MANIPULATIONISM AND CAUSAL EXCLUSION

Mark Pexton

ABSTRACT

A new way of avoiding the causal exclusion argument in the context of manipulationism is proposed. In manipulationism, causal explanations are defined by counterfactual information accessed through manipulations. It is argued that the property of manipulability can be an emergent property of aggregate systems. Therefore, some causal explanations are non-reducible and causal exclusion is avoided. This emergentist notion of causal explanation addresses the question of how the special sciences can be based upon causal reasoning, even if fundamental physics is absent of causal relations.

1. Introduction

The causal exclusion argument due to Kim (1999), is a powerful anti-emergentist argument. Originally developed as an argument against mental causation, it is also applicable to any system that is realised by a microphysical supervenience base. Recently, there have been several attempts to avoid the exclusion argument from within the framework of a particular account of causal explanation: Woodward's manipulationism (see Woodward 2011, List and Menzies 2009, Shapiro 2010, Raatikainen 2010).

In this paper, I propose a new manipulationist strategy for avoiding exclusion. In this, we do not have causal overdetermination, since only the higher-level explanation is a causal explanation, whilst the microphysical explanation is non-causal. This is because for such explanations, changes at the microphysical level cannot be interpreted as manipulations but macro-level changes can be. This proposal is motivated by contrasting the microphysical and macrophysical explanations in astrophysics of white dwarf stability. It is argued that the property of manipulability is related to a confluence of other well-defined physical properties. Those properties can be emergent at the macro-level from the interactions of other, non-manipulatable, microphysical properties. The causal-explanatory relation itself then, if characterised by manipulability, can be an emergent phenomenon.

The structure of this paper will be as follows. Firstly, we will briefly review the exclusion argument and the manipulationist framework of causal explanation. We will then look at an application of manipulability to the problem of exclusion by Raatikainen. We will then look at a case study of non-causal explanation (orbital stability) to establish what counts as a non-causal explanation in manipulationism. Finally, we will look at both the macrophysical and microphysical explanations of white dwarf stability, arguing that only the macrophysical explanation is causal

in manipulationist terms. Thinking of causal explanation as an emergent feature can potentially reconcile the widespread use of causal explanation in special sciences with a Russellian view of fundamental physics as non-causal.

1.1 Kim's causal exclusion argument

Kim's downward causation argument is sketched in figure 1. Let's imagine we have a non-fundamental level L at which an event M occurs at time t_1, and this event causes event M^* at time t_2. Let M^* have reduction base P^* at a lower level L-1, which nomologically necessitates the occurrence of M^* at t_2. For Kim, we now have two causes of M^*: M at t_1 and P^* at t_2. Therefore, to say that M causes M^*, we must also say that M causes P^*. In other words, we can only cause a higher-level property M^* by causing its lower-level base P^*.

L   ·   ·   ·   M(t_1) ⟶ M^*(t_2)

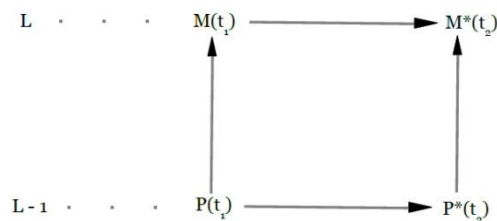L-1   ·   ·   ·   P(t_1) ⟶ P^*(t_2)

Figure 1: The causal exclusion argument

So, M causes M^* and P^*. But M itself has a reduction base at t_1, P, which nomologically necessitates M and therefore P^*. We have causal overdetermination: all causes at the level L work downwards to cause events at level L-1, and the events of L-1 nomologically necessitate all the higher-level events. Kim's conclusion is that reduction is established. There are no higher level causal relations that are not translated into

relations between lower level events. We have causal overdetermination: two apparently competing causal stories for one set of events. Hence since we can reduce the higher-level causal story to the lower level story we exclude the higher-level causal explanation.

1.2   Woodward's manipulationist framework

Woodward's manipulationist framework (Woodward 2003) can be summarised as follows:

- Manipulationism defines causal explanation as a matter of providing modal/counterfactual information (**M**) in order to answer *What-if-things-had-been-different?* questions (w-questions)
- Any given system is partitioned into input and output variables, which are assigned a binary numeric value: either 0 or 1. (In Woodward's scheme the variables can be continuous, but the subsequent discussion will be simplified by using just two values.)
- **M** is extracted by imagining performing *hypothetical* interventions, **I**, on system input variables to change their value from 0 to 1, in order to see the effect this has on the output variable value
- **I** are themselves causal, so manipulationism is not a reductive account of causation. Interventions are causal interactions with a system used to identify other causal, rather than correlational, connections between variables in that system
- **I** are not limited to just the interventions humans can actually perform. Instead, **I** are a wider hypothetical class which are continuous in some sense with actually performable interventions

This last point is important to stress. For example, we cannot intervene in practice to substantially change the orbit of Halley's Comet, but we could conceive of the hypothetical outcome of such an intervention. So, in manipulationism changing the position of Halley's Comet is an allowed intervention. Often manipulations are

categorised as changes which are *logically physically possible,* not merely actually possible physical changes. But this should *not* be confused with manipulations being any *logically possible* change. Woodward is clear that interventions themselves are *physical causal* changes:

> The notion of an intervention is itself a causal notion–among other things, it involves the idea of an intervention variable I that causes a change on [a variable] C. (Woodward, 2003 p22)

So, not every change in a variable is interpretable as an intervention.

In manipulationism, a given system is partitioned into input and output variables. The input variable is changed and the resultant change in the output variable is calculated. For example, imagine Billy throws a rock and breaks a window. We have the input variable Billy (B) which takes values B = 1 if he throws a rock, or B = 0 if he doesn't. The variable B is altered hypothetically from 1 to 0 to see effect on the variable Window (W): W = 1 when the window breaks, and 0 when it doesn't. So, the counterfactual (or modal) structure of our explanation is: {B = 1 & W = 1} or {B = 0 & W = 0}. Since a change in B results in a change in W we therefore identify that Billy caused the window to break.

I suggest a fruitful way of graphically representing **M** structure is in terms of the possibility space of variables. Here we will define the colour black as corresponding to a variable having value 1, and the colour white to the variable having value zero. This is shown in figure 2. Of course, many real variables are continuous and not discrete in this way, but for a given explanandum we can impose a cut off on a continuous variable, so that above the threshold the variable value is black, and below the threshold it is white.

Take our toy example of Billy breaking a window. The direction of the thrown rock and its momentum are each continuous variables but we can discretise them by saying that the variable B = 1 (black) means "Billy

throws a rock with enough momentum so that the rock is capable of breaking a window along a vector that intersects with the window". Similarly the B = 0 (white) means "Billy doesn't throw a rock.", but it also means "Billy throws a rock vertically in the air.", "Billy throws a rock without enough momentum to break the window." etc. These are all folded into the general variable value 0. So "Billy doesn't throw a rock", is a placeholder term for: Billy either does not throw a rock, or throws it in such a way it isn't capable in principle of breaking the window. Hence, we have bisected a continuous possibility space into two discrete zones.
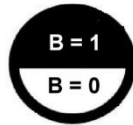

Figure 2: we discretise the possibility space of any variable into black for B = 1 or white for B = 0. When we have a circle with both colours, this represents our input variable.

We can then graph the effect each value of our input variable has on the possibility space of the output variable. At the centre of our diagram is the input variable which is split into two, one half black, the other half white, representing the change from 0 to 1. Each hemisphere, black or white, is then connected to the allowed possible values of the output variable given that input value. This 'modal graph' (**MG**) is shown in figure 3 for the rock/window example. The **MG** displays the ways a possible state of the input variable connects to the possible states of the output variable (for that value of input).
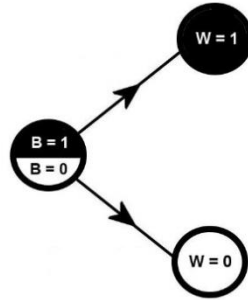
Figure 3: Example of a modal graph (__MG__). Possibility space for Billy breaking the window. The central circle is the variable B, the outer circles represent the variable W.

If B = 0 (white) (Billy doesn't throw the rock) then the possibility space of the window is constrained to W = 0 (white = the window does not break). (This assumes the usual "all other things being equal qualifications"). If B = 1 (black) then this possibility limits the possibility space of the window to W = 1 (black), it breaks. In such a way we have mapped the modal structure of the system (white connects to white + black connects to black) and can answer w-questions, such as: Q: If Billy had not thrown the rock would the window have broken? A: No. Having now established the basics of manipulationism, in the next section we will look at one particular manipulationist response to causal exclusion, due to Raatikainen.

2. Raatikainen's argument

Raatikainen's argument (2010) against causal exclusion relies on the idea that manipulationist causal explanations are always implicitly contextualised by contrast classes. In order to make unambiguous what it means for a variable to change, we must pick a contrast. So, for

example, Raatikainen says of a causal explanation involving variables X and Y:

> If, for example, X could take as its value either x1 or x2, and Y either y1 or y2, the relevant causal claim, with contrasts made explicit, could be:
> X's being x1 (rather than x2) causes Y's being y1 (rather than y2). Note that different choices of contrasts, say x3 and y3, for the same x1 and y1, for example, lead to different causal claims, some of which may be false, some true. The most natural 'default' contrast is though...that the presence rather than the absence of the property (or whatever) at issue is caused by the presence of another appropriate property (or whatever) rather than the absence of it. (Raatikainen, 2010, 7)

Let us call Raatikainen's default contrast class the negation class: *p/not-p*. By using the negation contrast class, a mental state can be identified as the cause of a physical event, whilst the brain state associated with that mental state can fail to be the cause of that same event. Raatikainen uses an example to illustrate this:

- *John has a desire (he wants a beer)*
- *John has a mental state (he forms a belief that there is a beer in the refrigerator)*
- *Hence a physical event results (John goes to the refrigerator to get a beer)*

Now John's mental state can be intervened upon. John's flatmate, Pete, tells him that there is no beer left in the refrigerator, hence John's mental state changes. A different physical event results: John goes to the shop instead of the refrigerator for beer. As with our broken window example earlier, we can construct a modal diagram of this situation. Let B be the mental state of John, such that $B = 1$ means John believes there is beer in the refrigerator, and $B = 0$ means he thinks there isn't. Let $L = 1$ mean John

goes to the fridge, and L = 0 mean John goes to the shop. The modal graph of this situation is given in figure 4 and is the same shape as figure 3.

Figure 4: Modal graph (dropping the labels). Independent variable is at the centre which changes from white to black (0 to 1). The value of the dependent variable is on the branches. In this case John's mental state is the variable at the centre, which changes from 0 to 1 (white to black). As a result the physical state changes from white to black, meaning that counterfactually John's mental state is a causal difference maker for whether John goes to the refrigerator or the shop for beer.

Let's say that John's mental state B corresponds to a brain state b. So now b = 1 means John has the brain state associated with B, and b = 0 means John does not have that brain state. It is straightforward to see that an intervention to change b from 1 to 0 does not change L from 1 to 0. This is because John's mental state is multiply realisable by different brain states. The contrast, between having the *particular brain state* John had when he thought there was beer in the refrigerator, and not having that *particular brain state*, does not pick out a change in John's actions. For instance, we could change John's brain state to b = 0 by changing the position of one molecule in his brain, and his mental state would be unaffected. We can see this in the modal graph shown in figure 5. No matter what the value of b, this variable still connects to both possible values of L. Hence there is no difference made by changing b and no causal dependence established.
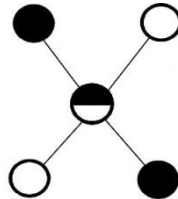
Figure 5: John's brain state (b) is the variable at the centre, which changes from 0 to 1 (white to black) in the contrast class: has b/doesn't have b. But this change does not screen off the possibility space of John's actions. Counterfactually, for this contrast class, John's brain state is not a causal difference maker for whether John goes to the refrigerator or the shop for beer.

So, for the negation contrast class (p/not-p) there is a well-defined difference made by John's mental state but there need not be a difference made by John's brain state. This is because for each mental state there are many different brain states. Since an intervention on the mental makes a difference to behaviour, but an intervention on the brain state does not, we have no possibility of causal exclusion. The only causal relationship established (for the negation contrast class) is the mental one.

Raatikainen's argument is not limited to mental states. Any properties studied by the special sciences which are multiply realised will fail to be excluded causally within manipulationism by this 'default' negation contrast class of p/not-p. So for example, consider a thermodynamic explanation of how a piston is driven by hot steam. There will be a variable corresponding to some well-defined bulk property of the steam which we can change, say gas pressure. This variable change will result in the piston not moving. But at the level of the molecular motions that make up the steam there are many different changes which will not alter the movement of the piston. In the contrast of a particular concrete

microphysical distribution, d, and its negation, not-d, the change from d to not-d does not have to bring about a change in the piston movement.

Raatikainen's argument clearly captures something correct about the way manipulationism identifies causal relations. But we should be cautious about a casual use of contrast classes. Indeed, as Raatikainen himself admits, the *choice* of contrast class greatly affects whether a causal relation is established by interventions. This means that there will be contrast classes for which John's brain state is a causal difference maker. Now as far as refuting causal exclusion is concerned this in and of itself may not be a problem. We are free to stipulate that to causally exclude a variable we must be using equivalent contrast classes at the higher and lower level. So what is really at issue is whether the negation contrast class is the contrast class we should be using at both the higher and lower levels, and I suggest that it isn't.

The first thing to note is that in manipulationism, in order to test a causal dependence, one must perform a special type of intervention: *a testing intervention.* A testing intervention is one in which the intervention is strong enough to possibly change an outcome. The reason for testing interventions is to rule out trivial interventions. For instance, imagine someone claims that flipping a particular switch turns a light bulb on. We can test this claim by flicking the switch to see if the bulb comes on or not, and by doing so can map modal relations, and (in manipulationist terms) establish a causal explanatory relation. But we must depress the switch enough that it is possible for a circuit to be made. If for instance, we merely brush the switch with a feather this is an intervention, but it is *not a testing intervention.* The feather cannot depress the switch enough to *test* the relation between the switch/bulb systems.

So, imagine now we have a variable P: P = 1 corresponds to switch not depressed, and P = 0 corresponds to the negation of this (switch depressed). Now this change in P from 1 to 0 doesn't pick out a change in our light bulb, since the negation (switch depressed) is multiply realisable. The switch could be depressed a tiny amount (as by the

feather) or by enough to make it click and test the circuit. Similarly changing one molecule in John's brain is not a testing intervention: *it is not a change that we would expect could ever track a change in his behaviour*. Nothing much rests on this trivial example of a bulb and switch, but the notion of testing interventions in Woodward's scheme should immediately make us suspicious of the idea that the appropriate contrast in manipulationism is merely the negation of a property.

The negation contrast is not the standard contrast of manipulationism. Rather manipulationism always involves either an explicit, or implicit, modelling step. In this step, the parameter space of a variable is bisected. We almost never simply use the negation of a variable as the appropriate contrast class. Consider our rock window example from figure 4. The different values of the variable B do not correspond to rock thrown/rock not thrown. Instead we have implicitly modelled the system and bisected the parameter space by clumping together many different possibilities. For example, B=1 means that the rock is thrown towards the window with enough momentum to reach the window. In other words, contrast, even for a simple example like a thrown rock, is a highly disjunctive set of circumstances which are abstracted to appear as a simple bisected contrast of thrown or not-thrown. This abstraction is implicit in all manipulationist claims and is resultant from the modelling necessary to link causal relata by well-defined interventions. In many cases this modelling step is so trivial it is not made explicit.

To split a parameter space, we always abstract across disjunctive states to create binary contrasts: the rock must be thrown towards the window, the rock must be thrown with momentum above a certain value X, etc. This is done for both John's mental state, and in principle could be done for John's brain state. We can group together (in principle) all the disjunctive multiple brain states that connect to the belief state "beer in refrigerator", and all those that connect to the belief state "no beer in

the refrigerator". Then our parameter space is split such that a change in b *does* counterfactually link to a change in L.

Note that even the simplest default contrast of John's mental states also involves abstracting across disjunctive parameter regions. For example, for a change from B = 1 to B = 0 to change L in the right way, B = 0 *cannot* include the belief state "the beer is on top of the refrigerator". If it does then changing B from 1 to 0 will *not* make John go to the shop. Note that the default contrast does not automatically rule this out. So Raatikainen's simple default contrast class is illusionary. All contrast classes implicitly collect together disjunctive elements to bisect the parameter space. There is no *principled* problem with multiple realisability, since there is no principled difference between how the contrast classes are defined for higher and lower states relative to the bisection of parameter spaces.

Now this is not to say that there isn't a *pragmatic* difference between the contrast classes of the mental and brain parameters. It is much more difficult to abstract across the disjunctive parameter space of the brain states. It may be so difficult a task that neurologists are never able to do it. For the purposes of prediction and control, the correct bisection of brain states may not be possible. But this pragmatic difficulty merely identifies the usual utility of *epistemic* emergence. Causal exclusionists do not deny the practical utility of higher-level causal relations as epistemic constructs. Nor do they claim that one must abandon using such relations for prediction. Rather, causal exclusion is a metaphysical claim about what is necessary for the world to be a certain way. It is not a claim about how it is useful for us to represent the world. There is no principled distinction between multiply realised higher level manipulationist contrast classes, and the contrast classes of their lower level supervenience bases. This means that the only difference that thinking about contrast classes can identify is the pragmatic epistemic difficulty of tracking the modal relations at the lower level. But this is not enough to avoid the important challenge of causal exclusion, since it is already

well established by the *practice* of science that we cannot *pragmatically* reduce all causal relations to microphysical relations.

The reductionist hypothesis is not a constructivist one in which special sciences are reduced because every calculation should start with, say, quantum field theory. All sides of the debate accept a pragmatic failure of reduction; what is at stake is a principled failure of reduction. All Raatikainen's argument does is establish the epistemic utility of manipulating mental states for control, it does not establish that brain states cannot be the ultimate seat of the relevant modal relations, or indeed that the modal relations captured (in principle) by thinking about brain states are a wider set than those captured by thinking about mental states.

3. Failure of reduction of manipulability

In this section, we will examine another way in which manipulationism may be able to avoid causal exclusion. It is possible for some higher-level determination relations (accessed by causal manipulations) to reduce to lower level determination relations (which cannot be accessed by manipulations). In order to justify this claim we will look closely at two case studies. Firstly, a dimensional explanation of planetary orbital stability will provide an exemplar of an explanation which is non-causal (within the manipulationist framework). Secondly, we will examine both the macro-level and microphysical explanations of white dwarf stability. It will be argued that only the macroscopic explanation is causal by manipulationist standards. It will then be argued that the white dwarf case is indicative of a wider puzzle: how can causal special science theories be related to non-causal theories of fundamental physics in light of the exclusion argument?

3.1   *Non-causal explanation: Spacetime dimensionality and two-body stability*

Consider a classic two-body problem in physics where we have one large stationary body and another small body that is in motion and interacts with it at a distance *r*. The two bodies might represent different masses interacting gravitationally, such as the Sun and Earth; or the bodies might be charges interacting electrostatically, such as a positive nucleus and a negatively charged electron. Let the number of space dimensions in the universe in which our two bodies are interacting be *n* and the number of time dimensions be *m*, so a universe with spacetime such as ours is described by: $(n, m) = (3,1)$.

Both electrostatics and Newtonian gravitation can be described canonically by Poisson's equation $\nabla^2 \phi = \rho$, which relates a potential ϕ (electrostatic/gravitational) to a source ρ (charge/mass). Poisson's equation allows one to derive how this potential varies with distance. For a point particle the potential is given by r^{2-n} for $n > 2$. The force law for such system is related to the potential by taking the gradient. Hence, the force felt by the small test body from the large stationary body is proportional to r^{1-n}. In our universe $n = 3$ so these forces go as the inverse square of distance. But the inverse square laws of Newtonian gravitation and Coulomb electrostatics become inverse *cube* laws if $n = 4$. Ehrenfest was the first to notice that that neither classical atoms, nor planetary orbits, can be stable with *n* > 3 (Ehrenfest, 1917). Traditional quantum atoms cannot be stable either (see Tangherlini 1963). When $n > 3$ the two-body problem no longer has any stable solutions. This case is illustrated in figure 6 below.
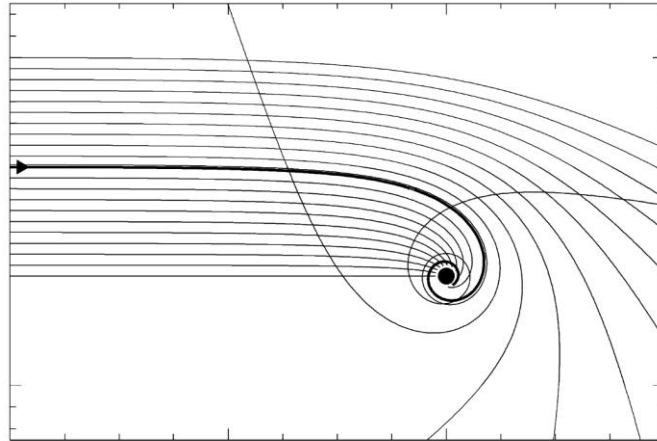
Figure 6: (after Tegmark 1997). The two-body problem in four-dimensional space: light-weight particles approach from the left towards the heavy body at the centre. They either escape to infinity or get captured and collide with the central body; there are no stable orbits.

In figure 6 we have an array of test particles (each with the same momentum) moving from the left towards a massive point particle represented by the black dot. The test particles either fly away to infinity, or they spiral inwards towards the central particle and annihilate. This is in contrast to $n = 3$, which gives stable bound elliptic orbits (or non-bound parabolic and hyperbolic orbits). Therefore, our 3-dimensional universe has no annihilation solutions, except for head on collisions.

A similar situation occurs in quantum mechanics. A system governed by the Schrödinger equation, such as the hydrogen atom, has no bound states for $n > 3$. There is also an equivalent result for gravitation in General Relativity (Tangherlini 1963, also see Tegmark 1997 for a thorough discussion of problems such as these in the context of string theory where both n and m can vary, and what bounds this places on the possibility of having observers in a universe.)

A case like this seems to be a form of non-causal explanation following this pattern:

Explanandum: why are planetary orbits stable?
Explanans: $n \leq 3$

Note that the explanans does not specify sufficient conditions for planetary orbits, there may be many ways in which planetary orbits do not form, e.g. we may not even have planets! But the explanans does allow us to answer w-questions. For example, Q: How many planets would the solar system have if $n > 3$? A: None; and so on.

This allows us to identify an *asymmetrical* explanatory dependence: the number of space dimensions helps to explain how we come to have a stable solar system, yet the stability of the solar system does not explain how we came to have 3 spatial dimensions.[1] Furthermore we do not access the modal information contained in the explanans by virtue of I (Woodward's causal interventions). We do not perform I on n to see how things change, since we cannot *causally* change the number of dimensions space has. Indeed, this is Woodward's own interpretation of this case:

> Does the dimensionality of space-time explain why the planetary orbits are stable? On the one hand, this suggestion fits well with the idea that explanations provide answers to what-if-things-had-been-different questions on one natural interpretation: we may think of the derivation as telling us what would happen if space-time were five-dimensional, and so on. On the other hand, it seems implausible to interpret such derivations as telling us what will happen under interventions on the dimensionality of space time. (Woodward, 2003, 220)

---

[1] This ignores anthropic explanations for simplicity. Anthropic explanations, if they are explanations, are of a very different character than ordinary explanations since they do not characterise physical dependence relations at all.

> [T]he common element in many forms of explanation, both causal and noncausal, is that they must answer what-if-things-had-been-different questions. When a theory or derivation answers a what-if-things-had-been-different question but we cannot interpret this as an answer to a question about what would happen under an intervention, we may have a noncausal explanation of some sort. This accords with intuition: it seems clear the dependence of stability on dimensionality...is not any sort of causal dependence. (Woodward, 2003, 221)

We have an explanation because we are able to map the counterfactual effects of changes to n. But a change of n cannot be viewed as a manipulation, since such a change would be non-local. (Later on we shall try and be more precise about what physically constitutes a manipulation.) There are some points raised by this case worth making explicit.

Firstly, one might object that this case is not explanatory at all. As stated, the number of spatial dimensions is a necessary but not sufficient condition for stable orbits. Lots of other (causal) processes are required to get planetary orbits (such as having matter, planets forming etc.). So how is a necessary background condition explanatory of anything?

The answer to this lies in the contrastive aspect of manipulationism. Whenever we give an explanation in manipulationism, we track a pattern of modal relations. That pattern does not have to be fully exhaustive of all possible relations. That is, in any explanation (causal or non-causal) there will be many difference makers, but we will usually pick out a small subset of those for our explanans. This selection process is defined relative to a contrast class. Usually we compare two different systems which are the same in all aspects except for some change in a potential difference maker.

For any given event, the full exhaustive list of difference makers will be enormous and no scientific explanation ever cites them all. For example, consider the asteroid impact explanation of the extinction of

the dinosaurs. An exhaustive version of that explanation would have to cite the big bang and all the processes that lead up to the formation of asteroids, planet Earth, dinosaurs etc. The full explanation, citing all difference makers, might be an example of Railton's ideal explanatory text (Railton 1981). Railton suggests that in science all explanations are partial in nature. Each explanation science gives is merely a part of a larger, never realised, ideal-explanatory-text which would be the exhaustive explanation of a phenomenon.

Clearly no biologist needs to give such an ideal text explanation, even if they could in principle. When we give an explanation for a phenomenon, our interests pick out some difference makers rather than others. So, in the asteroid explanation we are implicitly contrasting an Earth with dinosaurs and an asteroid impact, to an Earth with dinosaurs without an asteroid impact. We shuffle into background conditions all other difference makers to dinosaur extinction, (such as the processes by which the asteroid was formed). By doing this we highlight one particular difference maker because it's salient to the contrast class implicitly defined by our explanandum. This highlighting of a difference maker does not mean that other factors are not difference makers though. Let us call the difference maker picked out by a particular contrast class the *contrastive difference maker.*

In our spacetime dimensionality example, of course it is true that there are many other difference makers to stable planetary orbits in addition to the number of dimensions. But this does not mean that the number of dimensions is not explanatory. In the contrast class of two worlds where everything else, apart from dimensionality, is kept the same, the contrastive difference maker is the number of dimensions. Just as all other explanations do not need to cite the whole ideal text of every difference maker to count as an explanation, so too our dimensional explanation does not have to state all other difference makers to be a genuine explanation. To give an explanation we must highlight at least one difference maker, but we are not obliged to highlight all the other

difference makers, relative to a particular contrast class. (If we were obliged then probably nothing in science would count as a genuine explanation.)

A second possible objection to this spatial dimensionality explanation might be as follows: *even if this case is explanatory, it is actually just a causal explanation.* This is because there are lots of causal processes (interactions of planets with gravitational fields etc.) that make the explanans true. These difference makers can be manipulated, hence are causal. Again though, one must keep the contrastive nature of explanation in mind. For our particular contrast class, the contrastive difference maker (space dimensionality) is one for which no manipulation is possible. All the causal, manipulatable, difference makers are excluded by the contrast class which fixes all the other variables.

There will be accounts of explanation (such as Skow 2014) for which this dimensional explanation is still causal. Note that, in this paper, we are only concerned with what is a causal explanation *in manipulationist terms.* Skow's account defines causal explanations as those explanations that cite information relevant to the causal histories of particular events. This is not the manipulationist criterion for causal explanation though. In manipulationism, to be causal the process by which a variable is changed must be interpretable as a manipulation. The dimensionality of space is not interpretable as a manipulatable variable, hence counterfactual changes to it give us non-causal explanations.

We can summarise the lessons of this case study as follows:

- Explanations involve contrast classes which pick out a particular contrastive difference maker from the ideal exhaustive set of all difference makers
- If a change in our contrastive difference maker can be interpreted as an intervention then (in manipulationist terms) we have a causal explanation

- If a change in our contrastive difference maker cannot be interpreted as an intervention then we have a non-causal explanation. (Regardless of whether other difference makers are causal or not.)

In the next section we will apply this notion of a contrastive difference maker to another case study: white dwarf stability. I will argue that, when we compare the macrophysical and microphysical contrastive difference makers, we get causal explanations at one level; yet non-causal explanations at the other.

3.2 White dwarf physics

White dwarf stars are stellar remnants left over when a star has exhausted all of its nuclear fuel. The star forms a planetary nebula (which is nothing to do with planets) and sheds its outer layers while the core, usually made of carbon and oxygen, contracts to form a white dwarf. Since white dwarf stars are not hot enough to continue fusing elements together, they have no source of energy generation. Normal stars are in a state of balance known as hydrostatic equilibrium in which a thermal pressure gradient counterbalances the inward pull of the star's self-gravity. In white dwarf stars this is not possible. As the star contracts, it becomes extremely dense (approximately 1 tonne of white dwarf material would barely fill a matchbox). The remaining core is highly ionised, and is made of heavy nuclei and free electrons. The star reaches such densities that it becomes electron degenerate, meaning that all available lower energy states (below the top one) are filled by electrons. This degeneracy results in a pressure (the degeneracy pressure) which stabilises the white dwarf from further collapse.

The Indian astrophysicist Subrahmanyan Chandrasekhar (1957) was the first to theoretically predict the relativistic upper limit on the mass

of a white dwarf star.[2] He calculated that for masses above about 1.4 times the mass of the Sun ($1.4M_\odot$) degeneracy pressure is not strong enough to halt stellar collapse. Subsequently it was discovered that for stars above this Chandrasekhar limit, but below about $3M_\odot$, collapse continues until neutron degeneracy pressure counterbalances gravity. The difference in the two cases is that above the Chandrasekhar limit the collapsing star has enough gravity to produce a process called electron capture, whereby electrons are "absorbed" by protons, producing neutrons (the resulting neutron star is essentially a giant nucleus in some respects). If the collapsing star has a mass above $3M_\odot$ then even the neutron degeneracy pressure cannot balance gravity and a black hole is formed.

So we have three different regimes. White dwarf stars form when the original collapsing star has a mass below $1.4M_\odot$. Neutron stars form when the star has a mass between $1.4M_\odot$ and $3M_\odot$. Finally, a black hole is formed if the star's mass is above $3M_\odot$. If a white dwarf accretes enough matter after it has formed to breach the Chandrasekhar limit, say from a binary partner, then it tends not to form a neutron star, instead it explodes as a type Ia supernova, the violence of the explosion blowing the star apart. By contrast, if a neutron star accretes enough matter to tip it above the limit of neutron degeneracy pressure, then it will form a black hole.

It is important to stress that this degeneracy pressure results from the restriction on energy level filling due to the exclusion principle; it is not due to collisions between electrons. In the kinetic theory of ordinary gases, pressure is an averaging of the bumpings of atoms against a surface, but this is not the case here. Degeneracy pressure is present in all matter. In ordinary situations it is minuscule in comparison with thermal pressure, but in white dwarf stars the lack of thermal pressure

---

[2] Chandrasekhar did this at just 20 as an effectively self-taught physicist, his calculation brought much scorn and ridicule form the eminent astrophysicist Arthur Eddington; ultimately though Chandrasekhar was shown to be correct.

and extreme densities mean that it is large enough to halt the collapse and make the star stable.

The microphysical explanation of this stability is rooted in the Pauli Exclusion Principle (PEP). PEP is named after Wolfgang Pauli who proposed it in 1925. It states that no two identical particles with half integer spins, "fermions", can occupy the same quantum state. Electrons are fermions so PEP applies to them. PEP is a consequence of the indistinguishability of quantum particles. Quantum systems are described by the Schrödinger equation:

$$H\Psi(x,t) = i\hbar \frac{\partial \Psi(x,t)}{\partial t}$$

Where H is the Hamiltonian, an operator related to the energy of a system, and Ψ(x,t) is the wavefunction whose square modulus gives a probability that the particle is in a given state. Since the wavefunction is a function of time, the probability of finding a particle at a particular point changes with time. In classical mechanics two identical particles can be distinguished because we can follow their separate trajectories, but when we have two identical particles in quantum mechanics we cannot label them as individuals in this way. We cannot distinctly follow the path of each particle, as the particles have no distinct position between measurements. Rather, when a measurement occurs at some time we can only say that one of the particles has been found at one point and the other at another, but we cannot say *which* of them is which in each given measurement. The wavefunction must remain valid regardless of whichever particle has been measured at a point. If one particle is described by Ψ_a and the other by Ψ_b then there are two possible ways of combining the wavefunctions to describe a system with two particles.

$$\Psi(r_1, r_2) = A[\Psi_a(r_1)\Psi_b(r_2) \pm \Psi_b(r_1)\Psi_a(r_2)]$$

Where r_1 and r_2 are the possible positions of our particles. Fermions have antisymmetric wavefunctions, that is:

$$\Psi(r_1, r_2) = A[\Psi_a(r_1)\Psi_b(r_2) - \Psi_b(r_1)\Psi_a(r_2)]$$

If both particles are in the same quantum state, $a = b$, so:

$$\Psi(r_1, r_2) = A[\Psi_a(r_1)\Psi_a(r_2) - \Psi_a(r_1)\Psi_a(r_2)] = 0$$

So the wavefunction disappears, which is a physically uninterpretable solution. From the indistinguishability of quantum particles, and the requirement of an antisymmetric wavefunction, we arrive at the exclusion principle: *no two fermions can be in the same quantum state.* In practice, since we can have a spin up or a spin down state, this means only two electrons can occupy a given energy state. Both electrons and neutrons are fermions, so the exclusion principle is very important for understanding both white dwarf and neutron stars.

One can imagine the electrons in the star as being confined to boxes of smaller and smaller size as the star collapses. Quantum mechanics says that particles such as electrons are also waves in certain circumstances (wave-particle duality). The exclusion principle means that the wave of each electron must stay inside the confines of its own box. As the boxes get smaller and smaller the wavelength must get smaller to stay inside. A smaller wavelength means a higher frequency and this implies a higher energy (blue light is a shorter wavelength, and hence more energetic, than red light, so when metal is heated to higher temperatures it will glow red first). As the star continues to collapse, the electrons are forced into higher and higher energy states and this generates a resistance to the collapse known as the electron degeneracy pressure. A full derivation

would take up too much space here, but we can do a very rough estimate of the resulting degeneracy pressure.

Let us say we have n electrons, which do not interact with each other in any way, confined to a volume V. We can fit two electrons into each box (with opposite spins) but no more due to the exclusion principle, so each electron occupies a volume of approximately V/2n, and occupies a lateral space of approximately:

$$\Delta x \sim \left(\frac{V}{2n}\right)^{1/3}$$

The Heisenberg uncertainty relation states that uncertainty in position, Δx, and uncertainty in momentum, Δp, multiplied together cannot be less than a constant, $\hbar$, divided by 2:

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

So:

$$\Delta p \geq \frac{\hbar \Delta x}{2} \sim \frac{\hbar n^{\frac{1}{3}}}{2^{2/3} V^{1/3}}$$

Let m be the mass of the electron. In the low temperature limit, assuming the electrons are not moving near the speed of light, the average kinetic energy is given by:

$$E = \frac{\Delta p^2}{2m} \sim \frac{\hbar^2 n^{2/3}}{2^{7/3} V^{2/3} m}$$

The total internal energy U is then:

$$U = nE \sim \frac{\hbar^2 n^{5/3}}{2^{7/3} V^{2/3} m}$$

In statistical mechanics, pressure, P, is defined as a partial derivative of internal energy with respect to volume under constant entropy. In other words, pressure is a measure of how internal energy changes in response to changes in volume. Hence our pressure due to degeneracy alone is given by:

$$P = -\frac{\partial U}{\partial V} \sim \frac{\hbar^2 n^{5/3}}{2^{4/3} V^{5/3} m}$$

The above derivation is extremely simplified. A rigorous calculation would involve doing quantum statistics and calculating the so called "density of states": the number of electrons with momentum between two infinitesimally near values. The density of states provides a statistical weighing of which energy values are used and allows a calculation of the average momentum flux across a surface, and hence, pressure. In our derivation we have shown that degeneracy pressure scales as the electron density (n/V) raised to the power of 5/3. As the electrons are squeezed into smaller volumes their velocity approaches the speed of light and we have to use a relativistic expression for the relation between energy and momentum. If we do this, we find that the pressure scales as the electron density raised to the power 4/3. This is important: as more and more energy is given to the electrons a shift occurs in how they can partition that energy. At low speeds, as in everyday life, if we add more energy to a moving body we increase its speed. But when we reach the speed of light we cannot keep going faster, and instead the extra energy has to be added to the mass of the electrons,

that is, their inertia. In other words, the way energy is distributed changes the pressure we have: at low speeds pressure increases with density faster than at higher speeds.

Degeneracy pressure does not result from forces acting upon the electrons. They do not push each other away electrostatically for instance, and as such it is a very different sort of pressure from thermodynamic pressure imagined as an average effect of many particle collisions. When we derived our degeneracy pressure expression we were not concerned with electrons colliding off of one another. We might be tempted then to say that degeneracy pressure is a fictitious pseudo-pressure, just a folk tale for physics students. A pressure in name only since it is so different from the collisions described by kinetic theory.

However, Skow argues that such fictionalism about degeneracy pressure is not consistent with how pressure is actually used in other areas of physics (Skow 2014). It is not the case that pressure is straightforwardly the summation and averaging of the forces felt by atoms bouncing off container walls. In statistical mechanics, properties such as pressure or temperature are defined independently of these causal mechanical force interactions and apply quite abstractly to some systems. Temperature is not always merely a measure of mean molecular kinetic energy. For example, the kinetic energy in a paramagnet is small, but the temperature can be very high. This is because temperature is defined quite abstractly as the partial derivative of internal energy with respect to entropy. Similarly, as we have seen, pressure is defined as minus the partial derivative of internal energy with respect to volume (at fixed entropy). In other words, in a system there are many different degrees of freedom that can contribute to these abstract statistical definitions of pressure and temperature.

In white dwarf stars, the exclusion principle means that electrons must occupy higher and higher energy states as the density of the dwarf increases. These electrons contribute to the internal energy of the star and hence the pressure defined in this way. To claim the fictionality of

degeneracy pressure because of a lack of causal mechanisms is erroneous. Just as temperature in a paramagnet is not determined by kinetic energy, so pressure in the unusual degenerate state of matter found in white dwarf stars is not dominated by mechanical "bumpings". Once pressure is seen as this abstractly defined ensemble property, we can say that degeneracy pressure is no more illusionary than any other type of pressure. It merely reflects the capacity of the internal energy of a system to respond to changes in volume.[3] Furthermore, the shift in the dependency of degeneracy pressure on density, from the classical to the relativistic regimes, shows that the way in which internal energy is distributed matters for degeneracy pressure. At relativistic speeds the electrons use energy differently and so the ensemble property of pressure behaves differently.

3.3   Causal and non-causal difference makers

The white dwarf case has been discussed in the philosophical literature before. Lewis (1986) contends that nothing halts the collapse of a white dwarf, rather the star simply has nowhere else to go physically:

> It's not that anything caused [the star] to stop [collapsing]—there was no countervailing pressure...There was nothing to keep it out of a more collapsed state. Rather, there just was no such state for it to go into. The state-space of physical possibilities gave out. (Lewis, 1986, 222)

---

[3] There is an analogy here with pseudo-forces in Newtonian mechanics. Forces such as the centrifugal force are not real, we introduce them to make up for the lack of an inertial reference frame. Yet if we look at forces from the point of view of general relativity, then gravity itself is a pseudo-force, hence what is or is not a "real" force is determined by the abstract framing definitions used to describe a physical system. There is no straightforward way of simply pointing to the real forces of the world.

Yet for Lewis this is still a causal explanation, since the information that the halting of collapse has no cause is itself "causally relevant" information. For Lewis, providing causally relevant information is all an explanation has to do to count as causal. (Skow 2014 has provided a similar argument for the causal characterisation of this case.) Notice that Lewis does not argue that the star is stable because PEP itself provides a cause of the collapsing. Rather, it is that being able to say something is *uncaused* is itself a form of causal explanation, since we have ruled out the possibility of any causal history making a difference.

However, Lewis' characterisation of the state space running out seems incongruent with the astrophysical picture we have described. There are physical possibilities for the system to explore, the white dwarf could have became a neutron star (or the neutron star a black hole) or a supernova, depending upon the time at which it accreted more mass. It is simply not true that the state space of the system runs out. Instead the collapse is halted by the degeneracy pressure. Colyvan (1999) argues along these lines, in my opinion quite correctly stating:

> [Lewis' characterisation] seems rather odd though. The oddness stems from the conjunction of the assertion "the stopping had no causes at all" and the claim that this is a causal explanation. There is only one way to make sense of this, and that is if Lewis really does see this case as analogous to that of ordinary space giving out. This analogy, however, seems entirely inappropriate since, as we have seen, the Pauli Exclusion Principle prevents stars of certain masses from collapsing further; it does not prohibit further collapse, simpliciter. Presumably if a white dwarf had a greater mass at the crucial second red giant stage its collapse would have continued.
>
> The case seems more analogous to a person trying and failing to break a door open by charging it with their body. It is not that physical space has given out; it is just that the person's momentum isn't great enough. In the latter case a causal story of why the door couldn't be broken open can be provided in terms of the door

> providing a resisting force, and it is precisely the lack of such a
> story that makes [this a case] of non-causal explanation. In effect,
> I am denying that Lewis's causal story is a satisfactory explanation
> of the phenomenon, since it fails to give an account of what
> prevents some stars and not others, from collapsing down to more
> compact configurations. The non-causal explanation (i.e.
> appealing to the Pauli Exclusion Principle) has no such
> shortcoming. (Colyvan 1999, 3)

A full examination of the positions of Lewis and Colyvan is beyond the
scope of this paper. Recall though that here we are interested in causal
explanation as defined by manipulationism. As described,
manipulationist explanations always implicitly define contrast classes.
We judge a particular explanation as causal or non-causal depending
upon whether the contrastive difference maker is manipulatable or not.
In the white dwarf case, the microphysical contrastive difference maker
is the PEP, or equivalently the symmetry constraints on the electron
wavefunctions. It is the PEP that counterfactually tells us whether the
white dwarf is stable (within the Chandrasekhar mass limit). If electrons
did not have to have antisymmetric wavefunctions then they would not
be forced into separate energy levels and the star would continue
collapsing. In Lewis' terms, if the state space does run out, it is because
the system is constrained by a non-causal difference maker: PEP.

The PEP is non-causal because we cannot imagine manipulating the
asymmetry of the electron wavefunction. This is analogous to our
planetary stability case discussed in section 3.1. In both cases there are
manipulations possible of many difference makers, *but not of the
contrastive difference maker* defined by our particular explanandum. In
both cases, counterfactual information is what is doing the explaining. In
each case we access that counterfactual information by changing the
variable picked out as our contrastive difference maker (spatial
dimension or asymmetry of electron wavefunctions). By imagining how
those difference makers constrain their respective systems we can

answer w-questions. But crucially, we cannot think of changing the values of those difference makers through interventions. So in conclusion, for certain contrast classes, the microphysical explanation of white dwarf stability is non-causal since it has a non-manipulatable contrastive difference maker.

What happens if we look at the macro-level explanation? At this level of description, the contrastive difference maker is the degeneracy pressure, a statistical property of the ensemble of fermions. Now the degeneracy pressure seems as open to manipulation as ordinary gas pressure. Degeneracy pressure is a function of particle number and volume, and we could imagine manipulating either of these to change degeneracy pressure and observe a change in the properties of the white dwarf. In a sense, with type Ia supernovae, nature does this manipulation for us when white dwarfs accrete matter in a binary system, tipping them over the 1.4 solar mass limit. This leads to a huge increase in degeneracy pressure such that the star explodes violently.

At the microphysical level, our contrast picks out a non-manipulatable variable as the contrastive difference maker (asymmetry of wavefunctions). But at the macro-level the contrast of our explanandum picks out an ensemble variable which is manipulatable: the degeneracy pressure. This is an unusual situation. We have physical supervenience in as much as the degeneracy pressure is ontologically dependent on the antisymmetric properties of fermion wavefunctions. But we do not have causal exclusion. This is because, in manipulationist terms, the macro-level explanation is causal whilst the microphysical explanation is non-causal. Therefore we do not have *causal* overdetermination. This is a different manipulationist strategy for avoiding exclusion than suggested by Woodward and Raatikainen.

3.4   *Interpretation and application*

The white dwarf stability case suggests that for the manipulationist there is another way of avoiding causal exclusion. If we have ontological dependence, but only have manipulability at one level then we do not have causal overdetermination. This approach is in sympathy with previous manipulationist ways of resisting exclusion (Raatikainen, 2010, Woodward 2011, List and Menzies 2009) but is distinct from them. In cases like the white dwarf example, we do not avoid causal exclusion because of multiple realisability in the lower level variables (as suggested by Raatikainen). Nor do we have overdetermination but in an unproblematic way (as suggested by Woodward). Instead we avoid causal overdetermination because we only have a causal explanation at one level.
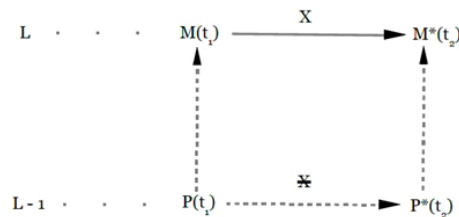

Figure 7: Causal emergence. X is a causal relation between manipulatable variables M and M*. X̶ is a non-causal relationship between non-manipulatable variables P and P*. M may supervene on P but the causal relation X is not excluded by X̶ since we do not have competing causal stories.

In the white dwarf case, the macro-level causal relations are emergent. This is because the property of manipulability is only a property of the ensemble of electrons. If one grants that there can be cases of causal emergence such as white dwarf stability, what is the nature of that emergence? One obvious way of categorising it is as epistemic. Once can say   that   the   'emergence'   of   causation   is   evidence   that   the

manipulationist notion of causal explanation depends on representation. As we move from a microphysical representation to a macro-level representation, we change the contrast class which picks out the contrastive difference maker. It is the shifting of contrast between levels that allows one level to be causally explanatory, while another level isn't.

This would certainly lend itself to a criticism of manipulationism as a theory of causal explanation: manipulation is not a well-defined or objective thing. It is circular with respect to causality, since manipulations are themselves causal interactions used to describe other causal interactions. No wonder that causation defined in these terms fails to reduce, since it is not a thing in the world at all, just a representational construct that is sometimes helpful (and sometimes not). I do not wish to explore this possibility here too much, partly because there is already a large literature devoted to eliminative notions concerning causation. I will therefore acknowledge the epistemic interpretation of causal emergence but leave it alone, as I wish to explore an alternative. That alternative is that manipulability, and hence causal-explanation explicated in manipulationist terms, can be ontically emergent. (Note that I do not claim that the epistemic interpretation is inferior to the ontic interpretation, I merely wish to explore whether an ontic interpretation is intelligible.)

An ontic interpretation then contends that causation/causal-explanation is real and objective. In this case, the fact we can manipulate certain physical systems suggests a certain kind of physical interaction which is the basis of causal-reasoning (in manipulationist terms at least). Being manipulatable is then a capacity, or power, that a system can possess in respect of one of the variables we can define it in terms of. Recall that Woodward's manipulationism does not offer a reductive account of causation. Manipulations are a type of causal interaction used to define other causal relations. As such, defining what precisely counts as a manipulation is as difficult as defining what causation is. I therefore offer only a sketch of some of the kinds of properties that must be part of

the mix of manipulability. I do not claim the suggestions here are exhaustive or in any way constitute a reductive account of the property of manipulability. I do claim that:

- Manipulability (**I**) is a physical property possessed by some variables and not others
- **I** can be a property of an aggregate of entities only. So properties a, b, c in relation R(a,b,c) allow **I** but a, b, and c on their own do not allow **I** (for a particular contrast class)
- To be manipulatable, a property must be local in space-time. A manipulation must be a process that is temporally ordered and takes time for a change to occur; it cannot occur instantaneously or extend backwards in time
- Changes in a manipulatable variable must be isolatable from other systems not directly causally linked to it
- A manipulatable property must be capable of well-defined change; a given determinable can take many different values when represented as a variable

To be manipulatable requires a certain confluence of physical attributes which not all variables of a physical system will possess. For instance the number of dimensions of spacetime in our universe is not a local variable or isolatable in any sense, so we cannot ever imagine manipulating the dimensionality of space. Equally the intrinsic symmetry condition of fermions does not seem like the kind of property which is isolatable from the rest of physical law. One way of thinking about this is that we can have circumstances where a confluence of physical interactions produces a new aggregate variable which is open to causal manipulation.

This emergentist way of viewing causal relations has a potential application. Many philosophers of physics hold a Russelian view that there is no causal explanation in fundamental physics. Russell's view was

that causation was a mere projection onto the world; an "anthropomorphic superstition" (1917 [1913]) based upon the asymmetry of memory. Many since have proposed similar views (see Redhead 1990, Norton 2007, and Ladyman 2008). For example, Norton states:

> Mature sciences...are adequate to account for their realms without need of supplement by causal notions and principles. The latter belong to earlier efforts to understand our natural world or to simplified reformulations of our mature theories, intended to trade precision for intelligibility. (Norton, 2007, 12)

While not everyone subscribes to this non-causal interpretation of fundamental physics (see Shrapnel 2014), it is a "majority view" according to Ladyman (2008). Yet, this way of thinking about physics seems at odds with the way many other sciences use the causal-explanatory relation. Causal explanation plays an active role in special sciences as well as in other areas of physics. Given standard reductionist intuitions though, this use of causation in some sciences, but not in fundamental physics, raises an incongruence: *where does causation come from as we move up the hierarchy of nature?*

This also relates to a form of the concern Ned Block has raised with the causal exclusion argument (Block 2003). Block has argued that the exclusion argument relies on the assumption that there is a fundamental level. But if instead there is no fundamental level then causal exclusion means that, instead of being merely reduced, causation drains away from the world altogether. We can reformulate Block's concerns in a different way. If there is a fundamental level, but that level is non-causal, then causal exclusion implies that there is no real causation in the world. We can avoid this in three ways.

- Firstly: say that all causation is epistemic
- Secondly: make causation broadly construed so that fundamental physics is causal after all. This is not possible for the manipulationist unless any change to a variable is called a manipulation. (See Pexton 2013 for an examination of this.)
- Thirdly: hold that manipulability can be an emergent macrophysical property. So it is possible to have no manipulation/causation at the fundamental level but have manipulation/causation at higher levels.

Space prevents a thorough exploration of such a broad topic as the causal status of fundamental physics. My aim is merely to suggest that the third option above is potentially viable, and worth exploring further.

4. Conclusion

The argument presented in this paper can be summarised as follows.

- Manipulationist explanations always involve an input control parameter. The parameter space of this input must be bisected such that we can assign either 1 or 0 to some set of states. This bisecting process can involve abstracting across many disjunctive states. Extreme multiple realisability only provides a pragmatic challenge to this process and so does not avoid the metaphysical consequences of causal exclusion
- Manipulationist explanations are always contrastive. The particular contrast class an explanandum defines can change how the input parameter space is bisected. The contrast class also picks out which variable should be chosen as our input variable
- When the contrastive difference maker is manipulatable, we have a causal explanation. But for some explananda we can select contrastive difference makers which are not manipulatable. In these

cases, we have non-causal explanations, regardless of how many of the other non-contrastive difference makers are causal

• There are systems for which the microphysical difference makers for a particular contrast are non-causal, while the macrophysical contrastive difference makers are causal. In these cases, the property of manipulability may be an ensemble property only

• In situations like this, there cannot be causal exclusion since there is no causal overdetermination. We do not have competing causal explanations. Instead, we have a macrophysical causal explanation and a microphysical non-causal explanation, which are compatible and complimentary

Clearly, suggesting that causation itself can be an emergent phenomenon is a very different approach to causation than many extant accounts. I do not claim that construing manipulability as well-defined physical property (sometimes possessed only by macro-level systems) does not raise many questions. Nor do I claim to be able to answer all of those questions here. Rather, the aim of this paper is to map out the starting point of thinking about manipulability in such a way. In manipulationist accounts, the notion of manipulation is very rarely described in exhaustive terms. In this paper I have attempted to state some of the physical properties a manipulatable variable has to have, but this list is by no means exhaustive. My claim is merely that if manipulability is a confluence of some set of physical properties, then it is possible that those properties are sometimes possessed by macro-level systems alone.

In future work, I hope to be able to make an ontic interpretation of emergent manipulability more precise. It is also possible that by pinning down a necessary set of physical properties that define a manipulation we move too far from Woodward's initial conception, and are unable to capture the rich variety of causal reasoning manipulationism excels at. However, I believe this is ground worth exploring. This approach may offer a way of avoiding the causal exclusion argument. It might also

provide an account of how fundamental physics can primarily deal with non-causal structural explanations, yet special sciences are very often concerned with causal explanations.

Durham University
Email: mpexton1979@gmail.com

REFERENCES

Block, N., 2003, "Do causal powers drain away?" *Philosophy and Phenomenological Research*, 67(1), 133-150.

Cartwright, Nancy, 1983, *How the Laws of Physics Lie*. Oxford: Clarendon.

Chandrasekhar, S., 1957 [1939]. *An Introduction to the Study of Stellar Structure*. New York: Dover

Colyvan, M., 1999. "Causal Explanation and Ontological Commitment". In *Metaphysics in the Post-Metaphysical Age: Papers of the 22nd International Wittgenstein Symposium* (Vol. 1, pp. 141-6).

Dowe, P., 2000, *Physical Causation*, Cambridge, Univ Cambridge Press.

Ehrenfest, P., 1917, "In what way does it become manifest in the fundamental laws of physics that space has three dimensions". *Proc. Royal Netherlands Acad. Arts Sci*, *20*, 200-209.

Kim, J., 1999, "Making Sense of Emergence". *Philosophical Studies*, 95: 3–36.

Ladyman, J., 2008, "Structural realism and the relationship between the special sciences and physics". *Philosophy of Science*, 75(5), 744-755.

Lewis, D., 1986, "Causal Explanation", *Philosophical Papers, Volume II*, Oxford: Oxford University Press, 214-240.

List, C., & Menzies, P., 2009, "Nonreductive physicalism and the limits of the exclusion principle". *The Journal of Philosophy*, 475-502.

Norton, J., 2007, "Causation as Folk Science". In Huw Price and Richard Corry (eds.), *Causation, Physics and the Constitution of Reality*. Oxford: Oxford University Press, 11–44.

Pexton, M., 2013, *Non-causal explanation* (Doctoral dissertation, Ph. D. thesis, University of Leeds).

Raatikainen, P., 2010, "Causation, exclusion, and the special sciences". *Erkenntnis*, 73(3), 349-363.

Railton, P., 1981, "Probability, explanation, and information". *Synthese*, 48(2), 233-256.

Redhead, M., 1990, "Explanation". In Dudley Knowles (ed.), *Explanation and Its Limits. Cambridge.* Cambridge University Press, 135–154.

Russell, B., 1917 [1913], "On the Notion of Cause", in *Bertrand Russell, Mysticism and Logic.* London: Unwin, 173–199.

Shapiro, L. A., 2010, Lessons from Causal Exclusion1. *Philosophy and Phenomenological Research*, 81(3), 594-604.

Shrapnel, S., 2014, "Quantum causal explanation: or, why birds fly south". *European Journal for Philosophy of Science*, 1-15.

Skow, B., 2014, "Are There Non-Causal Explanations (of Particular Events)?" *The British Journal for the Philosophy of Science*, 65(3), 445-467.

Strevens, M., 2011, *Depth: An Account of Scientific Explanation.* Cambridge, MA: Harvard University Press.

Tangherlini, F. R., 1963, "Schwarzschild field in dimensions and the dimensionality of space problem". *Il Nuovo Cimento*, 27(3), 636-651.

Tegmark, M., 1997, On the dimensionality of spacetime. *Classical and Quantum Gravity*, 14(4), L69.

Woodward, J., 2011, '*Manipulationism and causal exclusion*', PhilSci-Archive, URL:<http://philsci-archive.pitt.edu/8651/>.

Woodward, J., 2003, *Making things happen: A theory of causal explanation.* New York: Oxford University Press.