws etc. they are only physically necessary, there is room for them to have been different, so we can easily make sense of that being the case. However, mathematics is at a much higher level of necessity and so *could not* have been different so the interventions we would need do not make sense. In a recent paper, Baron, Colyvan & Ripley (2019) discuss similar issues and their thoughts can be applied here.

Mathematical truths are necessary. As such, people may assert that alternate mathematics or interventions on mathematics (mathematical twiddles in their terms) are impossible and therefore inconceivable. But this is not the case. Take an example of a proven theorem, e.g. Fermat’s last theorem. Before it was proven, presumably the idea of it being false had to be entertained. The idea of mathematics in which this theorem did not hold was certainly considered conceivable (the proof would have been much easier otherwise) even though such mathematics is impossible. This is also the case with theorems yet to be proven, e.g. Goldbach’s conjecture. There is a definite answer as to whether or not Goldbach’s conjecture holds (even if it is unprovable and/or unknowable). But until we know, we can conceive of it as either holding or not, even though at least one of those things is impossible. Baron, Colyvan & Ripley allege that if we can conceive impossible mathematical situations in the cases in which their possibility status is unknown, then we can equally do this in cases where the possibility status is known. The fact that we know whether or not it is impossible should not affect our ability to conceive of such things (Baron, Colyvan & Ripley, 2019, 8-9). In terms of interventions on mathematical relationships, as discussed we can certainly conceive of these relationships having been different. So we could consider interventions on this in a similar way to Lewisian small miracles whereby we consider the closest world in which the value is different and as much else as possible remains the same. Alternatively, we could model a mathematical system (perhaps governed by a paraconsistent logic) and see what effects this has. Thought experiment reasoning along these lines could allow us to reach conclusions about what would or would not be the case at such worlds.

4.6 How much of mathematics does constraint get us?

A concern at this stage is that, so far, mathematical constraint has only shown to involve the natural numbers, as in **Strawberries**. If constraint is supposed to give us epistemological access to mathematics, one might worry that the natural numbers are only a very small portion. Thankfully, mathematical constraint is common, and it does not only concern the natural numbers.

Consider the example of Kirkwood gaps as discussed in Colyvan (2010). Kirkwood gaps are regions in the asteroid belt between Mars and Jupiter. Kirkwood gaps contain relatively few asteroids (Colyvan, 2010, 302). The scarcity of asteroids in these gaps is a fact in need of explanation. For each asteroid that enters a Kirkwood gap, there will be a complicated causal explanation, in terms of the asteroid's specific causal history, as to why that asteroid in question did not form an orbit inside the Kirkwood gap. However, these individual explanations do not tell us why *no* asteroid can form a stable orbit in a Kirkwood gap (Colyvan, 2010, 302). If all we have are individual causal explanations, then it is a coincidence that there are no asteroids in the Kirkwood gaps (Colyvan, 2018, 29). But there is more to this than individual causal explanations. The reason that asteroids that would otherwise form orbits inside the regions covered by Kirkwood gaps cannot form orbits is mathematical. The resonance of the system means that some orbits are unstable, i.e. for any object which is heading for an orbit in a Kirkwood gap, the object is dragged off course (Colyvan, 2010, 302). This can be explained mathematically by appealing to an eigenanalysis. As Colyvan (2010, 302) puts it: "The explanation of this important astronomical fact is provided by the mathematics of eigenvalues (that is, basic functional analysis)". That is to say, this explanation will appeal to, at a minimum, vectors and real numbers. In short, then, this mathematical explanation involves more than just the natural numbers as in **Strawberries**.

There is the further question of whether the mathematical explanation of Kirkwood gaps is a constraint explanation. Consider that, without a mathematical explanation, the absence of asteroids is a mere coincidence. With the mathematical explanation, we can see that this is no mere coincidence, the mathematical explanation provides us with new information. Moreover, recall Lange's treatment of constraint. Constraint explanations are explanations which help us to generalise over apparent coincidences and give us the overarching reasons for any such apparent coincidence occurring. This is exactly what is happening in the mathematical explanation of Kirkwood gaps. Furthermore, the relation itself seems to share some structural similarities with the mathematical constraint examples we have seen already. An object is on a trajectory that would otherwise be stable, but due to various mathematical facts, the particular orbit it would enter into is unstable (i.e. impossible), as a result, the asteroid veers off course. This should sound familiar because this is functionally parallel to **Strawberries**. A mathematical fact has resulted in the impossibility of a particular event/fact. This is surely going to be a much more complex constraint than **Strawberries**, but it still seems to be essentially a matter of constraining possibilities. Highlighting the example of Kirkwood gaps as a case of constraint is important. It shows that constraint involves more than just the natural numbers. It also shows that it is promising that constraint explanations are going to be common across mathematics. There are likely to be many constraint explanations that involve different parts of mathematics. For an object-based account of mathematics, we want epistemological access to as many different mathematical objects as possible. Showing that constraint can give us access to the natural and real numbers is a significant portion of mathematics and although much more remains to be captured, this should serve as a promising start.

4.7 General Objections

4.7.1 Is constraint a 3-place relation?

One objection to discuss is that there is a significant difference we have not yet considered between causation and constraint; causation is a 2-place relation, whereas constraint seems to be 3-place. Causation takes place between a cause and an effect, and that is that. Constraint seems to involve three things: the constrainer, i.e. the factor doing the constraining; the initial range of options, some of which are 'ruled out'; and finally the range of options left by the constrainer, e.g. the open distributary in **River**. Even if causation and constraint are alike in many aspects, differing with respect to the number of relata would be a significant disanalogy. It is worth spelling out this claim more before we analyse it. Let us take a canonical report of a causal relation: "Event (or fact) **A**, caused event (or fact) **B**". Compare this with a canonical report of a constraint relationship: "Our initial range of options, **A**, was constrained by **X**, leaving a smaller range of options **B**". In the constraint case, we have to refer to three things to get a complete report, but not in the causal case.

It is worth considering one potential response which *will not* work. One might accept that constraint is a 3-place relation, but point to transitive causation examples as three-place relations as well and claim that this shows there is no distinction after all. Reframed in this way, a canonical report of transitive causation might take the form of: "**A** caused **C** *via* **B**". We might view the canonical report of constraint as actually taking the same form: "**X** resulted in **C** being the case, *via* constraining **B**". One might then claim that, because both relations can be three-place, there is no problem. As I have said though, this response fails, and talking about why will highlight interesting features of constraint.

Whilst constraint and causation *can* be 3-place, they do not both have to be. Discussions of transitivity aside, all agree that causation is at least sometimes not transitive, i.e. in our first case when **A** causes **B**. On the other hand, it seems that constraint *has* to be 3-place. If we were to simply say that "**X** constrained our range of options **A**" or "**X's** constraining resulted in **B**" we have incomplete accounts. In the first case, we need to talk about what the constrainer resulted in, and in the second we need to talk about what exactly was constrained. If we similarly cut out parts of a report of transitive causation we are left with complete causal reports, either "**A** caused **B**", "**A** caused **C**" or "**B** caused **C**". So one could say that constraint was always a subset of causation, now it is just a smaller subset, it is analogous to transitive causation only, not general causation. This would be a mistake though, instead we can solve the problems with transitivity more broadly and unite constraint and causation by taking a route that Schaffer (2012) takes by saying that both relations are contrastive.

Due to the transitivity issues discussed in 2.3.1, Schaffer thinks that rather than being a 2-place relation, causation is actually a 4-place contrastive relation. Rather than a canonical causal report being of the form "**A** caused **B**", Schaffer takes it that causal reports are actually of the form "**A** rather than **A*** caused **B** rather than **B***" (2012, 132). We can take a similar approach with constraint, a canonical report should be of the form "**A** rather than **A*** constrained it to be the case that **B** rather than **B***". This seems to fit well with how we view constraint. We can illustrate this with the case of gravity. As Lange (2017, 49) points out, gravity is not, according to general relativity, a force at all. However, discussing it *as if it were* a force is helpful in illustrating how constraint works. The (Newtonian) law of gravity being an inverse square law (**A**) as opposed to an inverse cube law (**A***) constrains it so that the gravitational force conserves energy (**B**) as opposed to not conserving energy (**B***). A constrainer being in place instead of another makes it that case that things are a certain way when they might have been otherwise. This seems in line with what we already thought about constraint and Schaffer (2012) makes a compelling case that it is actually how we already view causation. With that in mind, a commitment to a contrastive account of constraint[39] does not modify the view too much.

4.7.2 Should the level of necessity matter?

Having seen how the relationship is structured, how it would behave under interventions and having considered some of the constraint relationship's features, it is time to consider some objections to the account in general. An immediate concern is that the different cases of constraint seem disanalogous in an important way. In **River**, both options are physically possible, i.e. the river *could* have gone the other way, but due to a contingent factor, did not. In **Strawberries** this is not the case. Instead, we are not really dealing with two options, we only have one, what actually happened. The other option was not an option in the first place because it is mathematically impossible.

[39]This will be discussed more in 6.4 where we discuss how a move to contrastivity can help us avoid problems with transitivity, asymmetry and irreflexivity that we discussed in 2.3.1 and 2.3.2.

In as much as this is a disanalogy, it is not significant. The degree of necessity has certainly shifted, that is true. But as I have already argued, this is less important in determining a relationship to be causal and instead we should look to things like counterfactual dependence and the relationship structure. Another way of responding to this is that in an absolutely deterministic universe, everything follows from the events that preceded it, necessarily. For example, in **River** the tree had to fall, that was always going to happen. But it would still be informative to talk of what would happen if it had not fallen, and talk about the constraint relationships in a counterfactual sense. So I allege it is still informative to think of these mathematical constraint cases in the same way. A further allegation is that in **River** we are discussing straightforward contingent matters which might easily not have held. In **Strawberries** we are dealing with necessary truths, things it seems could not have been otherwise. In this case it is less clear that we are dealing with causation rather than another relation. The response here is that we can help ourselves to counterpossible statements to take the place of the contingent statements such as "if the tree had not fallen". Counterpossibles like "if 23 had been divisible by 3" can be used in the mathematics case and we can see whether certain situations would have held if this was the case. Whilst the modality we are dealing with is different, the abstract structure of these relations is the same and how they behave under interventions is unchanged, this seems more important in determining a relation to be causal; the modality can be considered merely a difference in degree rather than kind.[40]

4.7.3 Constraint interventions

4.7.3.1 Intelligibility of interventions

Another potential issue is the intelligibility of interventions. We can make sense of preventing the tree from falling but not what it means to intervene such that 23 be divisible by 3. It is easy enough to make sense of counterfactuals like "if the tree had not fallen". We know what that world in which the tree did not fall would look like. We can imagine holding everything fixed except for the tree falling in that world. But with counterpossibles, what we mean by "if 23 had been divisible by 3" is less clear, what that world would look like is mysterious. If we change that mathematical fact then one might naturally think we have to change more, that

[40]An appeal to counterpossibles is of course controversial, 4.7.3.1 and 4.7.3.2 will deal with this briefly and the issues will be more fully addressed in chapter 5.

there would be a massive ripple of changes. The objection is that we simply cannot track these changes enough to make sense of these counterpossibles. So even though we could make statements about imagined interventions on mathematical facts, they would not be meaningful. Moreover, we could not make conclusions about what would lead from them because we do not know what else would be true or false at that world.

Lange (2019a) responds to similar issues concerning his theory of mathematical objects that my theory can borrow. As Lange sees it, the essence of the problem is that counterpossibles 'ripple out' into the world at such a magnitude that we cannot make sense of what is true and false in that world. His response is that this is the case with a lot of ordinary counterfactuals that people would use. For example, we tend to think that for counterfactuals like "if Julius Caesar had been alive today…" we can hold everything else fixed. But this is not the case, because we can ask how it is that Julius Caesar came to be alive today, whether he time travelled for example. That would require a lot of differences, for example the existence of time machines. Perhaps instead of time travel it is just that the baby which was born and grew up to become Julius Caesar was actually born in 2007. But then unless he has the same upbringing (which seems impossible) he will in fact not be Julius Caesar. Lange's (2019a) point is that although it might seem like ordinary counterfactuals hold with 'everything else fixed' this is not the case. All counterfactuals ripple out because the world is fundamentally connected. Seemingly simple changes will be connected to other things that would require changing. It would not be possible to simply change one thing whilst holding everything else fixed.

With this in mind, it would seem disingenuous to criticise counterpossibles due to the number of changes required. If it is not a problem for counterfactuals it should not be a problem for counterpossibles. As discussed in 4.5.1.1, although it might be difficult to conceive of mathematical counterpossibles, it is not beyond our ability. We might well have to do some work in spelling out what exactly is held fixed in counterpossible worlds and what remains unchanged but the fact this work needs to be done is not a problem[41]. Alternatively, it seems in ordinarily rippling counterfactual cases we can merely stipulate that such counterfactuals are non-trivially true or false, so this should also be available to us in the counterpossible case (Lange, 2019a). Lange has got this right, just because a counterfactual seems simple to

[41]See chapter 5 and particularly 5.7.3 for more discussion of this.

understand does not mean it is and just because a counterpossible seems impossible to understand does not mean it is. Given this, it does seem we have grounds to say that we can understand impossible interventions. A problem that remains however is whether every statement we make about such an intervention will be vacuously true.

4.7.3.2 Are counterpossibles trivial?[42]

One may still object that counterpossibles are useless in virtue of being vacuously true. We will not be able to form a useful account of interventionism concerning mathematical objects if all counterpossibles are trivially true, as on such a view, “If 2+2=4 were false, then 2+3=5 would be false” and “If 2+2=4 were false, then 2+3=5 would be true” are both equally true. This is a particular problem for my account of causation because although I appeal to these counterfactuals, the orthodox view states that both “Had 23 been divisible by 3, then Jane *would* have divided her strawberries equally between her 3 children” and “Had 23 been divisible by 3, then Jane *would not* have divided her strawberries equally between her 3 children” will both be true. If both are true then it does not seem that we can make these interventionist assessments because any “**A** caused **B**” claim and its opposite would turn out to be true. My account thus requires the non-triviality of counterpossibles. It is worth considering an argument that intuitively suggests that counterpossibles can be non-trivial.

As discussed in 4.5.1, one might worry that we cannot merely suppose that 23 was divisible by 3 (or that 2+2=4 was false). Wilson (2018b) has a response to these sorts of worries. Wilson advocates that at least some counterpossibles are non-trivial. The argument he offers to support this goes as follows:

1. The interventionist analysis of grounding is correct. (Premise.)
2. The fact that Socrates exists fully grounds the fact that Singleton Socrates exists,
but not vice versa. (Premise.)
3. If the interventionist account of grounding is correct, then if A fully grounds B
(but not vice versa) then an intervention on A would alter the truth-value of B,
but not vice versa. (Definition of interventionism.)
4. It is false that if an intervention had prevented Singleton Socrates from existing,

[42]This section includes a very quick motivation for the non-triviality of counterpossibles. This discussion is complex and not settled here and will be handled in much more detail in the next chapter.

then Socrates would not have existed. (From 1, 2, 3.)
5. 'If an intervention had prevented Singleton Socrates from existing, then Socrates would not have existed' is a counterpossible. (Premise.)
6. Not all counterpossibles are trivially true. (From 4, 5.)
Wilson, 2018b, 724-725

This argument seems plausible, **1** is a reasonable claim to make. The interventionist account of causation is very useful and productive, and equally so in the grounding case. Grounding relations do seem to possess a specific kind of counterfactual dependence that is captured by the interventionist perspective. For those unhappy to unite causation and grounding as Wilson (2018a, 2018b) is aiming to do then we could maintain a suitably altered version of interventionism. **2** also seems a reasonable claim to make, but if one rejects the specific example it can be substituted for any purported case of full grounding. **3** of course follows from **1** by being the definition of an element. **4** simply follows from the previous premises. **5** is an additional premise but quite a plausible one. If sets exist, they do so necessarily. So, if Socrates exists, then necessarily, singleton Socrates exists. Given this picture, preventing the existence of singleton Socrates (in a world at which Socrates exists), is impossible, so **5** refers to a counterpossible. Moreover, it does genuinely seem the case, as per **4**, that the counterpossible referenced in **5** is false. This is because worlds at which there are no sets could be worlds in which there could still be the members, i.e. Socrates could exist without the singleton but not vice versa. This falsity is more than merely trivial, concluding this counterpossible is false seems correct in a way that concluding it is true does not. Although intuitive, this argument only takes us so far, a full defence of the idea that counterpossibles can be non-trivial takes place in chapter 5.

4.7.4 Are constrainers full or partial causes?

We might ask if **Strawberries** is a case of full grounding/causing or partial grounding/causing. If it is full, a problem emerges that the mere fact of 23 being indivisible by 3 will not mean that 23 strawberries are indivisible by 3, because the antecedent makes no mention of strawberries it makes no demands on the world. But if we want to conclude that it is partial, then it is mysterious what will fill in the gap. If whatever fills in the gap mentions strawberries, it seems that would result in problematic circularity. We will consider both options.

4.7.4.1 Constrainers as full causes

To say that 23 being indivisible by 3 is a full cause in **Strawberries** is to say that it is "all that is needed" to cause Jane to fail to divide her strawberries. As stated though, this makes no actual demands on the world. 23 being indivisible by 3 makes no mention of anything in the world. One might be concerned that it does not seem like this alone will do any causing, so the allegation is that, in fact, it is not a full cause. Thankfully, we can respond to this worry.

It follows from 23 being indivisible by 3 that 23 objects are indivisible into 3 groups. This is not mentioned in "23 is indivisible by 3" but is in a sense an 'instance' of it, whilst being a distinct claim. Perhaps we could use a Hume's principle-like reasoning process to understand divisibility. Hume's principle can be put as "The number of Fs = the number of Gs iff the Fs and Gs are equinumerous". If we take this to be an analytic truth, then we can derive facts about equinumerous collections from facts about numbers. For example, "23 is indivisible by 3 iff any 23 objects are indivisible into 3 groups". This definition would allow us to 'jump' from the mathematical fact to the object fact (or vice versa). We could even jump from the mathematical fact to the fact that 23 strawberries are indivisible into 3 groups, because this is just another example of a collection with 23 members. I assert that this is the connection between the mathematical claim and the claim about the world, and I have already argued that this connection is causal.

There might still be some resistance from people who will want to say that 23 being indivisible by 3 and the fact that 23 objects cannot be divided into 3 groups will still not *do* anything to the world, because they make no direct mention of things *in* the world. Firstly, this response is slightly question-begging and only arises if you start from the view that mathematical objects are acausal and do nothing to the world. Even putting this to one side, we can respond. In any case of causation, to "get things working" we need all the relevant causal claims and connections, but what we also need are *things* to be connected. It is surely the case that 23 being indivisible by 3 and the fact of 23 objects being indivisible by 3 will not *do* anything to the world unless there are *things* to obey these rules. In short, we treat this like an equation and plug in objects (e.g. strawberries), with this assumption that there are such things then we will get the constraint we are looking for. The assumption of there being objects in existence is not an additional cause in the situation, it *is* the situation. It is helpful to think of constraint relations as rules. With things to follow the rules, effects will be produced. Having to plug in the fact that there are objects in existence to obey constraints firstly does not mean the original alleged cause was not a full cause, and secondly, this additional fact is not thereby a cause itself.

4.7.4.2 Constrainers as partial causes

Alternatively, we might wish to say that 23 being indivisible 3 is a partial cause. Naturally, if we want to say this then we need an account of what 'fills in' the cause and we need to make sure that whatever it is, we need both it and the mathematics. I see two candidates. The first is the fact that 23 objects are indivisible into 3 groups. Couple this with 23 being indivisible by 3 and Jane will be unable to divide her strawberries. Here, nominalists would caution us that we have a third option. Rather than asserting mathematics as either full/partial, we could treat it as no cause, dispensing with it. Treating 23 being divisible by 3 and 23 objects being indivisible into 3 groups as equally causally important means that end up doing the same work. This makes one of them redundant, and in this case, the obvious candidate for redundancy seems to be the mathematical objects. Nominalists would say we should take this route, but we can resist this for two reasons.

Firstly, it is not clear that this strategy of eliminating mathematical facts in favour of nominalistic ones is always going to work. It may not always be possible to eliminate the (nominalistically) problematic constrainer. We run into probability frequently within quantum physics and some argue that this ties it inextricably with mathematical objects that cannot be nominalised away (Bueno, 2002). If this remains the case, then nominalist theories will struggle to account for this probability nominalistically. Whilst we may be able to take this nominalistic strategy in simpler cases, we may not be able to in all cases and so will still be appealing to mathematics.

The second reason that this strategy is unattractive is that a choice between this and taking 23 being indivisible by 3 to be a full cause is essentially the choice that Lange (2017, 49-58) discusses when he talks about treating energy conservation as either a constraint or a coincidence. The reason we should not take this route is that in treating mathematical facts as mere coincidences we lose a lot of explanatory power and multiply the number of potentially unexplained facts. It is worth pausing here to draw out what is meant by this.

Let us assume that energy conservation is a brute fact of the world, not explained by anything else, and is a constraint, i.e. it is the reason why force laws conserve energy. For simplicity, assume there are three force laws. These are explained by energy conservation, which is itself unexplained, in this situation we are committed to one unexplained fact. Alternatively, we might assume that energy conservation is in fact a coincidence explained by each force law conserving energy independently. The fact that each force law conserves energy is a brute fact, so there are three brute facts. It seems that in terms of theoretical virtues this second situation is worse off by having more unexplained brute facts. Likewise, holding the mathematical fact as a coincidence and not a constraint is going to increase the number of brute facts we have to accept.

Given this, it seems reasonable to resist the nominalist demand and try to find another candidate for a completing cause. We might naturally turn again to something we considered in the previous section, the fact that objects exist in the first place to be constrained. This, plus 23 being indivisible by 3, could lead us to the fact that 23 objects cannot be divided between 3 groups and then onto the strawberries fact etc. To expand on my point from 4.7.4.1, this will not work because such a fact is not a cause. Admittedly my account is aiming to describe things not normally taken to be causes as causes so it is worth highlighting the difference.

Even people who object to my account should agree that the fact that things exist seems even less like a cause than any constrainer we could care to name. This suggests a fundamental difference between these things. Consider questions of the form “What caused x?”. Let us take **Strawberries**, “What caused Jane to fail to distribute her 23 strawberries evenly between her 3 children?”. An informative answer to this could be mentioning the mathematical constrainer of 23 being indivisible by 3. What would not be informative is to mention that the strawberries exist. This is not a cause of the situation because it is part of the situation which needs explaining. The fact that things exist seems more like an initial condition. This is not merely a matter of question-framing either, any way we phrase such a question we are tacitly assuming that the things involved exist. We are not always assuming the constrainers in place, which is why it is informative to point to them. This is a useful distinction to bear in mind. We should agree that whatever constrainers are, they are not the same sort of thing as the assumption that objects exist. At this stage, we can dismiss this as a candidate-completing cause. However, I see no other available candidates. In absence of another candidate, we should take the conclusion of 4.7.4.1 on board and hold that 23 being indivisible by 3 is a full cause in **Strawberries**, albeit an indirect one. Constrainers, then, are full causes.

4.8 Comparison with Lange's Account

Lange (2017) asserts that constraint models such as **Strawberries** are not causal, but I will resist this claim. If we think of mathematical facts as being constrainers then we need to say that they stand in some relation to the world. They need to *do* something to be constrainers rather than coincidences. This *doing something* will have to take a form and if it is not causation or grounding then it will have to be a distinct relation of its own. We would then need to have at least 3 fundamental relations in our ontology, causation, grounding and constraint. What I propose is that rather than these 3 distinct relations at play we simply have causation and grounding, with constraint seeming to be an interesting overlap. Admittedly, the idea that constrainers, particularly mathematical constrainers, can be causes is proposed here as an intuition and so it does need to be supported by argument. I will do this in several ways:

As pointed out in 4.2, there are two distinct claims to make about constraint: that it is a causal relation (**1**), and that constraint explanations are causal explanations (**2**). I have already given my arguments for **1**. But I also said that if something is causal, it should form part of a causal explanation, and the only candidate for this in the case of constraint is the constraint explanation. If **2** is false, then this might suggest that **1** is false too. As such I will now discuss Lange's (2017) account of why these explanations are non-causal and argue against it.

For Lange, an explanation is causal if it explains "by virtue of describing contextually relevant features of the explanandum's causal history or, more broadly, of the world's network of causal relations" (Lange, 2017, 16). Interestingly, for Lange, there are cases in which an explanation can be causal without referencing any causes. In the explanation of the uniform motion of a body, we cite the absence of forces on it. But Lange thinks that this explanation is causal because it works “by citing a law that specifies how a body must move in the absence of any cause influencing its motion” (Lange, 2017, 13). In this way, the explanation still cites the causal structure of the world because it implicitly describes what the causal structure would have to be like for it to be different. In the case of a square peg being unable to fit in a round hole, Lange emphasises that the peg’s squareness is not a cause of the failure, instead, it ensures there are such causes of the failure, i.e. that the surfaces will come into contact (Lange, 2017, 14). Concerning how such an explanation does work, Lange asserts “The peg’s squareness explains not just by describing the actual causes of the outcome, but also by telling us that had different initial conditions prevailed, similar causes would have produced a failure to fit through the hole” (2017, 14). The peg explanation is a program explanation in line with the account proposed by Jackson and Pettit (1990), but contrary to them, Lange thinks that the way it works is sufficient to make it a causal explanation.

Now we understand what it takes for an explanation to be causal on Lange’s account, we can discuss why he says that distinctively mathematical explanations are non-causal. Of **Strawberries** he says “this explanation does *not* work by describing the various causal relations that would have obtained if she had tried other ways of distributing the strawberries, and then showing that each of these causal processes would have failed to distribute the strawberries evenly” (Lange, 2017, 14). What this explanation does is to show that the outcome is inevitable to a stronger degree than mere causal relations, i.e. the actual ones such as the peg hitting the edges of the circle or counterfactual ones about similar pegs and circles in the same size-size ratio coming into contact. This sort of explanation, rather, shows that the outcome is mathematically inevitable. It is important to recognise that Lange accepts that distinctively mathematical explanations may happen to cite causes, but that this is not where they derive their explanatory power from. For example, Lange accepts that both Jane having 23 strawberries and having 3 children were causes of her failure to evenly distribute the strawberries (Lange, 2017, 19). Although these may be causes, and figure in an explanation, that does not suffice to make the explanation causal, because the power of the explanation does not come from citing the causal structure of the world. At this stage we have a picture of Lange’s account, explanations by constraint are non-causal because, whilst they may appeal to causes, they do not appeal to them *as* causes. Constraint explanations derive their explanatory power from describing how the phenomena could not have been otherwise, with a variety of necessity stronger than physical necessity (Lange, 2017, 10). We are now in a position to assess this account.

Lange makes an interesting point which deserves highlighting “Benacerraf argues that it is mysterious how we could know about mathemabstracta because we are not in causal contact with them. By comparison, I am arguing that it mysterious [sic] how mathemabstracta could constrain (and thereby explain) physical facts when there is no causal, nomological, constitutive, or similar connection between them” (Lange, 2021, 50). The point is that mathematical objects can only be constrainers if they *do* something. Without asserting that they are *actually doing something*, then any theory of mathematical objects is going to suffer from Benacerraf challenges. If mathematical objects are going to play his proposed constraining role, then constraint has to consist in them *doing something*. The quote above highlights a few places this *doing something* might lie, the constraint relationship is either:

- **a.** Causal
- **b.** Nomological
- **c.** Constitutive, or
- **d.** ‘Similar’ to all these.

In terms of **c**, it is pretty clear that constraint is not of this kind, nor could mathematical objects ever be constitutive. Little argument needs to be made for this, it should hopefully suffice to say that no ‘making up’ claims are being made and indeed if any theory of mathematical objects would be committed to mathematical objects ‘making up’ reality then it would have to do a lot of groundwork to be plausible, and such a task seems impossible.

Claim **d** has to be spelt out in terms of grounding, as it is not clear what else might be appealed to. Although Lange (2019b, 1) has claimed that constraint is not grounding, in 4.1.3 I claimed that we should consider constraint to be a form of grounding. My further claim is that constraint is also a kind of causation, which I believe I have argued for successfully. Given how Lange discusses things such as energy conservation being constrainers, it seems that he would want to say that this constraint relationship is nomological. But given that the sorts of constraint Lange talks about seems to exhibit a pattern of counterfactual dependence that suffices for causation, **b** will end up being causal as well. In short, I claim that **a**, **b** and **d** are all, in fact, the same. Since mathematical objects are not constitutive, the only way for them to *do* something is to be causal in one of these ways. Unless they are causal like this then it is not clear that Lange's account is going to allow room for mathematical objects/facts to do the very work that he says they need to do.

Lange (2017, 2019a, 2021) wants to emphasise that the principles he mentions are constraints rather than coincidences. But the issue is that unless the constraints actually *do* something, then they are simply coincidences. To be constraining rather than coincidental there has to be some sort of interaction with the world, they have to *actually constrain* the world and affect what can happen in it. But Lange wants to say that these constraint explanations are not causal explanations. This would suggest that the constraint relation is not a causal relation. But without being a causal relation then it is difficult to see how these constraints are more than mere coincidences. Without the power to *do* anything to the world, these constrainers are just as idle as traditional abstract objects. If we grant that the constraint relationship is a causal one, and takes part in causal explanations then this worry is avoided because we are asserting that these objects are exotic and interact with the world, elevating the above mere coincidences.

Lange locates his distinction between constraint and causation in the level of necessity involved in the respective cases. A constraint explanation works by showing that something is necessary to a greater degree than the laws of nature could make it. For example, Lange (2017) asserts that **Strawberries** is a non-causal explanation because regardless of what the laws of nature are, the divisibility would still be impossible due to the 'higher' or 'stronger' force of the mathematical fact at play. Constrainers then, act on their 'effects'[43] in a way 'above' causation. I argue that Lange is mistaken to focus too strongly on this. The jump from natural necessity to some other form of necessity is not as significant as he believes. Constraint, as discussed earlier, has a particular form and structure with a specific pattern of counterfactual dependence. Given that we can construct parallel cases like **River** which are straightforwardly causal, this suffices to make constraint a causal relationship. This structure of counterfactual dependence is what is important in assessing a relationship. Shifts in the level of necessity are not important in this way and do not signify a difference in kind.

It will be possible to construct an explanation with the structure of a constraint explanation (**Javelin**) for a given thrown Javelin's trajectory in terms of air resistance and gravitational

[43]I am using "effect" here not to prejudge that the relationship is causal, merely because it is a useful term for what the constrainer results in.

forces. This would explain why it had one trajectory rather than another. This would show that the explanandum arises from certain features and is necessitated by the laws of nature, i.e. gravity. The explanation would show how, in the absence of other causes (e.g. being thrown harder or softer), the javelin *must* move. Given Lange's account of causation, it seems he would want to say that this explanation would be causal, given that it appeals to the past causal history of the javelin (who it was thrown by, how hard etc.). But once again, it seems constraining. The difference between **Javelin** and, for example, **Strawberries** seems more a matter of degree than of kind. The only difference here, in Lange's terms, is that **Javelin** involves a degree of necessity at the level of the laws of nature. It does not seem to be a particularly useful distinction to hold that this explanation is causal, whilst **Strawberries** is not causal.

To strengthen this idea, let us consider the fact that we can construct causal explanations (which have the structure of constraint) with varying levels of necessity. **Javelin** can be thought of as constraint in terms of physics. It should be fairly easy to construct a parallel explanation in terms of chemistry, for example, the way ionic bonds form and behave will restrict behaviour in certain ways. Going 'down' further an explanation could be constructed in terms of biology, we can imagine an explanation which shows that organisms are constrained in terms of the size they can grow to. For example, there are upper and lower limits on how tall human beings can grow, given surface area to volume ratios and the square-cube law. All these explanations will differ in necessity but they will be equally constraining, and equally causal. To go 'above' the level of laws of nature, there will be stronger and weaker forms of constraint. As Lange says, the principle of energy conservation constrains what laws there might be. This is one level of necessity. But the principle of energy conservation arises from, and is constrained by, the principles of supersymmetry, which are at a higher level of necessity. The ultimate point I want to make is that just because an explanation points to a higher or lower level of necessity, does not mean that it is not causal. Where Lange draws the line between constraint and causation seems arbitrary.

4.8.1 Do constraint explanations reference causes *as* causes?

I also disagree with Lange's claim that explanations by constraint do not invoke causes *as* causes (Lange, 2017, 20). This seems particularly strange given that Lange asserts that things such as Jane's having 23 strawberries are causes of the phenomena involved (2017, 19). As in

4.2, if something is causal then it must in principle form part of a possible causal explanation. The only potential explanation of this kind could be the constraint explanation, so this must be a causal explanation. It seems mysterious how we could assert something is causal but that it does not play a part in any causal explanation.

I have argued that mathematical objects are causal. But they are causal in a very specific and restricted sense. Because how they interact with the world is so limited, though significant, the scope of potential causal explanations that they can be a part of is limited. The only explanation I see as being a candidate for the causal explanation is the constraint explanation. Lange (2017, 13) claims that causal explanations work in virtue of describing relevant features of the explanandum's causal history, or the world's network of causal relations. He claims that constraint explanations do not work in this way but it seems arguable that they do. Constraint explanations describe the world's network of causal relations in virtue of the fact they describe why some causal relations can or cannot take place. In **Strawberries**, the explanation seems to show that there are no causal relations which would lead to collections of 23 being divisible by collections of 3. Moreover, it seems to work by telling us that if different initial conditions had prevailed, then there would have been a similar indivisibility. The explanation generalises away from strawberries and children to show that with any so-numbered collections, there will be a similar failure. **Strawberries** seems to be very close to a program explanation in that the numbers of items involved encode a property of indivisibility, such that regardless of cutting method or starting point, the divisibility will fail. This should suffice to make constraint explanations a kind of causal explanation.

4.8.2 Lange's account of mathematical objects

It will be helpful to explain Lange's motivations for his view and his overarching view of mathematics so that we can illuminate his account of explanation with some context. Lange (2019a) starts at the same point as I did in the introduction, with the puzzle that if mathematical objects are non-spatiotemporal, acausal entities, then it is mysterious how they can constrain the world, given that they enter into no relationships with it. Given this, mathematical objects cannot be abstract objects or they would be subject to **MND**. At this stage, Lange and I diverge on views. Lange is still operating within the abstract/concrete dichotomy, so, given his rejection of nominalism, his only option is to conclude that mathematical objects must be like Aristotelian universals. I, however, do not see mathematical objects as fitting well into the Aristotelian conception. A characterisation of them as exotic objects is much better.

Lange's (2019a) view is that numbers are a relation between an aggregate of concrete objects and some unit-making universals. For example, the aggregate of concrete objects stands in the '23' relation to the unit-making universal 'strawberriness'. The 23 relation stands in the 'indivisibility' relation to the '3' relation (which relates a further aggregate to the unit-making relation 'childness'. In this way, Jane cannot divide the strawberries because the relations at play constrain the possible actions she can take. This account and explanation generalises to any aggregate which stands in the 23 relation and any aggregate which stands in the 3 relation because these two relations stand in the indivisibility relation; this is why the constraint explanation is explanatory.

This Aristotelian account is unsatisfying for several reasons. For example, the Aristotelian view leaves some features of mathematics mysterious. The account seems to suggest that there is nothing more to, for example, the number 2 than its instantiation in collections of exactly two objects. But someone might press that numbers seem to have properties that are above and beyond their instantiations, i.e. 2 seems to have the property of 'being the unique even prime'. It also seems that classifying numbers as a relation is mistaken. The mistake centres on the use of 'aggregate', with two ways to understand its use, both being problematic. The first route involves interpreting 'aggregate' as already appealing to a number of objects. The explanation would then run something like "the aggregate of 23 objects stands in the 23 relation to strawberriness….". This generates problems immediately, if we already have 23 objects then the 23 relation is redundant as it would be an unnecessary part of the explanation. But if we remove the 23 relation then it cannot stand in the indivisibility relation to the 3 relation, which we would also remove. Instead the non-relational 23 stands in the indivisibility relation to the non-relational 3. This looks like we have arrived back at something like a traditional platonic conception of mathematical objects that Lange (2019a) wants to rule out. But as Lange points out, and as I argued in the introduction, there are significant problems with this view.

So, we must take the other interpretation of the use of the word 'aggregate', that it refers to one object, something akin to a unified mass which is differentiated by its relation to a unit-making universal. Given the name 'unit-making universal', this is the conception that most naturally fits with Lange's view. The unit-making universal literally makes units out of aggregates. This conception, however, is not without its issues. Firstly, given that the aggregate is conceived of as a singular object, it seems that we have numbered it; i.e. "there exists one aggregate such that…". But given that numbers are relations on this view, we must instead say something like "the aggregate stands in the '1' relation to the unit-making universal 'aggregateness'". A problematic regress immediately forms.

It is also a pertinent question whether or not this account extends to all mathematical objects. We might grasp the notion that numbers are a relation, but it is not clear if this extends into other areas, e.g. geometry. In the often-cited example of why bees arrange their honeycombs into hexagonal cells, a theorem is cited which proves that the most efficient way of dividing a plane whilst minimising the perimeter and dividing the plane into equally sized segments (Lyon & Colyvan, 2008). In Lange's terms, this would be a distinctively mathematical explanation so, presumably, it would involve aggregates in a relation to unit-making universals. It seems that "honeycombness" would be taking the place of the number in the explanation and its relation to a plane. So it seems that the explanation will go via way of saying that an aggregate stands in the honeycomb relation to the unit-making universals of cell walls. But similar problems to the ones just discussed emerge, e.g. what shape the aggregate is 'prior' to its relation with the unit-making universal. By viewing mathematical objects as relations we place them 'outside' the objects in question, though closely tied to them. The problem is that being outside in this way will simply not work as an account because the notion of object that this relation attaches to is a deeply unsatisfying and unclear one. Nor can we work around this by placing numbers 'inside' the object as Aristotelian properties, as this will simply generate the same issues. Aristotelianism is an unsatisfying account for mathematical objects. Ultimately I agree with Lange (2017, 2019a) that abstract objects are too remote to fulfil the role that mathematical objects do as constrainers. But this need not wed us to Aristotelianism or Nominalism. The exotic objects framework can give us theoretical room to develop new accounts of mathematical objects. For example, we take the constraint relationship at face value and explain that it is a causal relation.

4.9 Section Summary

In the introduction, I discussed significant problems for platonist accounts of mathematical objects. Mathematical objects need to *do* something to play a role in our theories. But abstract objects do not seem to do anything. I, therefore, proposed the exotic objects framework in chapter 1. I claimed that, as a way to avoid the problems of platonism, we might view mathematical objects as non-spatiotemporal but causal. This chapter discussed physical constraint cases which we should readily describe as causal. These cases are causal and constraining in virtue of the same elements. In other words, constraint is a kind of causal relationship. These physical constraint cases have a distinctive structure and pattern of counterfactual dependence that is indicative of causation as treated under an interventionist account. Importantly, mathematical constraint shares the same structure and pattern of counterfactual dependence under interventionism. I argue that this justifies us in describing it as a causal relationship. This requires us to expand our notion of causation somewhat, but there is historical precedent for doing so. Given that constraint fits with all the different kinds of causation discussed in chapter 3, we should be put off from doing this. I argue that the best view of mathematical objects would be as non-spatiotemporal exotic objects that are causal in virtue of constraining.

Some problems for *Exotic Realism* remain, however. Straightforward counterfactual causal claims require the non-trivial truth of counterfactuals like "Had Julius Caesar been alive today, then he would have…". Producing non-trivial counterfactuals of that form is easily done. In the mathematics case that I propose, we similarly need non-trivial counterfactuals to back up the causal claims. The issue is that statements of the form "Had 23 been divisible by 3 then…" are not simply counterfactual, they are counterpossible. 23 *cannot* be divisible by 3, and the orthodox view of counterpossibles is that they are all trivially true. If I want to back up my causal claims I need to dispute this. Chapter 5 will do this by arguing against vacuism. There are also many problems sidestepped temporarily in this chapter along with previous ones. For example, those concerning transitivity, indeterminacy and synchronicity to name a few. Responses to these issues will be delayed until chapter 6.

Chapter 5 - Counterpossibles[44]

5.1 The counterpossible problem

In chapters 2 and 4, I appealed to counterfactual analyses of grounding and constraint. I aimed to make a parallel with counterfactual analyses of causation. Counterfactual analyses of causation work because we can determine the truth value of the relevant causal counterfactuals (e.g. via interventionism). I want to make a similar claim about counterfactual statements regarding constraint. A problem emerges however because these are not merely counterfactuals, but counter*possibles*. Their antecedent is metaphysically impossible, e.g. "had 23 been divisible by 3…" or "had sets not existed…"., I will symbolise these as $\boldsymbol{\varphi}_i > \boldsymbol{\psi}$. By being metaphysically impossible, it seems that there are no worlds at which $\boldsymbol{\varphi}_i$ will be the case. Orthodoxy states that a counterfactual ($\boldsymbol{\varphi} > \boldsymbol{\psi}$) is true when the nearest $\boldsymbol{\varphi}$-worlds are also $\boldsymbol{\psi}$-worlds. If the antecedent of a counterfactual logically implies its consequent, the counterfactual will come out true, regardless of the content of either part. If there are no $\boldsymbol{\varphi}$-worlds as described by the antecedent, then trivially *all* $\boldsymbol{\varphi}$-worlds are $\boldsymbol{\psi}$-worlds, i.e. the counterfactual will come out as true (Stalnaker, 1968). So, orthodoxy tells us that counterpossibles are trivially true, regardless of subject matter. If all counterpossibles are true, then I cannot point to the true ones as being indicative of causal dependence, as with counterfactuals, because even opposite pairs of counterpossibles would both be true. Therefore, I would not be able to back up the claims about dependence relations that I need to make. As such, this chapter will challenge the orthodox view, known as vacuism, and in particular Williamson's (2007, 2018) conception of it. Ultimately I will argue that his arguments in favour of vacuism fail and that we actually have compelling reasons to think that non-vacuism is true. Therefore, the relevant counterpossibles that I appeal to are available to back up my account of mathematical causation.

Of course, no non-vacuists need to say that all counterpossibles are non-trivial, so many restrict the non-triviality thesis to specific domains. One place it might be difficult to imagine the occurrence of non-trivial counterpossibles is in mathematical reasoning. Proofs by reductio seem to typically involve making impossible suppositions and then reasoning from them, ultimately proving that the supposition is impossible and necessarily false. For these to

[44]Large portions of this paper have been written into an independent paper and published as 'Against Vacuism' (Dickson, 2022).

work, it seems that all the statements in these proofs need to be true[45]. This is exactly as the vacuist prescribes and so one might view this as a compelling argument to agree with vacuism. I disagree, the reasons we can give for believing in the non-triviality of other counterpossibles are extendable to the case of non-trivial countermathematicals (counterpossibles with impossible mathematical antecedents). The basic argument I will offer is as follows: We have compelling reasons to think that there are non-trivial counterpossibles in the sciences, some scientific counterpossibles come out as false (and some true). This datum is significant enough to override the prescriptions of logical orthodoxy. Two things might be going on at this stage, either: we are implicitly using a non-standard semantics for counterfactuals, allowing them to come out with differing and non-vacuous truth values; or we are working within a standard semantics but still delivering this verdict, contra orthodoxy. It seems most likely that a vacuist would say such counterpossibles are true because there are no $\boldsymbol{\varphi}_i$ worlds. It further seems that what might actually be going on in the cases of scientific counterpossibles is that we are genuinely considering an impossible world, and because the truth value of $\boldsymbol{\psi}$ is up for grabs at these $\boldsymbol{\varphi}_i$ worlds, the truth value of the counterpossible as a whole can change. I will discuss how this is applicable to countermathematicals.

This is the strategy I will be considering in this chapter. We should genuinely consider the closest world at which any $\boldsymbol{\varphi}_i$ is the case. Considering impossible worlds, on some minimal level, allows us to deliver the verdict from science, it also shows us that vacuism is false. This chapter aims to make a number of unique contributions to this debate. As above, I aim to show that if we genuinely consider an impossible world/suppose that $\boldsymbol{\varphi}_i$, then different counterpossibles will have different truth values. This is illustrated with examples of scientific counterpossibles. I also aim to show that vacuism about countermathematicals is a redundant thesis. Further contributions to the literature are made by distinguishing between two kinds of projects that one might undertake in counterfactual form.

In the first case, one may wish to use counterfactual form to either work out the truth value of the statement which forms the antecedent, I will call such a usage 'investigative'[46]. The second case involves reasoning from the antecedent to potential consequents to see what

[45]This is something explicitly endorsed by Williamson (2018, 363) and will be discussed more in 5.7.1.

[46]It might turn out that all counterpossibles in an investigative project turn out to be true, or there might turn out to be some non-vacuous investigative counterpossibles. For the purposes of this chapter, I am happy to concede to the vacuists that all such counterpossibles will be vacuous because, importantly, the counterpossibles I am interested in are *not* of this investigative type and so these concerns need not apply.

would be the case, if the antecedent were true, I will call this usage ‘exploratory’. Importantly for the exploratory process, this is done *regardless* of the actual truth value of the antecedent, one has to genuinely imagine it to be true[47]. This distinction is a close companion of the distinction between a consensus and non-consensus context given by Yli-Vakkuri & Hawthorne[48] (2020). I aim to show that Williamson is engaged in an investigative process, rather than an exploratory process. Even if all counterpossible statements in investigative processes turn out to be true, it is not the case that counterpossibles used in exploratory ones will, so vacuism is false. What Williamson (2007, 2018) does is to determine the truth value of a statement ($\boldsymbol{\varphi}_i$), which he does by embedding it as the antecedent in a counterfactual form (an investigative process) . But this is different from genuinely considering what would be the case if $\boldsymbol{\varphi}_i$ were true (an exploratory process). Importantly, this genuine consideration is what Brogaard & Salerno (2013) are engaged in when responding to Williamson and this is the core reason that Brogaard & Salerno and Williamson appear to be in disagreement. They each think the other side is performing the same reasoning task and producing a different result, when in fact they are engaged in different enterprises. So this chapter aims to provide a methodological explanation of why the disagreement between vacuists and non-vacuists has arisen. It is also worth noting that, in the literature on counterpossibles, it is often the case that non-vacuists will provide examples of counterpossibles that are non-vacuous (e.g. Jenny 2018), but not necessarily provide a general overarching explanation for their non-vacuity. They say *that* the counterpossibles in question are non-vacuous, but not always *why*. I aim to start providing an answer to that question by pointing to the use of non-vacuous counterpossibles in scientific explanations, and showing how the mathematical cases mirror this.

As a final prelude, it will be worth clarifying some assumptions at play. I am implicitly assuming some variation of a Lewisian conception of worlds[49], which includes an account of impossible worlds. Although I have reservations about the specific account, Yagisawa’s (2010) extended modal realism is an interesting take on impossible worlds and the general spirit of that account can be kept in mind when impossible worlds are mentioned. Given an account of both possible and impossible worlds, one could maintain the standard semantics of

[47]How this is done is the topic of section 5.7.3.
[48]This also seems close to the suppositional procedure that Williamson describes in Suppose and Tell (2020). However, as I will argue later on, Williamson fails to properly engage in the suppositional procedure, and that is why he believes the counterpossibles to all be true.
[49]Along with the associated semantics.

Lewis-Stalnaker with minimal modifications. If there are impossible worlds, then we can assess counterpossibles on the basis of the closest one. However, it is worth noting this is not an inherent commitment of non-vacuism, one can be a non-vacuist without believing in this specific conception of impossible worlds, or in them at all. Perhaps one way to do this is to alter the standard Lewis-Stalnaker semantics instead.[50] Although one might say that an appeal to either a different ontology of possible worlds or a different semantics is problematic, it is worth noting that the only reason that Williamson can achieve a vacuist result is by assuming a specific semantic account/a specific conception of worlds, so if this is problematic for non-vacuists, it is equally so for vacuists. One way to read the following arguments about scientific and mathematical practice and the treatment of counterpossibles is that they provide reasons to think that experts in those disciplines make assumptions close to the ones described above, and that provides us a reason to make them too, rather than the ones that vacuists make. With these clarifications in place, we are in a position to begin considering counterpossibles.

5.2 Intuitive case

Counterfactuals are a useful tool in philosophy. Many theories have attempted to reduce causation to counterfactual dependence, elucidated in terms of true counterfactuals. As we have seen, interventionism is not reductive, but uses counterfactuals as a test for causation. Theories of counterfactuals are often backed up by theories of possible worlds; we can make sense of saying things could have been otherwise because there are worlds at which things are otherwise. Along with counterfactuals, there are counterpossibles; a subset of counterfactuals which have impossible antecedents, e.g. "if it had been raining and also not raining…"[51]. Vacuism states that counterpossibles are vacuously true. It is often said that if we allow an impossibility then 'anything follows' and so the truth value becomes meaningless. I dispute this, arguing that, whilst undoubtedly there are likely to be some counterpossibles which have vacuous truth (or falsity), not all will; just as with ordinary counterfactuals. The vacuist concern might come from the concern that counterpossibles must be backed up by the existence of worlds, just as counterfactuals are. But

[50]Another way to do this would be to adopt some appropriate form of non-classical logic. Whilst I do not wish to rule out this route, there is no space to consider it here.
[51]At exactly the same point, at exactly the same time, at the same world etc. There are situations in which it could be true to describe it as raining and not raining, where the 'it' you're referring to would be different places. Ignore such circumstances and focus on the impossible ones as these are the counterpossibles.

counterfactual-backing worlds seem unproblematic, we can easily make sense of a world in which one drinks coffee rather than tea. But counterpossible-backing worlds are not possible worlds but *im*possible worlds. For example, the counterpossible antecedent "had 23 been evenly divisible by 3..." could only be backed up by a world at which the necessary mathematical truths are otherwise. Assuming a nominalist account of properties to be correct and metaphysically necessary, the counterpossible antecedent "had platonic universals existed…" will have to be backed up by a world at which platonic universals exist. Similarly one could construct a counterpossible from the platonic perspective to talk in terms of "if tropes had existed…". Some may worry that there are no such worlds to back these claims, so all counterpossibles will be vacuously true. Others may worry that we have no sense of what such impossible worlds would be like, or a way to talk about how such statements could be true or false. These seem to be key intuitions behind vacuism.

I, and many others, argue that vacuism is too hasty. A large part of philosophical discussion involves reasoning counterpossibly. As mentioned, to at least some nominalists, the platonist is talking about matters which are impossible, and vice versa. If platonic universals exist, they exist necessarily, so a world at which they do not exist is an impossible world. Thus, the nominalist is talking counterpossibly when she says things like "if the universal redness had not existed, tomatoes would still have appeared to us as they do". Similarly, when a platonist makes claims starting "if tropes had not existed…." she is making a claim that is a counterpossible according to the nominalist. To an Aristotelian about such matters, both could be construed as counterpossibles. Clearly the fact that fruitful discussion can and does take place between metaphysicians shows that we can understand these counterpossibles in a meaningful way. Moreover, if these statements are underpinned by the existence of worlds, then the possibility of meaningful discussion about these counterpossibles would suggest that we can understand some impossible worlds. Indeed much argument against metaphysical theories requires that we give non-trivial truth values to such counterpossibles. We have to reason as if such a theory were true and then show it would lead to undesirable consequences. Clearly we are gifting counterpossibles such as "if theory x were true, it would lead to consequences y" with more than merely trivial truth values. Indeed we tend to assert such counterpossibles are non-trivially true in order to reject the theory in question. Such intuitive judgements are perhaps the reason we view different counterpossibles as having different truth values.

What this intuitive and basic line of reasoning shows is that we do seem to treat some counterpossibles as non-trivial, I will call these relative-counterpossibles. Such statements are only counterpossible relative to a theory of metaphysics, to other theories they are merely counterfactuals. What I have not shown is that the kind of counterpossibles I need to be non-trivial, namely countermathematicals, either belong to this class, or are independently non-trivial. The archetypal example of the sort I need is "Had 23 been divisible by 3…". One could accept that there are certain non-trivial counterpossibles whilst denying that all counterpossibles need to be. One way of doing this would be to point out that relative-counterpossibles are not *really* counterpossibles, because they are only counterpossible relative to a specific viewpoint or domain of discourse. Their antecedents are not impossible *tout-court*.[52] Countermathematicals might seem to be of this tout-court-counterpossible kind. 23 being divisible by 3 is not merely impossible relative to something, it is simply impossible. Given this, one might assert that the truth value of such a counterpossible is merely trivial and we cannot use them to, for example, test causal relationships as I want to. This would be bad news for my account of mathematics if this is where the discussion ended, but thankfully it is not. We have, then, two ways that a statement might be a counterpossible. A statement could have an antecedent which is only impossible relative to something; or a statement has an antecedent which is unrestrictedly impossible. One task is to find out what class of counterpossibles the mathematical kind falls into. There is a hollow sense in which we might say that countermathematicals are of the relative kind. If we consider countermathematicals as referring to different mathematical systems, then of course the impossibility is relative. 23 cannot be divisible by 3 in *this* system but could be in *this other* system. Of course, if this is the case, then these countermathematicals are not counterpossibles at all. This point will come up in 5.7.3 so for now I put it to one side and assume that countermathematicals are indeed absolutely impossible, rather than relatively so. At this stage, it is worth laying some more intuitive ground for the non-vacuous treatment of counterpossibles. We will then consider the merits of vacuism as proposed by Williamson.

Take the following case of a countermathematical that seems to be unrestrictedly impossible, but nonetheless seems to be non-vacuous. 23 is a prime number, its only factors are 1 and itself. This is a mathematical fact, a necessary truth. So one might say that a counterpossible

[52]This does not need to lead one to anti-realism. One could maintain there is a correct answer to the problem of universals so technically some statements are counterpossibles, but until we know the answers we cannot say which are the counterpossibles. Perhaps a more fitting name would be assumed-relative-counterpossibles.

antecedent starting “Had 23 been divisible by 3…” is unrestrictedly impossible. But this does not require us to say any counterpossibles involving it are trivially true. Compare the following:

A - Had 23 been divisible by 3, then 3 would have been a factor of 23
B - Had 23 been divisible by 3, then 3 would not have been a factor of 23.

It seems clear that **A** should strike us as true, whilst **B** should strike us as false. These are not merely trivial statements. Admittedly, this judgement may come from the fact that in **A** the consequent is merely a rewording of the antecedent, and in **B** it is a straightforward contradiction. That may well be a good point but we can still construct non-trivial counterpossibles:

C - Had 23 been divisible by 3,[53] then a calculator would have been able to perform the division operation on 23 and 3, yielding a whole number as a solution.
D - Had 23 been divisible by 3, then a calculator would not have been able to perform the division operation on 23 and 3, yielding a whole number as a solution.

Again, **C** should strike us as intuitively true, whereas **D** should strike us as intuitively false. Compare both these with a further counterpossible:

E - Had 23 been divisible by 3, then Paris would be in France.

A counterpossible such as **E** seems importantly vacuous in a way that **A-D** do not. This is not intended as a knockdown argument to prove non-vacuity, but what the above should do is pull on intuitions and at least open people up to the idea that countermathematicals, even if unrestrictedly impossible, need not be trivial.

[53]Perhaps the antecedent should be lengthened to include “and had mathematicians known this fact” to avoid makes-no-difference style complaints about the inefficacy of mathematical truths on our practice. This would still produce a counterpossible.

5.3 Vacuism

5.3.1 Williamson's view

Williamson is a strong proponent of orthodoxy. Standard Lewis/Stalnaker semantics are taken to result in all counterpossibles coming out as true, as outlined in 5.1 and 5.2. However, Williamson accepts that there are strong intuitive pulls towards judging that counterpossibles are non-vacuous. However, he thinks this is simply a pre-theoretic or pre-reflective intuition, and that once theory kicks in, we will see that they are vacuously true. Although there are situations in which we want to follow intuition rather than theory, there are limits to this. As Williamson (2007, 173) points out, there is a strong comparison with the treatment of empty universal quantifications as true. Our intuitive judgements may be that the following is false:

F - Every golden mountain is a valley

One can see how this intuitive judgement is attractive, mountains cannot be valleys, so golden mountains cannot be. We might intuitively view **F** as false, but a theory which rejected these pre-theoretic judgements in favour of making all empty universal quantifications vacuously true turned out to have significant advantages. Indeed, Williamson thinks this development proved vital to logic; "we should not allow the logic of counterfactuals to be similarly confused by unwillingness to recognize vacuously true counterpossibles" (Williamson, 2007, 175).

Let us consider two specific counterpossibles:

G - If Hobbes had squared the circle in secret, then sick children in the mountains of South America would have cared.
H - If Hobbes had squared the circle in secret, then sick children in the mountains of South America would not have cared.

Our instincts tell us that the first is false whilst the second is true. But Williamson (2018, 359) says this is purely an intuitive judgement. When we analyse it through the lens of theory we will, Williamson thinks, see that this judgement is incorrect. For Williamson, this is a legitimate move to make because nothing much hangs on whether counterpossibles such as **G** and **H** are trivially true. We can still get by doing philosophy if the claim that all counterpossibles are true (the vacuity thesis) turns out to be itself true. Given that we can get by with counterpossibles being trivial, and the theory seems to point in that direction, Williamson thinks we should accept it.

McLoone (2020) identifies two axioms which Williamson may be implicitly appealing to in order to reach this judgement, **CLOSURE** and **EQUIVALENCE**. McLoone uses '□→' to symbolise the subjunctive conditional i.e. counterfactuals, contrasted to his usage of a simple '→' to symbolise an ordinary indicative/material conditional. With this in mind, **CLOSURE** can be expressed as:

If $\vdash (B_1 \wedge ... \wedge B_n) \rightarrow C$ then $\vdash ((A \;\Box\!\!\rightarrow B_1) \wedge ... \wedge (A \;\Box\!\!\rightarrow B_n)) \rightarrow (A \;\Box\!\!\rightarrow C)$

(McLoone, 2020, 12164).

Conjoining **CLOSURE** with two axioms, that both Williamson and a non-vacuist would accept, results in any contradiction counterfactually implying any other. These further two axioms are **REFLEXIVITY** and **MP** (modus ponens).

REFLEXIVITY: $\vdash A \;\Box\!\!\rightarrow A$
MP: IF $\vdash A \rightarrow B$ AND $\vdash A$ THEN $\vdash B$

Since (1) $\vdash (P \wedge \neg P) \rightarrow (Q \wedge \neg Q)$ we can apply **CLOSURE** and get:

(2) $\vdash (A \;\Box\!\!\rightarrow (P \wedge \neg P)) \rightarrow (A \;\Box\!\!\rightarrow (Q \wedge \neg Q))$

We can sub in $(P \wedge \neg P)$ for A and get:

(3) $\vdash ((P \wedge \neg P) \;\Box\!\!\rightarrow (P \wedge \neg P)) \rightarrow ((P \wedge \neg P) \;\Box\!\!\rightarrow (Q \wedge \neg Q))$

REFLEXIVITY tells us that the antecedent of (3) is true so:

(4) $(P \wedge \neg P) \;\Box\!\!\rightarrow (P \wedge \neg P)$

Then from simple **MP** we can derive:

(5) $\vdash (P \wedge \neg P) \Box\!\!\rightarrow (Q \wedge \neg Q)$

So any contradiction counterfactually implies any other.
(McLoone, 2020, 12164)

The axiom of **EQUIVALENCE** can be expressed as:

$\vdash A \equiv A^*$ then $\vdash (A \Box\!\!\rightarrow B) \equiv (A^* \Box\!\!\rightarrow B)$

This essentially states that the antecedent of any counterfactual is not hyperintensional, i.e. it can be substituted with a logically equivalent sentence whilst preserving truth value. Since in the case of a counterpossible antecedent we are assuming that A and A* are both necessarily false, it is true that $A \equiv A^*$. So it seems that we can show **EQUIVALENCE** to hold by simply using **MP**.

With all this in mind, we can then see why Williamson thinks that counterpossibles should be vacuously true, this result follows from some logical principles that he holds. Having now considered Williamson's view, we can consider some reasons why one might reject his conclusion.

5.4 Against orthodoxy

5.4.1 McLoone

McLoone (2020) challenges Williamson's support of the vacuity thesis. Williamson (2018) thinks that our judgement that some counterpossibles have a non-vacuous truth value is pre-theoretic and pre-reflective. As he says, "once theory kicks in" (Williamson, 2018, 364), we will see that counterpossibles are vacuous. McLoone does not think that "pre-theoretic" and "pre-reflective" are interchangeable terms. We can call a judgement "pre-theoretic", if we have not assigned a truth value to it by considering a particular semantic theory (McLoone, 2020, 12162). But a judgement that is pre-reflective is instead one that is not robust: on further reflection the truth value we have assigned to it is likely to change.

We might ask if our judgements about counterpossibles are pre-theoretic or pre-reflective. They certainly seem pre-theoretic because we have not yet treated them under an explicit semantic theory. But if Williamson is alleging that these judgements are bad because we have not treated them under his preferred theory of semantics, this would clearly be an inadequate criticism, for we do not necessarily need to accept his theory. There are other semantic theories out there that treat counterpossibles non-vacuously. Moreover, our judgements about counterpossibles do not seem pre-reflective. After all, even Williamson admits that our judgements about counterpossibles say they are non-trivial unless we apply theory. Merely thinking about them more in depth and what they mean will not lead us to think they are vacuous.

Stalnaker (1968) explicitly tells us that if a theory contradicts such intuitions and usages in natural language then we should jettison the theory over those initial judgements. Let us think of this with an example from standard Lewis-Stalnaker semantics. On this account, both material and strict implication allow for antecedent strengthening. Essentially this means that one can add any conjunction you would like to the antecedent and the conditional will stay true. But this does not work for counterfactual conditionals. As Stalnaker (1968, 48) points out with the following example:

I - If this match were struck, it would light.
J - If this match had been soaked in water overnight and it were struck, it would light.

Clearly, our reflective judgements on **I** and **J** will tell us that **I** is true, whilst **J** is false. These judgements, however, are pre-theoretic (McLoone, 2020, 12163). Specifically, these judgements are made prior to the theory of monotonicity which led to antecedent strengthening in the first place. McLoone points out that rather than insisting our judgements must always follow the theory, Stalnaker allows that in some circumstances the reflective but pre-theoretic judgement must be kept and theory must be modified. If we want to make a semantics which accounts for the relative truth and falsity of **I** and **J** then we need to maintain our pre-theoretic judgements.

So we have two options when our reflective judgements conflict with our judgements under theory. We can jettison our theory in favour of maintaining our reflective judgements or we can jettison our intuitions in favour of a theory, provided it has sufficient advantages or theoretical virtues. This is exactly the situation we are in with counterpossibles. McLoone favours the first option, he thinks we should maintain our reflective judgements on the non-vacuity of counterpossibles. In order to support this, he wants to show that Williamson's preferred theory, and treatment of counterpossibles, does not have any specific advantages, and, moreover, that treating counterpossibles non-vacuously does not have any disadvantages.

McLoone (2020, 12164) alleges that Williamson's criticisms of non-vacuity are not specific. In the case above of universal quantification, there were marked disadvantages to holding onto our pre-theoretic judgements. McLoone (2020) picks apart exactly the things which he believes lead Williamson to this conclusion. Treating counterpossibles as non-trivial would require the rejection of at least one of the two axioms discussed above which Williamson sees as central, **CLOSURE** and **EQUIVALENCE**. Williamson does allude to this option when he says that being a non-vacuist would require rejecting elementary features of logic and that this makes the theory unattractive. McLoone identifies a problem with this, however. Williamson is not specific in *how* or *why* such rejections would make non-vacuism unattractive. He merely says that to do so would be "confused". The problem, as McLoone (2020, 12164-12165) points out is that only vacuists are committed to maintaining **CLOSURE** and **EQUIVALENCE**, non-vacuists would be happy to reject them. A non-vacuist might reject **CLOSURE** if they thought that some statements, such as the liar paradox could be both true and false (McLoone, 2020, 12164). Non-vacuists would also certainly reject **EQUIVALENCE** because they are committed to the antecedent position in a counterfactual being hyperintensional. Take $\mathbf{A_i}$ to be "Had Hobbes squared the circle" and take $\mathbf{A_i^*}$ to be "Had Hobbes squared the circle in secret". As McLoone says:

"Since both $\mathbf{A_i}$ and $\mathbf{A_i^*}$ are necessarily false, $\mathbf{A_i} \equiv \mathbf{A_i^*}$ is true. But if **B** is "Hobbes would have been a famous mathematician," then it seems $\mathbf{A_i} \;\square\!\!\rightarrow \mathbf{B}$ is true and $\mathbf{A_i^*} \;\square\!\!\rightarrow \mathbf{B}$ is false." (McLoone, 2020, 12164-12165)

For the non-vacuist, the following will come out as true, based on the imagined surprise and amazement of mathematicians:

"Had Hobbes squared the circle, he would have been a famous mathematician"

On the other hand, the following will be false, because if he had done just that in secret then of course nobody would know and so he would *not* be famous:

"Had Hobbes squared the circle in secret, he would have been a famous mathematician"

Of course, if one accepts **CLOSURE** and **EQUIVALENCE** then these will both come out as true, rather than having differing truth values. But the only reason offered to accept these axioms, is exactly so that counterpossibles like these come out as true. Someone who wants to judge these counterpossibles as differing in truth value need not accept the axioms in the first place. One can reject them and not be doomed to an unattractive theory.

As McLoone says, non-vacuists are at liberty to simply reject **EQUIVALENCE** precisely because it leads to vacuism. One might think that a motivation for **EQUIVALENCE** is that there is only one impossible world, i.e. all impossible antecedents have the same logical consequents. Again though, this is not something the vacuist has to accept. Sendlak (2021, 7288-7289) provides a good reason to reject this account. Sendlak argues that counterpossibles can be paraphrased into story prefixes, i.e. "Had $\mathbf{A_i}$ been true then…" is equivalent to "According to the story of $\mathbf{A_i}$…". It is simply not the case that all stories in this sense have the same consequences because they range over a limited number of claims concerning the subject matter. Take Priest's (1997) story 'Sylvan's box' as an example. Sylvan's Box is an impossible tale and according to it, it is false that Richard was at the farmhouse, it is not also true that Richard *was* at the farmhouse (1997, 79). Non-vacuists are in a position to make this claim (which should seem true), whereas vacuists are not and have to make counter-intuitive claims about the kinds of truths entailed by fictions.

The rejection of **CLOSURE/EQUIVALENCE** is only "confused" if we already assume that vacuism is true, and we have no reason yet to do so. The disadvantages that Williamson alludes to are simply not present unless one already accepts his preferred semantic theory. On this point then, there does not seem to be a significant advantage to accepting vacuism for counterpossibles, and given that it conflicts with some of our valuable intuitions to the contrary, we have some licence to in fact reject it.

So far we have considered an intuitive argument for non-vacuous counterpossibles, we have seen some reasons why one might want to reject this intuitive judgement, offered by Williamson. However, these reasons are not entirely persuasive. A non-vacuist can simply deny them, insisting on non-vacuity, just as the vacuist can affirm them, insisting on vacuity. To try to push this in the direction of non-vacuity I will consider some more concrete examples of non-trivial counterpossibles. These will be more than mere intuitive judgements as their truth or falsity will be based on, and essential to, good scientific reasoning. Showing that there are such non-trivial counterpossibles will provide strong support to non-vacuity. After all, arguments based on intuition and semantic theory in support of either side of the debate do have a place, but a concrete example of counterpossibles being appealed to outside of philosophical debate as vacuous or non-vacuous is much better.

5.4.2 Counterpossibles in science

Although the intuitive cases above can seem persuasive, the vacuist can simply say this is the mere appearance of distinct truth values, but the logical form tells us we are actually mistaken. The vacuist cannot appeal to this response in the case of scientific counterpossibles. If good scientific practice leads us to say some counterpossibles are false, we need to account for this. The results from science outweigh philosophical/logical inclinations we may have. Compare this with how developments in quantum mechanics have led some to consider alternative quantum logics to account for the discrepancies (e.g. Putnam, 1969), of course such usages are controversial and by no means the norm, but this shows that orthodox theories are not universal. The usage of counterpossible reasoning in the sciences is documented by a number of people (e.g. McLoone, 2020, Wilson, 2021).

5.4.2.1 Scientific explanation

One such discussion takes place in Tan (2019), in which he presents examples of the use of non-trivial counterpossibles in science. Not only are there multiple examples of counterpossibles used in science, but they are used in different ways and for different purposes. Tan (2019) focusses on their use in: scientific explanation; in reasoning about superseded scientific theories; and in idealised scientific models. In each of these cases, he offers an archetypal example of a counterpossible and discusses why viewing it as counterpossible and non-vacuous is the correct verdict. In the case of scientific explanation, the counterpossible offered is:

K - "If diamond had not been covalently bonded, then it would have been a better electrical conductor." (Tan, 2019, 40).

Tan claims that this is a scientific explanation of the fact that diamond cannot conduct electricity whereas solid carbon in some other forms can. The reason the covalent bonding explains this fact is because covalent bonds do not leave free electrons, they 'use up' all the electrons forming the strong bond. In other substances, free electrons allow for electrical conductivity (Tan, 2019, 40). Diamond's property of poor conductivity is brought about as a result of these bonds, i.e. the microphysical structure. So **K** provides an explanation in virtue of highlighting that dependence relation. But one might wonder if this is indeed a counterpossible; one may wonder whether diamond could have been otherwise bonded, leaving this as a mere counterfactual.

One can approach this in two ways, we might consider whether something is called diamond in virtue of its microphysical structure or in virtue of its theoretical role in science (Tan, 2019, 41). The first approach straightforwardly delivers the result that this is a counterpossible, because if something is only diamond in virtue of its microphysical structure, then something possessing a different microphysical structure would not be diamond. As a matter of metaphysical necessity, diamond has the structure that it does. So it is metaphysically impossible for diamond to be differently bonded. Alternatively, one may think that we define diamond by its theoretical role, the diamond-stuff is the stuff that does *x, y and z*. But the reason diamond is distinguished from other substances, and the reason it does the things it does, is because of its microphysical structure. In other words, nothing else could do the things diamond does without its microphysical structure. Nothing could fill the diamond role without actually being diamond. Again, it is metaphysically impossible that diamond could have been differently bonded than it in fact is. It seems, then, that **K** is a counterpossible. Tan (2019) goes further than this, he insists that this is also a counterpossible which is non-vacuously true. This is because it describes an empirical fact, that the poor conductivity of diamond physically depends on its microphysical structure. So science relies on non-vacuous counterpossibles in scientific explanation (2019, 42). Nor is this an isolated case, many scientific explanations of why substances have the properties they do will rely on a similar explanatory structure.

If we were to contrast **K** with another counterpossible, we can motivate this non-vacuist verdict more:

K* - If diamond had not been covalently bonded, then it would have been no better at conducting electricity.

K* is false because the covalent bond is the reason for the poor conductivity of diamond. The reason that we ought to treat **K** as non-vacuously true is because it is being used in an exploratory process to illuminate and illustrate a dependence relation. We do not merely want to assess the truth of **K**'s antecedent, we want to work out what would be the case if it *were* true. This is precisely why **K*** is false, because, given the dependence relation between covalent bonding and electrical conductivity, its consequent would not be the case were diamond not covalently bonded[54]. In this kind of case, we are not performing an investigative procedure, we are performing an exploratory one and **K*** is a failure of that process. We are exploring what would be the case were the impossible antecedent to hold. We are not merely trying to assess the truth value of the antecedent because we know it to be false.

The vacuist may point out that orthodox philosophical practice leads us to conclude that **K** and **K*** are vacuously true. But there is nothing to stop the scientist from pointing out that *scientific* practice leads us to conclude **K** is true, and **K*** false. In short, the scientist need not be persuaded by what the vacuist has to say. If we are to base our judgements on the views of either, it seems we should base them on the views of the scientists regarding these scientific matters, rather than what the philosopher thinks about the truth/falsity of these statements.

5.4.2.2 Superseded scientific theories

Tan also thinks that we need to make use of non-vacuously true counterpossibles when reasoning about superseded scientific theories. Sometimes, we need to reason about scientific theories using counterfactuals; “If Jupiter were a point mass then…” and “If classical

---

[54]See 5.7.3 for a more detailed discussion of the exploratory process.

mechanics had been true…" are examples of each of these (Tan, 2019, 48). As Tan points out, we might counterfactually reason about a false theory to describe its empirical content, e.g. "Had the geocentric Ptolemaic system been correct, celestial spheres would be unobservable entities". Counterfactual reasoning is also used in order to explain the falseness of a false scientific theory. Tan considers a straightforward example of this concerning Bohr's theory of the atom:

$\mathbf{B_1}$ - If Bohr's theory of the atom had been true, then an electron's angular momentum, *L*, in the ground state would have been observed at $L=h$ (the reduced Planck constant).
$\mathbf{B_2}$ - It is not the case that the electron's angular momentum, *L*, in the ground state is observed at $L=h$.
$\mathbf{B_3}$ - Therefore, Bohr's theory of the atom is false.
Tan, 2019, 48

Bohr's theory of the atom requires that the angular momentum of an electron is observed in the above way, that is to say that $\mathbf{B_1}$ is correct. As this is a result of the theory, if the theory were correct, it would be the case. Repeated experimentation and observation has shown that the angular momentum of an electron is actually zero in the ground state, i.e. $\mathbf{B_2}$ is true. Given that both $\mathbf{B_1}$ and $\mathbf{B_2}$ are true, it then simply follows that $\mathbf{B_3}$ is true. This is a substantial result, and Tan argues that $\mathbf{B_1}$ is true more than merely trivially. As he puts it: "in order for this commonplace pattern of reasoning to be epistemically fruitful, theory-evaluating conditionals must describe genuine relations of counterfactual dependence and implication. They must, in other words, be non-vacuously true." (Tan, 2019, 49).

This seems to be correct, the above essentially takes the form of "if *that* were right, we would see *this*. We don't see *this*, so *that* must be wrong". We want such arguments to produce truth that is not merely trivial, because the process Tan talks about seems like an example of good scientific reasoning. There are many examples of this process being used in the sciences for all manner of theories. As a method of theory falsification, it is a good one, and we need it to produce substantive, non-trivial results. One may be willing to accept this but unwilling to extend it to the counterpossible case, because of a commitment to vacuous counterpossibles. However, $\mathbf{B_1}$ is already a counterpossible. This archetype of non-vacuous scientific reasoning turns out to involve counterpossible reasoning. If one wishes to trivialise all counterpossibles then one is going to have to trivialise a lot of scientific reasoning, and this seems an unattractive feature of any account. The reason that **B_1** is a counterpossible is that Bohr's theory of the atom is an inconsistent theory. It rests on both classical and quantum assumptions, therefore some aspects of the theory represent orbiting electrons as radiating energy as they move around; other aspects of the theory represent electrons as non-radiative (Tan, 2019, 49). In other words, the theory as a whole contains a contradiction, as it represents electrons both as radiating energy and as not radiating energy. **B_1** does not merely refer to one aspect of Bohr's theory, it refers to the theory as a whole, and the theory as a whole contains this contradiction. So it is simply logically impossible that Bohr's theory of the atom be true, it is impossible that Bohr atoms could exist. **B_1** then, is a counterpossible. But we have already established that **B_1** is non-vacuously true. A potential response from vacuists could be that we can maintain vacuism because we accept that **B_1** is true (vacuously) and also accept that **B_1*** is true:

B_1* - If Bohr's theory of the atom had been true, then an electron's angular momentum, *L*, in the ground state would *NOT* have been observed at *L=h* (the reduced Planck constant).

B_1* negates the consequent of **B_1**, but as it is a counterpossible, is also true (vacuously so). The vacuist might respond that the reason we appeal to **B_1** rather than **B_1*** is because the former has proved useful for scientific progress and prediction due to the way the world happens to be, whilst the latter has not.

Much turns here on what we are aiming to do with the counterpossible, whether we are trying to perform an investigative or exploratory process. Let us imagine that we are using **B_1** in order to assess the truth of Bohr's theory, i.e. an investigative process. In that case, I may well agree with the vacuist, **B_1** and **B_1*** might both be vacuously true (contra Tan), and we need to assume that to use the line of reasoning above and show that Bohr's theory is necessarily false. Alternatively, we might, for whatever reason, be using **B_1** in an exploratory process, to see what the world would be like if Bohr had been right[55]. If this is the case, then **B_1** and **B_1*** might turn out to be non-vacuous. This is why it is crucial to acknowledge the difference in these processes, because it does seem to matter for the truth status of a counterpossible how it is being used[56].

---

[55]Again, see 5.7.3 for a discussion of this process.

[56]In 5.6 we will discuss Williamson's arguments for vacuism from the usage of counterpossibles in reductio arguments. If we are trying to use them in a reductio argument, we might need to assume that all involved

5.4.2.3 Idealised models

Another place that Tan (2019) alleges science makes use of counterpossibles is in reasoning with idealised models. Science often treats planets as points for the purposes of performing calculations on their gravitational effect. Sometimes, planes are often treated as if they are frictionless and liquids as if they are continuous. The use of such idealised modelling is prevalent throughout science, and again arguably essential. For example, the sheer complexity of modelling liquids as a series of discrete but bonded particles makes performing such calculations so difficult as to be unproductive, if not downright impossible. Scientists do tend to model things *as if they were* these idealised things. Tan (2019) alleges that these idealised things could not exist and could not fill the role of the substance being tested/investigated. For example, a continuous incompressible liquid *could not do* the things that water does, it could not *be* water. Yet we model water as if it were such an idealisation. Tan's claim is that we are modelling an impossible situation. Furthermore, reasonings based on such impossibilities constitute counterpossibles e.g. "had water been a continuous incompressible medium..." (2019, 46). Such modelling is useful because the behaviour of water as it actually is closely approximates that of a continuous incompressible medium. The antecedent of this counterfactual model, i.e. "had water been a continuous, incompressible medium…" is metaphysically impossible. This makes the statement, as a whole, a counterpossible.

One could argue that this statement is a counterpossible in similar ways to the diamond case. It is held that, necessarily, water is identical to H_2O. As such, water has to be built up out of H_2O, and nothing that is made of anything else can be water. H_2O is not a continuous, incompressible medium, it is a series of bonded but discrete particles. So if something was such a strange medium, it would not be H_2O (and so not water). It would be metaphysically impossible for water to be a continuous, incompressible medium. But people may not be convinced because they wish to define water by its theoretical role, rather than its chemical composition. Tan (2019, 46) thinks that even this view would lead to the statement in question being a counterpossible. One might allege that perhaps some continuous, incompressible medium can fulfil the role of water by acting exactly as actual water does.

---

counterpossibles are true, and this will be the case if we are performing an investigative process on Bohr's theory of the atom.

The problem is that this cannot be the case, a continuous, incompressible medium *cannot* fulfil the role of water. For example, a key property of water is that it is a solvent for particulate solids. No continuous, incompressible medium could ever act as a solvent for particulate solids, so no continuous, incompressible medium could ever fulfil the causal role of water (Tan, 2019, 46). However we define water, it is metaphysically impossible that it be a continuous, incompressible medium. Yet we model it as such, so such models employ counterpossibles.

One may be willing to accept this but deny that this counterpossible is non-vacuously true. Tan's answer to this is to point to scientific practice and how things are actually done (and indeed how they have to be done). He alleges that such practices require us to treat these counterpossibles as non-trivially true. Tan uses the example of two competing models about the behaviour of water, $\mathbf{M_1}$ and $\mathbf{M_2}$. They both represent water as an idealised continuous fluid but they differ with respect to the viscosity they ascribe to water (2019, 47). To test these models, scientists will see how close the behaviour of water is to each model. Let us imagine they discover the predictions of one theory, $\mathbf{M_1}$ to be very close to the behaviour of water, whilst the predictions of $\mathbf{M_2}$ are further off. Scientists would judge $\mathbf{M_1}$ to be a true (or approximately true) theory, whilst $\mathbf{M_2}$ would be false. Furthermore, they would take the following counterpossible to be false:

$\mathbf{L_1}$ - "If water were a continuous, incompressible medium, then it would behave as $\mathbf{M_2}$ predicts"

Whilst taking this one to be true:

$\mathbf{L_2}$ **-** "If water were a continuous, incompressible medium, then it would behave as $\mathbf{M_1}$ predicts"
(Tan, 2019, 47).

As we have already established, both are counterpossibles, and yet they seem non-vacuous. $\mathbf{L_1}$ would be false because $\mathbf{M_2}$ fails to model the behaviour of continuous, incompressible mediums. So, were water such a medium, it would not behave in the way that $\mathbf{M_2}$ predicts. Orthodoxy might dictate that both of these are vacuous, but this does not constitute an argument for that being the case. Furthermore, the fact that it seems a worthwhile endeavour to reason using such counterpossibles is in fact evidence *against* the orthodoxy. If scientists were unable to reason so, then a large swath of scientific practice would disappear. Scientists need to use models like this and do so fruitfully, this would not be possible from vacuous counterpossibles, so we need to hold them to be non-vacuous.

Tan argues that these counterpossibles are non-vacuous. But interestingly, it is not clear if these statements are counterpossible in the first place, or which process they are being used in. One might think that scientific reasoning from models takes the form of "continuous substances behave like *x*; water is importantly similar to a continuous substance, so we should expect water to behave in an *x*-like way" which does not seem to be counterpossible at all. There is something to such an objection. But what this does is highlight the importance of delineating the investigative from the exploratory process. Because it is not clear what is going on in this case. One can see a case for it being investigative/exploratory. For example, we might be trying to work out if water is indeed a continuous, incompressible medium. It seems that we might be doing this because it seems that models are aimed at determining the properties of their target object, in this case water. If this is the case, maybe $\mathbf{L_1}$ and $\mathbf{L_2}$ are vacuous. However, we might instead be doing an exploratory process. We might be trying to work out how continuous, incompressible mediums work (because it turns out to be useful for studying the behaviour of water). In this case, it will turn out that $\mathbf{L_1}$ and $\mathbf{L_2}$ are non-vacuous. It is not clear which is correct in this instance, but again this highlights the importance of distinguishing between these processes and acknowledging their differences.

Scientific practice seems to require us to treat counterpossibles non-vacuously, and this is important. The vacuist may have to say that scientists are simply mistaken, but this is an unattractive position. Nor is it a position that scientists are likely to accept. If our semantics conflict with successful scientific practice then this might indicate a flaw in the semantics rather than the scientific practice. Given this, non-vacuism may seem preferable. It will be helpful to consider one line of response the vacuist might make which, I argue, would fail. A vacuist could respond that scientific practice indeed requires us to *treat* some counterpossibles as non-vacuous, but that this is not because such counterpossibles *are* non-vacuous. Instead, perhaps what matters is that some scientific counterpossibles are *assertable* and some not, these are the ones we treat as non-vacuous.

5.5 Assertability

A vacuist might say that the counterpossibles I want to describe as false are merely unassertable (as discussed by Grice, 1975) while the ones I want to describe as true are assertable. This could be the case whilst all of them are true, and so I have not shown the vacuist thesis to be false, I have merely shown that some counterpossibles are unassertable. Perhaps, the class of 'true' counterpossibles are assertable because they point to some underlying non-counterpossible truth, with the 'false' counterpossibles failing to do this. For example, take the following pair of counterpossibles:

1a - If Obama had had different parents, he would have had different DNA.
1b - If Obama had had different parents, he would have been two inches tall.
(Emery & Hill, 2017, 136).

1a is assertable because it points to the following underlying fact that Obama's parents were the cause of him having the DNA that he has (Emery & Hill, 2017, 138), whereas **1b** does not. Because they fail to do this, such counterpossibles are unassertable, and we mistake this intuition and say that they are false (Emery & Hill, 2017, 137-138). However, as orthodoxy shows us, this is mistaken because all counterpossibles are true. The assertability of a statement, ***s***, such as **1a** and the unassertability of its converse ***s****, such as **1b**, does not imply that ***s*** is true and ***s**** is false. One could then account for the views of non-vacuists whilst maintaining vacuism.

This is an interesting point, but it need not threaten my view. Firstly, if it is the case that, for a given conflicting pair of counterfactuals, the assertability of one and the unassertability of the other does not imply that one is false, then it is also the case that it does not imply that they are both true. As Sendlak (2021) argues, the same pattern of assertability can be found in non-counterpossible counterfactuals. Perhaps this pattern fails to imply that one is false, but it also fails to imply that both are true. If it did, we could be vacuists about counterfactuals. Given that vacuism about counterfactuals is false, we should be suspicious of this supportive move for vacuism about counterpossibles. Take the following example:

"the assertion of 'If Christopher Columbus had reached the place he was planning to reach in 1492, he would have arrived in India' can be explained by the fact that this allows one to indirectly express a more substantial

proposition that is related to the asserted proposition in subject matter, e.g., 'Christopher Columbus was planning to reach India.'"
Sendlak (2021, 11)

The converse "If Christopher Columbus had reached the place he was planning to reach in 1492, he would not have arrived in India", should intuitively be false. However, under the Emery & Hill analysis, the truth value of the first sentence should not affect the truth value of the second. We think it is false but crucially we can explain that the reason we intuitively think it is false is due to its unassertability and maintain that it is true regardless. Sendlak claims that if this is indicative of vacuism in the counterpossibles case, it is indicative of vacuism in the counterfactuals case, and this is problematic. Given the intuitive falsity of vacuism about counterfactuals, this represents a problem for a vacuist account of counterpossibles that would endorse this (Sendlak, 2021, 11). Whilst it is true that a statement can fail to be assertable whilst failing to be false, it does not mean that each and every statement which fails to be assertable also fails to be false. Emery & Hill (2017) try to introduce a gap between the unassertability of a statement and its falsity. But this creates a total disconnect between assertability/unassertability and the truth of a statement. In doing this, they miss the target they aim for.

Vacuists may wish to say that the counterpossibles appealed to in exploratory scientific processes do not *need* to be non-vacuous, instead we should appeal to the distinction between assertability and unassertability. But the kind of arguments used for this could also be used to show that ordinary counterfactuals are vacuous. This is clearly false, so something must be faulty with the argumentation. Assertability/unassertability will not help the vacuist. At this stage, the most we can have shown is that at least *some* counterpossibles are non-vacuous, the ones used in exploratory processes which plausibly captures a large portion of scientific practice. This is a significant result, if scientists need to treat counterpossibles as non-vacuous, then our accounts of counterpossibles need to treat them as non-vacuous. This of course does not show that *all* counterpossibles are non-vacuous. As we noted in 5.1, non-vacuous countermathematicals intuitively seem to be a difficult case. Given that we need all counterpossibles in proofs by reductio to be true, the vacuist seems to be in a strong position. I argue that we can and should extend the spirit of why the exploratory scientific counterpossibles are non-vacuous to the case of countermathematicals and show that there are also non-vacuous examples. First, I will discuss Williamson's (2018) discussion of why countermathematicals should be vacuous, as it highlights some important points.

5.6 Williamson against countermathematicals

Williamson (2018) discusses the use of counterpossibles in mathematical proofs by reductio. As a hallmark example of this, he uses the proof that there is no largest prime number, known as Euclid's theorem. Williamson stresses that one does not necessarily need to phrase mathematical proofs in terms of counterfactual conditionals, but that it is a legitimate and natural way of doing so. So regardless of particular views on counterpossibles, all parties need an explanation of why this reasoning is legitimate and works. Williamson borrows the example from Lewis (1973, 25):

M - if there were a largest prime, *p*, *p!* + 1 would be prime
N - if there were a largest prime, *p*, *p!* + 1 would be composite

Williamson (2018, 363) helpfully summarises this proof: of **M** he explains that it holds because "if *p* were the largest prime, *p!* would be divisible by all primes (since it is divisible by all natural numbers from 1 to *p*), so *p! + 1* would be divisible by none" (2018, 363). Of **N** he points out that it holds because "*p!* + 1 is larger than *p*, and so would be composite if *p* were the largest prime" (Williamson, 2018, 363). Given that both these conditionals have the same antecedent, we are entitled to conjoin their consequents, resulting in:

O - If there were a largest prime *p, p!* + 1 would be both prime and composite.

Given that the consequent of this counterfactual is a contradiction, we can deny the antecedent, and conclude that in fact there is no largest prime. Quite obviously these are counterpossibles as well, because there *cannot be* a largest prime, that is a mathematical impossibility. Williamson and other vacuists, along with non-vacuists, will accept this as a good mathematical proof. In other words, everyone should accept all of **M-O** as true. Williamson's strategy is then to offer another proof by contradiction, using vacuous counterpossibles, which he says vacuists can accept easily, but that non-vacuists cannot accept, and cannot reject without rejecting Euclid's theorem. If non-vacuists deny the truth of the premises in Williamson's proof, he alleges they must also deny the truth of the premises in Euclid's theorem. Since rejecting such a proof would be unacceptable, we have a strong reason to doubt non-vacuism[57]; so Williamson's argument goes. Before explaining why this argument fails, I will spell out Williamson's second proof.

Williamson asks us to consider someone who answered '11' to 'What is 5 + 7?' but who mistakenly believes that they answered '13', and utters the following counterpossibles, for the non-vacuist, **P** is false, whilst **Q** is true:

P - If 5 + 7 were 13, I would have got that sum right
Q - If 5 + 7 were 13, I would have got that sum wrong
(Williamson 2007, 172)

Williamson is not persuaded by the initial intuitiveness of such examples:

> "... they tend to fall apart when thought through. For example, if 5 + 7 were 13 then 5 + 6 would be 12, and so (by another eleven steps) 0 would be 1, so if the number of right answers I gave were 0, the number of right answers I gave would be 1." (2007, 172)

If the number of right answers the person gives is 0, i.e. they give a wrong answer, then the number of right answers they give is 1, i.e. they get the sum right. So both counterpossibles turn out to be true. Williamson then asserts that this is a result that the vacuist achieves, but that the non-vacuist cannot. He claims this points in favour of vacuism about counterpossibles. However, there is room for debate here. In particular, Brogaard & Salerno (2013) develop a series of objections against Williamson's reasoning.

5.6.1 Brogaard & Salerno

Brogaard & Salerno (2013) analyse Williamson's argument in depth and draw out the extra steps Williamson alludes to. The conclusion Williamson draws is that "if the number of right answers I gave were 0, the number of right answers I gave would be 1", hence, both **P** and **Q**

[57]Although see 5.7.3 for an unproblematic case in which we might wish to say that Euclid's theorem was false on non-vacuism.

are true. The steps that Williamson abbreviates will be something akin to, if not exactly the following:

(**i**) If 5 + 7 were 13, then 5 + 6 would be 12

(**ii**) If 5 + 7 were 13, then 5 + 5 would be 11

...

(**xi**) If 5 + 7 were 13, then 5 + - 4 would be 2.

(**xii**) If 5 + 7 were 13, then 5 + - 5 would be 1.

(Brogaard & Salerno, 2013, 649)

One reading of Williamson's argument is that he is saying that worlds in which 5+-5=1 are also worlds in which 0=1, because we can substitute 5+-5 for 0. So we can conclude that:

(**xiii**) If 5+7 were 13, then 0 would be 1.

Hence, we get to Williamson's (2007, 172) conclusion that "if the number of right answers I gave were 0, the number of right answers I gave would be 1", with **P** and **Q** both being true. Brogaard & Salerno object that we can reject Williamson's proof because he does not do a good enough job in establishing that the closest impossible world in which 5+7=13 is also one in which 5+6=12 (2013, 650). At this stage, we can return to Williamson's (2018) argument against non-vacuism.

Williamson's allegation is that *if* non-vacuists reject his proof on the grounds that it does not establish that the described world is the closest impossible world, then they must reject Euclid's theorem for the same reason. Mathematicians will not concern themselves with the relative closeness of impossible worlds when producing proofs by contradiction, they just produce the proof. So there is no evidence that the closest impossible world in which there is a largest prime, *p*, is also a world in which *p!*+1 is both prime and composite (Williamson, 2018, 363). Non-vacuists will then be compelled to either reject Euclid's theorem, or to find a way of showing that the closest impossible world in the prime number case is indeed the world that Euclid's theorem describes. However, there is no guarantee that the same process cannot be performed for Williamson's proof, which would seem to tell against the non-vacuist. Therefore, we should be viewing both counterpossibles in both cases as true, this is exactly as the vacuist describes, but contra the non-vacuist (Williamson, 2018, 363-364).

Having seen Williamson's argument we are in a position to respond to it. We will do this by making some clarifications about vacuism and what Williamson has established so far, and also by building upon Brogaard & Salerno's objection, because whilst it might not work in its current form, there is an important idea contained within it.

Williamson claims that the counterpossibles used in Euclid's theorem and in his own proof are all true, because they follow from mathematical reasoning. Williamson claims the vacuist can obviously account for this, but the non-vacuist cannot. Perhaps the non-vacuist intuition that, for example, **P** and **Q** have different truth values stems from some commitment that for *any* pair of counterpossibles that have contradictory consequents, but the same antecedent, at least one must be false. Williamson will claim that this failure to deliver the verdict of mathematics is a significant drawback of the non-vacuist account, and so we should reject such an account. I assert that non-vacuists can respond to this however. Not only has Williamson failed to successfully establish vacuism, these mathematical proofs fail to even constitute an argument for it.

5.7 Responses to Williamson

5.7.1 Redundancy of vacuism

One could say that non-vacuists do not *have* to reject Williamson's proof. Certainly, Brogaard & Salerno did so under the banner of non-vacuism, but there is nothing inherent in non-vacuism that says one cannot accept Williamson's proof. Perhaps Williamson has shown that all *those* counterpossibles are true, but that does not mean he has shown that vacuism is true, or even that all counterpossibles are true. Vacuism is the thesis that all counterpossibles are vacuously true, because their antecedents are necessarily false[58]. For the vacuist, the necessary falsity of the antecedents is what makes the counterpossibles true. The problem is that vacuism plays no role in making **M**, **N** or **O** true in Euclid's theorem. As Williamson himself says, they are true because they are mathematical results; "(31)–(33) [**M-O**] should be true, for they are soundly based on valid mathematical reasoning" (2018, 363). But this is independent of vacuism. Williamson correctly points out that a semantic theory needs to produce this result, and indeed vacuism does, but for one it is unclear that it does so for the

[58]Non-vacuism of course being the thesis that there are at least some non-vacuous counterpossibles.

correct reasons, and two, it is not the only semantic theory that does this. The truth of **M-O** is a mathematical result, they are true for reasons stronger than the mere impossibility of the antecedent. Compare this with **M***:

"If there were a largest prime *p*, *p!*+1 would be a set"

Or **M****:

"If there were a largest prime *p*, *p!*+1 would be an infinite set"

I think mathematicians would want to reject **M*** and **M**** as false, as would non-vacuists. They would be false because they would be based on faulty mathematical reasoning. However, on vacuism, they would come out true. Consider a world in which mathematical practice was systematically wrong[59]. For whatever reason, mathematicians just get the wrong verdict when talking about these matters. In such a world, clearly some counterpossibles would be described as false by the mathematicians, e.g. they might misunderstand factorials and so say that **M** above were false. However, given that we would still be dealing with countermathematicals, the vacuist would describe all such statements as true. In this way, the vacuist result separates itself from the verdict of mathematical practice. I believe this in itself constitutes a criticism of vacuism because we ought to want our semantic verdicts to be at least somewhat reflective of the relevant practice, but vacuism has a complete disconnect from mathematical practice. As we discussed earlier, it is a mark against vacuism that it conflicts with scientific practice. We should also take conflict with mathematical practice as a mark against the theory.

In 5.4.2 we discussed cases in science that seem to require non-vacuous counterpossibles, so it seems vacuism about all counterpossibles might be false. But restricting vacuism to countermathematicals is a redundant thesis, this amounts to a claim that all the mathematically proven statements are true. Or, if mathematical practice told us that a particular countermathematical was false, it would amount to disagreement with mathematical practice. This second alternative is exactly what the vacuist charges the

[59]This is a possible world, so I am merely appealing to a counterfactual here.

non-vacuist with as a significant problem, and yet it seems they are vulnerable to exactly the same criticism.

But the point against the vacuists is not merely that it produces the wrong results in certain cases, that would merely be a reframing of the intuitive arguments for non-vacuity. Instead, the point is that in the mathematical cases appealed to, although it gets the right result, the result is obtained regardless of vacuism. We can see this by the fact that non-vacuists accept the result that both statements are true in the case of Euclid’s theorem, and they do so on non-vacuist grounds; because it is a mathematical result. Williamson’s mistake comes from the fact that he assumes that, to take the counterpossibles in his proof as being true, the non-vacuist would have to subscribe to some form of vacuism; but this is not the case. One can take **M-O** to be true without being a vacuist[60], and this is because, as Williamson points out, they follow from mathematical reasoning. Our intuitions led us to think that **P** and **Q** had different truth values, but mathematical reasoning showed us this was wrong. That is something the non-vacuist can accept, just because non-vacuism is committed to some counterpossibles being non-vacuous, it does not mean that on each occasion that our intuition points to counterpossibles having different truth values, we are right. Importantly again, even if all the countermathematicals we happen to have considered in this chapter are true, this is because of mathematical practice, this is a separate result from vacuism.

We have seen how Williamson’s proof works and how the non-vacuist can equally accept this result. Williamson’s proof does seem to fall out of standard mathematical definitions of addition, the successor principle, etc. But another point to be considered is whether or not Williamson has genuinely evaluated the truth value of the counterpossible in the way it should be, this is my ultimate argument against vacuism and will be developed in 5.7.3. Baron, Colyvan & Ripley (2017) presents a good expansion on the Brogaard & Salerno (2013) objection which builds towards my objection to vacuism. As such, it will be worth considering their strategy first.

[60]Indeed, it seems that one can understand and reach this result, without *any* view on the vacuity/non-vacuity of counterpossibles.

5.7.2 Baron, Coylvan, Ripley

Baron, Colyvan & Ripley (2017) provide an alternative strategy for responding to Williamson than that offered by Brogaard & Salerno (2013) above. They do this by addressing a slightly different variant of Williamson's proof and attempting to diagnose what they think his mistake is. Their allegation is that Williamson has held too much fixed in the counterfactual case he describes.

"What would happen if 5 + 7 were 13? Well, since 5 + 7 is 12, 12 would be 13. Subtracting 12 from both sides, 0 would be 1. But then all numbers would be identical: for any m, n, m × 0 = n × 0, and since 0 would be 1 under the supposition in question, we have it that m = n. And if all numbers were identical, any number of horrible things would follow."
(Baron, Colyvan & Ripley, 2017, 3).

Williamson goes wrong by keeping too much of how mathematics actually is the same. To highlight this, Baron et al. draw up a parallel case with the same reasoning as Williamon's example:

"What would have happened if John had worn a blue jacket, instead of the green one he actually wore for the photo shoot? Well, since he wore a green jacket, the blue jacket would be a green jacket, and this would mean that green is blue."
(Baron, Colyvan & Ripley, 2017, 3).

This is clearly an unconvincing argument because too much has been held fixed. In the jacket case, we clearly should be changing facts about what jacket John did or did not wear, in order to assess the counterfactual properly. In the mathematics case, we should likewise change some mathematical facts, we should not assume that everything is still as it is at the actual world. Primarily because in assuming something to have been different, that is at least one thing that is the case at the described world that is not the case at the actual world. If we assume everything to be the same, we are of course going to run into contradictions and strange results. This is the heart of my objection to Williamson. Williamson has failed to appreciate the distinction between genuinely conceiving of a distinct world, and considering a conjecture at the actual world, i.e. he has not differentiated the exploratory from the investigative process.

5.7.3 Genuine consideration of a world

Brogaard & Salerno (2013) along with Baron, Colyvan & Ripley (2017) have captured something with their objections. They charge Williamson with not correctly conceiving of the closest world in some way. Williamson says that rejecting his proof on these grounds would mean we also have to reject *any* mathematical proof by contradiction, such as Euclid's theorem. Because this is unattractive, we should not reject his account. This objection has targeted something important, albeit in the wrong way. Williamson's proof does not work by describing the closest world (in which the conjecture is true) to the actual world, but this is because his proof does not consider a distinct world at all. What Williamson has done is show that given what we already know, the actual world cannot be a particular way. This is a point worth spelling out in some detail.

Recall the difference between investigative and exploratory counterfactual processes from 5.1. In investigative processes, the truth value of a statement/hypothesis might be unknown, and so we want to find out/demonstrate whether it is true or false. To do this, we use the hypothesis to derive a prediction and make a counterfactual using the hypothesis as the antecedent and the prediction as the consequent. If the prediction turns out not to be the case, we can use this to show that the antecedent was false. This is what we are doing in the example of Bohr's theory and in Williamson's proof. We say if one thing were the case, a second thing would also be the case, as the second is not the case, we can say that neither is the first. If 5+6 were 13, then 0 would be 1, 0 is not 1, so 5+6 is not 13. In both these cases, the antecedents turn out to be necessarily false, and so the counterfactuals involving them are actually counterpossibles. The vacuist says that as counterpossibles are trivially true, these particular ones are trivially true. However, these particular counterpossibles are useful. The counterpossibles that non-vacuists wish to call true, (**B_1** and **xiii**) contain consequents that contradict our experience, as such these are the ones which can actually be used to show the antecedent to be false. This is the process one might engage in to show that the antecedent of a counterfactual is false, and this is the process that Williamson is engaging in. Perhaps this means that (**B_1**) and **xiii** are vacuous. But if so, that is merely because of the kind of process they are being used in, and does not support the more global vacuist thesis that Williamson is describing. Regardless, there are situations where we already know the truth value of the antecedent, and these are the cases I want to focus on.

There may be cases when we know that a statement is false, perhaps even necessarily false, but we want, for whatever reason, to explore what *would* be the case if in fact it were true[61]. This is what we are doing when explaining why diamond does not conduct electricity (5.4.2.1), potentially what is happening in the case of modelling water as a continuous medium (5.4.2.3), and certainly what Brogaard & Salerno (5.6.1) are trying to do. In these cases, we know that the antecedent is false, we know that diamond is not otherwise bonded, and we know that 5+6 is not 13, but we want to find out what would be the case if they were true. In order to find out what would be the case if they were true, we have to imagine a scenario in which they are true. To do that, we need to sacrifice some assumptions to avoid contradictions, e.g. that diamond is covalently bonded and that 5+6 is *not* 13. Doing this would prevent us running into contradictions. The counterpossibles are non-vacuous because they would be false if we made an incorrect statement about what *would* be the case if the impossible antecedent were the case. It is worth pointing out that we already make the distinction between investigative and exploratory projects in the case of ordinary counterfactuals. Let us take the case of a crime scene investigation. In conjecturing how the victim was murdered, the detective will make hypotheses. Perhaps one of these hypotheses is that the victim was shot. The detective may then form a counterfactual of the form "if it were the case that the victim was shot, there would be a gunshot wound on the body". If no gunshot wound is found, the detective can conclude that the antecedent was false. In such a case, counterfactual reasoning has been used to discover that something is false, the detective embarked on an investigative process.

Alternatively, sometimes we know that a statement is false, but we want to work out what would be the case were it true. If you cycle to work and your tyre bursts, resulting in you being late to work, you can usefully say "if I had driven to work, I wouldn't have been late". We know the antecedent is false, but we assume it to be true, and reject assumptions such as you actually having ridden your bike in order to make non-vacuous statements. If we did not reject assumptions, we would simply run into contradictions and end up proving that you had in fact cycled to work, but this is not what we wanted to do. What we need to do in such cases is embark on an exploratory process. Given that the investigative/exploratory distinction is present in the case of ordinary counterfactuals, it is not clear why it should not

[61]A similar idea to what follows occurs in Sendlak (2021, 16-18). However, this idea presents an important critique of vacuism concerning counterpossibles and so I think it warrants more attention and exploration than it has been given there and elsewhere in the literature.

be present in the case of counterpossibles. With this distinction more clearly in mind, we can assess Williamson's account of counterpossibles. We can now diagnose some of the dialectic between non-vacuists and vacuists. Put simply, Williamson is engaged in an investigative process, whilst Brogaard & Salerno are engaged in an exploratory process.

Williamson's proof is simply a proof that $5+7\neq13$. That is a perfectly legitimate thing to do and might be useful in some circumstances. But the reason that proof works, is the same reason the Euclid proof works. It works because we hold fixed everything we know about the world (in this case mathematics), and then show that *given that,* a particular fact could not be the case. In Euclid's proof, we hold fixed facts about prime numbers, where in the number sequence they tend to appear for example. We then want to show that the assumption that there is a largest prime number is inconsistent with this. In doing this, we have not considered a different world, we have not moved from our world; because we are showing that something cannot be the case, *at our world*. In Williamson's proof, he has held the mathematical facts fixed and then shown that *given these things*, $5+7\neq13$. But note, this is not to consider a world in which 5+7 *is* 13. Because if we are considering a world in which $5+7=13$, this cannot be a world in which it is also the case that $5+7\neq13$. Williamson has not considered a different world, he has considered the actual world and shown that a certain statement is false here. All the countermathematicals Williamson invokes might be true, but that is only the case if he is involved in an investigative process. Indeed, it is not clear that he is genuinely considering a counterpossible at all[62]. Moreover, we can show that Williamson is engaged in an investigative rather than exploratory process by considering his notion of a suppositional procedure (Williamson, 2020).

Williamson (2020, 18) describes the Suppositional Procedure (SP). In order to assess the truth of any conditional $\varphi\rightarrow\psi$, one has to suppose that φ and then judge whether, on the basis of that, it is also the case that ψ. Importantly, this simple form of the SP makes no mention of the possibility of φ or ψ, simply that one must suppose φ. One intuitive claim about supposing is that we have to suspend our disbelief in some way, perhaps just as in the case of make-believe games. Leng (2010) talks about make-believe in mathematics and describes the process as representing real objects in some way. Specifically it is to

[62]See 5.8.3 and 7.4.3.

"...imagine of real objects that they are other than they really are. It is clear in these cases that we are sometimes being required to imagine something *false* concerning the nature of such objects: we know that the tree stumps aren't *really* bears; that the fluids aren't *really* continuous." (Leng, 2010, 159 [author's emphasis]).

If we know **φ** to be false, but want to suppose it for some purpose, we have to reject other facts which would rule **φ** out. My allegation should be clear now, in Williamson's proof he has failed to suppose[63] that 5+7=13. Let us consider a more in depth spelling out of a true suppositional procedure.

Take any proposition, **R**, if one is to consider a world at which it is the case that **R**, then the world considered must also be a world in which ¬¬**R**. This is not to say that there cannot be worlds which contain contradictions. If we are considering a world at which it is raining and not raining, it seems that we are considering a world in which **R** and ¬**R** (neglecting to include ¬¬**R**). But this is not quite right. We are considering a world in which it is the case that it is raining and it is not raining. This is a proposition, **S**. If we need to consider that world, then we also need to be sure that it is a world at which ¬¬[it is raining and it is not raining], i.e. that it is also a world at which ¬¬**S**[64]. Williamson fails to consider a world at which 5+7=13, because he does not ensure that it is also a world at which it is the case that ¬¬[5+7=13]. Holding fixed the background mathematical facts is key to the proof working, because the proof aims to show that, at the actual world, something is not the case. Once again, that is not something peculiar to counterpossibles. This is exactly the process we need to engage in for ordinary counterfactual scenarios.

Let us take the ordinary counterfactual "If Julius Caesar were alive today then…". We have a number of assumptions that we are committed to at this world. For example, the average lifespan of a human being currently sits at around 81 in the UK. Given this, we probably assume that anyone who was alive at the time of Julius Caesar is now dead, including Caesar himself, i.e. we assume that ¬[Julius Caesar is alive today]. In order to embark on an exploratory process to imagine what might be the case if the antecedent were true, we need to genuinely consider a world at which it is the case that Caesar is alive today. We need to reject the implicit assumption for these purposes. We need to explicitly make sure it is a world at

[63]More support for my claim can be found in section 5.8.3.
[64]I make this concession to contradictory worlds because I do not want to rule them out by definition. In any case, regardless of whether there are worlds containing a contradiction, I want to focus on the impossible but internally consistent worlds. Such worlds might contain a metaphysical impossibility, but not be deductively closed so they are distinct from the explosion world.

which ¬¬[Julius Caesar is alive today]. If we do not do this, then we will of course run into inconsistencies and prove that our conjecture (that Caesar is alive) is incorrect. But this is not an exploratory process and it is not genuinely conceiving of a distinct world, this would be an investigative process. If we were to consider the closest world in which 5+7=13, then we would have to jettison some mathematical assumptions. In doing so, it is not clear that all of Williamson's statements would follow mathematically, and so be true. In fact, it actually seems more likely that Williamson's proof will fail, some statements will come out false. In just the same way, if we genuinely considered a world at which there was a largest prime, likely Euclid's theorem would not work. But this should not be surprising, a world with a largest prime is a world where Euclid's theorem is false[65]. This does not threaten mathematical practice, because this is not the aim of mathematical practice.

There are of course limits to how far this process can go, both in terms of unavoidable contradictions and in terms of the considered scenario being so distant from our own as to be irrelevant. But such things can be assessed on a case by case basis, Baron et al. (2017) propose a method in this style for "chasing out" contradictions from the immediately relevant vicinity of the counterpossible scenarios, in some cases the relevant vicinity will be much larger than in others, but the process is the same[66]. One may be concerned that such a process will in fact have no end and, by dealing with metaphysical necessity, there will always be contradictions in the counterpossible scenario we imagine. Alternatively, the concern may be that the process takes so long that in rejecting background assumptions we end up with a completely different arithmetic system in which everything works so differently that we cannot retrieve any useful conclusions from consideration of the scenario. Baron et al (2017, 8) address such concerns by pointing out that a similar process occurs in the consideration of ordinary counterfactuals.

[65]Berto et al (2018, 704). discuss a similar point related to Euclid's theorem. In the context of a reductio proof we should hold everything fixed, but in other contexts it might make sense to jettison some assumptions and in such cases not all statements would mathematically follow (i.e. some counterpossibles would be false).
[66]One concern I have with the specific way Baron et al. (2017) go about this process is that they might in fact no longer be considering counterpossibles because they redefine what various mathematical operators mean, specifically addition. At that stage, rather than considering impossible ways for the specific mathematical system we have to be, they might simply be considering a different mathematical system, and this seems like a counterfactual. Compare this to a counterfactual like "Had the queen in chess not been able to move diagonally, then…". This does not seem to be a claim about the specific set of rules we have for chess currently, but rather about a different set of rules.

In ordinary counterfactuals, we may run into contradictions in considering the scenario, but we simply reject all and only those relevant for whatever our purposes may be. For example, in considering the case of whether Suzy's throwing of the rock caused the window to break, we may consider counterfactuals beginning with "If Suzy hadn't thrown the rock…". In such cases, there are of course inconsistencies, in the scenario we are considering it may be the case that Suzy indeed moved to throw the rock but that the rock did not move for some unspecified reason. Or it could even be that Suzy made the decision to move her arm but that it simply did not happen (Baron et al., 2017, 8). Baron et al (2017) explicitly point out that it simply is not the case that we go back through the entirety of history to make this scenario consistent. In fact, they contend, we tend to ignore the inconsistencies and just conceptualise Suzy failing to throw the rock, without necessarily filling in the background details as to how that failure was realised. It seems that the same process should take place with countermathematicals. Baron et al. (2017, 9) think that when we dispense with the immediate contradictions in the mathematical case we can simply ignore the rest, even if actually addressing all the contradictions would be an infinite process. Of course, it might be the case that addressing the immediate contradictions in an ordinary counterfactual case is a much simpler and smaller job than in a countermathematical case. But there is no reason to think that this is anything more than a difference in degree. If we perform this process then we can consider internally consistent (but impossible) scenarios and try to determine what would/would not be the case, were these scenarios to take place.

5.7.4 Countermathematicals in explanation

So far we have discussed why it is that we should judge scientific counterpossibles to be non-vacuous. We have also shown how there are different uses of counterpossibles depending on which sort of reasoning we are engaged in (investigative or exploratory). It is time to extend this to the mathematical case. We have already seen how countermathematicals play a role in investigative processes. When we aim to test a mathematical hypothesis we hold everything else fixed and see if we run into contradictions. If we do, then the antecedent is false. In such cases, it might turn out that all the countermathematicals involved are true. But importantly, they are not true because vacuism is correct, they are true because they follow from mathematical reasoning. But countermathematicals can also be used in an exploratory process to explain something in the world.

There are many examples of this, but a key one is the discussion by Lange (2017) about constraint explanations[67]. Although Lange does not immediately enter into the counterpossible debate with these points, they are relevant in a number of ways. One example of a constraint explanation is **Strawberries** as considered in 4.1.2, "The reason that Jane cannot divide her 23 strawberries equally between her 3 children, is *because* 23 is indivisible by 3". In countermathematical terms we can say "Had 23 been evenly divisible by 3, then Jane would have been able to divide her 23 strawberries evenly between her 3 children.". In this case, we are not engaged in an investigative process because we know the antecedent is false (impossible in fact). What we are trying to do is work out what would happen if it *were* true and ultimately highlight the dependence relationship (constraint) that is at play. We are engaged in an exploratory process. In scientific explanation, we might use counterpossibles to highlight the dependence relationship between covalent bonds and poor electrical conductivity. In mathematical explanations, particularly constraint cases, it is parallel. We are simply trying to highlight the constraint relation between the mathematical fact and facts in the world. To do this, we have to suppose the antecedent to be true and explore the consequences of this. In order to suppose it to be true, we simply cannot hold everything else fixed. When we start to jettison assumptions (for starters, the fact that 23 is prime), we will no longer run into a straightforward contradiction between the antecedent and consequent.

Yli-Vakkuri & Hawthorne remark when discussing mathematics that "...'$\vdash$' expresses provability in mathematics—by which we mean pure mathematics. $\mathbf{\Gamma} \vdash \boldsymbol{A}$ only if both $\boldsymbol{A}$ and all of the statements in $\mathbf{\Gamma}$ are pure mathematical statements." (2020, 560). When we are discussing counterpossibles which contain a non-mathematical consequent, the consequent will not follow mathematically from the antecedent. As such, the counterpossible as a whole may well turn out to be false. The mistake of the vacuist is in thinking that the investigative process is the only one that can be performed with counterpossibles, or that it is the most important one. *If* it is the case that all the countermathematicals used in investigative processes are true, it is not because of vacuism, it is because of mathematical practice and its results. In exploratory cases, it is simply not the case that they all turn out true, their truth value will vary from world to world, just as with counterfactuals, i.e. vacuism is false.

[67]This kind of explanation is parallel to the usage of counterpossibles to explain the poor conductivity of diamond described in 5.4.2.1. We might *know* that the antecedent is false, but we want to suppose it to be true to highlight some sort of dependence relation.

5.8 Problems

5.8.1 Do mathematicians use counterpossibles?

One general point to bring up is whether or not mathematics does indeed use counterfactuals, as opposed to merely appearing to use them through language choice but actually relying on something else[68]. Non-vacuists about countermathematicals clearly think that mathematics makes use of them. But it is important to point out that many prominent vacuists also think this. For example, as Yli-Vakkuri & Hawthorne say, "...we will argue, mathematics makes use of the counterfactual conditional…" and that this usage "...is by no means a marginal feature of mathematical discourse." (2020, 552). Indeed they themselves ultimately view it as indispensable. Perhaps the most vocal vacuist, Timothy Williamson, also concedes that we must account for the use of counterfactuals in mathematics as it is a legitimate practice (2018, 363). Reutlinger et al. (2020) began a more formal study of mathematical language and, from those preliminary results[69], it seems to be the case that mathematicians frequently use counterfactuals. One could maintain a commitment to this choice of language being a facade, perhaps disguising material conditionals. However, given the apparent prevalence of counterfactuals in mathematics, and given that mathematicians seem to be *taking themselves* to be talking in counterfactual terms, this would be a very revisionary view of mathematical practice. As such, it would require extensive independent justification to be considered as a serious objection. Whilst both vacuists and non-vacuists seem to be taking counterfactual usage for granted, we can simply assume the usage is genuine for the purposes of this debate.

5.8.2 How do mathematicians use counterpossibles?

Even granted that mathematicians genuinely appeal to countermathematicals in their writings, it is unclear *how* they are appealing to them, i.e. if they are appealing to them as vacuous or not. Williamson would clearly disagree with my claims that the judgements of mathematicians about specific countermathematicals would match the non-vacuist judgement. It is worth discussing some evidence in favour of non-vacuism. Yli-Vakkuri and

[68]Credit should be given to an anonymous reviewer of Dickson (2022) for bringing up the importance of clarifying this point.
[69]Available in section 5 of that paper.

Hawthorne say that, in conversations with mathematicians, they will tend to assert counterpossibles like the following:

(**TB**): "If AC were false, then the Tarski-Banach theorem would not be provable from the truths of set theory."

Whilst denying counterpossibles like:

(**TB**[1]): "If AC were false, then the Tarski-Banach theorem would be provable from the truths of set theory"
(2020, 567).

This is exactly as the non-vacuist should accept (and indeed as I assert), and confusing only for the vacuist. **TB** is true because the consequent would follow if the antecedent were true. Part of what is for AC to be false is for the Tarski-Banach theorem to fail to be provable from the truths of set theory[70]. Thus, **TB**[1] is false because if the axiom of choice were false it would not be possible to prove the Tarski-Banach theorem from the truths of set theory, the proof requires the truth of the axiom of choice. This element of mathematical practice is an anomaly for the vacuist, as noted by Yli-Vakkuri & Hawthorne (2020, 567-8). This practice also extends to logicians discussing counterlogicals (counterfactuals with a logically impossible antecedent). This practice which seems to contradict vacuism is a problem for vacuists to solve. If this practice is stable then vacuists will have to be quite radically revisionary about mathematical/logical practice, an obvious weakness. Non-vacuists, however, have a prima facie explanation of this phenomenon; the reason that mathematicians deny such counterpossibles is because such counterpossibles are false. In these cases, the consequent does not follow from the relevant antecedent.

Further support for the non-triviality of countermathematicals can be found in Jenny (2018). Jenny proposes that mathematical practice implicitly relies on the assumption that

[70]Another version of this idea is presented in Sendlak (2021). Sendlak argues that counterpossibles such as "Had paraconsistent logic been true at the actual world then…" are paraphrases of statements like "According to the story of paraconsistent logic…". If the consequent in the paraphrase makes the statement false overall, then the counterpossible equivalent should also be false. By Sendlak's argument, Williamson (2020, 129-130) is simply wrong when he accepts "...*if* the Bible is to be believed, there are angels" and also accepts "if the Bible is to be believed, there are no angels" as both being true. Believing the antecedent to be false does not justify accepting the second statement as true, because that is simply false according to the story of the Bible.

countermathematicals are non-trivial, specifically in the case of relative computability theory. This is important work. Jenny also proposes (2018, 552) a project going forward whereby non-vacuists should aim to find counterpossibles in other areas, such as the sciences, to defeat vacuism on multiple fronts. As Jenny (2018, 552-553) says:

> "Once we have a clearer picture of the areas where non-vacuous counterpossibles are indispensable and once we have model theories for these various classes of counterpossibles, we may then investigate to what extent we can integrate these model theories to come up with a unified and fully general theory of non-vacuous counterpossibles"

This chapter can then be seen as a continuation of the Jenny project, an attempt to bring counterpossibles in these distinct areas together. This is also where I aim to go further than Jenny. I have aimed not merely to show individual cases of non-trivial counterpossibles in distinct areas, but also to show *why* these are non-trivial. I aim to show that exploratory processes are what we need to engage in in order to get non-vacuous counterpossibles, along with beginning to form the fully general theory that Jenny is looking for. I assert that the only way for non-vacuists to make a start on this general theory is to highlight the mistakes that vacuists make by making clear the requirements to *genuinely* conceive of something, and show how vacuists fail to do this.

5.8.3 Is Williamson in fact genuinely conceiving of a distinct world?

One may object to my criticism that Williamson (2018) has not considered a distinct world, and has simply considered the actual world. One way to do this can be drawn out from the work of Yli-Vakkuri & Hawthorne (2020). Yli-Vakkuri & Hawthorne invite us to take a standard proof by reductio in maths, e.g. Euclid's theorem. In this proof, one initially supposes that there is indeed a largest prime. Then, given this claim and other established truths, they deduce various other statements and eventually show that the hypothesis in question was false. Vacuists could claim that I have unfairly characterised the mathematical process because the above describes a situation in which one does suppose the false hypothesis to be true, i.e. this is in fact an exploratory process. This is mistaken however, and, in fact, Yli-Vakkuri & Hawthorne provide the material to explain why. They make the distinction between a consensus and a non-consensus context, as they say:

"In a consensus context the relevant axioms are taken for granted, it is common ground that they are being taken for granted, and no one is interested in challenging any of the axioms or in exploring the ramifications of giving up some but not all of the axioms… In a non-consensus context one is not entitled to assume that all of the axioms are true and hence also not entitled to assume that they are provable, since provability entails truth" (Yli-Vakkuri & Hawthorne, 2020, 566)

What it takes to genuinely suppose a statement is to be in a non-consensus context, as described in 5.7.3. For it is only in a non-consensus context that you drop the assumptions you have that will immediately contradict the hypothesis. In a consensus context, the countermathematicals may all turn out to be true because they follow from the relevant mathematics. In a non-consensus context, this is not the case. When one jettisons assumptions, one will not immediately run into contradictions, so the truth value of the counterpossibles will be up for grabs. To decide whether or not Euclid's theorem is a case of a consensus/non-consensus context, let us reiterate that example. As Yli-Vakkuri & Hawthorne (2020, 558) say, we take a set of assumed axioms, $\mathbf{\Gamma}$, e.g. the Peano axioms, and $\boldsymbol{\varphi}_i$, which is the claim that there is a largest prime, and ultimately conclude $\boldsymbol{\psi}_i$, our desired contradiction which shows us that the claim, $\boldsymbol{\varphi}_i$, was false. We should be able to see that, in their own terms, this is a consensus context because the set of assumptions, $\mathbf{\Gamma}$, has not been modified. This matters because $\mathbf{\Gamma}$ will either directly contain the proposition $\neg\boldsymbol{\varphi}_i$, or $\neg\boldsymbol{\varphi}_i$ will be a logical consequence of $\mathbf{\Gamma}$. In this way, consensus contexts fail to be a genuine conception/supposition of $\boldsymbol{\varphi}_i$, because they implicitly assume that $\neg\boldsymbol{\varphi}_i$ is the case.

To make clear the implications for Williamson's argument, my allegation is that Williamson stays within a consensus context. This is insufficient for a genuine conception of $\boldsymbol{\varphi}_i$. One cannot explore the consequences of a statement, $\boldsymbol{\varphi}_i$, if one implicitly assumes that $\neg\boldsymbol{\varphi}_i$ is the case. At best, one can undertake an investigation to show that $\boldsymbol{\varphi}_i$ is false. Investigative processes have to take place within a consensus rather than a non-consensus context. As I have claimed, an exploratory process is the one which can produce false counterpossibles and this is precisely because it operates within a non-consensus context. There is further support for this later in the paper when Yli-Vakkuri & Hawthorne describe a fictional community of mathematicians (2020, 566), "For example, if *A* [$\boldsymbol{\varphi}_i$] is the claim that there is a largest prime number, the point, if any, of a Boxer's assertion of $A \;\Box\!\!\rightarrow B$ [$\boldsymbol{\varphi}_i > \boldsymbol{\psi}_i$] will be to contribute to an explanation of why there is no largest prime number". In order to show that $\boldsymbol{\varphi}_i$ is not the case, they have to keep in place the assumptions that will contradict it. Plainly this will be a consensus context which fails to be genuinely considering a situation in which φ_i is the case.

A stronger response to Williamson is available. Rather than even dealing with counterpossibles, Williamson might instead simply be working with disguised indicative conditionals with necessarily false antecedents. One might think that, in the case of counterfactuals, one cannot stay within a consensus context. In order to assess a counterfactual, one has to enter a non-consensus context and engage in an exploratory process. After all, if one does not consider something 'counter to the fact', then at best one is considering an ordinary indicative conditional hidden in superficial counterfactual form. The same ought to be true for counterpossibles. If you fail to exit the consensus context, then in fact you are simply not considering a counterpossible at all and instead dealing with an indicative conditional with a necessarily false antecedent. All such indicative conditionals may well turn out to be true, but importantly we would no longer be dealing with counterpossibles. I do not commit to this stronger option here but I mention it again in 7.4.3 as it represents an important avenue for future research.

5.8.4 Is counterpossible usage a fringe phenomenon?

One of Williamson's key arguments in favour of vacuism is that counterpossibles are a fringe phenomenon. This seems to be implicit in his discussion in a number of places:

[In a discussion of counterlogicals] "... it would be naive to take appearances uncritically at face value in a special case so marginal to normal use of language, for example by offering them as clear counterexamples to a proposed semantics of conditionals… it is good methodological practice to concentrate on conditionals with less bizarre antecedents in determining our best semantic theory of conditionals…" (2020, 60).

"After all, once the impossibility of a supposition is recognized, continuing to work out its implications is typically a waste of time and energy." (2020, 234).

"In linguistic practice, counterpossibles are a comparatively minor phenomenon, which is one reason why it is implausible to complicate the semantics of modalized conditionals in natural language just to achieve a desired outcome for them…" (2020, 262).

However, I would simply deny that these are in fact fringe cases of counterfactuals. As we have seen, vast portions of scientific reasoning contain counterpossibles; mathematicians/logicians seem to use countermathematicals/counterlogicals respectively; and to engage in meaningful debate in metaphysics, it seems we might need to use countermetaphysicals. Given the wide usage of counterpossibles in all these domains, it makes little sense to describe these as fringe cases. Counterpossibles are a significant datum, and a semantic theory needs to account for their usage in a way that is not revisionary to the vast areas of practice which employ them. If, as Williamson says, such counterpossibles present a problem for a standard semantic theory, then that is simply a reason to reject that particular semantic theory, rather than be revisionary to all this practice.

5.9 Section summary

Vacuism conflicts with a lot of intuitions we might hold. Of course, intuitions only take us so far. But there is also a strong precedent in the sciences to treat counterpossibles non-vacuously. One reason to do this is that we sometimes need to partake in an exploratory process about the antecedent. Within a non-consensus context/exploratory process, we can use counterpossible form to highlight the counterfactual dependence at play, for example, that the microphysical structure of diamond is responsible for its poor electrical conductivity. We do this by reasoning about what *would* have been the case if something impossible were the case. This reasoning can go wrong when we make a mis-ascription as to what would have been the case, resulting in false counterpossibles.

Despite apparent surface level difficulties, we can also extend the same reasoning process to intuitively non-vacuous countermathematicals because the same distinction between investigative and exploratory processes exists there. This also gives us space to have non-vacuously false countermathematicals, when this exploratory reasoning process goes wrong. To engage in this exploratory process in either case we may need to, on some level, genuinely conceive of an impossible world. To consider a counterpossible, $\varphi_i > \psi_i$, we have to genuinely conceive of a world in which φ_i is the case, in doing so we have to reject our assumptions to the contrary. When we do this, some counterpossibles will turn out true, and some will turn out false. In other words, vacuism about counterpossibles is false. As such, despite the apparent problem for my theory of mathematical causation, the relevant true/false countermathematicals are in fact available to back up the causal claims I need to make in the right way[71].

[71]In 6.8 I turn again to some concerns surrounding the semantics of counterpossibles. In that section I go into more explicit detail about how the truth value of the specific counterpossibles (e.g. "Had 23 been divisible by 3…") are up for grabs.

Chapter 6 - General problems

Throughout the thesis, many issues arose but were sidestepped to avoid overly long tangents that detracted from the discussion. This chapter is dedicated to answering these problems and offering solutions. In physical constraint cases, the constraint relation seems to straightforwardly be a causal relationship. In mathematical constraint cases, constraint intuitively seems to be grounding but shares the features that show physical constraint is a causal relationship. Chapter 4 pointed this out and argued that this licences us to endorse **Con=G&C**. In chapter 2 I discussed some potentially problematic differences between causation and grounding. There is a worry that similar differences will emerge between mathematical constraint and causation so it is worth considering these issues again. For example, grounding seems synchronic whilst causation seems diachronic and the fact that grounders are more fundamental than their groundees, whilst this is not the case for causation. This chapter will argue that in fact, there is no difference between mathematical constraint and causation with regard to these and other key features. Having done this, I hope to have strengthened the analogy between mathematical constraint and causation enough to support **Con=G&C**.

6.1 Synchronicity, hasteners & delayers

6.1.1 Synchronicity

Mathematical constraint seems to be synchronic rather than diachronic. This is because mathematical constraints hold of necessity, they hold at every point in time. For any constrained entity or event at a time, ***t***, it is constrained as a result of something that holds at ***t***. In **River**, the explanation of the water not travelling down distributary **B** at ***t*** will appeal to the tree blocking **B** at time ***t***. Similarly, the explanation of why Jane cannot divide the strawberries at time t_1 will appeal to certain mathematical laws that hold at t_1. I think this is the right conclusion to draw. However, causation seems to be paradigmatically diachronic. This, therefore, threatens the analogy between constraint and causation. The unity of causation and constraint rests on showing that the relationship in **River** is clearly causal. Given that mathematical constraint is clearly synchronic, I need to claim that **River** also shows a synchronic relationship. I am happy to make this claim, in fact I think this is the correct result. **River** is a case of constraint precisely because it is synchronic. The water being unable to flow down **B** is a result of the fact that **B** is blocked at that very moment, it is not because **B** was blocked in the past. This is why the relation is synchronic. Whether or not the river was blocked in the past is irrelevant when talking about the constraint for our purposes. Constraint is always going to be synchronic because when explaining why something cannot be otherwise, we point to why its being otherwise is impossible at that time. In other words, I argue that constraint is an example of synchronic causation. The disanalogy that constraint might be synchronic whilst causation is diachronic is not a problem, because **River** is an example of causation that is synchronic.

6.1.2 Hasteners & delayers in constraint

In 2.3.9.2, I argued that **G=MC** faces no threat from issues surrounding hasteners and delayers. It is worth saying something here to show that **Con=G&C** fares similarly well. Recall **Driving** from 4.1.1. In the UK one cannot (legally) drive until the age of 17 *because* it is illegal to drive under that age; the law in place (the constrainer) constrains the kind of actions we can (lawfully) take. An additional constraint in place is that one cannot (legally) drive under the influence of alcohol. So, if an agent either drives whilst underage, or drunk, they are acting illegally. Let us assume an agent is doing both and is pulled over by the police. Under normal circumstances, the officers would have questioned the driver and eventually arrested them for failing a breathalyser at time t_2. However, they first check the driver's ID and see that they are underage, therefore arresting them for that, at time t_1. The constrainer resulted in an arrest. The arrest would have happened at t_2 but, due to the constrainer in place at t_1, the arrest was hastened and happened at time t_1. Therefore, constraints can act as hasteners.

One might raise the concern that the above case seems importantly different from **Strawberries** and the mathematical constraint that I want to talk about. I claim that the important difference between these cases is that **Driving** involves two constraints, whereas **Strawberries** involves merely one. A hastener is something that makes an effect happen sooner *than it would otherwise have done*, at least by standard definitions. In **Strawberries**, the constrainer does not make Jane fail earlier than she *would otherwise have done*, because she *would **not** otherwise have failed*. But this need not be a significant problem. My claim is that constraint is a subset of causation. It is not necessary to show that every case of constraint admits of hastening/delaying because not every causal relationship does. However, it is worth showing that at least some cases of mathematical constraint can straightforwardly involve hastening/delaying.

Let us consider an altered version of **River**, **Modified River**. In **Modified River** the structure is broadly the same as in 4.4. Before the tree falls, the river flows down distributary **B**, which winds and snakes at a gradual incline and eventually arrives at a waterfall. Distributary **A** is a much straighter path, and at a steeper incline, but rejoins **B** at the same waterfall. Let us imagine a kayaker, ***k***, has chosen to travel down this river in a world in which the tree does not fall. At ***t***, ***k*** follows the river along and eventually comes to the juncture between what would be the two distributaries. Naturally, ***k*** follows the way the river flows, **B**, and due to the winding river and the gradual incline eventually arrives at the waterfall at t_2. At this point, ***k*** has to stop and get out of the boat, lest they go over the waterfall. Now, let us imagine a situation in which the tree has fallen. The same kayaker[72], travels down the same river. At the point at which ***k*** reaches the fallen tree, they have to go down **A**, where the water now flows, due to the constrainer. Because **A** is a much straighter route and at a steeper incline, ***k*** can reach a higher speed and travel more directly towards the waterfall. They eventually reach the waterfall earlier than they otherwise would have done, and so have to stop and get out at t_1 rather than t_2. It seems that the tree, acting as a constrainer on the flow of the river, has hastened ***k***'s arrival at the waterfall. ***k***'s getting out of the river has happened sooner than it otherwise would have done were the constrainer as not in place, this seems to be a case of hastening constraint. In this case, ***k***'s getting out of the river would still have happened without the constrainer. We can also produce a similar mathematical case.

Let us consider the case of periodical cicadas, often used in indispensability arguments., e.g. Baker (2005). These periodical cicadas have life cycles ranging from 13-17 years due to the local environmental conditions, food supply etc. Because 13 and 17 are prime numbers it is advantageous for cicadas to have life cycles of these lengths, as this minimises intersections with predators. This is an example of a constraining relationship, albeit potentially a more complicated one than **River/Modified River**. For the purposes of this point, I will merely sketch the explanation. The fact that 13 and 17 are prime constrains the number of factors

[72]Or a counterpart k_1, whichever is preferred.

they have to two each. This constrains it so that life cycles that have these lengths intersect with fewer predator life cycles. This then constrains how these cicadas will evolve because it reduces the range of what is biologically advantageous. This results in cicadas emerging on specific dates, every 13 or 17 years. For example, a group of these magicicadas emerged in May 2020. It is worth noting that there is a different sense in which the constrainer makes other outcomes impossible than in previous cases. It is not biologically impossible that cicadas have 14-year life cycles; it is just that such a life cycle would be disadvantageous and not the 'fittest' (in terms of survival of the fittest) possible life cycle. As such, there is certainly a sense in which it would be impossible for cicadas with that life cycle to survive, perhaps due to over-predation. Whilst this is not a problem for the illustration, it is worth highlighting. Now let us imagine a counterpossible situation in which there were no prime numbers between 12 and 18. Given that there would be no life cycle lengths more advantageous than another, in terms of minimising predation, many different lengths would arise within this range. For example, we might imagine that a group of cicadas has 14-year life cycles, so there would be an emergence in May 2021. Given that 13 and 17 are indeed prime and this constrainer is in place, cicadas (this subspecies at least) will exhibit life cycle lengths of 13 or 17 years. We might say that the emergence of the cicadas that would have been due to occur in May 2021, actually happened in May 2020. We can diagnose this as follows: because the mathematical constrainer is in place, i.e. 13 is prime, the relationship that governs life cycles has been affected. This has resulted in a cicada emergence happening sooner than it otherwise would have done. It seems we have a case of constraint which admits of hastening[73].

6.2 Productivity

Causation is most often (at least on the folk understanding) thought of as a productive force, e.g. dominoes falling. Both physical constraint and mathematical constraint do not seem to be productive in this way. Constrainers do not produce anything, instead, they restrict the possible behaviours of systems. Constraint also seems to lack a mark transfer between the constrainer and the things caused by it. A common suggestion for a mark transfer in cases of

[73]Of course, conversely, we can imagine how we might formulate such an example to be a delaying case in would-be-16-year cicadas that actually have 17-year life cycles.

causation is energy. **Strawberries** does not seem to be a case of energy transfer[74] between the constrainer and the things being constrained, indeed it would seem a category mistake to even suggest this was happening. **Strawberries** involves a condition holding of mathematical necessity, which then restricts some features of the world. One may point to this as the reason why we in fact *have* to distinguish constraint from causation.

It might seem that I am using "causation" in a non-standard and illegitimate way. In chapter 3 I argued that process accounts of causation fail to deal with some cases of alleged causation, e.g. omission, where counterfactual theories can. I would now like to talk about the precedent in the literature for discussing cases of causation which are not of this process form. Consider Hume's description of causation:

'Or, in other words where, if the first object had not been, the second never had existed' (Hume, 1995, 54)

This makes no mention of energy transfer, indeed the most common interpretation of Hume's account of causation is as a mere regularity thesis. The idea of a productive force or mark transfer between cause and effect is typically rejected on such Humean accounts.

There are also non-Humean accounts that are nonetheless compatible with my usage of constraint as causal. Take the account of structuring causes proposed by Dretske (1988). Hoffman & Schulte (2014) work with this notion of a structuring cause, although with a different target in mind to Dretske. Dretske makes a distinction between two different types of causes of behaviour (Dretske 1988). First, there are triggering causes. We can think of triggering causes as exactly like this folk productive sort of causation. Dretske is interested in the causes of behaviour, and he defines behaviour as the process by which some internal state of the brain, ***C***, causes some set of movements, ***M***, symbolised as ***C→M*** (Dretske, 1988, 15). Triggering causes are whatever is directly responsible for the internal state, ***C***. Let us take ***C*** to be a feeling of a sharp pain in one's hand, and ***M*** to be a swift retraction of the hand. The triggering cause here could be a sharp pin pressing into the hand. A triggering cause triggers the behavioural process, it explains why ***C→M*** is occurring at the very time at which it occurs, and this is indeed straightforwardly productive. But Dretske discusses the notion of a

[74]See 6.3 for a discussion of whether this violates the principle of the completeness of physics in a problematic way.

structuring cause concerning behaviour as well. In contrast to a triggering cause, a structuring cause does not directly cause a specific instance of behaviour. A structuring cause is responsible for the relationship between ***C*** and ***M*** within a defined system. It explains why the system is so set up that ***C's*** bring about ***M's***. It explains why it is that, upon the occurrence of ***C***, the process of ***C→M*** is happening rather than some distinct process in which ***C*** brings about some distinct movement, ***K*** (Hofmann & Schulte, 2014). This should sound familiar to us because this seems to be exactly how Lange (2017) defines constraint. Lange says that the relevant causes in **Strawberries** are the fact that Jane has 23 strawberries and the fact that she has 3 children. The constrainer, that 23 is indivisible by 3, explains why, in each instance, Jane will fail to distribute her strawberries evenly between her 3 children. This seems exactly parallel to the triggering vs structuring cause distinction. This should suffice to show that my description of constraint as a causal relationship is not too unusual. It is not productive but there are already examples of non-productive causation appealed to across the literature. The fact my theory does this is not a unique problem, if it is even a problem at all.

6.3 Completeness of physics/causal closure

It might seem that *Exotic Realism* is threatened by arguments from claims about the completeness of physics/causal closure of the physical. I view mathematical objects as exotic, they are non-spatiotemporal but with causal power. This sounds like a prime example of something non-physical and so any causal influence would be ruled out by the completeness of physics. These claims are worth addressing. First, we will consider a general account of the completeness of physics and of causal closure. With that in place, I can respond to these concerns. I will offer two responses to these worries: that the notion of 'physical' used by proponents of causal closure is not well-defined; and a property causation response. We will then consider Papineau's more specific causal closure claim, that causation by non-physical entities would violate the conservation of energy, and show some ways my account could respond.

A helpful way of understanding the completeness of physics is offered by Noordhof (2020, 263):

Completeness of Physics [hereafter "**Completeness**"]: At any time at which a physical event, holding of a physical fact or instantiation of a physical property, e, has a cause, then, without any non-physical causes at a time, physical events, facts, and properties either are sufficient causal circumstances of e or causal circumstances that fix the probability of e.

The core of claims like this is that, in giving a causal account of a given physical event, we do not need to go beyond the physical. Physics can give us a complete account of the goings on.

Similarly, causal closure principles are intended to introduce a gap between the ordinary physical world and non-physical metaphysical entities. Noordhof again provides us with an initial formulation of causal closure by combining two conditions:

Causal Closure [hereafter "**Closure**"]
Cause Condition: For every narrowly physical cause, there are only narrowly physical effects, and only narrowly physical effects are nomologically possible.

Effect Condition: For every narrowly physical effect, there are only narrowly physical causes, and only narrowly physical causes are nomologically possible.

Proponents of **Closure** tend to be physicalists, and so are committed to the idea that there are only physical events/things. **Closure** straightforwardly falls out of this physicalist commitment, if there are no non-physical causes, then all causes are physical. There are many different ways to formulate causal closure principles, for example, see Lowe (2000) for a review of a number of different options. But as we have said the spirit of causal closure is to bar non-physical stuff from having a causal effect on the physical world.

6.3.1 Explaining causal closure

Justin Tiehen (2015) is a physicalist who nonetheless sees some problems with causal closure principles. Tiehen says that **Closure** cannot be true independently of a claim of the form of *P** (Every event is physical) for various reasons. Firstly, this could potentially lead to problematic overdetermination (2015, 2407-09). But it would also mean that physicalists would be committed to the falsity of a key counterfactual: “if there had been some nonphysical events, the physical realm would not have been causally closed.” (Tiehen, 2015, 2409). If one thinks that **Closure** is true because *P** is true, i.e. true derivatively, then one should think the above counterfactual would be true. However, if one thinks that **Closure** is true independently of *P**, then one should judge the counterfactual to be false, even had there been non-physical events, they would not have been causes. Tiehen (2015, 2409) alleges that this latter route is less plausible because a world with epiphenomenal non-physical events is more distant than a world with actively causal non-physical events. For example, unicorns do not exist at our world, so there are no unicorn-causes of events, but had there been unicorns then there would have been unicorn-causes (2015, 2406-09) This counterfactual is true because a world with epiphenomenal unicorns is more distant than a world with causal unicorns. Tiehen thinks the case should be parallel with non-physical, e.g. mental, events.

Tiehen thinks that in order to support **Closure** we could offer some sort of inductive argument in support of *P** that allows us to conclude **Closure**. The problem is that if you accept such an inductive argument, then you have to reject certain counterfactuals. If we consider the case of experiments performed on a good sample size of copper wire, we might make inductive claims about all copper wire. However, we cannot then accept counterfactual claims about what would happen if the wire were different in various ways, because if we do then we have no reason not to think that the copper wires we are generalising over are in fact of this different sort (Tiehen, 2015, 2419). Likewise, if we accept an inductive argument for *P**, we can no longer accept counterfactual claims about worlds in which non-physical events do exist. We would thus be in the problematic situation that Tiehen criticised earlier. As Tiehen says:

> “If it is agreed that nonphysical events would violate (Closure) if they existed, how on earth could induction inform us that (Closure) is true while leaving it open that there may be such events?” (2015, 2419).

Tiehen goes on to instead offer an *abductive* argument for *P** directly, which then goes on to support **Closure**. But as Tiehen (2015, 2423) says “...according to the abductive argument, our reason for believing that the physical realm is causally closed is just our prior warranted belief that every event is physical…”. The problem is that, firstly, this argument will not be persuasive to those who believe in non-physical events precisely because their warrant in believing that every event is physical will be close to, if not, zero. But more than that, these kinds of arguments might work against Tiehen's target of dualism about the mind, but it is less clear that they will work against non-physical mathematical objects. As Tiehen points out, if we keep investigating the mind and finding fully sufficient physical causes, then this tells against the dualist hypothesis and will increase the warrant in the relevant claim to support *P**. But indispensability arguments in the philosophy of mathematics can be used to avoid this. They are meant to give us independent reason to believe in mathematical entities (sometimes as non-physical things), and so they directly undermine warrant in the claim that every event is physical, and thus in **Closure** itself.

One might wonder if Tiehen could support **Closure** by appealing to **Completeness**. The success of science and the progress it makes without appeal to the non-physical is cited as evidence for **Completeness** (Yates, 2009, 125). If it is the case that, as per **Completeness**, we simply do not need to appeal to non-physical causes, we can justify belief in **Closure**. If we never need to appeal to the non-physical events, then we can reject that they exist. However, there might be parallel concerns with **Completeness** to those that affected **Closure**. In appealing to the success of science to support **Completeness**, we seem to be making an inductive argument in support of it. I refer again to Tiehen's copper wire example (2015, 2419). Inductive arguments for **Completeness** leave open the possibility that physics might one day be shown to be incomplete, it is therefore unclear how it is that inductive arguments can ever be used to show that **Completeness** is true. It seems that the only way to justify **Completeness** or **Closure** is by appeal to a statement like *P**. But once again, this will fail to convince anyone who is not already a physicalist.

It is not clear how **Closure** and **Completeness** are supported. This in itself might be a sufficient response to these concerns. However, perhaps support for these principles can be found elsewhere. If one can present sufficient support for **Completeness**, then one can justify **Closure** and so my account would be threatened. It is therefore worthwhile taking these concerns seriously. It is a strength of my account that, even if **Closure** and **Completeness** are true, I can respond to such concerns. This is what I will do in the remainder of 6.3. I will show that my claims about mathematical causation are in fact compatible with **Closure** and **Completeness**.

6.3.2 What does 'physical' mean?

One might respond to the closure/completeness worries by pointing out that the notion of "physical", be it an event or object is not well defined. We will focus on the notion of a physical object for simplicity and because, presumably, an event can only be classed as physical if its component objects are physical. When we produce a list of paradigmatically physical objects, we will list things such as tables, chairs, trees and dogs. In short, we will list those things that we tend to class as concrete. Folk understanding tends to conflate these terms. Indeed, when we are using 'physical' in philosophy, we commonly mean this to be synonymous with 'concrete'[75]. But it isn't clear that the notion of concreteness is well defined. For example, Lewis (1986c, 83-84) criticised the abstract/concrete distinction as being vague. A rough definition of what we mean by concrete is that something is concrete if:

It exists in space;
It exits in time;
It causally influences/is causally influenced by the world[76];

However, these terms themselves are not always well-defined. For example, by saying that something is spatial we tend to mean things such as:

It occupies a volume;
It has definite boundaries;
It has a definite location;
It has mass;

However, we can find objects which will seem to fail to do at least one of these things, and so in that sense fail to be physical/concrete. For example, in terms of volume, particles such as quarks and electrons are said to be *point* particles. This means that, in themselves, they occupy zero volume, so if volume is a prerequisite for something being spatial (hence

---

[75]There is another common usage of 'physical' that is also used in philosophy, which we will come back to shortly.

[76]To justify the allegation that 'concrete' and 'physical' are often conflated, see the following quote from Dainton (2021, "Homogeneity and Simplicity", paragraph 13). "What does seem (highly) plausible is that so far as membership of the physical realm is concerned, taking up room in physical space and causally interacting with other physical objects and our sensory systems are of paramount importance.". This is very close to the definition of concrete that we standardly work with.

concrete, and hence physical), quarks would not be physical. Likewise, photons are said to be massless particles, and so if mass was required for physicality, they would be non-physical.

In terms of boundaries, if this is our standard then very little (if anything) will meet it. At the microscopic level, it is difficult to determine whether or not something should be counted as part of the object in question. At a certain stage an arbitrary boundary has to be drawn to distinguish the object from the rest of the world, as is illustrated in this quote from Varzi (2015):

> "On closer inspection, the spatial boundaries of physical objects are imaginary entities surrounding swarms of subatomic particles, and their exact shape and location involve the same degree of idealization of a drawing obtained by 'connecting the dots'"

The idea here is that we draw lines between the dots on the page and point to that line as the true boundary of the object, as something which we have discovered. This is not the case, rather the boundary is something we have projected onto the page. One could view any material objects similarly. Material objects are constituted from atoms connected together with various forces. In that sense, objects are much like a connect-the-dots drawing. There are conglomerations of separate atoms but there is no line drawn between them in nature, there is no boundary. The boundary is something we project onto material objects when we come into contact with them and experience them as solid. But we mistakenly believe objects to have this definite solid boundary. If having a solid boundary in that way is what constitutes something being spatial, then in fact nothing will turn out to be spatial.

In terms of definite location, we can return to the discussion from 1.2.2 where we discussed conceptions of particles like electrons. If we look at modern physics, our everyday notions of spatiality do not seem to apply at the quantum level. Consider another quote:

> "According to Democritus, atoms had lost the qualities like colour, taste, etc., they only occupied space… but in modern physics, atoms lose this last property, they possess geometrical qualities in no higher degree than colour, taste, etc…"
> (Heisenberg, 1952).

Our notion of a definite location as being a hallmark of the spatial and so the concrete/physical is simply not applicable in the physical sciences.

What the above shows is that these are not how we should categorise things as being physical, because most things would turn out to be non-physical. But there is a sense of 'physical' we have not considered which seems stable enough. We could define the physical, roughly, as whatever is revealed/studied by physics. As we will shortly see, in terms of properties this is similar to how Noordhof defines physical properties. A property is narrowly physical if it is recognised by our physics (or something like it) (2020, 252), and a property is broadly physical if it is recognised by a science other than our physics (2020, 265). This notion of 'physical' distances itself from being synonymous with concrete. Something is physical if science recognises it, but, perhaps in a weird world, some of those things will fail to be concrete. As I argued in 1.2.2, we may in fact be in such a weird world. I am happy to say that, for the reason considered here, some of the objects of physics fail to be concrete. But, of course, that should not mean they are non-physical, because the physical is simply the stuff that science recognises. If this is how we conceive of the physical, then the worries of **Completeness** and **Closure** will not threaten my view. Once again, indispensability arguments aim to show that science recognises the existence of numbers and mathematical objects, so on this looser definition of 'physical', mathematical objects would be physical (even if they were not concrete)[77]. If a defender of **Completeness/Closure** wishes to insist that we have taken a misstep somewhere and that we should really be conceiving of 'physical' as synonymous with 'concrete' then they will have to address the fact that we have no firm notion of exactly what 'concreteness' means and it is not clear that it will capture the things we want it to capture. We cannot form a causal closure argument effectively if we are not sure exactly which sort of entities it rules out.

6.3.3 A property causation response

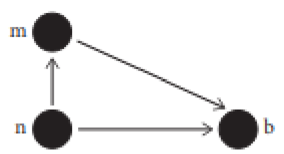
As we saw above, to respond to concerns around **Closure/Completeness**, we may wish to make a response similar to that used to defend property causation in Noordhof (2020) and say that the mathematical objects I want to talk about simply come out as physical anyway. We define property causation in the following way:

> "F is a property cause of G if and only if

[77]See the next section for how we might motivate this kind of response using property causation.

> Particularity: part of the (minimal) necessitation base for the instance of F causes part of the (minimal) necessitation base for the instance of G.
>
> Generality: (part of) each (minimal) necessitation base of F is such that all its instantiations would cause (or in the case of indeterminism, raise the - probability of) an instantiation of one of the (minimal) necessitation bases of G if they were in some causal circumstances C—where C may vary for each kind of necessitation base." (Noordhof, 2020, 255).

To explain how we could use this idea we should explore the notion of a narrowly physical property. A narrowly physical property is a property identified by some physics resembling our own (Noordhof, 2020, 252), contrasted with a broadly physical property that would be one not identified by physics, but potentially by the other sciences or just an everyday property such as 'being a table' (2020, 265). As Noordhof claims "It is the conviction that a proper causal explanation of narrowly physical instantiations of properties need not go outside instantiation of other narrowly physical properties that lies at the heart of the causal closure claim" (2020, 266). So for the world to be causally closed, there just needs to be a causal explanation of any narrowly physical property instantiation that involves only other narrowly physical property instantiations. An explanation of broadly physical properties that involves only broadly physical properties would technically fall outside of this, but not in a problematic way. We can see this with a diagram:



(Noordhof, 2020, 264).

We take *n* to be a narrowly physical property instantiation of the determinable *m*, and *b* to be a narrowly physical property instantiation of some other kind. *m* then is a broadly physical property (perhaps), that can provide an explanation of *b*. But we do not *need* to invoke that in explanation, if we want a purely narrowly physical explanation of *b*, we merely invoke *n*. In this way, *m* is not a violation of the causal closure of the physical, because causal closure only pertains to narrowly physical explanations. Noordhof alleges that "if we take the verdicts of a counterfactual theory seriously with regard to what they say about the efficacy of broadly physical properties, then they qualify as causes themselves with no implication of depending upon the efficacy of narrowly physical properties." (2020, 265). Given that all causal closure really demands is that for narrowly physical properties there should be a narrowly physical explanation, causal closure is not threatened by broadly physical explanations.

We can apply this to mathematics in a similar way. The fact that 23 is indivisible by 3 is (I assert) a mathematical cause[78] of the fact that Jane cannot divide her 23 strawberries between her 3 children. One minimum necessitation base of this fact is the existence of 23 strawberries. Take this strawberries fact to mirror the role of a narrowly physical property, similarly take Jane's inability to divide the strawberries as mirroring a narrowly physical property. The true spirit of causal closure, as above, is that in this instance there be a purely narrowly physical explanation of Jane's inability. There is, this is that she has 23 strawberries. Of course, there is also the mathematical fact playing a role at a broader level, but one does not need to invoke it to explain *this specific* scenario. One would need to cite the mathematical fact to explain the general occurrences of events like Jane's inability. But this would not violate causal closure as Noordhof sees it. Of course, I formulate my theory of mathematica as mathematical *objects* not properties. But one can see the analogy intended here. One can substitute 'narrowly physical property' for 'narrowly physical fact/event' and make the same arguments. Alternatively, although it is not something I am tempted by, one could take a property view of mathematics and stick more closely to the line of argument above. Either way, a response is available to these **Completeness/Closure** worries.

One might wonder if, by saying that one does not *need* to invoke the mathematical fact to explain Jane's inability to divide her strawberries, we have dispensed with any need to involve the mathematics. If the narrowly physical is enough, we can ignore the broadly physical. We can respond to this concern by pointing to Lange's distinction between constraint and coincidence from 4.1.2 and 4.7.4.2. The reason we appeal to the broadly physical mathematical property is because it is a constraint rather than a coincidence. The reason that the narrowly physical properties cause the relevant effects is because of the way the broadly physical properties are. We cannot dispense with them completely because then we would be unable to give a general account of why events like Jane's inability to divide

[78]Take mathematical cause in this sense to mirror the role of a broadly physical cause.

strawberries occur. Appealing to the mathematical fact provides us with additional explanatory power, but it also seems correct. 23 is not indivisible by 3 because, in all instances, we fail to divide 23 objects into 3 groups. We fail to divide 23 objects into 3 groups because 23 is indivisible by 3. Once again, it is a constraint not a coincidence.

We might have successfully responded to **Completeness/Closure** worries by safely bringing mathematical objects under the banner of physical objects[79]. However, perhaps a determined physicalist can come up with a clear conception of what 'concrete' means and so insist we use 'physical' as a synonym for it. This imagined definition might rule out mathematical objects from being physical and so it is worth considering what could be said in response to this. We will therefore assume, for the rest of 6.3, that mathematical objects would have to be non-physical. Moreover, whilst **Completeness/Closure** are open hypotheses, we know for a fact the universe is closed with respect to the total amount of energy. One might worry that mathematical objects being causal might violate principles of the conservation of energy because it is not clear how they affect the energy levels of the rest of the universe. It is therefore worth considering responses to these more specific kinds of closure arguments.

6.3.4 Papineau's argument from conservation

In her 2010 paper, Gibb gives a formalisation of Papineau's (2000, 2002) argument from the conservation of energy. The version she gives is as follows:

1. Every physical system is conservative or is part of a larger system that is conservative (*Conservation*).
2. There is (probably) no non-physical energy. (*Energy*).

Therefore,

3. (Probably) no physical effect has a non-physical cause (*Exclusion*).

(Gibb, 2010, 370)

---

[79]Although see below for an argument that even if mathematical objects were physical, there are only two ways for physical objects to be causal and it is not clear that mathematical objects do either, so violations of closure principles might still be present.

From this argument we can go on to derive *Completeness* - Every physical effect has a sufficient physical cause[80] (Gibb, 2010, 365).

Gibb (2010) alleges that this argument is not currently valid as the premises make no mention of causation. As such, she thinks two more premises need to be added in order to entail the conclusion. Firstly she adds a restriction on how it is that physical systems can be affected. She calls this restriction *Physical Affectability* (Gibb, 2010, 370).

Physical Affectability: The only way that something non-physical could affect a physical system is by (1) affecting the amount of energy or momentum within it, or (2) redistributing the energy and momentum within it.

The second premise that Gibb adds to Papineau's original argument is as follows:

Redistribution: Redistribution of energy or momentum cannot be brought about without supplying energy or momentum. (Gibb, 2010, 374).

Combining these two additional claims with the existing premises leads to an argument for *Exclusion* above. Gibb (2010, 378) points out, however, that if both *Physical Affectability* and *Redistribution* are simply hypotheses of physics, they are never mentioned explicitly. Moreover, if one conceives of causation as energy transfer, then one might find *Physical Affectability* convincing, but for other theories of causation, notably counterfactual theories, it is fine to deny it because causation is not conceived of as objects pushing and pulling each other (Gibb, 2010, 379). One might wish to say that denying *Physical Affectability* and *Redistribution* is problematic because it leads to an incomprehensible notion of mental causation. Gibb makes the point that this objection assumes a physical process theory of causation (2010, 380) and that this is exactly what is in question.

But there are other ways to respond to Papineau's argument. Within the mental causation debate, one might wish to argue that mental causation is not a violation of energy conservation. Lowe (1996, 65-67) discusses a property of causal networks he calls

[80]This is similar to the Noordhof completeness claim above, however, Gibb intended to derive her exact formulation of the claim, so I will adopt her italicising of key principles (e.g. *Completeness* rather than **Completeness**) to avoid ambiguity and make clear that the claims being argued about are exactly as formulated in Gibb (2010).

'convergence'. This is when a series of distinct neural events converge to produce certain bodily movements. As Gibb says "according to Lowe, a mental event is causally responsible for the fact that the maze of neural events has this convergence characteristic" (Gibb, 2010, 372). Intuitively, this might seem to be a kind of redistribution of energy. A mental event being responsible for this convergence simply is the mental event being responsible for this certain distribution of energy and momentum. Gibb uses the analogy of sand pouring down a funnel. The funnel makes sand converge in a narrow stream, it is causally responsible for the convergence. The funnel seems straightforwardly causally responsible for the distribution of energy in the sand-system.

> "One might interpret Lowe's claim that a mental event is responsible for the fact that a maze of neural events converge upon a specific bodily movement in an analogous way… a mental event is responsible for directing various electrochemical signals along certain neural pathways so that they converge upon a particular bodily behaviour" (Gibb, 2010, 373).

In this way, we might say that mental causation is not threatened by the above argument because mental causation is simply a form of redistribution of energy. However, as Gibb notes, this is not quite Lowe's point. Lowe thinks that convergence is a formal property of causal trees, i.e. a causal tree converges if multiple independent events come together to result in the occurrence of a distinct event (Gibb, 2010, 373). A mental event is responsible in the sense that it is responsible for this chain of independent events together bringing about a certain bodily movement. So rather than being an example of redistribution, it seems that this convergence could be some sui generis 'third way' to add to physical affectability. To see how this affects the debate on mathematical causation, we should first discuss why exactly mathematical constraint as a causal relationship might seem to be a problem.

6.3.5 Does mathematical constraint satisfy *Physical Affectability*?

Having suggested that mathematical constraint is causation, one might wonder whether or not it satisfies *Physical Affectability*. One might see parallels between mathematical constraint and Gibb's initial formulation of what Lowe meant by 'convergence'. The sand funnel analogy seems very close to **River**. Recall that in **River**, the fallen tree prevents the river from flowing one way, thereby causing it to flow another. In the funnel case, the funnel prevents sand from flowing one way, thereby causing it to flow another. Gibb alleges that the funnel case is a simple redistribution of energy, and so violates no conservation principles (2010, 373). If this is the case with funnels, it will be the case with **River**. The pertinent question then becomes whether or not it is parallel to mathematical constraint, if it is, mathematical constraint does not violate conservation, if it is not, then it might seem that we have found a disanalogy between mathematical constraint and causation. Therefore, if we want to respond to this argument, there are two options, we can show how *Physical Affectability* does not rule out mathematical causation, or we can insist that we need to modify/reject *Physical Affectability*. We will consider both.

If I want to say that mathematical constraint *is* a matter of redistributing energy, then I will have to argue against *Redistribution*, because plainly it would not do so by supplying energy. If instead, I want to say that mathematical constrainers cause via some route other than redistributing energy/directly supplying energy, then I will have to argue that other constraint cases are also like this, and spell out exactly what the other way of causing physical effects is. We will consider these in turn.

6.3.5.1 Mathematical constraint is redistribution of energy

In order to argue against *Redistribution* to make this claim I can refer again to Gibb (2010). It does not seem clear that *Redistribution* is a basic principle of physics, as it is in fact never stated. Moreover, as mentioned we can deny it without contradiction because it seems to implicitly assume a process-based account of causation. I am implicitly relying on a counterfactual notion of constraint. In chapter 3 I justified a counterfactual approach towards causation generally and it is not clear that such a stance requires any mention of energy transfer in order to judge that a relationship is causal. As Gibb points out (2010, 380), someone who takes this stance simply does not need to answer the question of 'how' the mathematical constraint happens, because it is not a process/mechanism. But it is worth saying something. There is a sense in which mathematical constraint is an instance of redistributing energy within a system. This is so because, had the relevant mathematical facts been different, the system would have been arranged differently. So there is counterfactual dependence of the particular distribution of the system upon the mathematical facts. This sounds like redistribution, and perhaps that is all that is needed to show that conservation principles are not violated. There might be a remaining intuition that something is wrong here, that this just does not seem to be redistribution of energy. I am sympathetic to this, and assert that this is because the mathematical constraint as described above simply is not *re*distribution of energy.

6.3.5.2 Mathematical constraint is not redistribution of energy

To argue that mathematical constraint is not redistribution of energy, but is still causal, I would have to deny *Physical Affectability*. An intuitive view of what redistribution of energy amounts to is taking some of the energy in a system and moving it to another place in the system. This is not the mechanism of mathematical constraint though, nothing is *re*distributed. Mathematical constrainers simply affect how energy can be distributed in the first place, and this does not seem to be redistribution. But nor is this directly putting energy into a system. In short, it seems that mathematical constraint, if genuine, is a counterexample to *Physical Affectability*. It seems that mathematical objects affect physical systems in some sui generis third way. In short, we should modify *Physical Affectability* as follows:

Physical Affectablity+: The only way that something non-physical could affect a physical system is by (1) affecting the amount of energy or momentum within it, (2) redistributing the energy and momentum within it, or **(3) affecting how energy can be distributed in the system in the first place.**

Once again, this is only mysterious if one has an energy transference account of causation. With a counterfactual view, we need not explain the *process* of how this happens (because there is no such *process*) but can still classify it as a causal relationship. This is not an ad hoc move. Although we initially characterised the funnel case (Gibb, 2010, 373) and **River** as cases of energy redistribution, this is not the case. Rather than primarily being cases of redistribution, they are instead restrictions that affect how it is that energy can be distributed in the first place. Adding a third condition for *Physical Affectability* may seem controversial if only noted in the mathematics case. But, relevant features of the funnel case and **River** seem to point to the need for this in other cases as well. **River** is not primarily causal in virtue of redistributing energy, it is primarily causal because it restricts how energy can be distributed within the system in the first place.

We have seen that there are a number of different problems for mathematical causation concerning causal closure/completeness. We have a number of different options: we can question support for causal closure in the first place; we can point out that the notion of 'physical' in these principles is often ill-defined and so when we refine it, it might capture mathematical objects anyway; or we can argue that non-physical mathematical objects do not violate energy conservation principles by either showing that constraint falls under the definition of causation outlined, or by redefining how it is that physical effects can be caused. All of these routes are open for my account of mathematical objects. My preference extends to the latter option and *Physical Affectability+*, but that is not an in-built commitment of *Exotic Realism*.

6.4 Transitivity, irreflexivity & asymmetry

As we saw in 2.3.1 and 2.3.2, there are important problem cases for the notions of transitivity, irreflexivity and asymmetry with regard to both causation and grounding. Constraint, however, seems to straightforwardly be transitive where causation/grounding might not be. Given the threat of a significant disanalogy, it is worth discussing this in some detail.

Transitivity seems to hold for constraint. In **River**, the tree's falling results in water being unable to flow down **B**, this results in water flowing down **A** which will lead on to other things. At first glance, **Strawberries** may not seem transitive in the same way but it is. The fact that 23 is indivisible by 3 constrains the world such that 23 objects cannot be divided into 3 groups, this then constrains the world such that 23 strawberries cannot be divided into 3 groups. Transitivity seems to hold in both cases. Irreflexivity likewise seems to hold for both relations, the idea of something constraining itself seems to be intuitively mistaken. The fallen tree in no way constrains anything about itself and similarly 23 not being divisible by 3 does not constrain itself. Both notions seem well-founded, the chain of explanation in **Strawberries** terminates at 23 not being divisible by 3 (or perhaps further mathematical facts which ultimately terminate in axioms), in **River**, it seems to terminate at the tree's having fallen. Both relations seem asymmetric in a relevant sense. In **River**, the direction of explanation must clearly flow from the tree's falling to the effects on the rivers. In the mathematics case, it seems likewise fairly obvious that the direction of explanation must flow from the fact that 23 is not evenly divisible by 3, rather than an explanation flowing from the fact that the strawberries cannot be so divided to the mathematical fact. Consider that Jane's failure to divide the strawberries is not the reason why 23 is indivisible by 3. In answering the question “Why is 23 indivisible by 3?” one should not turn to the fact that 23 strawberries cannot be divided between 3 people, nor should one explain it by the fact that in any instance of 23 objects, they will fail to be divided between 3 groups. Intuitively this seems wrong, something is missing. This is parallel to Lange’s discussion of energy conservation being a constraint rather than a coincidence from 4.7.4.2. Energy conservation does not hold because all the forces happen to conserve energy, all the forces happen to conserve energy because energy conservation holds. 23 is not indivisible by 3 because all groups of 23 individuals happen to fail to divide equally into 3 groups, all groups of 23 individuals fail to divide equally between 3 groups because 23 is indivisible by 3.

There might be those who deny that causation is unilaterally transitive, irreflexive or asymmetric. Wilson (2018a) highlights potentially problematic cases which show that causation and grounding do not possess these characteristics in each and every token relationship. These debates are interesting and important for the causation literature. However, what these debates do not show is that even if the problem cases prove to be genuine and unavoidable, that *all forms of* causation fail to be transitive, irreflexive or asymmetric. There is always going to be room for some kinds of causation to possess these characteristics, and that is all that my account requires. My account does not require that causation is all of these things in all cases; merely that a subcategory of causation exhibits these characteristics. **River**, along with similar cases, shows that there is a genuine and important sense in which we can talk about constraint as a causal relation. There is scope to describe it as its own unique subcategory of causation. As a subcategory, it is fine to assert that it is transitive, irreflexive and asymmetric, straightforwardly. I argue that this is the right result, given that this is how constraint seems to operate. This need not commit me to a view of general causation on these matters, the scope is restricted. Some people, however, might still be resistant. Given the prevalence of problem cases of causation (for transitivity etc.), people might worry that they will emerge in cases of constraint as well. This is an understandable worry but happily, one that could be accommodated. Instead of a straightforward understanding of these characteristics, one might take the route of contrastivity that Schaffer does in various places (2005, 2012).

Schaffer adopts a contrastive account for a number of reasons, but one key reason is that doing so, he alleges, solves issues surrounding absence causation. There are positives to accepting absence causation, Schaffer (2005, 329-331) lists the following: it is often intuitive (e.g. the failure to water a plant causes it to die); absences play predictive/explanatory roles (e.g. a pilot's failure to lower the landing gear could predict/explain the plane crash); absences play the legal and moral role of causes (e.g. not feeding one's child makes one morally and legally responsible for the child's starving; absences mediate causation by disconnection (e.g. decapitation causes death by preventing oxygenated blood from preventing brain starvation). However, absence causation presents some difficulties as follows (Schaffer, 2005, 330-331): absence causation is unintuitive in a lot of cases (e.g. it seems wrong to say that the queen of England's not watering my flowers caused them to die); absence causation is theoretically problematic (i.e. it is unclear what negative statements like 'the queen's not watering of my flowers' denotes); some find absence causation metaphysically problematic (e.g. absence causation involves no energy transfer, which is [as discussed] often treated as a hallmark of causation). Schaffer (2005, 331-332) claims that a move contrastivity can preserve our positive judgements whilst vindicating our negative judgements in the other cases. For these reasons, contrastivity is theoretically attractive as a way to reason about causation.

It is also particularly helpful that Schaffer gives a contrastive treatment of grounding which can easily be transferred over to the constraint case. Contrastive grounding can help deal with some of the issues we identified in 2.3 in the following ways:

"Differential Irreflexivity: It is not the case that the fact that φ rather than φ* grounds the fact that φ rather than φ*
Differential Asymmetry: If the fact that φ rather than φ* grounds the fact that ψ rather than ψ*, then it is not the case that the fact that ψ rather than ψ* grounds the fact that φ rather than φ*
Differential Transitivity: If the fact that φ rather than φ* grounds the fact that ψ rather than ψ*, and the fact that ψ rather than ψ* grounds the fact that ρ rather than ρ*, then the fact that φ rather than φ* grounds the fact that ρ rather than ρ*"
Schaffer, 2012, 132

It seems we can simply replace "grounding" with "constrains it to be the case that" or similar. This will be a fairly simple modification to the account, and one that does not result in any loss. Hopefully, such a move would satisfy transitivity sceptics.

As we discussed in 4.7.1, a move to a contrastive understanding of causation and constraint helped to avoid problems concerning the number of relata in causation and constraint. A contrastive understanding of both relations results in them both being 4-place relations. With the discussion here, we can now see that a contrastive spin on these accounts also allows us to avoid any lingering problems with transitivity, irreflexivity and asymmetry. Given the advantages it conveys, this is a move worth making. Moreover, this is not an ad hoc move, we saw in 4.7.1 that our intuitions about constraint seem to describe it in contrastive terms so this is a natural evolution of the constraint account. We should, therefore, take constraint and causation to be contrastive relations.

6.5 Fundamentality

6.5.1 Grounders are more fundamental than groundees

In 2.3.4 we discussed that in grounding relations, the grounder is more fundamental than the groundee, whereas this seems not to be the case for causes/effects. Causation can operate between elements at the same level of fundamentality. When billiard balls hit each other, neither is more fundamental than the other in any relevant sense. We are now in a position to see how this need not be a problem in the constraint case because in all constraint cases, the constrainer is indeed more fundamental than the constrainee, including physical constraint cases like **River**. This in fact seems to be a defining feature of constraint because if the constrainer is not more fundamental then it cannot *be* a constrainer, and is simply a coincidence.

There might be those who object that this connection with fundamentality suggests that constraint is not causation. This is a little hasty though. Let us remember that what we did was take a straightforward causal relationship, show that it is also constraining and that it has a certain structure and pattern of counterfactual dependence. We then showed that mathematical constraint possesses these same features and argued that this makes it causal. If one wants to object that this connection to fundamentality makes something acausal then we might end up concluding that a straightforwardly causal case (**River**) is acausal because it has this connection. Concluding that such obviously causal cases are in fact not causal seems to be worse than conceding that some cases of causation have a connection to fundamentality.

Still, the objection might be pressed that in the mathematical cases there is a sense of differences in fundamentality that is not present in **River**. The mathematical fact about the primeness of 23 seems more fundamental than the facts about strawberry division in a way that the fallen tree is not more fundamental than the flowing river. Given that causation does not seem to track fundamentality, it suggests a disanalogy with constraint. We might think that **River** is causal but **Strawberries** is not. To respond to this I want to draw out what exactly we mean by fundamentality in these cases.

6.5.2 Fundamental in what way?

In spelling out why **River** does not involve differences in fundamentality we might say that the case is "operating at the same level of reality". I argue that what this points to is that when we are talking about fundamentality in these cases we are talking about a "building up" kind of fundamentality (Bennett, 2017, 137). The tree is not more fundamental than the river because the river is not "built up out of" the tree, in the way that atoms are more fundamental than molecules. Of course, this kind of fundamentality is not present in **River**, but neither is it present in **Strawberries**. The fact that the strawberries cannot be distributed in the relevant way is not built out of the mathematical facts (although it follows from them, but as does the river flowing the way it does follow from the relevant facts). But this is not what is being claimed when we say that the strawberries fact is more fundamental. The sense in which the mathematical fact is more fundamental can also be applied in **River**.

That 23 is indivisible by 3 is more fundamental because it has what I am going to call 'primacy' over the situation. It is 'more important' in a relevant sense. I suspect primacy is going to turn out to be a primitive notion but we can nonetheless talk about such cases more and get a better idea of it. In metaphorical terms, if something with primacy is in place then it is 'listened to' above all else first. Obviously, we can talk about different levels of primacy. Compare this with the idea of computer programs. Computers may have core programs that have primacy over other programs. These programs need to be executed and obeyed before other programs. This idea is encapsulated perfectly by the fictional concept of Asimov's three laws of robotics. In these fictional scenarios, robots have to obey their master's commands, but if such commands would endanger human life, they get ignored, because the three laws have primacy. This is an intuitive example that perfectly highlights the idea of primacy.

Programs with primacy, like these core programs, are ones that, if in place, restrict what other programs can operate. In the same way, I assert, part of what it is for mathematical constraint to hold is for the constrainer to have primacy. If it is the case that 23 is indivisible by 3, then, given its primacy, this will restrict what things can happen in the world. A useful way of viewing constraint in general is as a kind of programming for the world. The various constrainers operating on the world restrict the kinds of things that can happen/be true at the world. This is the sense of fundamentality that 'primacy' is aiming at. Constrainers are more fundamental because they program how things can be. I argue that this is the sense in which **Strawberries** involve a link with fundamentality. However, I claim that this sense of primacy is in place in **River** as well. Given that the tree has fallen, it has primacy and will prevent the water from being able to flow. Now it is surely true that the fallen tree does not have *as much* primacy as 23 being indivisible by 3 does. But this is only a difference in degree. The principle of energy conservation has primacy over the laws of nature, in virtue of constraining them. But the principles of supersymmetry arguably have more primacy in virtue of governing the principle of energy conservation. This does not signal that one of these relationships is constraining and one is not. So I argue it should be the same in the cases of mathematical constraint and **River**.

6.6 (In)determinacy

6.6.1 Is constraint deterministic?

In 2.3.3.1 we sidestepped the issue that grounding is deterministic whereas causation is probabilistic/indeterministic. It is now time to address this. This issue is not a significant one for my unification of constraint as a kind of causation. This is because both physical constraint and mathematical constraint are deterministic. If energy conservation holds then force laws must obey it, the probability for non-energy conserving laws is 0. Similarly, if 23 is not divisible by 3 then groups of 23 objects cannot be evenly divided (without cutting) into 3 groups, the probability of groups actually being divided in this way is 0. Likewise, if the tree falls in the place it does then the river cannot flow down distributary **B**, the probability of water flowing down **B** is 0. **River** seems deterministic and seems causal, so we should not dismiss **Strawberries** as being non-causal because it is deterministic. Instead, we should be interested in the fact that we seem to have found genuinely deterministic causation.

The idea that there might be probabilistic/indeterministic constrainer is a point worth exploring because it could bring up some latent issues in the account that represent a disanalogy. One suggestion of a probabilistic/indeterministic constrainer is a hypothetical scenario in the case of spin and entangled electrons. For our purposes, we can think of the property of spin as the direction the electron would deflect if fired at a magnetic field. It seems that the electron in question should be able to take on any value from a continuous range, but in fact they only ever take on one of two values, 'up' or 'down' (or any opposite directions, dependent on the original angle of the field) and they will do so in a 50/50 ratio. A series of experiments known as the Stern-Gerlach experiments show the unusual behaviour of electrons and the spin property (Sklar, 1992). Interestingly, if a beam of electrons is fired at such a horizontal field and so split, then one of these beams, say the one consisting of 'up' electrons is fired at a vertical field, they will split 'left' and 'right'. If these left and right beams are recombined and then fired at yet another field, horizontal again, then one would expect them to all fire 'up', but instead the ratio is once again 50/50 'up' and 'down' (Sklar, 1992, 169-171). What is interesting is that the ratio is always either 100/0 or 50/50. It seems that there is a constrainer operating in such situations. Let us imagine a more indeterminate version of such a law in a scenario, **Electrons**. In **Electrons**, the 'ratio law' stipulates that the ratio must always be either 100/0, or must not exceed 70/30 in either direction. This is purportedly probabilistic because it does not stipulate what the ratio is, merely that it cannot exceed a certain limit, so what the ratio turns out to be is indeterministic. Plainly **Electrons** also seems to be a case of constraint as I have so far described it. Putting aside issues of whether such a law is physically possible (broadly understood), this is certainly a logically consistent law so it is worth taking seriously. What I want to say about such cases is that they are in fact *not* indeterministic. I argue that the confusion that they are arises from the fact that things can be deterministic but have indetermin*ate* effects.

6.6.2 Indeterministic vs. Indeterminate

I want to deny that **Electrons** would be indeterministic by drawing on a distinction between indetermin*ate* and indetermin*istic*. In the case discussed, the effect of the constrainer on the world, provided it is within the specified ratio limits, can be anything and the constrainer does not influence this; in this sense it is indetermin*ate*. Smoking/lung cancer cases are indetermin*istic* because the cause can occur without the effect occurring. Indetermin*istic* cases are cases in which it is true to say that ***φ*** causes ***ψ***, even though there are cases when ***φ*** occurs without ***ψ*** occurring. Indetermin*ate* cases are cases in which ***φ*** causes ***ψ*** but what exact form ***ψ*** takes is either not defined or a range. To highlight this point, let us assume, for simplicity, that when smoking causes lung cancer it is always the 'same' cancer that is caused, i.e. the same size of tumour, in the same location. It is indetermin*istic* whether or not in each case smoking would produce lung cancer, but as and when it did, the effect would be determin*ate*, it would produce the same thing in all cases. It is determin*istic* that when we throw a rock in a pool, it is going to produce a ripple, but the exact form of the ripple is indetermin*ate*.

To bring this back to **Electrons**, let us say the constrainer is something like "fermionic matter which exhibits spin cannot exceed a distribution of 70/30". The effect of this constrainer is that electrons cannot exceed this distribution. This is an aspect of the situation which is determin*istic*. If that constrainer is in place, then electrons will not exceed that distribution. The actual distribution they do take, however, will be indetermin*ate*, because it could take any value under that threshold. To put it in other terms, if ***φ*** constrains ***ψ***, there will be no circumstances in which ***φ*** occurs/is the case/holds and in which ***ψ*** does not occur/is not the case/fails to hold. Although the exact form of ***ψ*** need not be defined and may be a range of values.

Having made this case, it would only serve to strengthen the constraint-causation analogy if we could find a case of mathematical constraint exhibiting these same characteristics. Happily, we can, and indeed it is one already considered. **Strawberries** is likewise determin*istic* but indetermin*ate*. The fact that 23 is indivisible by 3 constrains it so that 23 strawberries cannot be divided between 3 people. If it is the case that 23 is indivisible by 3, it is the case that 23 strawberries cannot be divided into 3 groups. This is a determin*istic* connection. What is indetermin*ate* is that Jane can divide her strawberries in any way she sees fit, provided it is not equally between her children. ***φ*** constrains it to be the case that ***ψ***, the actual strawberry distribution (the content of ***ψ***) can take a range of values. Noting the indetermin*istic*/indertermin*ate* distinction and this case should be sufficient to show that constraint is a determin*istic* relation, although it can have indetermin*ate* aspects. Part of what makes constraint what it is, is that it is deterministic in this sense spelt out above. If constraint is indeterministic, it fails to *be* constraint. Compare this with the idea of a law of nature, or a physical constant. If a law of nature was only obeyed some of the time (even most) then that is not actually a law, or is misdescribed/missing details etc. A physical constant would not be a constant if it changed value. A constrainer is only a constrainer if it always constrains, otherwise, it is a coincidence.

6.7 Iterated constraint

Mirroring 2.3.10, we might wonder whether there is a problem of iterated constraint, i.e. what makes it the case that constrainers constrain. If we offer a constraint as the answer to this, we can ask the same question again and so a problematic regress emerges. One possibility to respond to this is to take the Sider (2020)/Kovacs (2022) route and find an appropriate answer in each case, as in 2.3.10. Alternatively, we might deny that there is a regress. We might stipulate that it is part of the nature of constrainers that they constraint, they are deterministic[81]. One might think of constraint as a kind of computer programming for the universe. There is no question about why computers obey their programming and function in the ways they are dictated because that is to misunderstand what the programming is. In the same way, to ask why the universe obeys the laws of nature is to misunderstand what the laws of nature are; they are constraints on how the universe can behave, so it is not a fact in need of explanation that the universe behaves in line with them.

Constraint relationships are distinguished from coincidences by the fact that they are deterministic, because to postulate this reduces our commitment to brute facts (Lange, 2017, 49-58). One does not need to appeal to a meta-constraint, because it is simply part of how constrainers work that they are ‘obeyed’ in this way. If we have reached explanatory bedrock, there is simply no more to say about why the constraints are the way they are. We have yet to reach this bedrock so it might seem that we can always ask questions about these meta-constraints but that is simply because we have yet to reach the bedrock. Indeed, as Lange points out (2017, 49-58), we ought to treat energy conservation as a constrainer, but energy conservation is itself a product of a higher level constrainer, the principles of supersymmetry. But perhaps it will turn out that these higher level constraint relations are simply brute facts, this need not pose a problem for the constraint view. Indeed, given that

[81]As discussed in 6.6.

there is a problem of iterated causation, a parallel problem for constraint merely strengthens the analogy between them.

6.8 Semantics of impossible worlds

It is worth saying something about the semantic account needed to back up *Exotic Realism*. As I mentioned in 5.1, there are three ways to approach the counterpossibles problem. One can start by altering the ontology, i.e. commit to impossible worlds. One can also simply start by changing the semantics in order to deliver the correct result. As a final option, one might adopt a non-classical logic[82], perhaps paraconsistent, in order to avoid issues with explosion. My primary interest is in altering our ontology as the main project, whilst maintaining much of our standard semantics. However, some more of the semantic picture needs to spelt out. For example, how we might assess the closeness of impossible as well as possible worlds. This section will discuss some options available which intuitively gel well with the aims of my account, specifically with the nature of exploratory processes. There is much to say on these issues but I hope that this section will suffice to show that much can be said in favour of my account.

One might worry that in appealing to impossibility, we will lose all sense of closeness between worlds. We might be able to judge how far away a world in which we were a foot taller than we in fact are by an intuitive estimate of *how different* that situation is to the actual one. The concern is that with impossible scenarios, we cannot gauge how far away impossible worlds are from our own. They might in fact seem *infinitely* far away. We should temper such intuitions however. Whilst it might be strange to think about impossibilities, we can still make sense of relative similarity between our world and impossible worlds. For example, as Nolan (1997, 544) points out:

---

[82]This option will not be discussed in more detail, but it is worth saying something about it here. Priest (2016, 2657) takes this option and explains his thoughts as follows: “If we consider counterfactuals such as ‘If intuitionist logic were correct...’, the context requires us to consider worlds in which intuitionist logic is true; so we need to extend the realm of possibility to those worlds that are intuitionistically possible.”. For Priest, if we want to make statements about what would have happened if intuitionistic logic were correct, we should do so from the perspective of intuitionistic logic. This option is available to anyone unhappy with a change in the semantics, although note that this seems to have the same consequence as Baron et al’s (2017) method for chasing out contradictions (see 5.7.3 and footnote 66). If we switch logic we might no longer be dealing with impossibilities and counterpossibles at all, but rather something that is possible under a different logic. We might term this as ‘differently possible’ rather than impossible.

> "some impossible worlds are more similar, in relevant respects, to our actual world than others. The "explosion" world - the impossible world where every proposition is true - is very dissimilar from our own. Indeed it seems to be one of the most absurd situations conceivable. On the other hand, the world which is otherwise exactly like ours, except that Hobbes succeeded in his ambition in squaring the circle (but kept it a secret), is far less dissimilar."

Or consider a further example. An impossible world in which water is not H_2O but all else remains standard is plainly closer than a world at which water is not H_2O *and* diamond is not made from carbon atoms.[83] Even if it becomes trickier to think about, relative similarity between the actual world and impossible worlds can still be assessed. As a restriction on this, Nolan employs the 'Strangeness of Impossibility condition' (SIC). As Nolan (1997, 550) puts it "any possible world is more similar (nearer) to the actual world than any impossible world.". After stating the principle, Nolan goes on to offer intuitive support for it (1997, 550). For example, if I am conjecturing about a scenario in which I lack legs, the natural conclusion to draw is that I would not be able to walk. Hence the counterfactual "had I lacked legs, I wouldn't have been able to walk". But, presumably, one could conceive of an impossible world in which I have legs, despite lacking them. For the purposes of assessing the truth of the above counterfactual, we plainly ignore such worlds. Nolan takes this to show that we all tacitly commit to something like SIC. Indeed, SIC is often taken up as an important element of semantics by those who commit to impossible worlds.

Brogaard & Salerno (2013) offer a variation of Nolan's account of impossible worlds and make some suggestions on how to further modify the closeness component. Furthermore, they say that this method does not need to abandon classical logic, nor does it need to move away from a Lewisian similarity semantics (2013, 640). For Brogaard & Salerno (2013, 654-655), closeness of worlds is a partially epistemic matter, depending on what is a priori implied in a given context. Brogaard & Salerno modify a notion of a priori from Chalmers & Jackson (2001). For Chalmers & Jackson, "a priori entailment is just an a priori material conditional" (2001, 316). That is to say, if one can know the material conditional 'P→Q' without experience, then Q is an a priori entailment of P (2001, 316). Brogaard & Salerno differ in their usage in that, contra Chalmers & Jackson it is not the case that all logical and mathematical truths are a priori (2013, 654). Were they to be, then, assuming that 'ex falso quodlibet is valid' is a logical truth, 'Paraconsistent logic is correct' would a priori imply 'ex

[83]See 5.4.2.1 from an explanation of why this later situation would be an impossible one.

falso quodlibet is valid'. Since non-vacuists are committed to such an inference being *false* (e.g. see footnote 70 in 5.8), Brogaard & Salerno make this modification. This might sound too hasty on the part of Brogaard and Salerno but some informal justification can be offered. Depending on one's philosophical/logical leanings, one may be persuaded to think that paraconsistent logic is not only a good logical system but the one true logic. Clearly for such an agent, it cannot be a priori that ex falso quodlibet is valid, because this directly contradicts commitments they have. This is perhaps why Brogaard & Salerno (2013, 654) are keen to say that what is a priori is incredibly context sensitive and will vary from agent to agent and situation to situation. Given their departure from Chalmers & Jackson, they call refer to their notion as 'a priori*' and define it as follows:

"For a speaker s in a context c, P a priori* implies Q iff for s in c, Q is a relevant a priori consequence of P." (2013, 655).

Using this notion of a priori* implication, Brogaard & Salerno go on to define closeness with impossible worlds as follows:

> "For any two impossible worlds w_1 and w_2, w_1 is closer to the base world than w_2 iff
> (a) w_1 does not contain a greater number of sentences formally inconsistent with the relevant background facts (held fixed in the context) than w_2 does.
> And if w_1 and w_2 contain the same number of sentences formally inconsistent with the relevant background facts (held fixed in the context):
> (b) w_1 preserves a greater number of a priori* implications between sentences than w_2 does."
> (2013, 655)

Consider an agent to whom it is a priori* that H_2O is not a monkey, but to whom it is not a priori* that water is H_2O, for whatever reason. Now consider two worlds, w_1 and w_2. At w_1, the following sentences are true: "water is not H_2O" and "H_2O is a monkey". At w_2, the following sentences are true: "water is not H_2O" and "water is XYZ" (Brogaard & Salerno, 2013). Both these worlds contain the same number of impossibilities, but our intuition should be that, even if water were not H_2O, it would not be a monkey, and this account ratifies that result because w_2 will be closer than w_1. In this case, w_2 preserves a greater number of the a priori* implications between sentences that w_1.

Likewise, with this account of the semantics, the results I want will be ratified. Recall the following pair of counterpossibles from 5.4.1.2.

K: "If diamond had not been covalently bonded, then it would have been a better electrical conductor." (Tan, 2019, 40).

K* - If diamond had not been covalently bonded, then it would have been no better at conducting electricity.

Consider a world, w_k at which the following sentences are true: "diamond is not covalently bonded", "covalent bonds prevent electrical conductivity". Consider also w_{k*} at which the following sentences are true "diamond is not covalently bonded" and "covalent bonds do not prevent electrical conductivity". If it it a priori* for an agent that "covalent bonds prevent electrical conductivity", then w_k will turn out to be the closer world, and so **K** will turn out to be true over **K***. Consider also a mathematical case. To those convinced by Lange's account of constraint, it may become a priori* that "23 being indivisible by 3 prevents Jane from dividing her 23 strawberries amongst her 3 children"[84]. On the basis of this, consider two worlds, w_s and w_{s*}. At w_s, the following sentences are true "23 is divisible by 3" and "*J*". At w_{s*}, the following sentences are true "23 is divisible by 3" and "¬*J*". Given this, the counterpossible "Had 23 been divisible by 3, then Jane would have been able to evenly divide her strawberries between her 3 children" will come out as true because w_s is closer than w_{s*}, which is as I and the non-vacuists want. The converse counterpossible "Had 23 been divisible by 3, then Jane would not have been able to evenly divide her strawberries between her 3 children" will be false exactly because w_{s*} is further away.

Brogaard & Salerno's account works well with my distinction between investigative and exploratory processes. In investigative processes, we hold everything we know fixed, so all the impossible worlds a priori* conflict with what we know. But in an exploratory context, we dispense with some intuitions/assumptions so then it makes sense to talk of worlds which conflict with fewer of the a priori* intuitions that we hold fixed. Of course, this is just the beginning of a consideration of semantics concerning the impossible. Brogaard & Salerno provide a good start to such an account, but others are available. Priest (2016) offers his own

[84]I will abbreviate this sentence as '*J*' for brevity.

account of impossible worlds, as do Berto & Jago (2019). As long as such accounts are available, my own account of conceiving of impossible worlds is able to appeal to them.

6.9 Section Summary

With the discussion of constraint more generally, and this discussion of its characteristics, we are in a good position. We have seen strong positive reasons to view constraint as causation and responded to potential problems. To summarise this section so far we have considered some potential differences between causation and constraint (and grounding more generally). I believe that these differences are easily explained and in fact, more often than not, do not present a problem for constraint viewed as a kind of causation. The traditional problems with uniting grounding and causation do not apply in the constraint case so there is nothing inherently wrong with uniting the two. Indeed as we have seen in the discussion so far, there are positive reasons why we should do exactly that.

Having considered *Exotic Realism* and the arguments for it through chapters 1-4, we moved on to assess some problems for the account over chapters 5 and 6. Having responded to these issues, we are now in a position to conclude the thesis and note avenues for future research and the directions in which *Exotic Realism* could continue to be developed.

7 - Conclusion

7.1 Summaries

Having considered the arguments presented throughout the thesis, we are finally in a position to conclude. I have argued for the strong claim that constraint is a kind of causal relationship (**Con=G&C**). But for those not persuaded, at least my arguments should have highlighted the closeness of constraint and causation and made the weaker claim that they are simply examples of the same kind of broader relation (**weakCon=G&C**) more plausible. It is worth briefly summarising the chapters before moving on to discuss this conclusion more fully and flag some interesting avenues for future research.

7.1.1 Chapter 1 summary

In chapter 1 I introduced the notion of an exotic object. I argued that many theories already appeal to a notion of in-between objects that are neither abstract nor concrete. I argued that we needed a way to talk about these in-between objects systematically and that is what the exotic objects framework can provide. I briefly discussed some of the objects of physics that we are already committed to, but that seem to fit the criteria for being exotic objects. I discussed the intuition that, if they were to exist, mathematical objects should affect the world in some way, their existence should make a difference. Taking this intuition seriously, I developed the initial account of what sort of exotic object mathematical objects might be most suited to being, i.e. non-spatiotemporal but causal.

7.1.2 Chapter 2 summary

In chapter 2 I began the initial work of arguing that mathematical objects might be causal. I did this by arguing for Wilson's (2018a) **G=MC** claim, that grounding is simply metaphysical causation. By arguing for **G=MC**, I made the weaker claim (**weakG=MC**), that grounding and causation are simply both kinds of a broader class of dependence relations, more plausible. By arguing for such a close similarity between grounding and causation, I made it much easier to discuss and argue for my strong claim that constraint is a crossover case between grounding and causation (**Con=G&C**). We surveyed some potentially problematic differences between grounding and causation and made progress by aiming to dissolve or respond to a lot of these differences. This showed that, although people may be sympathetic to saying that grounding and causation are similar, they are much more similar than normally acknowledged.

7.1.3 Chapter 3 summary

Chapter 3 discussed the kinds of causal theories that are compatible with any analogy between causation and grounding/constraint. I argued that counterfactual theories are the most natural fit, but are not the only option. Nevertheless, given that there seems to be something counterfactual in grounding or constraint, I suggested that a counterfactual account of causation is the most plausibly compatible. Given the involvement of interventions, interventionism is a particularly difficult theory to match with grounding/constraint. Any interventions on grounding/constraint relationships would be metaphysically impossible. To show the strength of my account of constraint, I chose to adopt the interventionist approach. By showing that my theory can adapt to very tricky versions of counterfactual theories, it is made more plausible and shown to have flexibility.

7.1.4 Chapter 4 summary

Chapter 4 focused specifically on constraint in both physical and mathematical cases. I explained what constraint is and how it operates. I gave an intuitive case of physical constraint which seems genuinely constraining and causal in virtue of the very same elements. I showed that it exhibits a particular pattern of counterfactual dependence and how it behaves under interventions. I then considered a case of mathematical constraint, showing that it possesses the same pattern of counterfactual dependence, and behaves in the same way under interventions. In the physical case, this licences us to say that the relationship is causal, and I argued that this conclusion is equally licenced in the mathematical case. Mathematical constraint is causation. I considered some preliminary issues with this account and responded to them. I also considered whether or not constraint explanations were causal explanations. I compared my account of mathematical objects to that of Lange (2017, 2019a) and argued that we would be justified in saying that constraint explanations were in fact causal. I also argued why we should reject Lange's Aristotelian property view of mathematics, arguing that my account is much clearer.

7.1.5 Chapter 5 summary

Chapter 5 turned to one of the significant problems affecting a causal analogy to grounding/constraint; counterpossibles. I argued against the orthodox view that counterpossibles are trivially true. I pointed to examples of the usage of non-vacuous counterpossibles in science, and how this usage could be backed up by distinguishing between two kinds of reasoning processes, investigative and exploratory. I argued that exploratory processes can result in non-vacuous counterpossibles. I applied the investigative/exploratory distinction to the case of countermathematicals, arguing that the situation is parallel. By arguing that counterpossibles are non-vacuous, I showed that it is possible to have the relevant true and false counterpossibles available to back up the counterfactual claims I want to make about grounding/constraint.

7.1.6 Chapter 6 summary

Many issues from previous chapters were sidelined to avoid detracting from the discussion. Chapter 6 dealt with those issues. One problem with **Con=G&C** is that causation seems to be diachronic whilst constraint is synchronic. I argued that constraint is simply a kind of synchronic causation, and that whilst causation is often diachronic, it does not have to be. This is linked to the problem of whether constraint can involve hasteners/delayers. I argued that this was dependent on the circumstances but did not present a disanalogy with causation. Another concern with **Con=G&C** is that this might be thought to violate the completeness of physics/conservation of energy principles. I responded by showing that mathematical constraint does not violate either principle and can instead be accommodated under them. I also spelt out some reasons why we should take constraint to be a contrastive relation just like causation. Finally, constraint seems intimately connected to fundamentality in a way that causation is not and constraint seems to be deterministic whereas causation is indeterministic. I responded by affirming the connection with fundamentality that constraint has, but showing that this is also present in straightforwardly causal cases of physical constraint. I also affirm the deterministic nature of constraint, this turns out to be a key feature of constraint. But this is not a problem as causation can also be deterministic, again I pointed to straightforwardly causal physical constraint cases. By responding to these issues I hope to have made the case for **Con=G&C**, but at the very least I have made **weakCon=G&C** plausible as a weaker alternative.

7.2 Summary of contributions

It is worth discussing some of the main contributions of the thesis before summarising the conclusion.

The exotic objects framework would be a valuable tool for metaphysicians. Philosophers talk about objects a lot, and they talk about different objects differently. We need a framework to be able to make sense of all this talk. The exotic objects framework forms part of a project to give us the tools to do that. The bare work offered in this thesis is a good start but is far from complete. We need to talk about objects and we need to start doing so in a more systematic way, the framework could allow us to do that.

Platonism is an attractive theory to some. But it suffers from some significant issues and is inherently mysterious. *Exotic Realism* is a way of preserving the spirit of platonism but giving mathematical objects some actual role in the world to avoid these problems and this mysteriousness. I have suggested one way of doing that but other options are available. Exotic objects present a particularly useful tool for mathematical realists to strengthen their theories.

I have also made contributions to the counterpossibles debate. I have aimed to explain not only *that* counterpossibles are non-vacuous, but also *why*. I did this by distinguishing the investigative from the exploratory process when assessing counterfactuals. This is an important distinction for this debate and I believe it needs to be taken seriously.

7.3 Exotic Realism

With the conclusions of chapters 1-6 in mind, we can give a full statement of *Exotic Realism*. Mathematical objects are exotic objects in virtue of the fact that they are non-spatiotemporal but constraining. I have endorsed **Con=G&C** so this constraint is viewed as a kind of causation. In that way, mathematical objects are causal. This causation can be described in terms of counterfactual dependence as explicated by the non-vacuous counterpossibles we invoke.

At this stage, it is worth flagging some interesting avenues for future research that have emerged from the discussion throughout the thesis. Firstly, I will discuss the choice between **Con=G&C** and **weakCon=G&C** that I alluded to in the introduction. After that, I will discuss a structuralist modification of *Exotic Realism* and the work that needs to be done in that area. Finally, I will turn back to the counterpossibles debate in 7.4.3 and discuss the additional research that needs to be done to settle matters further.

7.4 Future research

7.4.1 A weaker claim

In 0.1 I mentioned that, although this thesis would be arguing for the strong claim that constraint is a kind of causation (**Con=G&C**), a weaker alternative is available. The weaker option (**weakCon=G&C**) is to say that, rather than being a kind of causation, constraint is merely *very like* causation. It is worth exploring that option here as a potential development of *Exotic Realism*.

One may be concerned that mathematical objects are the archetypal acausal object, or simply not be convinced by the arguments I have offered here. If this is so, one should nonetheless note the striking similarities between constraint and causation that have been discussed, particularly across chapters 4 and 6. We tend to treat causation as an important dependence relation. Causation is seen as so important that people often endorse an eleatic criterion (e.g. Armstrong, 1978, 132) to say that only causal things exist. Given the similarity between causation and constraint, we might think that constraint should inherit some of that importance. We might switch out the eleatic criterion for a directed dependence relation criterion (DDR) and say that only causal or constraining things exist (more dependence relations could be added if needs be). We could then view mathematical objects as constraining (but not causal) things. This shift from causation would be liable to appease people with these concerns. It would keep causation 'pure' in some sense and maintain the intuition that mathematical objects are not causal.

In 1.2 I classified mathematical objects as a specific kind of exotic object, but if constraint is not causation, then mathematical objects would be a different kind of exotic object. Given the similarities to causation, it seems that we should at least say that constraint is a form of impingement. This would result in a new classification of exotic objects as below:

Objects	Spatial	Temporal	Causal	Constraining
(i.) Concrete objects	1	1	1	1
ii.	1	1	1	0
iii.	1	1	0	1
iv.	1	1	0	0
v.	1	0	1	1
vi.	1	0	1	0
vii.	1	0	0	1
viii.	1	0	0	0
ix.	0	1	1	1
x.	0	1	1	0
xi.	0	1	0	1
xii.	0	1	0	0
xiii.	0	0	1	1
xiv.	0	0	1	0
xv.	0	0	0	1
(xvi.) Abstract objects	0	0	0	0

This vastly expands the kinds of exotic objects. If we endorse **weakCon=G&C** then we will conceive of mathematical objects as being the kind described by row xv, i.e. exotic objects which are non-spatiotemporal and acausal but constraining. Needing to expand this table is not problematic, as noted in 1.3, the initial table was not intended to be complete. Much more work needs to be done on the exotic objects framework, there will be other kinds of impingement and other relevant characteristics. Those resistant to saying that mathematical objects should be considered as causal might be more willing to say that they are nonetheless constraining things. After all, in 6.5.2 I appealed to an analogy of constrainers as a computer program for the universe. One might extend this analogy to mathematical objects, viewing mathematics as a kind of programming language for the universe. It seems plausible to say that treating the constraints of mathematics as things which restrict possibilities is how we already see them.

Although this would be a move away from *Exotic Realism* as I initially proposed it, this is not a problem. In 1.6.1, I spoke of the need to investigate objects neutrally and to be willing to re-conceptualise them if our initial conception proves not to capture all the facts or does so poorly. My idea was to re-conceptualise mathematical objects from being abstract to being non-spatiotemporal but causal exotic objects. If it turns out that the reconceptualisation needs to go further, then that is still in the spirit of my account. What is worth noting though is that the decision between the strong and the weak claim is not settled yet and whether this weaker claim is better is an interesting point for future research.

7.4.2 Structuralism

Another avenue for future research would be a structuralist modification of *Exotic Realism*, *Exotic Structuralism*. I have frequently noted the useful analogy between constraint and computer programming (as in 6.5.2). In a sense, we can view mathematics as a program for the universe. But one way to read this analogy is that it is not the individual mathematical objects that mathematically constrain the world, but the entire program of mathematics. I have been working under the assumption that mathematics is underpinned by mathematical objects, but perhaps the way to look at this is to say that mathematics is underpinned by *structures*.

Shapiro (1997, 72) concisely explains the disagreement between platonism and structuralism. Here we can take platonism as a stand-in for *Exotic Realism*. Platonism says that the numbers

2 and 6 are ontologically independent and so individual objects. Shapiro points out that platonists think that numbers exist necessarily, therefore:

“...we cannot say that 2 could exist without 6, because 6 exists of necessity. Nothing exists without 6. To be sure, there is an epistemic independence among the numbers in the sense that a child can learn much about the number 2 while knowing next to nothing about 6 (but having it the other way around does stretch the imagination). This independence is of little interest here, however.”
(Shapiro, 1997, 72).

For Shapiro, this points to a *lack* of ontological dependence between the natural numbers. Perhaps what the platonist means is that the essences of the numbers 2 and 6 are independent of one another. But this is where the structuralist would disagree. For a structuralist, there is nothing over and above the essence of a number than its position in the number sequence. As Shapiro (1997, 77-78) puts it:

“Individual numbers are analogous to particular offices within an organization. We distinguish the office of vice president, for example, from the person who happens to hold that office in a particular year, and we distinguish the white king's bishop from the piece of marble that happens to play that role on a given chess board. In a different game, the very same piece of marble might play another role, such as that of white queen's bishop or, conceivably, black king's rook. Similarly, we can distinguish an object that plays the role of 2 in an exemplification of the natural-number structure from the number itself. The number is the office, the place in the structure.”

For Shapiro, one cannot even talk of objects without structure:

“It makes no sense to “postulate one real number,” because each number is part of a large structure. It would be like trying to imagine a shortstop independent of an in field, or a piece that plays the role of the black queen's bishop independent of a chess game. Where would it stand? What would its moves be?” (Shapiro, 1997, 76).

All there is to being the number 2, is to be the first prime, the successor of 1 and so on. It is not ontologically independent in the relevant sense. One might naturally wonder if structuralism dispenses with the notion of objects completely. But Shapiro’s brand of structuralism does not make this further claim. Just as the notion of ‘object’ makes no sense without the structure that the object is a part of, the notion of ‘structure’ makes no sense without the objects of which the structures are composed. Shapiro (1997) defines a system as a collection of objects with certain relations. A structure is then the abstract form of a system

(Shapiro, 1997, 73-75). In this sense, a system can be seen as a particular instantiation of the abstract structure. Shapiro's approach leaves open the possibility of the existence of both structures and objects to occupy positions in them. Neither is *reduced* to the other, they are both fundamental. In the particular case of mathematics Shapiro points out that:

> "This office orientation presupposes a background ontology that supplies objects that fill the places of the structures… "In the case of arithmetic, sets—or anything else—will do for the background ontology" (Shapiro, 1997, 82).

For Shapiro, the positions in the offices *are* the objects of mathematics (1997, 83). This kind of picture could work well with my account of mathematical constraint. Recall the distinction between constraint and coincidence discussed in 4.7.4.2. A constraint explanation works not by merely explaining why, in a particular case, a result happened. A constraint explanation works by explaining why, in every case, the particular result happens. One can see the parallel with Shapiro's view of structure vs system. The particular mathematical system explains why various things were constrained to be the case, but the abstract structure explains why this constraint would hold in all cases. In this way, we can see the objects and structure as co-fundamental and both constraining. One cannot object to this by invoking overdetermination because, as Shapiro argues, the notions of object and structure are interwoven. It is also worth pointing out Shapiro's discussion of quasi-concrete objects as defined by Parsons (2009, 37). Shapiro explicitly endorses these objects as being capable of forming parts of structures, indeed he thinks they are ineliminable (1997, 105). On my view, quasi-concrete objects are simply another kind of exotic object. Given that Shapiro's structuralism is compatible with the quasi-concrete, it is compatible with the exotic in general.

An interesting consequence of *Exotic Structuralism* is how such a view affects our view of science. If we point to the mathematical structure as the thing doing the constraining, then we should make a parallel judgement in the case of science. We should point to scientific structures as the things doing the constraining. Such a view seems to be a form of, if not at least in the spirit of, ontic structural realism (**OSR**). McKenzie (2017, 2) characterises the core of **OSR** as the claim that "all we can have in any theoretical confidence is the structural relationships that unobservable objects bear to each other, not the intrinsic nature of those objects themselves". In this sense, what is important to science is the structure of nature itself.

McKenzie (2017, 5) defines a particular form of **OSR** that is analogous to the *Exotic Structuralism* described above:

"Priority‐based structuralisms hold that relations and other structures of physics have a claim to fundamental status. To recap, according to the strong (sometimes called the "radical" position), structures enjoy a one‐way priority over objects; according to the "moderate" position, the so‐called "priority" is reciprocated."

A moderate form of priority-based structuralism (referred to as moderate structural realism [**MSR**])[85] like this would hold that objects and structures both exist. In this respect, it is similar to Shapiro's 'office view' of mathematical structuralism. If we commit to structuralism in the mathematics case, we could commit to it in the scientific case. Perhaps we simply claim that the objects/structures of physics are constraining in much the same way as it is in mathematics.

It is worth noting that French (2010) explicitly discusses an analogy between **MSR** and Shapiro's (1997) structuralism. However, French takes the more 'extreme' elements of structuralism discussed in Shapiro (1997) (that certain conceptions of mathematical structuralism hold that mathematical objects do not exist, only the structures) to be comparable and illustrative of his own 'extreme' version of **OSR** (that there are no objects, only structures). I am instead appealing to the more moderate elements of Shapiro's theory, i.e. that mathematical objects and structures are co-fundamental, to be analogous and illustrative of **MSR**. French (2010, 104-106) believes that **MSR** simply collapses into his eliminative **OSR**, but Esfeld & Lam (2011, 148) argue that not only is it not clear that this is the case, but eliminative **OSR** is more revisionary than it needs to be whilst **MSR** preserves more of our intuitions and maintains the advantages that structuralism in general offers.

All of this discussion serves to point out that the particular brand of structuralism that *Exotic Structuralism* would appeal to is available and *Exotic Structuralism* represents an interesting avenue for future research. It promises to perhaps provide an interesting development that is in keeping with popular metaphysical theories elsewhere, without a loss in ontology.

However, we may not wish to commit to the existence of structures as well if we can make do with just objects. An interesting point in Shapiro is when he says that "we cannot say that 2

[85]A version of this moderate thesis is developed in Esfeld & Lam (2011).

could exist without 6…". For Shapiro, all the natural numbers exist of necessity. However, it is not clear what this 'cannot say' amounts to. One might think that Shapiro means something similar to the Emery & Hill (2017) analysis from 5.5. The sentence "2 could exist without 6" is unassertable because it fails to acknowledge the ontological dependence between 2 and 6 as part of the same system/structure. If this is what Shapiro means, then plainly this begs the question against any object-based view of mathematics because it presupposes the dependence between the numbers that the platonists deny. In chapter 5 I argued that countermathematicals can be non-vacuous. What I focussed on in particular was extramathematical countermathematicals, i.e. countermathematicals with a non-mathematical consequent. I argued that we should treat such countermathematicals as non-vacuous, using the semantic method outlined in 6.8. However, I did not explicitly extend this account over to the case of intramathematical countermathematicals, i.e. those with a mathematical consequent. Indeed, throughout chapter 5 I noted that we might need to say that all the counterpossibles used in mathematical proofs need to be true. Those which are used in mathematical proofs invariably tend to be intramathematical. Much turns here on whether the investigative/exploratory process distinction maps onto the intramathematical counterpossibles. If it does, perhaps we can find a way of having non-vacuous intramathematical counterpossibles.

This represents an avenue for future research. If we can articulate the independence of 2 and 6 non-vacuously, then this would speak in favour of an object-based view of mathematics. If we cannot, structuralism is the way to go. So not only does *Exotic Structuralism* present an interesting area in its own right, but depending on the formulation it has further interesting links with intramathematical countermathematicals.

7.4.3 Counterpossibles and impossible worlds

In chapter 5, I aimed to continue the Jenny (2018) project of providing areas in which counterpossibles seem to be non-vacuous. I also aimed to provide the overall explanation for *why* counterpossibles are non-vacuous. I did this by offering a preliminary distinction between the explanatory and investigative processes. More work needs to be done here though. The explanatory and investigative processes need to be distinguished more thoroughly and this will be an interesting project for future research. My intuition is that investigative counterfactuals are not *actually* counterfactuals, they just appear to be in counterfactual form. What it takes to be a counterfactual is to be involved in an explanatory process. The "counter" prefix is crucial here. In an investigative process, everything else is kept the same, apart from the hypothesis in question, nothing is 'counter' to the fact. In an investigative process, we remain in a consensus context as described by Yli-Vakkuri & Hawthorne (2020, 566). For a statement to be properly assessed as a counterfactual, we need to engage in an exploratory process and I do not think that can be done outside of a non-consensus context. Further research is needed to support and explicate this and this research would constitute a fruitful project.

I also said that to assess counterpossibles through an exploratory process we needed to, on some level, consider impossible worlds. An account of impossible worlds would, therefore, be worthwhile considering. Some options are available. As noted in 5.1, Yagisawa's (2010) account could be a useful starting point. Yagisawa (2010) views worlds as points in modal space. To analyse counterpossibles, one might simply need to assess the closest impossible point in modal space and determine the truth value of the counterpossible there in the usual way. However, one of course does not need to be committed to Yagisawa's account to accept *Exotic Realism*. An ersatz approach to worlds, both possible and impossible, might also prove fruitful. Compare this with Sendlak's (2021) analysis of counterpossibles. Sendlak views counterpossibles as story prefixes, i.e. a counterpossibles "Had *x* been the case then *y*." is a paraphrase of "According to the story of *x*, *y*.". We might take this seriously. Perhaps impossible worlds are just descriptions of the world which are incorrect in some way significant enough to make them impossible descriptions. We can make non-vacuous statements of this form precisely because different stories have different truths within them, regardless of their possibility. Perhaps we do not need to appeal to impossible worlds as anything over and above these descriptions. We might be able to make judgments on the closeness of these stories to the actual description of the world and so be able to say which counterpossibles are true or false. Recall from 5.7.3 that Baron et al (2017) suggest a process of 'chasing out' contradictions from countermathematical scenarios. One way to look at this might be that we want to make sure the stories we tell are as close to the actual world as possible.

Regardless of whether one wishes to endorse realism or ersatzism about worlds, there will be a corresponding theory which is extendable to impossible worlds. That is all my account requires. Investigating the best fit for *Exotic Realism* is interesting but falls beyond the scope of this thesis.

To summarise, in this thesis I have argued for *Exotic Realism*, the idea that mathematical objects should be conceived of as exotic objects as they can play a causal role via constraining the world, whilst being non-spatiotemporal. The main contribution of this thesis lies in developing and expanding on this new conception of mathematical objects. I have argued for a strong claim, **Con=G&C**, and given an object-based conception of mathematics, but as alluded to in 7.4.1, there are various avenues of future research to change this conception in response to issues. Nevertheless, I have argued for the strong conception of *Exotic Realism* presented here as a viable alternative to mathematical platonism; one that also serves as a good counter to nominalism.

Bibliography:

Armstrong, D. 1978. Nominalism & Realism: Universals and Scientific Realism Volume I. Cambridge: Cambridge University Press.

Baker, A. (2003). Does the Existence of Mathematical Objects Make a Difference?. Australasian Journal of Philosophy. 81(2). p. 246-264

Baker, A. (2005). Are there Genuine Mathematical Explanations of Physical Phenomena?. Mind. 114(454). p. 223–238.

Balaguer, M. (1998). Platonism and Anti-Platonism in Mathematics. Oxford: Oxford University Press.

Baron, S; Colyvan, M; & Ripley, D. (2017). How Mathematics Can Make a Difference. Philosophers' Imprint, 17(3), 1-19.

Baron, S, Colyvan, M, Ripley, D. (2019). A Counterfactual Approach to Explanation in Mathematics. Philosophia Mathematica. 28(1). pp 1-34.

Benacerraf, P. (1973). Mathematical Truth. The Journal of Philosophy. 70(19). p. 661-679.

Bennett, K. (2017). Making Things Up. Oxford: Oxford University Press.

Bernstein, S. (2016). Grounding Is Not Causation. Philosophical Perspectives. 30(1). p. 21-38.

Berto, F & Jago, M. (2019). Impossible Worlds. Oxford: Oxford University Press.

Berto, F; French, R; Priest, G; Ripley, D. (2018). Williamson on Counterpossibles. Journal of Philosophical Logic, 47(4). 693-713.

Born, M. (1954). The Statistical interpretation of quantum mechanics. Nobel Lecture. December 11th.

Brogaard, B & Salerno, J. (2013). Remarks on counterpossibles. Synthese, 190(4), 639-660.

Bueno, O. 2002. Is it possible to nominalise quantum mechanics? Philosophy of Science. 70(5). p. 1424-1436

Camilleri, K. (2007). Indeterminacy and the Limits of Classical Concepts: The Transformation of Heisenberg's Thought. Perspectives on Science. 15(2). p. 178-201.

Chalmers, D. J. (1996). The Conscious Mind: In Search of a Fundamental Theory. Oxford: Oxford University Press.

Chalmers, D. & Jackson, F. (2001). Conceptual analysis and reductive explanation. Philosophical Review. 110(3). p. 315-360.

Colyvan, M. (1998). Can the Eleatic Principle be Justified?. Canadian Journal of Philosophy, Vol. 28(3). p. 313–336.

Colyvan, M. (2001). The Indispensability of Mathematics. New York: Oxford University Press.

Colyvan, M. (2010). There is No Easy Road to Nominalism. Mind. 119(474). p. 285-306.

Colyvan, M. (2018). The Ins and Outs of Mathematical Explanation. The Mathematical Intelligencer. 40(4). p. 26-29.

Coquand, T. (2022). Type Theory. The Stanford Encyclopedia of Philosophy. (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/type-theory/>.

Dainton, B. (2021). The Silence of Physics. Erkenntnis. Available at: https://doi.org/10.1007/s10670-021-00450-5

Davies, J. (2019). On the Varieties of Abstract Objects. Australasian Journal of Philosophy. Vol. 97(4). p. 809-823.

Dickson, S. (2021). Mathematical Causation. Analiza. (1). p. 85-102.

Dickson, S. (2022). Against Vacuism. Studia Semitoyczne (Semiotic Studies). 35(2). p. 7-33.

Dretske, F. (1988). Explaining Behavior. Cambridge: MIT Press.

Egg, M. (2021). Quantum ontology without speculation. European Journal for Philosophy of Science. 11(1). p. 1-26.

Emery, N., & Hill, C. H. (2017). Impossible worlds and metaphysical explanation: Comments on Kment’s modality and explanatory reasoning. Analysis, 77(1), 134–148.

Esfeld, M & Lam, V. (2011). Ontic structural realism as a metaphysics of objects. IN Scientific Structuralism. Bokulich, A & Bokulich, P (eds.). p. 143-159. London: Springer.

Field, H. (1989). Realism, Mathematics and Modality, Oxford: Basil Blackwell.

Field, H. 2016. Science without numbers. 2nd Ed. Oxford: Oxford University Press.

Fine, K. (2012). Guide to Ground. IN, Correia, F & Schnieder, B (eds.). Metaphysical Grounding: Understanding the Structure of Reality. P. 37-80. Cambridge: Cambridge University Press.

Forrest, P. (1982). Occam's Razor and Possible worlds. The Monist. Vol. 65(4). p. 456-464.

French, S. (2010). The interdependence of structure, objects and dependence. Synthese. 175. p. 89-109.

Friedell, D. (2020). Abstracta Are Causal. Philosophia. 48(1). p. 133-142.

Gibb, S. (2010). Closure Principles and the Laws of Conservation of Energy and Momentum. Dialectica. 64(3). p. 363-384

Grice, H. P. (1975). Logic and conversation. In: Cole, P & Morgan, J (Eds.), Syntax and semantics, 3: Speech acts (pp. 41–58). New York: Academic Press

Hall, E. J. & Paul, L. A. (2013). Causation: A User's Guide. Oxford: OUP.

Hall, N. (2004). Two concepts of causation, IN J. Collins, N. Hall & L.A. Paul (eds.), Causation and Counterfactuals,. Cambridge: MIT Press.

Heisenberg, W. (1926). Mehrkörperproblem und Resonanz in der Quantenmechanik. Zeitschrift für Physik. 38. p. 411-426.

Heisenberg, W. (1952). Philosophic problems of nuclear science: Eight lectures by Werner Heisenberg. London: Faber & Faber.

Hofmann, F & Schulte, P. 2014. The Structuring Causes of Behaviour: Has Dretske Saved Mental Causation? Acta Analytica. 29. p. 267–284

Hume, D. (1978). A Treatise of Human Nature. edited by L.A. Selby-Bigge, 3rd ed. revised by P.H. Nidditch. Oxford: Clarendon Press.

Hume, David. (1995) An Enquiry Concerning Human Understanding. Upper Saddle River: Prentice Hall.

Jackson, F & Pettit, P. (1990). Program Explanation: A General Perspective. Analysis. [Online]. 50(2). pp-107-117. [Date Accessed: 24th May 2017]. Available at: http://www.jstor.org/stable/3328853

Jenkins, C. (2011). Is Metaphysical Dependence Irreflexive?. The Monist. 94. p. 267–76.

Jenny, M. (2018). Counterpossibles in Science: The Case of Relative Computability. Noûs, 52(3), 530-560.

Kovacs, D. M. (2022). The Question of Iterated Causation. Philosophy and Phenomenological Research. 104(2). p. 454-473.

Lange, M. (2017). Because Without Cause: Non-Causal Explanations in Science and Mathematics. New York: OUP.

Lange, M. (2019a). What could mathematics be for it to function in distinctively mathematical scientific explanations?. Mathematics and its applications: Philosophical Issues. 23-24th September. University of Leeds.

Lange, M. (2019b). Ground and Explanation in Mathematics. Philosopher's Imprint. 19(33). p. 1-18

Lange, M. (2021) What could mathematics be for it to function in distinctively mathematical scientific explanations?. Studies in History and Philosophy of Science. (87). pp44-53

Leng, M. (2010). Mathematics and Reality. Oxford: OUP.

Lewis, D. (1973). Counterfactuals. Hoboken: Blackwell.

Lewis, D. (1986a). Philosophical Papers, vol. 2. Oxford: Oxford University Press.

Lewis, D. (1986b). Postscripts to 'Causation'. IN Lewis (1986a), p. 172-213.

Lewis, D. (1986c). On the Plurality of Worlds. Oxford: Blackwell.

Liggins, D. (2010). Epistemological Objections to Platonism. Philosophy Compass. 5(1). p. 67-77.

Linsky, B & Zalta, E. N. (1996) In defense of the contingently nonconcrete. Philosophical Studies. 84(2-3). p. 283-294.

Lipton, P. (2009). Understanding Without Explanation. IN Scientific Understanding: Philosophical Perspectives. Eds. De Regt, W, Leonelli, S & Eigner, K. pp43-63. Pittsburgh: University of Pittsburgh Press.

Litland, J. (2017). Grounding ground. Oxford Studies in Metaphysics. 10. p. 279– 316

Lowe, E, J. (1998). The Possibility of Metaphysics. Oxford: Oxford University Press

Lowe, E.J. (1996). Subjects of Experience. Cambridge: Cambridge University Press.

Lowe, E.J. (2000). Causal Closure Principles and Emergentism. Philosophy. 75(294). p. 571-585.

Lyon, A & Colyvan, M. (2008). The Explanatory Power of Phase Spaces. Philosophia Mathematica. 16(2). pp227-243.

Maddy, P. (1990). Realism in Mathematics, Oxford University Press.

McKenzie, K. (2017). Ontic Structural Realism. Philosophy Compass. 12(4). p. 1-11.

McLoone, B. (2020). Calculus and Counterpossibles in Science. Synthese, 198(12), 12153-12174).

Menzies, P. (1989). Probabilistic Causation and Causal Processes: A Critique of Lewis. Philosophy of Science. 56(4). p. 642–663.

Nolan, D. (1997). Impossible Worlds: A Modest Approach. Notre Dame Journal of Formal Logic. 38(4). p. 535-572.

Noordhof, P. (2020). A Variety of Causes. Oxford: OUP.

Oppenheimer, J. R. (2014). Atom and Void: Essays on Science and Community. Princeton: Princeton University Press.

Papineau, D. 2000, 'The Rise of Physicalism', in: M. W. F. Stone and J. Wolff, eds, The Proper Ambition of Science, New York: Routledge, pp. 174-208.

Papineau, D. 2002, Thinking about Consciousness, Oxford: Oxford Uni

Parsons, C. 2009. Mathematical Thought and its Objects. Cambridge: Cambridge University Press.

Priest, G. (1997). Sylvan's Box: A Short Story and Ten Morals. Notre Dame Journal of Formal Logic. 38(4). p. 573-582.

Priest, G. (2016). Thinking the impossible. Philosophical Studies. 173. p. 2649–2662.

Putnam H. (1969) Is Logic Empirical?. In: Cohen R.S., Wartofsky M.W. (eds) Boston Studies in the Philosophy of Science. Boston Studies in the Philosophy of Science, vol 5. Springer: Dordrecht. https://doi.org/10.1007/978-94-010-3381-7_5

Reutlinger, A; Colyvan, M; and Krzyżanowska, K. (2020). The Prospects for a Monist Theory of Non-causal Explanation in Science and Mathematics. Erkenntnis. Available online at: <https://link.springer.com/article/10.1007/s10670-020-00273-w>

Robinson, E, H. (2014). A Documentary Theory of States and Their Existence as Quasi-Abstract Entities. Geopolitics. 19(3). p. 461-489.

Robinson, H. 2020. Dualism. The Stanford Encyclopedia of Philosophy. (Fall 2020 Edition). Zalta, E, N. (ed.). Available at: <https://plato.stanford.edu/archives/fall2020/entries/dualism/>.

Saatsi, J. (2011) The Enhanced Indispensability Argument: Representational versus Explanatory Role of Mathematics in Science. The British Journal for the Philosophy of Science. 62(1). p. 143-154.

Schaffer, J. (2000). Causation by Disconnection. Philosophy of Science. 67(2). p. 285-300.

Schaffer, J. (2005). Contrastive Causation. Philosophical Review. 114(3). pp327-58

Schaffer, J. (2010). Monism: The Priority of the Whole. The Philosophical Review. 119(1). p. 31–76.

Schaffer, J. (2012). Grounding, Transitivity and Contrastivity, IN Grounding and Explanation, eds. Correia and Schnieder. p. 122-138. Cambridge: Cambridge University Press.

Schaffer, J. (2016). Grounding in the image of causation. Philosophical Studies. 173 (1). p. 49-100.

Sendlak, M. (2021). Counterpossibles, story prefix and trivialism. Synthese. 199. p. 7283-7301. Available online at: <https://doi.org/10.1007/s11229-021-03114-7>

Shapiro, S. (1997). Philosophy of mathematics: Structure and ontology. Oxford: OUP.

Sider, T. (2020). Ground grounded. Philosophical Studies. 177. p. 747–767

Sklar, L. (1992). Philosophy of Physics. Oxford: Oxford University Press.

Skow, B. (2016). Reasons why. Oxford: OUP

Skow, B. (2018). Causation, Explanation, and the Metaphysics of Aspect. Oxford: OUP

Stalnaker, R. C. (1968). A Theory of Conditionals. In W. Harper, R. Stalnaker, & G. Pearce (Eds.), (pp. 41–55). Oxford: Basil Blackwell.

Tan, P. (2019). Counterpossible Non-vacuity in Scientific Practice. Journal of Philosophy, 116(1), 32-60.

Thompson, N. (2016). "Metaphysical Interdependence", in M. Jago (ed.), Reality Making. Oxford: Oxford University Press. pp. 38–56.

Tiehen, J. (2015). Explaining Causal Closure. Philosophical Studies. 172(9). p. 2405-2425.

Varzi, A. (2015). Boundary. The Stanford Encyclopedia of Philosophy (Winter 2015 Edition). Zalta, E. N. (ed.). Available at: https://plato.stanford.edu/archives/win2015/entries/boundary/.

Williamson, T. (2007). The philosophy of philosophy. Oxford: Blackwell.

Williamson, T. (2013). Modal Logic as Metaphysics. Oxford: Oxford University Press.

Williamson, T. (2018). Counterpossibles. Topoi, 37(3), 357-368.

Williamson, T. (2020). Suppose and Tell: The Semantics and Heuristics of Conditionals. Oxford: OUP.

Wilson, A. (2018a). Metaphysical Causation. Noûs. 52(4). p. 723-751.

Wilson, A. (2018b). Grounding Entails Counterpossible Non-triviality. Philosophy and Phenomenological Research. 96(3), pp716-728.

Wilson, A. (2021). Counterpossible Reasoning in Physics. Philosophy of Science, 88(5). p. 1113-1124.

Woodward, J. (2003). Making Things Happen: A Theory of Causal Explanation. Oxford: Oxford University Press.

Yagisawa, T. (2010). Worlds and Individuals, Possible and Otherwise. Oxford: OUP.

Yates, D. (2009). Emergence, Downwards Causation and the Completeness of Physics. The Philosophical Quarterly. 59(234). p. 110-131.

Yli-Vakkuri, J & Hawthorne, J. (2020). The Necessity of Mathematics. Noûs. 54(3). p. 549–577.