1 Introduction

Our world abounds in phenomena that are 'temporally asymmetric', that is, directed differently towards the past and future. Bodies decay, but don't spontaneously rejuvenate. Smoke disperses in a room but doesn't naturally recoalesce. We remember the past but not the future. One of the most important and pervasive temporally asymmetric phenomena in our world is the temporal asymmetry of causation: the fact that (at least around here) causes always come prior in time to their effects. The causes of a gene mutation, a plane crash, or a failed dinner party only ever lie in the past of these events, never in their futures.

The temporal asymmetry of causation is so central to our thinking about the world that it is easy to take for granted. On reflection, one might take it to be a fundamental or primitive fact about how the world is – and so not something in need of further explanation. In Sections 1 and 2, we'll consider why there is a substantive empirical project underway to explain the temporal asymmetry of causation. In Sections 3–5, we'll then consider positive attempts to explain the temporal asymmetry of causation – including those using statistical mechanics (Section 3), features of agency (Section 4), and so-called fork asymmetry accounts (Section 5). While current explanations are incomplete, each programme provides resources for ultimately explaining why causes come before their effects.

Why, beyond curiosity, might we be interested in explaining the temporal asymmetry of causation? First, explaining its temporal asymmetry matters to understanding causation. A plausible constraint on any account of causation is that it can account for why causation is temporally asymmetric. As we'll see (Sections 1 and 2), not all accounts can, and some even conflict with the claim that causation is temporally asymmetric. Second, explaining the temporal asymmetry of causation in physical terms provides a model for how other temporal asymmetries might be explained, and sometimes provides a basis for explaining those asymmetries – including a record asymmetry (the fact that we have memories and other records of the past and not the future), an explanatory asymmetry (the fact that we typically explain events by reference to the past and not the future), and a deliberative asymmetry (the fact that we deliberate about what to do in the future but not the past). Insofar as understanding causation's temporal asymmetry helps us make sense of temporal asymmetries in general, we can also use resources from this programme to explain real and apparent features of time, particularly those that involve time being directed – such as the apparent flow of time and the apparent openness of the future and fixity of the past. Third, we learn lessons about the relation between fundamental physics,

higher level sciences, and metaphysics through attempting to make sense of causation's temporal asymmetry. As we'll see, part of what motivates an empirical project of explaining the temporal asymmetry of causation is an apparent conflict between the relations used in fundamental physics and those used in higher level sciences – a conflict that is particularly sharp, given certain intuitive views about causation. By resolving this conflict, we have a broader story to tell about the unity of science, how philosophy helps us negotiate that unity, and the role of intuition in that negotiation.

1.1 What Is the Temporal Asymmetry of Causation?

To begin, we need some terminology. We will call *the direction of time* the direction in time from past to future. We will call *the direction of causation* the direction in time from cause to effect. To begin, to claim that there is a temporal asymmetry of causation in our world is to claim that, in our world, the direction of causation aligns everywhere with the direction of time – causal and temporal directions are globally aligned. If causation is temporally asymmetric in our world, causes always come before their effects and there are never any cases of causes coming after (or simultaneously with) any of their effects. We may need to revise this strict definition – I'll consider possible revisions in this section and the sections that follow. While I will sometimes talk of 'past', 'present', and 'future', these uses are always to be read as 'before', 'simultaneously with', and 'after' a particular time, respectively – such talk does not presuppose a so-called A-theory of time in which past, present, and future are different regions of time.

The temporal asymmetry of causation is strictly stronger than three other asymmetries of causation. First, causation is *not* symmetric – **a** is the cause of **b** does not *imply* that **b** is the cause of **a** (unlike the sibling relation). Second, causation is *asymmetric* – **a** is the cause of **b** implies that **b** is *not* the cause of **a** (unlike the liking relation). Third, causation is *locally* temporally asymmetric – **a** is the cause of **b** implies that **a** comes before **b** in *some local* temporal ordering. The temporal asymmetry of causation implies these three asymmetries, but, in addition, claims a global alignment between causal and temporal directions – causes come before their effects in *all* temporal orderings.

Most philosophers accept that causation is temporally asymmetric in our world. Some have defended the claim that there is *backwards causation* in our world in the context of quantum mechanics (Price, 1984, 1996: Ch. 8). We might also think we can change the significance of the past. However, these claims haven't gained widespread support. Moreover, even those who defend backwards causation in some settings still typically take causation to be temporally asymmetric when the causes and effects are large-scale 'macroscopic'

events (Price, 1991, 1992a, 1992b). The temporal asymmetry of causation could be restricted to macroscopic causes and effects to allow for such views. However, for the most part, I will assume the standard view that there is no backwards causation in our world.

Some have suggested that causes often or sometimes occur *simultaneously* with their effects (Kant, [1781/1787] 1996: A203/B248; Carroll, 1994: 141–4). These proposals have also not received widespread support. Moreover, accepting simultaneous causation still leaves us with the problem of explaining why causes always come before their effects in cases that *aren't* simultaneous. Again, while we could restrict our investigation to non-simultaneous causation, I will assume the standard view that there is no simultaneous causation in our world.

1.2 What Might Explain the Temporal Asymmetry of Causation?

The temporal asymmetry of causation is robust and pervasive. It is taken for granted in much of our thinking about the world. For this reason, it may be tempting to think that causation's temporal asymmetry is a *necessary* feature of causation. Perhaps the fact that causes always come prior to their effects is *constitutive of* or *intrinsic to* the *nature* or *concept* of causation. If so, it might seem that the temporal asymmetry of causation warrants no further explanation. In the remainder of Section 1, I'll consider several proposals that treat causation's temporal asymmetry as necessary and not in need of empirical explanation. I'll argue that none of these proposals is successful and that the temporal asymmetry of causation is a contingent feature of our world – and so in need of empirical explanation.

One proposal takes the fact that causes always come before their effects to be part of the *definition* of the concept CAUSATION. The temporal asymmetry of causation therefore holds as a matter of conceptual necessity. Hume ([1739–40] 2000: Book I), for example, claims that the idea of causation derives in part from the idea of temporal priority.

However, even if the temporal asymmetry of causation is part of the *concept* CAUSATION, this fact merely shifts the explanatory burden. The question remains, why do we use the concept CAUSATION (in which causes always precede their effects) rather than the concept CAUSATION* (the same as CAUSATION, but in which causes don't always precede their effects)? One might argue that using CAUSATION rather than CAUSATION* is mere convention – in the same way that it is conventional to drive on the right in Sweden – and so is not in need of substantive explanation. However, because of the way causation is tied to our practices, the direction of causation doesn't appear to be merely a matter of

convention (see Section 2.4). We explain effects using earlier causes (and not vice versa) and we decide on causes to ensure their later effects (and not vice versa). If the temporal asymmetry of causation were merely conventional, we would expect these practices to be temporally reversible without too much awkwardness – such as when Sweden switched from driving on the left to driving on the right. However, it doesn't seem that we could switch the temporal order of our practices of explanation and control with the same success. Put otherwise, if causation is temporally asymmetric by definition, we will be unable to explain why these practices are, non-definitionally, temporally asymmetric or why causal relations are apt to play these temporally asymmetric roles (Mellor, 1998: Ch. 10; Field, 2003; Price and Weslake, 2009).

A further argument against conventionalism is that we can make good *conceptual sense* of the possibility of backwards causation, that is, cases in which causes come before their effects (Dummett, 1964; Lewis, 1976). It is conceptually coherent, for example, that our rituals influence the past or that Dr Who using her time machine in the future *causes* her appearance in the past. If these cases are coherent, the temporal asymmetry of causation is not conceptually necessary. For related arguments against conventionalism, see Papineau (1985: 273–4), Mellor (1998: Ch. 10), Field (2003), and Price and Weslake (2009).

A second proposal for why we might not need to explain the temporal asymmetry of causation is that it is simply part of the *nature* of causation that causes always come before their effects – either as a primitive feature of causation or due to the way causation relates to laws. This proposal is compatible with views that take causation to be a primitive irreducible relation (Anscombe, 1975; Tooley, 1987; Carroll, 1994), views that take causal dispositions or powers as primitives (Greco and Groff, 2013), and views that take causal relations to be relations of nomic dependence (Kim, 1973; Armstrong, 2004) or otherwise closely related to laws. What ties these views together is the claim that there is a direction of time established independently of the direction of causation and to which causal direction must correspond. Mackie (1974: 225–6), for example, argues that the temporal asymmetry of causation is due to a temporal asymmetry in the laws of nature, which itself presupposes an asymmetry in time. Understanding the temporal asymmetry of causation might seem to require understanding the nature of *time's* directedness.

To defend this 'primitivist' proposal, one needs to explain why the direction of causation is necessarily aligned to the direction of time. One way to do so is to argue that causation, laws, and the direction of time all are related to *production*. Present states of the world, through time, causes, and laws, generate future

states: 'what happens next flows from what is there already' (Mackie, 1974: 225). Laws (Maudlin, 2007: Ch. 4) or causal relations (Carroll, 1994) are the means by which the past produces the future.

However, this primitivist proposal is susceptible to a similar worry. If backwards causation or time travel are *metaphysically* possible (that is, possible given the nature of causation, laws, and time), then the nature of causation, laws, and time does not explain why causes come before their effects. The conceptual coherence of backwards causation and time travel may be enough to suggest that they are metaphysically possible. Results from physics provide further arguments. First, scientists and philosophers have entertained theories involving backwards causation as a genuine candidate for the physics of our world, including the Wheeler–Feynman theory of radiation, Feynman's theories of tachyons and positrons, and interpretations of quantum mechanics. Insofar as these theories are metaphysically coherent, the nature of causation, time, and laws does not explain the temporal asymmetry of causation (for discussions, see Earman, 1976; Horwich, 1987: Ch. 6; Friederich and Evans, 2019; Faye, 2021). Second, the equations of general relativity, that is, Einstein's field equations, have solutions that involve 'closed time-like curves'. These are possible trajectories such that objects travelling at velocities less than the speed of light could traverse these curves and find themselves previous in time to their starting point. The most famous of these solutions are from Gödel (1949). Assuming these features of general relativity remain in subsequent physics, backwards causation is compatible with the laws of our world and so is *physically* possible. If backwards causation is physically possible, this standardly implies that it is metaphysically (and conceptually) possible as well. For further discussions, see Horwich (1987: Ch. 7) and Arntzenius and Maudlin (2013).

A separate argument for rejecting the primitivist proposal is that, even if there is a primitive direction in time, this primitive direction needs to be manifest in physical phenomena if we are to be sensitive to it (Price, 1992a: 513, 2007: 264; Loewer, 2012: 132–6). We have no good model of how we could be sensitive to a direction of time that makes no difference to the kind of physical phenomena that we are sensitive to – such as the velocity and arrangement of matter. Even if there were a primitive direction, we would still need an account of whatever physical phenomena that primitive direction is manifest in and an explanation of why those phenomena are temporally asymmetric. Therefore, even accepting a primitive direction does not obviate the need for investigation into how various *physical* temporally asymmetric phenomena arise – which is precisely the project of those who reject a primitive direction of time.[1]

[1] A third argument against primitivism is that there is no direction of time independent of causal direction, as the causal theory of time suggests (see what follows in this section).

A third proposal for why the temporal asymmetry of causation is necessary is that the direction of causation *determines* or *defines* the direction of time – an implication of the causal theory of time (Kant, [1781/1787] (1996): A190–211/B233–256; Tooley, 1987: Ch. 9; Mellor, 1998: Chs. 10 and 11; Lowe, 2002: Ch. 18). If the direction of causation determines the direction of time, then it seems that causes must always come prior in time to their effects.

A first problem with this proposal is that the causal theory of time does not imply the temporal asymmetry of causation. Recall the fact that the temporal asymmetry of causation implies a global alignment between causal and temporal order. The causal theory of time only implies that the *local* direction of causation aligns with the *local* direction of time. The theory doesn't even imply that there is a *global* causal or temporal order. Compatible with a causal theory of time, there may be regions, even very large regions, where the directions of time and causation are counter to their directions in other regions. There may be causal and temporal loops.

One might respond by arguing that causal loops are metaphysically impossible or at least very rare. Lowe (2002: Ch. 18) argues that causal loops would violate conditions prohibiting circular explanation and so, at most, only small regions of backwards causation would be possible. Mellor (1998: Ch. 12) argues that causal loops would violate the logical independence of chances. These claims are countered by the arguments set out earlier in favour of the possibility of (widespread) backwards causation. If causal loops and backwards causation are possible, then explaining the temporal asymmetry of time using a causal theory of time will require explaining why, contingently, there are no causal loops or cases of backwards causation in our world.

A second problem with the causal theory of time proposal is that the causal theory of time does not explain why the direction of time and direction of causation go *this* way rather than *that* way, where *this* and *that* are defined by ostension or by reference to particular events. A causal theory of time does not explain, for example, why causal and temporal direction head away from the Big Bang and towards the direction in which the universe expands.

While the causal theory of time is plausible, it doesn't explain the temporal asymmetry of causation. For this reason, even those who accept something close to a causal theory of temporal *direction* and take the direction of causation (and other phenomena) to be the closest thing to what we mean by the direction of time (Reichenbach, 1956; Albert, 2000; Rovelli, 2018) don't take themselves to have explained the temporal asymmetry of causation. Indeed, their view about the origin of temporal direction only makes the business of explaining temporally asymmetric phenomena more pressing.

If we reject these proposals, explaining the temporal asymmetry of causation does not require us to look to the nature of time or the concept of causation – instead, it requires a, presumably empirical, investigation into how causation comes to be temporally directed at our world. The source of causation's temporal asymmetry will be not an asymmetry in time, but asymmetries in how phenomena are directed and arranged in time (Price, 1996: 16–21). Such an approach does not imply that causation is reducible – causation may be a primitive relation. Nor does it imply that we should disregard the nature of causation – we will still need some account of what causation is. However, such an approach does require that we investigate the conditions of the world, in broadly scientific terms, in order to explain how these conditions give rise to causal relations that are, contingently, temporally asymmetric.

However, perhaps you remain unconvinced. You might think that that there's an obvious fourth proposal that would explain causation's temporal asymmetry – a temporal asymmetry in the laws of nature. Such a view would not presuppose a primitive direction of time and so would avoid some of the arguments above.[2] The work of Section 2 is to argue against this proposal and use temporal features of laws and causation to show just how hard it is to fit causation into a physical view of the world.

2 Russell's Challenge

In *On the Notion of Cause*, Russell (1912–13: 1) argues for the wholesale elimination of causal concepts from philosophical vocabulary: 'The law of causality . . . is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.' While most have disagreed with Russell's eliminativist conclusions, his arguments have led many to reject the idea that causal relations can be identified with the laws or law-like relations of fundamental physics.

Russell gives three main arguments for the elimination of the *relation* causation.[3] These arguments take the following general form:

1. Causal relations will be found in the relations of fundamental physics, if they are found anywhere.
2. The relations of fundamental physics lack features that are essential to causation.
3. Therefore, there are no causal relations.

[2] However, it would not avoid the argument of general relativity, which suggests that the laws of our world don't imply a temporal asymmetry of causation.
[3] Russell also gives other arguments, some of which are directed at the *concept* or *law* of causality.

While only Russell's third argument directly concerns causation's temporal asymmetry, all are relevant for whether causation should be identified with the laws or law-like relations of fundamental physics. However, while Russell takes the first two arguments more seriously, I'll argue that the third provides the strongest challenge for understanding how causation fits into a physical picture of the world. For further discussion of Russell's arguments and limited endorsements, see Earman (1976), van Fraassen (1993), Field (2003), Eagle (2007), Hitchcock (2007), Ladyman, Ross, and Spurrett (2007), Ross and Spurrett (2007), Farr and Reutlinger (2013), and Blanchard (2016). For more critical discussions, see Smith (2000), Ney (2009), and Frisch (2012, 2014).

2.1 Russell's First Argument

Russell's first argument for the elimination of causation is that advanced scientific theories don't mention 'causes' (Russell, 1912–13: 1): 'All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word "cause" never occurs.' These theories don't mention causal relations and don't identify particular events as causes and others as effects. Therefore, it might seem that there is no place for causation in an advanced scientific view of the world. Russell's argument can be formalised as follows:

P1. Causal relations will be mentioned in or identified by fundamental physical theories, if they exist.
P2. Fundamental physical theories don't mention or identify causal relations.
C. Therefore, there are no causal relations.

Let's look a little deeper into *why* theories of the advanced sciences that Russell has in mind, that is, fundamental physical theories, don't mention or identify causal relations. Imagine a closed system consisting of twenty-six billiard balls bouncing off one other. For simplicity, we'll ignore friction and electrodynamics; take the collisions to be elastic (without loss of kinetic energy) and take the system to be described by simple Newtonian laws of motion.[4] Say billiard ball A knocks into stationary billiard ball B, and then billiard ball B moves off. It seems that the movement of ball A *causes* the movement of ball B. However, the fundamental physical laws don't imply this. What the fundamental physical laws imply is that, given the positions and velocities of all twenty-six billiard balls at one time, t_1, their positions and velocities will be thus and so at another

[4] While Newtonian mechanics is strictly false, the argument holds for more realistic candidates for fundamental theories. I return to this point in Section 2.2.

time, t_2. The laws relate states of affairs of the *whole system* at different times – global states. The laws don't relate individual local states of affairs to one another and so do not select particular events as causes and others as effects. Therefore, when we ask a question such as, "What caused ball B to move off at velocity *v* at time t_2?", the theory provides no direct answer. The theory can tell us which of the balls collided with ball B, but this doesn't tell us what caused ball B's motion, without some further analysis.

Assume for the moment that Russell is right – theories of fundamental physics don't mention or directly identify causal relations. Why would this claim imply there is no causation? After all, there exist many things (for example daffodils, pain, and the colour blue) that aren't mentioned in or directly identified by fundamental physical theories. However, in the case of daffodils, pain, and the colour blue, no one *expects* to find them mentioned in fundamental physical theories. In the case of causation, a very natural assumption is that causation *is* fundamental to the operation of the world and is precisely the sort of thing that *should* be mentioned or directly identified by fundamental physical theories, if it exits at all. P1 is at least prima facie plausible.

In response, some have rejected P2 by arguing that scientists (even physicists) *do* use causal concepts (Suppes, 1970; Earman, 1976: 5; Smith, 2000; Hitchcock, 2007; Ney, 2009; Frisch, 2012, 2014). However, while scientists certainly use causal reasoning in the practice of science, such as giving explanations, applying physical theories, interpreting equations, and justifying dismissing certain solutions to equations (Woodward, 2007: 69), it's unclear why the necessity of causal *reasoning* or *representing* in science and uses of the word 'cause' in these contexts shows that causation is part of the *content* of scientific theories or that scientific theories directly identify causal relations. Nor is the use of causal concepts in 'higher level sciences', including much of physics, relevant to Russell's argument, which concerns *fundamental* physics – the (as yet undiscovered) scientific theory that is universal in scope and can explain the success of other sciences.

A second response is to reject P1 by arguing that fundamental physical theories *presuppose* causation, even if they don't mention it. Perhaps the physical laws must have a causal force 'backing' them in order to direct how a system evolves over time. Russell's second and third arguments provide reasons to reject this suggestion – whatever 'backing force' that physical laws require, it lacks features that we take to be essential to causation.

A third response, and the one I recommend, is to reject P1 and Russell's eliminativist conclusion by giving up the assumption that causation is fundamental to the operation of the world. If causation is macroscopic, emergent, or

reducible, it is not something that we *would* expect to find mentioned in fundamental physical theories – so its absence shouldn't motivate eliminativism. Russell's first argument can thereby be accommodated by revising a natural assumption about causation. While removing causation from the workings of fundamental physics may be a surprising move, Russell's first argument provides no *additional* challenge for making sense of the place of causation in a physical view of the world. A fourth response is to reject P2 by arguing that causation is closely related to the laws of fundamental physics. We'll examine this response in Section 2.2.

2.2 Russell's Second Argument

Russell's second argument relates to his earlier observations about the laws of fundamental physics. Russell argues that that there is a conflict between satisfying three requirements on causation:

R1. Causes necessitate their effects.
R2. Causal relations are general (relating events of repeatable kinds).
R3. Causes and effects are separated by some time interval.

In defence of R1, causation, as it is commonly understood, seems to involve a *necessary* connection between two events (Russell, 1912−13: 2); once the cause occurs, the effect *has to* follow. In defence of R2, causation seems to involve a *general* relation between events – a relation that can hold of multiple particulars that we might actually observe (Russell, 1912−13: 7). While the causal relata might be particular events, they are events of a *type* that might plausibly be repeated, rather than events of a type that occur just once. Given that our universe is large and complex, this requirement suggests that causal relata are *local* events (or states), rather than global states of the universe. In defence of R3, Russell argues that there must be a separation in time between cause and effect (Russell, 1912−13: 5). Setting aside the possibility of simultaneous causation (the cause and effect overlap entirely in time; see Section 1.1), Russell argues that *contiguous* causation is not possible – the effect cannot follow *immediately* after the cause. Russell reasons that we can identify no 'last moment' before the effect begins that contains the cause. He also thinks that we should not allow for causes that exist for a time and then 'suddenly explode into the effect' (Russell, 1912−13: 5). Therefore, there will always be a time interval between the end of the cause and the start of the effect.[5]

[5] Russell's reasoning isn't convincing, but Russell turns out not to need R3.