Chapter 7 – Mental Causation

Contents

[7.1 The basic problem 1](#_Toc329942810)

[7.2 Is causal closure true? 2](#_Toc329942811)

[7.3 Does causal closure crowd out higher-level causation? 5](#_Toc329942812)

[7.3.1 Reductionism about causation 5](#_Toc329942813)

[7.3.2 The cosmic coincidence argument 7](#_Toc329942814)

[7.3.3 The argument from the rationality of induction 9](#_Toc329942815)

[7.3.4 Irreducible causation at the fundamental level does not crowd out reducible causation at non-fundamental levels 13](#_Toc329942816)

[7.3.4 Reductionism about mental causation as a sceptical scenario 14](#_Toc329942817)

[7.3.5 Irreducible mental causation is inconsistent with fundamental causal closure 16](#_Toc329942818)

[Appendix – The coherence of past-stable laws and past-stable kinds 16](#_Toc329942819)

# 7.1 The basic problem

Perhaps most pressing empirical issues when trying to find the place of consciousness in nature is the question of whether the bottom floor of reality is causally closed. Suppose that the World is carved up into distinct layers of being, like a wedding cake. At the bottom layer we have microscopic particles of matter interacting in various ways (to take a simple picture of how things might be). At progressively higher levels we have the chemical facts, the biological facts, the psychological facts and the sociological facts. The pressing empirical issue is whether the micro-level is causally closed: whether each micro-level event has a sufficient micro-level cause.

The reason this issue is pressing is that there is a worry that micro-level causation might ‘crowd out’, as it were, causal goings on at other levels. Suppose my close friend Fred is crying, and I put my arm around him. Presumably where parts of my body are at any particular moment supervenes on the location of the particles in my body at that moment: I can’t be in Japan if all the particles constituting my body are in New York. Similarly, at that moment when I have my arm around Fred, there is a micro-level fact M (concerning the location of particles in my body and in Fred’s body) such that M could not have obtained without it being the case that I have my arm around Fred. If the micro-level is causally closed, then there is some sufficient micro-level cause of M, call is M\*.

What caused me to put my arm around Fred? We would naturally want to say that certain mental facts about me caused this action, let us say my feeling sympathy for my friend. Assume for the moment, as seems natural, that my sympathy for Fred is a property of some higher-level entity, an organism or a brain, rather than a property of fundamental particle. But if M\* is a complete cause of M, and M entails that my arm is round Fred, then it seems that there is a complete micro-level explanation of the fact that I have my arm around Fred. If the higher-level fact of my feeling sympathy for Fred *also* causes me to put my arm around Fred, then the fact that I have my arm around Fred has too many causes.[[1]](#footnote-1)[[2]](#footnote-2)

This is not to be confused with the problem of free will. Firstly, the micro-level causes need not be deterministic for this problem to arise. It may be that M\*, rather than being causally sufficient for M, entirely determines the probability that M will occur. Still, if M\* completely determines the probability that M will occur, there seems to be no room for some higher-level event, such as my feeling sympathy, to play some role in the explanation of M’s occurrence. Furthermore, the problem is much more serious that the problem of free will. Suppose it turns out that we are not free in the way we pre-theoretically take ourselves to be. This may be hard to accept, it may even be in some sense tragic. But we can arguably carry on living reasonable lives in spite of this. But the supposition that my partner’s thoughts don’t play any role in causing her words amounts to a radically sceptical scenario; I am unable to accept this whilst flourishing in the life I seem to be living.

I take the causal efficacy of mental events, then, the be non-negotiable, and the rest of this book will be concerned with how we need to suppose the World to be in order for our thoughts and feelings to have impact on the world.

#  7.2 Is causal closure true?

There is a prima facie concern, then, that micro-level causal closure (from now on ‘causal closure’) would be inconsistent with mental causation. The obvious first question to address is: Is causal closure true?

The opposing view we need to have in mind is causal emergentism. According to causal emergentism, higher-level systems, perhaps biological systems (cells, organs, or whole organisms), or perhaps systems involved in mentality (neurons, neuro-transmitters, or whole brains), have their own distinctive causal powers. The employment of these higher-level causal powers makes changes at the micro-level, either by acting directly on the micro-level (‘downwards causation’), or by making changes at the higher-level which entail changes at the micro-level, e.g. a higher-level mental event makes my arm move up, which entails that the particles in my arm go up. It is clear that causal emergentism would secure the causal efficacy of higher-level properties.

One might think that the principle of the conservation of energy would ensure that causal closure is true, on the grounds that if higher-level events are making causal changes at the micro-level, either directly or indirectly, then this will result in additional energy being present in the system, contrary to the conservation principle.[[3]](#footnote-3) If my mind influences my behaviour by bringing about action potentials in a number of neurons, then one might suppose that this action will add energy to the brain. However, higher level forces, perhaps mental or biological forces, may act conservatively, taking away as much energy as they give. After all, physicists accept that there is a plurality of causal forces at the fundamental, which all respect the principle of the conservation of energy. Why not add one or two more at higher-levels?

It is pretty plausible that causal closure is true in manmade objects. The causal powers of a computer, for example, are completely determined by the causal powers of its micro-level components. The system as a whole does not develop distinctive causal powers, which alter the function of individual circuits. We know this about computers because we built them. But what reason to we have to think that causal closure is true in entities in the natural world which we did not build? It is beyond our perceptual powers to observe the micro-level happenings in, say, a whole tree, from moment to moment, and thereby observe that the micro-level happenings at later moments are entirely determined by the micro-level causal happenings at earlier times.

We may nonetheless have strong inductive grounds for accepting causal closure. Suppose that every time we investigated a complex system, we found that we could predict the behaviour of its components from their behaviour in more simple systems, e.g. we could entirely predict the behaviour of insulin in the body from an understanding of its behaviour in non-biological contexts. If we have found this to be so in enough cases, then this would count as strong evidence that the evolution of systems at the micro-level is wholly determined by micro-level causes.

David Papineu believes that a strong inductive case for causal closure was accumulated in the twentieth century:

During the first half of the century the catalytic role and protein constitution of enzymes were recognised, basic biochemical cycles were identified, and the structure of proteins analysed, culminating in the discovery of DNA. In the same period, neurophysiological research mapped the body’s neuronal network and analysed the electrical mechanisms responsible for neuronal activity. Together, these developments made it difficult to go on maintaining that special forces operate inside living bodies. If there were such forces, they could be expected to display some manifestation of their presence. But detailed physiological investigation failed to uncover evidence of anything except familiar physical forces (253-4).[[4]](#footnote-4)

One concern with Papineau’s argument is that it lumps together the question of whether there are special biological forces with the question of whether there are special mental forces. The brains of mammals are incredibly complicated biological systems. There are billions of neurons in an adult human brain, each neuron connected to ten thousand other neurons. It may be that we know enough about, say, cells in the liver to know that their causal workings do not involve special cellular forces. But it is much less clear that we know enough about the workings of the human brain to know that its causal workings do not involve special forces.

What do we know about the brain? At a lower-level, we have a good understanding of the structure of an individual neuron, and of how individual neurons communicate with each other. At a higher-level, we also have a pretty good understanding of the division of labour within the brain as a whole, of the overall functioning of different areas of the brain. But do we know enough about how the causal role of a given region of the brain, say, the frontal cortex, is built up from the interactions of several million neurons to know, or have good inductive reason to think, that there are no forces governing the interactions between these neurons which aren’t present in simpler systems?

Answering this question may depend on how we know the function of a given brain region. Much of what we know about the causal role of parts of the brain is from observing what happens when a certain area of the brain is damaged, or is stimulated. Knowing the causal properties of a certain area of the brain in this way, won’t provide us with detailed understanding of the nature of the dependence of those causal properties on their parts.

What we would need to demonstrate causal closure would be a detailed *neuro-anatomical* case, showing how the behaviour of large brain regions arises from the behaviour of neurons and neuro-transmitters, such that that behaviour is entirely predictable from knowledge of the behaviour of those entities of in simpler systems. If we could do this in enough cases, then this could constitute a strong inductive case against the existence of special forces operating in the brain. Papineau’s brief observation that we have now ‘mapped the body’s neuronal network and analysed the electrical mechanisms responsible for neuronal activity’ without finding special neuronal forces does not constitute such a case. I have never come across such a detailed neuro-anatomical case, and until such a case is made it seems that we have no good reason to think that there are no special forces operating in the brain which don’t operate in other causal systems.

Moreover, even at the cellular level, the thesis that there are no special forces operating is not as uncontentious as Papineau supposes. Stephan et al have presented a detailed empirical case that components of cells behave differently in the cellular context than they do in simpler systems.[[5]](#footnote-5) I am not claiming this case is entirely conclusive, but it has not as far as I know been responded to by proponents of causal closure. In general, the question of whether there are special biological or neurological forces is one philosophers tend to assume has been answered by scientists, when in reality it is not a question scientists tend to have much interest in.

In the end, then, I make no stand either way on whether micro-level causal closure obtains in the brain. I think there is a crucial, currently undecided, empirical question that philosophers and scientists should work together to address. Until this matter is settled, it will be useful to think about what our options are depending on how it gets resolved. I will proceed, therefore, on the assumption that causal closure is true, and try to formulate a theory of consciousness consistent with it.

If causal closure is true, then coming up with a theory of consciousness is less of an empirical enterprise than would otherwise be the case. If the behaviour of physical systems can be in principle predicated from the behaviour of their micro-level components in simpler systems, then the behaviour of physical systems we might want to associate with human consciousness will not be empirically distinguished in the way they would be if their presence was associated with special mental forces. Furthermore, it is already a hard question how to decide which physical systems are associated with human consciousness, given that we can’t directly observe another individual’s consciousness.

The correct starting point, I think, is to try to develop a view according to which human consciousness is able to causally affect the world in the way that it needs to if scepticism is to be avoided. How must the world be if causal closure is true, and yet my pain causes my screams, and my thoughts cause my words? In this way, we can begin by working out what options there are, and then narrow them down with empirical work.

# 7.3 Does causal closure crowd out higher-level causation?

## 7.3.1 Reductionism about causation

Whether or not causal closure crowds out higher-level causation depends on whether or not causation is irreducible. I shall understand a reductionist about causation to be someone who believes that the world is truly describable such that certain events cause certain other events, but that the world in and of itself is not such that certain events cause certain other events. Regularity theory is the crudest kind of reductionism. According to the regularity theorist, the World is not such that match strikings cause match lightings, but the World is such that match strikings are regularly followed by match lightings; in virtue of the fact that the World is such that match strikings cause match lightings, the world is truly describable such that match strikings cause match lightings. The regularities of the World make the world truly describable such that there are causal goings on, just as the people revelling in the World make the world truly describable such that there are parties.[[6]](#footnote-6)

There are more sophisticated forms of reductionism, but they all ultimately ground causal truths in the regularities of the World. Consider, for example, David Lewis’s counterfactual account of causation.[[7]](#footnote-7) For Lewis, the fact that the match striking caused the match lighting is grounded in the fact that, if the match hadn’t been struck it wouldn’t have lit. This counterfactual truth (that if the match hadn’t been struck it wouldn’t have lit) is grounded in facts about the nearest possible world: in the nearest possible world in which the match is struck it lights, and in the nearest possible worlds in which the match isn’t struck it doesn’t light. Which possible world is nearest is determined by facts about resemblance between worlds, both in terms of their intrinsic nature and in terms of their laws of nature. And laws, at least in the actual world, are a matter of the best systematisation of the regularities that obtain throughout space and time. So despite the much more complicated analysis, Lewis’s counterfactual theory, like the much simpler regularity theory, ends up with causal truths being grounded in the regularities of the World (or of Worlds).[[8]](#footnote-8)

The non-reductionist, in contrast, thinks that the World is such that certain events cause certain other events. Typically non-reductionists think that things causally interact in virtue of the causal powers of objects in the world. Salt has the power to dissolve in water, which grounds the fact that putting salt into water causes it to dissolve. Objects with mass have the power to attract other objects with mass, which grounds the fact that dropping a stone causes it to fall to the ground.

There are roughly three options for accounting for the causal powers of objects. The first is to take them as primitive, to view the essential nature of properties or causal powers as causally defined. Brian Ellis, for example, believes that what makes NaCl what it is is its membership of the nature kind whose members are essentially disposed to dissolve in water.[[9]](#footnote-9) The second option is to ground the causal powers of objects in the categorical properties of objects and the laws of nature, where the laws of nature are taken to be contingent, genuine additions to the world. On this view, perhaps most associated with David Armstrong, the laws of nature are taken to be contingent relations of natural necessitation between universals, for example, the laws F=MxA is constituted of a complex relationship between the universals force, mass and acceleration.[[10]](#footnote-10) The third option is to identify the causal powers of objects with their categorical properties. Mass, in and of itself, might turn out to be a quality which is intrinsically suited to endow its bearer with certain causal powers.

I don’t think causal reductionists need not worry about crowding out. Return to our original example. We seem to have two causal-explanatory routes to my arm being around Fred: (i) M\* causing me to put my arm around Fred, in virtue of causing M, (ii) my feeling sympathy for Fred causing me to put my arm around Fred. For the reductionist, there is clearly no causal competition between these two stories in the World, as there is no causation in the World. And there is no reason to think that there is any causal competition in the world as it is described. If the relevant regularities obtain at the micro-level, then the world is truly describable such that M caused me to put my arm around Fred. If the relevant regularities obtain between mental events and actions, then the world is truly describable such that my putting my sympathy for Fred causes me to put my arm around Fred. And there is no reason to think the relevant regularities obtaining at the micro-level inconsistent with the relevant regularities obtaining between mental events and actions.

It is only if we are non-reductionists about the causation that crowding out may become an issue. It cannot be a fundamental fact about reality that M brought about my putting my arm around Fred *and* be a fundamental fact about reality that my feeling sympathy for Fred brought about my putting my arm around Fred, where M and my feeling sympathy for Fred are distinct events (and not parts of a single causal chain). The very idea that event X brought event Y into being excludes it being the case that some other event brought Y into being (unless Y is an earlier or later link on a causal chain linking X and Y).[[11]](#footnote-11)[[12]](#footnote-12)

Therefore, in order to decide whether causal closure would crowd out causation at higher-levels, we need to decide whether or not causal reductionism is true. In 7.3.2 and 7.3.3, I will consider two arguments from the literature which attempt to establish non-reductionism about causation. In 7.3.4, I will show how the kind of irreducible causation these arguments, if sound, commit us to, does not entail that causal closure at the micro-level crowds out causation at higher levels. In 7.3.5 I will argue that there is irreducible mental causation at the mental level. In 7.3.6, I will argue that irreducible causation at the mental level does entail that higher-level mental causation is inconsistent with (is ‘crowded out by’) micro-level causal closure.

The overall conclusion will be that micro-level causal closure does entail the non-existence of mental causation as a higher-level phenomenon, but perhaps not the non-existence of mental causation per se.

## 7.3.2 The cosmic coincidence argument

If reductionism is true, then the fundamental regularities of the World are brute, in the sense of being unexplained. Perhaps regularities at higher levels can be explained in terms of regularities at lower levels, but the fundamental regularities admit of no explanation. The non-reductionist can potentially explain the regularity of the world in terms of the causal powers of objects. Why do match strikings always cause match lightings? Because matches have the causal power to burst into flame when struck. But for the reductionist, fundamental causal powers are ultimately grounded in fundamental regularities, and hence the regularities cannot be explained in terms of causal powers.

Some non-reductionists have rejected reductionism on the grounds that if fundamental regularities are brute, then it is a massively improbable coincidence that the world continues, moment to moment, to behave in a highly regular manner. Galen Strawson gives the argument as follows:

Suppose….there is an external world of objects that is highly regular in its behaviour, but there are no forces at all governing what goes on in this external world. (‘In nature one thing just happens after another.’) There are all those mind‐independent physical objects knocking about out there completely independently of us, persisting and interacting, and nothing—*nothing*—governs or orders the ways in which they do this. And yet they persist and interact in a highly regular fashion. How can this be so? It is really rather extraordinary. For, *ex hypothesi*, nothing constrains them to behave and interact in a way that exhibits any order or regularity at all.[[13]](#footnote-13)

For the reductionist the world is truly describable such that there are causal relations and hence causal explanations of why one event follows another, but in the World in and of itself there are no causal relations to explain why striking the match leads to the match lighting. There is nothing in Nature that guarantees, or makes it more probable, that striking the match will be followed by the match lighting. So why is it always the case, given the right circumstances, that striking of matches leads to lit matches? Why doesn’t the match sometimes light, sometimes turn blue, and sometimes do nothing at all?

There is an intuition here that the fact that the world is so regular requires an explanation, an explanation that the reductionist is unable to provide. Explanations have to end somewhere; the non-reductionist argument is that this is not a good place to stop. We might compare it to the intuition that taking it to be just a brute fact that particles of matter happen to have arranged themselves into complex organisms is not a good place to stop; hence the need for Darwin to provide an explanation of this fact.

Judging that certain hypotheses involving brute facts of certain sorts are unacceptable is just one way of judging a certain hypothesis to be theoretically vicious. I don’t have any general theory of how to assess such appraisals of theoretical vice, nor any account of why we should take them to be a guide to truth. Nonetheless, such judgements are indispensable to any metaphysical or scientific endeavour: where there are multiple hypotheses consistent with the data, we must choose between them on the basis of such judgements.

For what it’s worth, I feel the force of this argument. But I think there is a non-obvious metaphysical assumption in the set up, or at least I find I feel the force of the argument only when making this assumption. The assumption is that time is dynamic. If the World is such that new states of affairs are continuously coming into and passing out of being, as we ordinarily take to be the case, the fact that these brand new additions to reality keep coming in a highly regular manner seems extraordinary and demanding of explanation.[[14]](#footnote-14) There is a deep intuition that something must be guaranteeing this regularity, or at least making it highly probable.

But now consider the four-dimensionalist hypothesis, that is, the thesis that (i) the World is extended in four dimensions – three spatial dimensions and one temporal, (ii) the facts of time are unchanging, (iii) objects persist through time in virtue of having temporal parts. When I suppose that four-dimensionalism is true, it does not strike as remarkable that there are regular patterns throughout the static four-dimensional block of spacetime.[[15]](#footnote-15)

Consider the following analogy. It is relatively recent discovery that the universe is expanding. Suppose it had turned out that our universe was static and stable, as we thought was the case before the 20th century. And suppose it had turned out that our static universe was fairly well ordered across space: the planets and stars in the universe were arranged in a highly regular manner, and had always been so. Would we think it improbable, demanding of explanation, that space was so tidy? Well, perhaps we would in so far as we feel the pull of the Principle of Sufficient Reason, and so demand an explanation of *any* way the World might have turned out. But in so far as we are content to let explanation stop somewhere, the eternal existence of a nicely ordered universe seems as good a place as any. Indeed, in so far as simpler explanations are preferable, the hypothesis that arrangements of objects in space instantiate some kind of simple pattern might be very attractive.

On the four-dimensionalist hypothesis, the fourth dimension of time is to be understood analogously to the other three dimensions of space. If it is theoretically acceptable for there to be brute patterns throughout static (over the long term) space, then it is theoretically acceptable for there to be brute patterns throughout static time. Indeed, insofar as those patterns are simple, order might be more theoretically palatable than chaos.

I don’t know how to press this intuitive difference, any more than I know how to persuade someone who doesn’t feel the intuitive difficulty of leaving the emergence of organisms unexplained. But careful reflection leads me to think that it is brute regularity *across dynamic time* which is theoretically unacceptable. If I am right about this, then we have an argument not for non-reductionism about causation, but for the disjunction of non-reductionism about causation and four-dimensionalism. In so far as we find non-reductionism about causation problematic, we could even turn this into an argument for four-dimensionalism:

1. Either non-reductionism about causation is true or four-dimensionalism is true.
2. Reductionism about causation is true.
3. Therefore, four-dimensionalism is true.

Armchair reflection on our pre-theoretical conception of time reveals it to require that time is dynamic. According to our ordinary way of thinking about things, the future is yet to come into being and the past has ceased to be. The static conception of time is a strange and recent conceptual development. However, careful armchair reflection provides us with no reason to think that our folk conception of time is satisfied, and the non-armchair investigation of twentieth century physicists is arguably at odds with that conception.[[16]](#footnote-16) In the absence of some reason to believe the dynamic conception of time, either from our two sources of data (the empirical facts and careful reflection on the essential nature of consciousness), or from the need to avoid radically sceptical hypotheses, the cosmic coincidence argument cannot provide us with an argument for non-reductionism about causation.

(In chapter 8 (8.5. and the appendix) I shall explore in more detail considerations in favour of these different views on the nature of time).

## 7.3.3 The argument from the rationality of induction

We predict the regularity of the future based on the regularity of the past, and we generally take it to be rational to do so. To take a very simple example, we suppose that massy things will continue to attract each other in the future, because they always have done in the past. But what justifies this kind of inductive inference? Why should the regularity of the past have any implications for the future? We have here a classic philosophical problem.

Some non-reductionists about causation have offered an account of the rationality of induction, which we can call *the two-inference account*. The two-inference account construes induction as involving two sub-inferences. The first inference is *an inference to the best explanation* from the past regularity of the universe to certain facts *f* which best explain the past regularity of the universe; the second inference is *a deductive inference* from *f* to the future regularity of the universe.

David Armstrong defends the two-inference account, as a proponent of the non-reductionist view according to which the causal powers of objects are grounded in laws of nature.[[17]](#footnote-17) Armstrong looks to the past, and sees that in the past all *F*’s were *G*’s. He explains this past regularity by supposing that the universal *F* and the universal *G* are related by *natural necessitation*. Once Armstrong is justified in believing that *F* and *G* are related by *natural necessitation*, he can then deductively infer that all future *F*’s will be *G*’s; the fact that *F* and *G* are related by *natural necessitation* guarantees that all future *F*’s will be *G*’s.

Brian Ellis defends the two-inference account, as a proponent of the view that the essential nature of natural kinds is causally defined.[[18]](#footnote-18) Ellis looks to the past, sees that (all things being equal) electrons, whenever they come into proximity of each other, move away from each other. He explains this past regularity by supposing that electrons are members of a natural kind such that its members are essentially disposed to repel one another. Once Ellis is justified in believing that electrons are members of this natural kind, he can deductively infer that future electrons (all things being equal) will move away from one another; the fact that all electrons are members of a kind whose members are essentially disposed to repel each other guarantees this.

Most forms of reductionism are going to be unable take advantage of the two-inference account. If the fundamental regularities of the world lack an explanation, it is not clear that the first sub-inference can be made. Armstrong and Ellis have thus argued against reductionism in the following way:

Argument from the rationality of induction

*Premise 1 –* Reductionists are unable to adopt the two-inference account of the rationality of induction.

*Premise 2 –*There is no way of accounting for the rationality of induction other than by adopting the two-inference account.

*Conclusion 1 –* Reductionists are unable to account for the rationality of induction.

*Premise 3 –* Any view that cannot account for the rationality of induction ought to be rejected.

*Conclusion 2 –* Reductionism ought to be rejected.

My real concern with this argument is premise 3, as I shall explain at the end of this section. But I will spend most of this section arguing that, even if all the other premises of this argument are true, premise 1 is false, as there is a form of reductionism about causation which is able to adopt the two-inference account: namely reductionism combined with four-dimensionalist priority monism. Hence, causal reductionism in combination with four-dimensionalism and priority monism is untouched by the argument from the rationality of induction.

Consider a four-dimensionalist version of priority monism (from now on ‘4D monism’), according to which facts about the whole of spacetime ground events occurring at particular locations in time and space (6.4.4 explains priority monism). For the sake of simplicity let us suppose that the properties of spacetime are distributional properties. The causal reductionist who is also a 4D priority monist is able to adopt the two-inference account. The first inference is from the past regularity of the universe to the fact that the universe has ordered distributional properties. The second inference is from the fact that the universe has ordered distributional properties to the fact that the universe will continue to be regular in the future.

Let us consider an example. The priority monist wanting to give a two-inference account looks to the past temporal slice of the universe and see that it instantiates a distributional property in accordance with the principle of the conservation of energy – call this *a conservative distributional property*. She then explains this fact in terms of the fact that the universe as a whole instantiates a conservative distributional property. Once the priority monist is justified in believing that the universe as a whole instantiates a conservative distributional property, she can deductively infer that the future will instantiate a conservative distribution property. The 4D priority monist reductionist is thus able to give an account of the rationality of induction relevantly similar to that of the non-reductionist. Therefore, even if the argument from the rationality of induction has force against some forms of causal reductionism, it does not have force against causal reductionism in conjunction with 4D priority monism.

Let me clarify this reductionist strategy for avoiding the argument from the rationality of induction, by pre-empting an obvious objection to it. Consider the following distinction:

*Eternally-ordered distributional properties* – Distributional properties which are ordered throughout the whole of time.

*Past-ordered distributional properties*– Distributional properties which are ordered in the past, but disordered in the future.

The following objection may be made in terms of this distinction:

The hypothesis that the universe instantiates eternally-ordered distributional properties constitutes *one* explanation of the regularity of the past. However, the hypothesis that the universe instantiates past-ordered distributional properties constitutes an *equally good explanation* of past regularities. In the absence of a reason to favour the former explanation over the latter, the priority monist reductionist cannot make an inference to the best explanation to the postulation of eternally-ordered distribution properties, and hence cannot go on to deduce the regularity of the future. Because of this, the priority monist is unable to give a successful two-inference account of induction.

This is a legitimate concern, and the priority monist reductionist must respond to it with some kind of story as to why explanations involving eternally-ordered distributional properties are better than explanations involving past-ordered distributional properties. Presumably doing this is going to involve some kind of appeal to simplicity as a theoretical virtue, coupled with an argument that eternally-ordered distributional properties are simpler, and so theoretically superior, to past-ordered distributional properties. I think there is hope that the priority monist causal reductionist can successfully respond to this objection in this way. There is some intuitive pull to the idea that a distribution which is continually ordered is somehow simpler than a distribution which is half ordered and half chaotic; perhaps this intuition could be made into a rigorous defence of the causal reductionist priority monist two-inference account against this objection. Having said that, it might start to look like this requirement to appeal to simplicity is a cost of the priority monist causal reductionist account of induction, which might tilt the balance of considerations in favour of non-reductionism.

However, and this is the crucial point of dialectical significance, it looks like the causal non-reductionist is in exactly the same boat.[[19]](#footnote-19) Compare the following distinction:

*Eternally-governing laws*– Laws that govern throughout the whole of time.

*Past-governing laws* – Laws that govern in the past, but not the future.

Eternally-governing laws will explain the regularity of the past, but of course will not do so in a way which allows us to infer the regularity of the future. Therefore, if Armstrong is to have a good two-inference account of induction, he needs to get eternally-governing laws, rather than past-governing laws, out of the first inference. Just as the 4D priority monist reductionist needs to tell some story about why eternally-ordered distributional properties are superior to past-ordered distributional properties, so Armstrong is going to have to tell some strong about why eternally-governing laws are superior to past-governing laws. This will presumably involve an appeal to the greater simplicity, and hence theoretical superiority, of explanations involving eternally-governing laws.

Turning to Ellis, consider the following distinction:

*Eternally-stable kinds* – Kinds which dispose their members to behave in a single regular manner throughout all of time.

*Past-stable kinds* – Kinds which dispose their members to behave one way in the past, and a different way in the future.

Supposing that objects belong to past-stable kinds will explain the regularity of the past, but not in a way which allows us to predict the regularity of the future. Again, an argument for the greater simplicity, and hence theoretical superiority, of eternally-stable kinds is required to make the two-inference account work.

There may be difficulties facing the priority monist reductionist who wants to put together a two-inference account of induction, but these worries seem to be worries shared by the non-reductionist. (One might have concerns about the coherence of past-governing laws and past-stable kinds. In the appendix to this chapter, I defend their coherence).

Therefore, even if there is force to the argument from the rationality of induction, that force does not push us to the conclusion that non-reductionism is true, but to the conclusion that either non-reductionism or 4D priority monism is true. In the last chapter we raised concerns about the empirical plausibility of the view that subjects are grounded in the cosmos. However, we might combine priority monism with emergentism, such that the facts about the cosmos ground the micro-level facts, which in turn causally ground the phenomenal facts. We can then explain the regularity of the phenomenal facts in terms of the regularity of the micro-level facts, which is explained in terms of conservative distributional properties of the whole of spacetime. From the ordered distribuitional properties of spacetime, we can infer the regularity of future micro-level facts, from which we can infer the regularity of future phenomenal facts.

Hence, assuming the truth of premises 2 and 4, we have an argument not for non-reductionism, but for the disjunction of non-reductionism and 4D priority monism (perhaps combined with emergentism if we accept the arguments of the last chapter). Indeed, we might turn the argument from the rationality of induction into an argument for 4D priority monism:

1. Either causal non-reductionism is true or 4D priority monism is true.
2. Causal non-reductionism is false.
3. Therefore, 4D priority monism is true.

If we do take seriously the idea that we need to account for the rationality of induction, then we obviously have a delicate weighing of considerations. We must somehow work out whether it is a good idea to go for a world where micro-level facts are dependent on facts about the cosmos and phenomenal facts are dependent on micro-level facts, in order to give up on irreducible causation at the fundamental level with the potential to crowd out mental causation.

However, as I said earlier, I am not convinced that there is any real need to explain the rationality of induction. It may be that our commitment to the future resembling the past, like our commitment to the existence of other minds, has a basic pragamatic justification: it is impossible to plan any kind of action in the world without supposing that the future will resemble the past. And if this commitment has this kind of basic, pragmatic justification then we don’t need to show that it is rationally justified in some other sense.

## 7.3.4 Irreducible causation at the fundamental level does not crowd out reducible causation at non-fundamental levels

I have expressed concerns with both of the above arguments for non-reductionism. They are sound only if certain substantive metaphysical hypotheses can be ruled out – four-dimensionalism in the case of the first argument and 4D priority monism in the case of the second – and even then I have doubts about the soundness of the latter. But even if these arguments are sound, and hence establish the falsity of causal reductionism, I don’t think they establish enough irreducible causation to lead to problems reconciling higher-level causation with micro-level causal closure.

We have so far only been contrasting causal reductionism across the board with a commitment to irreducible causation across the board, as is common in the literature. But there is also room in logical space for *partial reductionism* about causation, the view that irreducible causation exists at some levels but not others. For example, a partial reductionist may hold that the World is such that certain micro-level events cause certain other micro-level events, but give some reductionist analysis of causation at higher-levels. On such a view, the World is such that electrons are disposed to repel each other, but the world is merely truly describable such that match strikings cause match lightings, in virtue of, say, the regular connection between match strikings and match lightings.

A partial reductionist need not worry about causal closure at the fundamental level crowding out mental causation at a higher level, so long as that higher level doesn’t involved irreducible causation. Suppose the partial reductionist thinks that the World is such that M caused me to put my arm around Fred, but the world is merely truly describable such that my feeling sympathy for Fred caused me to put my arm around Fred. In so far as the World is concerned, there is clearly no causal competition between these two levels, as causation exists only at the lower level. And we can assume that (i) the World being such that M caused me to put my arm around Fred is consistent with (ii) the relevant macroscopic regularities obtaining such that the world is truly describable such that my feeling sympathy for Fred caused me to put my arm around him.

Thus, it is only if we are non-reductionists about the causation at *both* the fundamental level and the mental level, that higher level mental causation is inconsistent with (is “crowded out by”) fundamental causation. To repeat the point made in 7.3.1, it cannot be a fundamental fact about reality that M brought about my putting my arm around Fred *and* be a fundamental fact about reality that my feeling sympathy for Fred brought about my putting my arm around Fred, where M and my feeling sympathy for Fred are distinct events (and not parts of a single causal chain). The very idea that event X brought event Y into being excludes it being the case that some other event brought Y into being (unless Y is an earlier or later link on a causal chain linking X and Y).

Although the argument from the rationality of induction and the cosmic coincidence argument are usually construed as arguments for non-reductionism across the board, they are really only arguments for irreducible causation at the *fundamental* level. In order to avoid cosmic coincidence, we need only have irreducible causation at the fundamental level, as once we have explained fundamental regularities we can then explain higher-level regularities in terms of them. Similarly, irreducible causation at the fundamental level is all we need to account for the rationality of induction, as the two inference account can be supplemented with explanations and inferences down and up levels: we can explain the regularity of higher levels in terms of the regularity of bottom level (which can then be explained in terms of fundamental causation there), and then deduce the regularity of higher levels from the regularity of the bottom level (once it has been deduced from the fundamental causation there).

Therefore, even if these arguments are sound, they do not give us reason to believe that higher-level causation is irreducible. Even if we are led, on the basis of these arguments, to the conclusion that there is irreducible causation at the fundamental level, we can render this quite consistent with higher-level mental causation, by giving a reductive account of higher-level mental causation.

The only problem is, as I will argue in the next section, to give a reductive account of mental causation is to embrace scepticism.

## 7.3.4 Reductionism about mental causation as a sceptical scenario

I said at the start of this chapter that the need to allow for mental causation is a non-negotiable constraint, on the grounds that the denial of mental causation amounts to sceptical scenario. In this section I will argue that further reflection reveals that what is required to avoid a sceptical scenario is not just mental causation, but *irreducible* mental causation.

I argued in chapter 6 that thought and talk about subjects of experience demands of the World the existence of subjects of experience, e.g. my thought that my partner is feeling sad demands of the World a subject of experience that feels sad. Now let us consider what thought and talk concerning subjects causing things demands of the World. To focus on an example that reflects something of central importance in an individual’s life, consider your partner telling you that he loves you and has decided to stay with you despite current difficulties.

When having a conscious experience as of hearing these worlds, there is always the possibility that you’re a brain in a vat and your partner doesn’t really exist. But in order to fully engage with the life you seem to be living, it is necessary to believe that that is not the case, that is, to believe in the real existence of your partner. What else is necessary? It is also necessary to believe that your partner’s words result from his thoughts and feelings; that it was his love for you and his decision to stay that caused his utterance.

Now reflect on what that means, on what exactly is required of the World in order to rule out the sceptical scenario in which your partner’s declaration of love and commitment was not caused by his love and commitment. Clearly it would not be enough that the feelings of love and the decision to commit exist, and were in existence shortly before the declaration; everyone would agree that that is not sufficient for the latter to result from the former. What else can the World provide in order for it to be the case that the love and commitment caused the declaration, in a way that avoids a sceptical scenario? What the reductionist adds are facts *outside of the immediate situation*. The regularity theorist adds that certain widespread regularities obtain. The counterfactual theorist adds that the possible situation that most resembles the actual situation but in which your partner doesn’t feel love and commitment is one in which he doesn’t declare love and commitment.

But it is evident that what is important in this situation, that what makes the difference between flourishing engagement with reality on the one hand and delusion on the other, is something *intrinsic to the situation*. Your focus of concern is solely on a relationship between yourself and another individual. The rest of the World is just not relevant to what is of importance here, or is relevant only in so far as it provides links on the causal chain between your partner and yourself.

Of course the reductionist will say that her analysis makes it true that your partner’s thoughts and feelings ‘caused’ his declaration. But what I am concerned with here is what is required of the World in order for this to be the case in way that preserves what is important. What is needed to work this out is a deep and thoughtful engagement with the situation. I recommend that the reader puts down the book for ten minutes, and spends that time considering an analogous situation in her own case when it’s personally important for communication to have taken place, for another’s thoughts and feelings to have impacted on your consciousness. Without doing this you are in no position to assess the claim I’m making. I believe that such careful reflection reveals that reductionist analyses of ‘my partner’s thoughts and feelings caused his words’ – or whatever analogous situation is being considered – do not plausibly capture what is required to avoid solipsistic despair.

In my experience, analytic metaphysicians are both too concerned and not concerned enough about intuitions. They are too concerned with intuitions in that they take a given pre-theoretical intuition to be of importance in metaphysics, without a clear account of why it is important. In contrast, for each intuition I have appealed to, I have explained why respecting that intuition is of importance. It is important either because it is an intuition about the essential nature of consciousness, the nature of which is directly revealed to us, or because it is an intuition about something which is required in order to fully engage with the life we seem to be living.

Analytic metaphysicans are not concerned enough about intuitions in the following sense: where they do draw conclusions from intuitions, they do not result from deep and thoughtful engagement with simple situations of everyday life. I say this because it seems to me the most plausible explanation of how so many intelligent, sensitive men and women working in analytic metaphysics claim to find intuitively adequate bizarre analyses of everyday phenomena.

Analytic metaphysicians have created beautiful systems. Most famously, David Lewis’s grand system, grounding so much in the commitment to possible worlds and the Humean mosaic of the actual world, is extremely aesthetically pleasing. But this system provides a caricature of the phenomena we encounter in daily life. I simply don’t believe that any mentally well person who sits in a quiet room and carefully and *seriously* considers the Lewisian thesis that *to suppose that someone is in pain is to the suppose that they have an inner state that plays the pain role* will find that thesis at all plausible. Or anyone who does has completely detached their metaphysics module from the module they use to engage with the real world.

Similarly, it is underappreciated how far the Lewisian counterfactual analysis of causation, and all other reductionist analyses, departs from our pre-theoretical understanding of causation. With respect to the Lewis’s analysis of pain, we know that it is inadequate to capture pain as it is out there in the world, because we know that our phenomenal concept of pain is satisfied and that our phenomenal concept of pain does not admit of causal analysis. With respect to Lewis’s analysis of causation, we don’t know for sure whether there is more to the relationship between pain and screaming, or match striking and match lighting, than can be captured with a reductionist account of the causal relation. But much of what is important in life depends on there being more to mental causation than can be captured in a reductionist analysis.

## 7.3.5 Irreducible mental causation is inconsistent with fundamental causal closure

Irreducible fundamental causation is consistent with reducible higher-level causation, even in a world where the fundamental level is causally closed, so long as the regularities required for the reducible causation at higher-levels are consistent with the regularities obtaining at the fundamental level. However, irreducible causation at a higher-level is not consistent with a causally closed fundamental level. This is because the fundamental facts ground the higher-level facts.

Return to the example of my feeling sympathy for Fred (at time T1) being followed by my putting my arm around Fred (at time T2). I take it that my having my arm around Fred supervenes on the micro-level facts at T2: in any counterfactual situation where I don’t have my arm around Fred, the particles in my arm differ from their actual location at T2. Therefore, the fact that I have my arm around Fred at T2 already has a sufficient ground, namely the micro-level facts at T2. If the micro-level is causally closed, there is no room for a non-fundamental fact at an earlier time to play any role in grounding the fact that I have my arm around Fred.

Matters would be different if fundamental causal closure where false, if higher-level mental events, such as my feeling sympathy for Fred, had an impact on the micro-level, either directly by exerting a ‘downwards’ causal impact on the micro-level, or indirectly by affecting higher-level facts in a way that entails micro-level changes. In the case of downwards causation, my feeling sympathy at T1 might cause the particles at T2 to be such that I have my arm around Fred at T2. In the case of indirect downwards influence, my feeling sympathy at T1 might cause me to put my arm around Fred (or more likely cause certain neuronal changes, which cause me to put my arm around Fred), which grounds certain facts about the location of particles at T2. Either way, my sympathy plays a causal role in making it the case that my arm is around Fred. But such (direct or indirect) downwards causal influence is not possible if the micro-level is causally closed.

We thus find ourselves in a sticky situation if causal closure turns out to be true. In order to avoid a sceptical scenario, we need to commit to irreducible mental causation. And yet irreducible mental causation is inconsistent with micro-level causal closure. The only solution, if causal closure turns out to be true, is to accommodate irreducible mental causation, of the kind we know from our everyday lives, as the micro-level. In the next and final chapter, this is exactly what I shall try to do.

# Appendix – The coherence of past-stable laws and past-stable kinds

It seems clear that past-ordered distributional properties are coherent, but one might harbour doubts concerning the intelligibility of past-governing laws, or past-stable powers. If past-governing laws and past-stable kinds are not intelligible, but past-ordered distributional properties are, then the non-reductionist and the priority monist reductionist are not in the same boat; the priority monist reductionist needs to make contentious appeals to simplicity as a theoretical virtue in order to rule out past-ordered distributional properties, whilst the non-reductionist can rule out past-governing laws/past-stable kinds on the much less contentious grounds that what is not intelligible is not possible (and hence not actual).

But I think we can make intelligible sense of past-governing laws and past-stable kinds. Helen Beebee suggests that we can make sense of the past-governing law that all *F*’s are *G*’s simply by imagining that *F* and *G* are related in the past but not in the future:

… imagine God watching the Universe unfold. At the beginning of time, he decides it would be nice for all *F*s to be *G*s, at least for the first few billion years, and the easiest way to guarantee that is to glue the universals *F* and *G* together [by which Beebee means relating them by the relation of natural necessitation]. (God sets a high premium on the Universe being law-governed, but he really can’t be bothered fixing the initial conditions and the other laws in such a way that they together deliver the result that all *F*s are, at least for the first few billion years, *G*s. That would way too complicated.) So he glues *F* and *G* together with *N*. After a few billion years...he gets a bit bored with the tedious regularity with which *G*s follow *F*s and he decides to make a change. (‘Those humans are getting a bit blasé about this whole science business’, he thinks. ‘They think they’ve cracked the secrets of the Universe. Well I’ll show them who’s boss!’) He removes the glue and – presto! – things down on Earth start getting really unpredictable, causing some serious confusion amongst the scientific and philosophical community until they eventually get their heads around the fact that the laws of nature have changed. I don’t see why God couldn’t do this. After all, as I have said, necessity, on Armstrong’s view, is contingent: it glues things together that are not glued together in other possible worlds. So why could God not *actually* separate them?[[20]](#footnote-20)

Pace Beebee, I don’t think this situation is coherent, as it implies a contradiction: that it is both the case that *F* and *G* are related by *N* and the case that *F* and *G* are not related by *N*. We would avoid contradiction if *F* and *G* were perduring objects, such that we could claim that the temporal parts of *F* and *G* in the past are related by *N*, but the temporal parts of *F* and *G* in the future are not related by *N*. But of course *F* and *G* are not perduring objects, but universals, identical in all their instances. There is, then, no way of removing this contradiction.

However, I think we can make sense of past-governing laws by supposing that *F* and *G* are related, not by the relation of *natural necessitation*, but by the relation of *natural necessitation before 2012*. This is obviously a pretty whacky relation, but looks to be intelligible. Things are slightly delicate here, as there are real concerns about the intelligibility of natural necessitation itself. Therefore, I will refrain from asserting that the relation of *natural necessity before 2012* is intelligible, and will instead assert the following conditional: if the relation of *natural necessity* is intelligible, then the relation of *natural necessity before 2012* is intelligible (or at least the onus is on the supporter of the Armstrong-type view to show why this conditional is false, if this is the line they wish to take).

It is worth noting that Armstrong himself allows for laws which are limited in their governance to certain epochs. Where I have suggested making sense of such laws by an alteration in the higher-order relation, Armstrong makes an alteration in the relata of that relation. Such laws, for Armstrong, relate not universals, but what he calls ‘quasi-universals’, which are universals except that they involve reference to a certain epoch.[[21]](#footnote-21) The past-governing law that all past *F*’s were *G*’s consists of *natural necessitation* relating *being F before 2011* and *being G before 2011*.

One might have several worries concerning the coherence of past-stable kinds. If a given disposition is essential to members of a certain natural kind, then surely members of that kind have that disposition at all times, not merely in the past, which might seem to rule out past-stable kinds. Or one might worry about the intelligibility of a disposition whose nature can only be characterised with reference to a specific time, for example 1st January 2012.

I think we can avoid both these worries. Imagine a chemical substance which has an essential disposition such that, when it comes into contact with water, it turns green for an hour, and then turns red for an hour. This seems like an intelligible disposition. We need not construe this as involving a change in the compound’s essential dispositions, we can instead suppose that the compound has a single unchanging disposition: *the disposition to turn red for an hour and then green for an hour after having come into contact with water*. Nor does the specification of the nature of this disposition involve reference to a specific time, although it does involve reference to periods of time, i.e. hours.

We could think about past-stable kinds in the same way. Electrons may be members of a natural kind such that, when its members are in condition *C*, they are disposed to behave in a certain way for *X* amount of time, and then to behave in a completely different way for *X* amount of time, where conditions *C* are the conditions of the universe when electrons first came into existence, and *X* happens to be the exact time span of the universe up to 2012. One might think this a pretty peculiar kind of disposition, but a case would need to be made that it is unintelligible. Of course observation would never lead us to suppose that a material object is of a past-stable kind, but it is the theoretical viciousness, not the unintelligibility, of the postulation of past-stable kinds which explains this.

1. Reference [↑](#footnote-ref-1)
2. If we are priority monists, a similar problem with arise if we think the cosmos is causally closed, such that, for every cosmic change, there is a sufficient cosmic cause. I shall mostly focus on the micro-level version of the causal closure worry for the sake of simplicity. [↑](#footnote-ref-2)
3. Reference [↑](#footnote-ref-3)
4. Reference [↑](#footnote-ref-4)
5. Reference [↑](#footnote-ref-5)
6. Reductionism about causation is more or less the view known in the philosophical literature as ‘Humeanism’. [↑](#footnote-ref-6)
7. Reference [↑](#footnote-ref-7)
8. Reference other kinds of reductionist views. [↑](#footnote-ref-8)
9. Ellis 2001, 2002. [↑](#footnote-ref-9)
10. Armstrong (1983). Tooley (1977) and Dretske (1977) have advanced similar views. [↑](#footnote-ref-10)
11. We sometimes talk of one event having two causes, e.g. we might say of a man who is shot and struck my lightening at the same time, that his death had two causes. However, I take it that this is just loose talk, and in the world as it is in and of itself, there is no event which has two distinct causes, each individually sufficient for the obtaining of that event (where the two causes are not links on a single causal chain). [↑](#footnote-ref-11)
12. Reference [↑](#footnote-ref-12)
13. Reference [↑](#footnote-ref-13)
14. Time would also be dynamic if the ‘growing block view’ were true, that is, if states of affairs were continuously coming into being and remaining in existence. On this view too, the regularities across time seem to call for explanation. [↑](#footnote-ref-14)
15. Time would also be static on an enduranist form of eternalism. As this is also a non-dynamic conception of time, the regularities across time do not seem to call for explanation. I argue against endurantist eternalism in 8.5. [↑](#footnote-ref-15)
16. I argue that that presentism is inconsistent with special relativity in the appendix of chapter 8. [↑](#footnote-ref-16)
17. Reference [↑](#footnote-ref-17)
18. Reference [↑](#footnote-ref-18)
19. Helen Beebee (reference) influenced my thinking here, although she aims to altogether reject the argument from the rationality of induction, whilst my main aim is to show that, even if we grant that the argument has force against most forms of causal reductionism, it does not have force against 4D priority monist causal reductionists. [↑](#footnote-ref-19)
20. Beebee Forthcoming: 10-11. [↑](#footnote-ref-20)
21. Armstrong 1983: 101. [↑](#footnote-ref-21)