Causal Explanation in Physics

Mathias Frisch

Are there causal explanations in physics? Answers to this question range from the claim that there are no causal explanations in physics, since the notion of cause plays no legitimate role in physics (and, perhaps, elsewhere) to the claim that all explanations in physics are causal in virtue of the fact that *all* explanations in general (or at least all scientific explanations) are causal. In addition to these two polar opposite positions some philosophers have argued for pluralist views that allow for both causal explanations and non-causal explanations in physics.

Some of the arguments concerning the place of causal explanations in physics appeal to general conditions that any account of scientific explanation ought to satisfy. Thus, according to Carl Hempel's deductive nomological (*DN*) model of explanation, there are no causal explanations in physics simply because there are no genuinely causal explanations anywhere (Hempel and Oppenheim 1948)(Hempel 1970). By contrast, on David Lewis's account of explanation any explanation of a specific event is causal (whether in physics or not) in virtue of providing us with information about the causal history of the event in question (Lewis 1986). If all explanations are causal, explanations in physics are causal as well.

Other views appeal to putatively distinct features of theorizing in physics, either to argue that physics is particularly well-suited for causal explanations or (perhaps more often) to argue that physics is especially inhospitable to causal notions and therefore to causal explanations. According to views of the latter kind, causal explanations may play a role in the so-called 'special sciences' but such explanations sit ill with how physical theories represent the world.

In what follows I will discuss both what influential general theories of scientific explanation imply for the particular issue of causal explanation in physics and whether there are arguments distinct to physics concerning the status of causal explanations.

1. The *DN* model

The theory of scientific explanation as a philosophical sub-discipline has its origins in Hempel's development of the *DN* model. According to the *DN* model a scientific explanation is a deductively valid argument from true premises (constituting the explanans), which has the explanandum sentence as its conclusion (Hempel and Oppenheim 1948)(Hempel 1970). This is the deductive part of the model. The nomological part states that the premises must contain at least one law essentially. The *DN* model was intended to provide an analysis of the concept of scientific explanation that answers to empiricist worries about causal and other modal notions and denies that there are properly causal explanations.

One of the many challenges the *DN* model faces is to provide an account of the concept of *scientific law* that allows us to distinguish laws from accidental regularities in a manner acceptable to empiricists (Hempel 1970).[1] Independently of whether and how this challenge can be met, physics with its generalizations of relatively broad scope appears to be particularly well suited for the *DN* model. Many explanations in physics involve derivations from equations that are taken to have broad or even universal validity, such as Newton's Laws, the Maxwell Equations, or the Schrödinger equation and no matter how one ultimately tries to flesh out the concept of law many of the basic equations of physics will clearly have to fall under it. It is less clear how well the *DN* model can be extended to other sciences and whether generalizations in the special sciences have enough of the requisite characteristic features to count as genuinely nomic constraints.

A number of widely discussed putative counterexamples to the *DN* model suggest that attempts to formulate a theory of explanation that avoids causal notions are ultimately unsuccessful. Perhaps the most prominent problem in this respect is the problem of explanatory asymmetries. There are many cases in which a derivation of an explanandum event E from laws L and initial conditions I strikes us as being explanatory, while the inverse derivation of I from L and E does not seem to be explanatory, even though both derivations satisfy the *DN* model. Consider Sylvain Bromberger's well-known example of the flagpole and its shadow

---

[1] On this question see also (Salmon 2006)(Earman 1986)(Cartwright 1983; Van Fraassen 1989).

(Bromberger 1966)(van Fraassen 1980): While a derivation of the length of the shadow from the height of the flagpole and the Sun's angle in the sky together with the law of the rectilinear propagation of light seems to constitute an explanation of the length of the shadow, a derivation of the height of the flagpole from the length of the shadow does not seem to explain the flagpole's height. A plausible diagnosis of why the second derivation is not explanatory is that it purports to explain the height of the flagpole in terms of its effect, the shadow. And while it may be possible to *derive* the presence of a cause from the occurrence of its effects, the effects' occurrence does not *explain* the cause. Thus, the fact that there are explanatory asymmetries suggest that it might not be possible to develop a general non-causal theory of explanation.

The problem of explanatory asymmetries puts pressure on the claim that the *DN* model provides sufficient conditions for explanation. Michael Scriven and others have argued that the *DN* model also does not provide necessary conditions for explanations (Scriven 1962). Scriven cites what appear to be paradigmatic cases of causal explanations, which do, however, not satisfy the *DN* model. For example, an adequate explanation of why the ink jar spilled might be that I bumped the jar with my elbow. This explanation is adequate, Scriven argues, even if we are not in a position to derive the ink jar's spilling from physical laws together with appropriate initial conditions. One might question the relevance of Scriven's case to the issue of scientific explanation, arguing that it is only an example of a common sense explanation. But as we will see below, even within physics there are many instances of inferences proceeding from less than a full specification of initial and boundary conditions and many of these inferences appear to be paradigmatically causal and explanatory inferences.

2. Conserved Quantity Accounts of Causation

The lesson some philosophers have drawn from examples such as the ones discussed in the previous section is that the *DN* model ought to be abandoned in favor of a causal account of explanation. We ought to put '*cause*' back into '*because*', Wesley Salmon urged (Salmon 1984; 2006). But what is it to causally explain a phenomenon? One answer to this question is given by the causal process account first proposed by Salmon (Salmon 1984) and developed further in Phil Dowe's conserved quantity account (Dowe 2000). Dowe distinguishes causal processes and causal interactions, which he defines as follows:

> CQ1. A causal process is a world line of an object that possesses a conserved quantity.
>
> CQ2. A causal interaction is an intersection of world lines that involves exchange of a conserved quantity. (Dowe 2000, 90)

Conserved quantities are those quantities, such as energy, momentum, mass, or charge, that are conserved according to our physical theories. Even more so than the *DN* account the conserved quantity account appears to derive its inspiration primarily from physics, where conservation laws play a fundamental role. In fact, according to Noether's First Theorem, there is a conservation law associated with each continuous symmetry property of a system (see Brading and Castellani 2003, especially the essay by Brading and Brown therein).

Even though process accounts of causation were designed partly with the problems of the *DN* model in mind, as an account of scientific *explanation* the conserved quantity account is arguably subject to many of the same counterexamples that plague the *DN* model (see Woodward 2017). A central problem for the *DN* model is that there can be nomic connections between an explanandum event and other events that do not capture features explanatorily relevant to the occurrence of the explanandum. Similarly, a quantity conserved in a causal process or exchanged in causal interactions also need not be explanatorily relevant to the phenomenon. Consider a collision of two billiard balls, during which some very small amount of electric charge is exchanged, and let us assume that the balls' charge is conserved both before and after the collision. The motion of the billiard balls constitutes two causal processes joined by a causal interaction. Yet charge conservation does not explain the billiard balls' motion. The relevant conservation law is energy and momentum conservation. But what makes it the case that it is energy and momentum conservation and not charge conservation that causally explains the motion? One possibility is to appeal to counterfactual information at this point: energy-momentum conservation provides the correct explanation, since the motion of the two balls varies with changes in the balls' initial momenta and energies but does not vary with changes to the balls' charges in the right way. But this requires that we supplement the pure conserved quantity account with counterfactual considerations.

The conserved quantity account does not, on its own, provide a distinction between cause and effect and hence, like Hempel’s account, is faced with the problem of explanatory asymmetries. Dowe's solution is to supplement the account by appealing to Hans Reichenbach's *fork asymmetry* (Reichenbach 1956). A *conjunctive fork*, as Reichenbach defines it, consists of three events A, B, and C, such that A and B are unconditionally correlated but conditionalizing on C renders A and B probabilistically independent. That is:

$$P(A\&B) > P(A)P(B) \tag{1}$$

$$P(A\&B\,|C) = \mathrm{P}(A|C)\ \mathrm{P}(B|C) \tag{2}$$

The event C, it is said, *screens off A from B*. Conjunctive forks can be temporally open or closed. If there is an event C occurring in past that screens off A from B, but there is no screening-off event in the future of A and B, then this constitutes an open fork. If there is an event C in the past and in addition an event C' in the future of A and B that screen off A from B, we have a closed fork. Now, Reichenbach's fork asymmetry thesis consists in the claim that all open forks are open toward the future: there are no conjunctive forks for which only a future screening-off event exists. Conjunctive forks allow us to introduce a direction for causal processes, and hence, allow a distinction between cause and effect. Dowe defines the direction of causal processes as follows:

> "The direction of a causal process is given by the direction of an open conjunctive fork part-constituted by that process; or, if there is no such conjunctive fork, by the direction of the majority of open forks contained in the net in which the process is found." (Dowe 2000, 204)

Two aspects of this definition are worth noting. First, adopting a disjunctive criterion allows him to distinguish causes from effects by their location in a causal net, even when a process itself does not involve a causal fork. Second, in a departure from Reichenbach, Dowe wants to allow for the possibility of backward causation and hence allows forks open toward the past. Dowe's motivation for this is that he wants to appeal to backward causal relations as providing a local causal explanation of the quantum mechanical correlations exhibited by entangled states, which cannot be given a common cause explanation.

In order to address the problems of explanatory asymmetries and irrelevancies the conserved quantity account has to take two additional kinds of explanatory information on board: probabilistic information and counterfactual information. This raises the question as to what ultimately is doing the explanatory work in the account: is it the appeal to a conserved quantity or rather the counterfactual and probabilistic dependencies? The significance of Reichenbach's conjunctive fork is that conjunctive forks instantiate what appears to be a fundamental explanatory relation: common cause explanations. Correlations between two distant events *A* and *B*, which are not related as cause and effect, are explained by an event *C,* the common cause of *A* and *B*, which, Reichenbach postulates, screens off *A* from *B.* Thus, one might want to bypass the appeal to conserved quantities and construct an account of causation and causal explanation directly in terms of counterfactual and probabilistic dependencies. I will take up this suggestion in section 3.2. First, however, I want to discuss a powerful and influential challenge to any causal account of explanation in physics, which has the same empiricist roots as the *DN* model but was developed independently of the literature on scientific explanation.

3. Causation in Physics

3.1. Mach's and Russell's Challenges to Causation

One answer to the problem of explanatory asymmetries has been to argue for the need of a causal account of explanation with an underlying notion of causation rich enough to underwrite a distinction between causes and effects. But there is another argumentative strand in the literature that denies the applicability or at least the usefulness of causal notions in physics. This literature is not primarily focused on the notion of *explanation*, but focuses directly on the notion of *cause*. Arguments pointing to an alleged incompatibility of causal thinking with theorizing in physics can be traced back to the writings of Ernst Mach (Mach 1900; 1905) and to Bertrand Russell's famous article "On the Notion of Cause" (Russell 1912). More recent defenders of such a view include Huw Price (Price 1997; Price and Weslake 2009), Hartry Field (Field 2003), John Earman (Earman 2011), and, to some extent, John Norton (J. D. Norton 2003; 2007; J. Norton 2009).[2] While Russell's attack was a broad attack on the notion of cause in general, recent neo-Russellians seem to follow Mach in focusing their attention on physics and argue that there is something distinct about physics that makes physics especially inhospitable to causal notions and, hence, to causal explanations.

Three Machian or Russellian arguments for why causal notions cannot have a legitimate place in physics have been particularly influential:

(i) The notions of cause and effect are inherently vague in contradistinction to the mathematical precision of derivations in physics. This vagueness infects especially metaphysically 'rich' notions of causal production, which also sit ill with a broadly empiricist outlook.

(ii) Causal explanations are legitimate in contexts in which we can isolate a small set of factors of interest as those responsible for a phenomenon or for the occurrence of an event and implicit in the notion of cause is a distinction between causes and background conditions. This distinction cannot be drawn in physics, where the character of basic physical equations requires that we take the complete backward lightcone of an event as its cause.

---

[2] See also the essays collected in (Price et al. 2007) and especially Woodward's essay (Woodward 2007). For a critical discussion of neo-Russellian arguments see (Frisch 2016). (Norton 2009) is part of a critical exchange on the role of causal notions in the derivation of dispersion relations (Frisch 2009b; 2009a). Russell himself, it is often forgotten, changed his mind about the role of causation. For example, in *The Analysis of Matter*, he says that “all science rests upon induction and causality" (Russell 1992)

(iii) The notion of cause is time-asymmetric, whereas the dynamical laws of the fundamental or established theories of physics have the same character in both temporal directions and are bi-deterministic.

The Machian and Russellian challenges imply that there can be no causal explanations in physics. Thus, neo-Russellians either would have to adopt a different, non-causal account of explanation in physics or would have to argue that explanation is an inherently pragmatic, context-dependent notion that may serve an important purpose as an external add-on to scientific investigations but has no place in physics proper.

3.2. Answering the Challenge: Structural Models and Interventionism

The first argument presents a challenge to any broadly empiricist theory of scientific explanation. Yet it is a challenge that arguably can be—and in recent years has been—met by the Bayes net or structural model accounts of causation developed by Peter Spirtes and his co-authors (Spirtes, Glymour, and Scheines 2000) and by Judea Pearl (Pearl 2000)(Pearl 2009). These accounts provide mathematically rigorous and precise representations of causal structures. On Pearl's account, a structural causal model (SCM) consists of:

(i) a directed acyclic graph (which can visually be represented in terms of a 'blobs-and-arrows' diagram) over a set of variables $V=\{X, Y, \ldots\}$ consisting of both endogenous variables V_i and exogenous variables U_i;

(ii) structural equations $x_i = f_i(pa_i, u_i)$, which specify the value of each variable x_i in terms of the value of the variable's causal parents pa_i and random exogenous disturbances u_i;

(iii) and a probability distribution $P(u_i)$ over the values u_i of the exogenous variables U_i, which induces a probability distribution over all variables.

There are two aspects of the structural model account that are particularly relevant to the issue of explanation. First, it is part of the definition of SCMs that the exogenous variables are probabilistically independent. From this together with the assumption that a causal model is complete one can derive the causal Markov condition, which states that for every variable X in V, X is probabilistically independent of the variables in the set (V – Descendants(X)) conditional on the parents of X. The causal Markov condition is a generalized common cause condition. SCMs, thus, underwrite common cause explanations.

Second, SCMs make perspicuous the tight connection between the notions of cause and intervention or manipulation. A causal model provides us with information on how the values of variables change under external interventions into a system.[3] And causal discovery algorithms allow us to construct causal models from information about probability distributions over the values of variables characterizing the system and from information about the effects of interventions. James Woodward has shown how this formal framework can be developed into a philosophical account of causation and of explanation (Woodward 2003). On Woodward's account, to explain a phenomenon is to exhibit systematic patterns of counterfactual dependency: explanations allow us to answer *what-if-things-had-been-different questions* or *w-questions* (Woodward 2003, 191). Since counterfactuals are in the first instance to be interpreted in terms of possible interventions into a system, w-questions are primarily requests for causal information. But Woodward's account allows for non-causal explanations as well. If the counterfactual changes in question cannot be interpreted as possible interventions, then an answer to a w-question can provide a non-causal explanation.

Like the *DN*-model and the conserved quantity account, Woodward's account of explanation seems to be motivated by features of explanatory practices in physics. While the *DN* model focuses on mathematical derivability and the conserved quantity account zeroes in on one particular albeit important feature of physical theories, Woodward's counterfactual account emphasizes the fact that physical equations constitute relations among *mathematical functions*, which provide us with information on how changing the value of one variable affects the

---

[3] There are various different notions of interventions that have been proposed in the literature: arrow-breaking or hard interventions, either involving intervention variables (in Woodward's account) or not (as in Pearl's *do-calculus*, see (Pearl 2000)) and non-arrow-breaking or soft interventions, as investigated in (Eberhardt and Scheines 2007).

values of other variables. Physical theories allow us to answer w-questions by representing a phenomenon as embedded into a pattern of functional dependencies.

(Frisch 2016) argues that many physical theories provide a mathematical machinery that appears to be tailor-made for the structural model account. Any linear differential operator L associated with an inhomogeneous differential equation $Ly=f(x)$ with constant coefficients possesses a *fundamental solution* or *Green's function G*, which is a solution to the inhomogeneous differential equation $LG = \delta(x)$. The Green function is quite naturally interpreted in causal or interventionist terms. The Green's function 'propagates a point-inhomogeneity' and thereby tells us what the contribution of introducing a disturbance or perturbation into a system at (x', t') is to the state of the system at some other point (x, t). Thus, Green's functions are a natural candidate in physics for the structural equations in SCMs.[4]

What about the two remaining Russellian challenges I distinguished above? An argument appealing to the considerations described in (iii) has been defended by Field (Field 2003). Paradigmatically causal explanations, Field claims, point to a small number of factors through which one could (at least in principle) manipulate a phenomenon. Field and others take this observation to suggest that the distinction between causes and background condition is an essential component of our concept of cause. But, the argument continues, this distinction appears to be obliterated in physics. For a large class of equations in physics it is the case that solving these equations requires as input initial data on a complete cross section of the backward lightcone of the spatial region occupied by the system of interest – that is, complete data in a region from which influences could reach the system by traveling at most at the speed of light – nothing less will do.

But if the set of an event's causes becomes too large, we seem to be committed to claims that conflict with central intuitions concerning the assertability of causal claims. Field asks us to consider a scenario in which Sara puts out a fire with a water-hose while Sam sits next to Sara praying for the fire to go out. It seems

---

[4] For a more critical view on the causal role of Green functions, see (Smith 2013), which is a criticism of (Frisch 2009b; 2009a). (Frisch 2016) contains a reply to Smith.

obvious that Sara's spraying the fire with water but not Sam's praying is a cause of the fire going out. Yet Sam's praying is in the backward lightcone of the fire's going out. Thus, if we interpreted an event's physical determination by cross sections of its backward lightcone causally, then Sam's praying next to a fire would come out as a cause of the fire's extinction just as much as Sara's aiming a hose at the fire does. This, Field maintains is absurd. Yet physics provides no additional means for distinguishing causally salient factors from other factors within an event's backward lightcone.

In reply to this argument Woodward (2007) has argued that it is important to keep distinct different representations at different levels of grain. In a putatively complete microphysical representation of the fire and its surroundings, the *precise microphysical realization* of Sam's prayer will come out as a cause of the precise microphysical realization of the fire's extinction. By contrast, a more *coarse-grained macrophysical representation* of the fire's extinction will be counterfactually independent of a broad class of changes to Sam's macrostate, including, for example, whether he prays or just sits and watches Sara's rushing to put out the fire. At both levels our intuitive causal judgments are preserved (see also Frisch 2014, ch. 3).

Taking a step back, Field's worry and the neo-Russellians' position more generally are motivated by a way of thinking about physics that is quite common in philosophy: physics, it is assumed, ultimately presents us with global dynamical models of fundamental laws. To many philosophers it seems difficult to conceive how causal models and, in particular an interventionist conception of cause, can get a foothold within this conception.[5] Yet there is an alternative conception that arguably fits much of the day-to-day practice of physics considerably better than does the globalists' picture and that can readily accommodate causal reasoning. On this second conception, the laws of physics are understood as rules governing localized subsystems of the universe (Ismael 2016). Viewed from this perspective

---

[5] (Frisch 2014, ch. 4) argues that interventionist causal notions can be introduced even within a globalist conception of physics.

the debate concerning the place of causal notions in physics is intimately tied in with debates concerning the aims of theorizing in physics.

Perhaps the most influential argument for the claim that causal notions cannot play a legitimate role in physics appeals to the asymmetry of the causal relation, which is often taken to coincide with a temporal asymmetry according to which effects do not precede their causes. Since, as it is claimed the basic laws of physics are time-reversal invariant and have the same character in both the past and future direction, time-asymmetric causal structures are an illegitimate and epistemically not justifiable add-on to our physical theories (Price and Weslake 2009; Earman 2011). Russell is often interpreted as making this claim when he says that "in the motion of mutually gravitating bodies, there is nothing that can be called a cause and nothing that can be called an effect; there is merely a formula" (Russell 2012, 141).

The argument has an analogue in discussions of the proper interpretation of the Green's function formalism. At least in the case of systems governed by so-called *hyperbolic equations*, one can represent one and the very same system in terms of a "causal" Green's function and in terms of its temporal inverse, an "anti-causal" Green's function. The first representation suggests that disturbances in a system propagate into the future, while the second representation suggests causal propagation into the past. Interpreting both representations causally threatens to result in a contradiction. But nothing can legitimately distinguish between the two representations. Hence, neither ought to be interpreted causally.

Its popularity notwithstanding, the argument can be challenged. First, the argument applies only to deterministic theories and among these arguably only to time-symmetric theories.[6] Theories with probabilistic state-transition laws are inherently time-asymmetric, as shown by Satosi Watanabe (Watanabe 1965).[7] Thus, if quantum mechanics is understood as a fundamentally probabilistic theory,

---

[6] Dynamical equations with a damping term, which are common in linear response theory, possess a unique, causal Green's function. (See Frisch 2016, ch. 6)

[7] See also (Callender 2000)

the argument's scope is limited to what by the argument defenders' own lights are the less fundamental theories of classical physics.

Second, in concluding that there is no place for time-asymmetric causal relations in a theory with time-reversal invariant laws, the argument presupposes that the content of physics is exhausted by its dynamical equations and ignores the role of initial or final conditions. Generally there is an asymmetry between prevailing initial and final conditions: initial conditions are random while final conditions are not. And this asymmetry, some philosophers argue, is intimately related to the causal asymmetry (Arntzenius 1992),(Maudlin 2007)(Frisch 2016, ch.5). For an initial randomness assumption allows us to engage in common cause reasoning (see, e.g. Pearl 2001). Thus an asymmetry between prevailing initial and final conditions allows us to introduce a causal asymmetry and allows us to distinguish between the causal and anti-causal Green's functions: Since a putatively causal model constructed from a representation of a system in terms of the anti-causal Green's function would violate the independence assumption—an anti-causal world would have to satisfy a final independence assumption instead of an initial independence assumption—this assumption allows us to pick out the causal Green's function as providing the causally correct representation.

Arguably causal reasoning, and in particular common-cause reasoning, is a central and ineliminable inference pattern in physics (and elsewhere), since it allows inferences based on local data rather than on full knowledge of the state of the world on a full initial or final value surface to which we often do not have access. As a particularly stark example consider the detection of gravitational waves in 2016. The extremely strong correlations between the signals detected in the two LIGO detectors in Washington and Louisiana are part of the evidence for the colliding black holes as the signals' common cause. Implicit in the inference from the detected signals to the collision event as their cause is the assumption that there was no 'carefully calibrated' gravitational wave coming in from past infinity, converging on the location of the black holes and re-diverging, thereby mimicking a wave produced by the collapsing black holes. We rule out this alternative explanation of the signals detected by LIGO as utterly implausible, because a source free gravitational field that mimicked the field associated with the black hole event would have required absurdly strongly correlated initial conditions, in violation of the randomness assumption. By contrast, we do not find 'absurdly strongly correlated' final conditions implausible: correlated final conditions are just what we would expect as joint effects of a common cause such as the collapsing of the black holes. If we wanted to derive the black hole event from knowledge of the data on a complete final value surface, we would have to know the precise state of the universe in a sphere with a diameter of many light-years -- something that is obviously impossible.[8]

Deterministic laws appear to undercut time-asymmetric common cause inferences for another reason, however. Under determinism, if there is an event *C* in the past of two events *A* and *B* that screens *A* and *B* off from each other, then there will also be an event *C** that occurs after *A* and *B* and renders the two events conditionally independent (Arntzenius 1992). This threatens our ability to apply Pearl-style SCMs to physics. In particular, if the existence of earlier screening-off events also implies the existence of later screening-off events, then we cannot rely on causal discovery algorithms to infer causal relations from probabilistic dependencies and independencies. A common reply to this worry is to point out that future screening-off events, unlike those in the past, will in general be highly non-natural and non-localized (see, e.g., Woodward 2007). Demanding that appropriate physical variables represent localized and not highly gerrymandered events allows us to preserve the asymmetry induced by the initial randomness assumption.

The initial randomness assumption is the very same assumption that underwrites the temporal asymmetry of statistical mechanics. Does this mean that the causal asymmetry and the thermodynamic asymmetry have the same origin? Two viewpoints seem possible: one can take the initial randomness assumption as fundamental and as the common origin of both the causal and thermodynamic

---

[8] See also the discussion in (Albert 2015), even though Albert does not identify the kind of inferences that an initial randomness assumption makes possible as causal inferences.

asymmetry. Alternatively, one can understand the initial randomness assumption to reflect a fundamental causal asymmetry: initial states are distributed randomly precisely because (and just in case when) these states do not have common causes in their pasts that result in correlations.

One open question for an account of causal explanations in physics is to what extent structural causal models can be applied to quantum systems. To be sure, common cause explanations and a microscopic randomness assumption also play a prominent role in explaining many quantum phenomena. For example, we explain why pure absorptions of a photon by an atom are much more rare than pure emissions, by appealing to the fact that the former would require photons that are finely tuned to one of the atoms excitation energies. By contrast, the energy of an emitted photon is of course always 'finely tuned' to the excitation energies of an atom as the emission's cause (Atkinson 2006). But there are quantum phenomena that apparently cannot be represented in terms of local causal model. In particular, quantum entanglement poses a special challenge for causal explanation. Outcomes of measurements on entangled states that are spatially separated from each other are correlated, but these correlations cannot be explained by a localized common cause model satisfying the causal Markov condition, as Bell's theorem shows. One version of the theorem says that there are quantum phenomena for which there is no model satisfying local causality (Wiseman and Cavalcanti 2015). Local causality is a screening-off condition that states that the outcome B at one wing of the experiment is independent of the outcome A and measurement settings a at the other wing, conditional on the state preparation c, the measurement setting at the first wing b, and any hidden variable λ:

$$P(B|A; a; b; c; \lambda) = P(B|b; c; \lambda)$$

One can respond to Bell's theorem by giving up the Markov condition and allow for quantum common causes that do not screen off their effects from each other. Another response is to give up the prohibition against superluminal causation. This could take the form of positing either a direct causal link between the two wings of the experiment or a partially retro-causal connection that 'zigzags' down and up the lightcone centered on the preparation event c (Price 2012).

How to causally model entangled states remains an unsolved question, but recent years have seen an increasing number of attempts to extend the framework of SCMs to quantum mechanics (Wood and Spekkens 2015).

4. Causal Imperialism

Neo-Russellians deny that causal explanations play a fundamental role in physics. We have just seen that it is possible to resist their arguments. If we allow that there are causal explanations in physics, should we conclude that all explanations are causal? David Lewis thought so. In (Lewis 1986) he proposes that to explain is to provide information about the causal history of an explanandum (Skow 2014).

Despite their stark disagreement, neo-Russellians and "causal imperialists", as we might call them, share a commitment to what Woodward has called "the hidden structure strategy" (Woodward 2017). Both views are committed to the existence of what Peter Railton has called an "ideal explanatory text" (Railton 1981) that contains all the information relevant to a complete explanation of some phenomenon. While actual explanations may fall short of providing us with the complete information contained in the ideal explanatory text, they are explanatory in virtue of providing us with some information about the text.

For the neo-Russellian, the fundamental explanatory structures consist of microphysically complete dynamical models of the backward lightcone of a given explanandum. While the neo-Russellian view is compatible with the claim that in some non-fundamental domains and for pragmatic reasons information about the ideal explanatory text may fruitfully be presented in causal terms, the view holds that ideal physical explanations are not causal. Causal imperialism turns this picture on its head and maintains that the underlying ideal explanatory structures are causal structures. Hence all explanations are causal in virtue of the fact that they provide information about this structure, even though the information provided in an actual explanation may not be presented in explicitly causal terms.

As Woodward has argued, a problem for the hidden structure strategy is to explain how hidden structures that are epistemically inaccessible to us can account for the explanatory import of the explanatory accounts we actually give (Woodward 2017). For the neo-Russellian the problem is that we seem to be able to provide successful causal explanations of phenomena even when the complete initial data that are part of the ideal explanatory text are in principle inaccessible to us.

The causal imperialist's version of the hidden structure strategy faces an analogous problem. There are apparently successful explanations of phenomena that do not identify causes of the phenomenon. How does pointing to a hidden causal structure make perspicuous the explanatory import of such an explanation and what accounts for the difference between such an explanation and one that does explicitly identify a phenomenon's causes? One may demand that an account of causal explanation be able to relate the explanatory role of such explanations to the function of causal information more generally. As we have seen, Pearl and Woodward's accounts of causation emphasize two features as the characteristic function of causal notions. First, knowledge of causal structures allows us to identify relationships amenable to manipulation and control; and second, common cause reasoning enables us to draw inferences from one time to another even when we possess only incomplete knowledge of the state of a system at a time. Yet as Woodward also points out, not every explanation fulfills either of these functions.

Take an explanation of the heat capacity of metals and, in particular, of the fact that the heat capacity is much lower than predicted classically that appeals to the Pauli exclusion principle. This explanation embeds its explanandum into patterns of functional dependencies in a manner that allow us to see how the heat capacity depends on particle statistics. In order to get the correct result, we need to model free electrons in the metal as satisfying the quantum-mechanical Fermi-Dirac statistics and the exclusion principle. The explanation appeals to properties of the phase-space available to the electrons and allows us to answer how the heat capacity would change if the available phase space were different. Thus, the explanation allows us to answer w-questions but not by specifying counterfactuals that can be interpreted in terms of interventions or manipulations of the electron states.

Arguably, by classifying explanations such as this as causal, the causal imperialist obliterates what is an important distinction between different explanatory functions and epistemic goals. That the value of the heat capacity of metals follows from features of the available phase space and is not something that, even in principle, is open to manipulation or control seems itself to be explanatorily relevant, just as it is crucial to an explanation of the length of the shadow that it can be manipulated by changing the flagpole's heights. This distinction is lost if we classify both these explanations as causal by virtue of the fact that they provide information about the causal history of a sample of metal or of the flagpole.

5. Conclusion

Are there, then, genuinely causal explanations in physics? For many decades the majority response to this question among philosophers of physics appears to have been 'no'. By contrast, after the end of the hegemony of the *DN* model there was considerable support for causal theories of explanation among philosophers of science and metaphysicians more generally. In recent years we may be witnessing a rapprochement of the opposing camps. On the one hand, structural causal models introduced a formally precise, arguably metaphysically 'thin' yet non-reductive notion of cause into philosophy that may be acceptable even to empiricist-minded philosophers of physics. On the other hand, there has been a growing interest in varieties of non-causal explanations in general philosophy of science challenging monolithic causal accounts of explanation (Lange 2016). These developments make room for pluralist positions that accord causal notions and causal explanation a legitimate and even crucial role in physics while at the same time allowing that there are genuinely non-causal explanations in physics and elsewhere. Explanatory pluralism raises several questions, however, which point in directions for future research. Allowing for different models of explanation reopens the problem of explanatory asymmetries: if it is not a necessary condition for explanations to identify causes of the explanandum, does the problem of explanatory asymmetries reemerge? Can the problem be solved if we take causal explanations to occupy a privileged position in a theory of explanation, as Woodward's counterfactual account appears to do? Do different explanatory strategies reflect different epistemic goals? Is it nevertheless possible to give a unifying account of different explanatory accounts or at the very least to identify features shared by different types of explanation? Here the close conceptual links between the notions of explanation and understanding (Regt and Dieks 2005) may provide some clues that may help in identifying common features of different models of explanation.

References

Albert, David. 2015. *After Physics.* Harvard University Press.

Arntzenius, Frank. 1992. "The Common Cause Principle." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992: 227–37.

Atkinson, David. 2006. "Does Quantum Electrodynamics Have an Arrow of Time?" *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 37 (3): 528–41. https://doi.org/10.1016/j.shpsb.2005.03.003.

Brading, Katherine, and Elena Castellani, eds. 2003. *Symmetries in Physics: Philosophical Reflections*. 1 edition. Cambridge, U.K. ; New York: Cambridge University Press.

Bromberger, Sylvain. 1966. "Why Questions." In *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, edited by Robert Garland Colodny and Herbert Alexander Simon. University Press of America.

Callender, Craig. 2000. "Is Time 'Handed' in a Quantum World?" *Proceedings of the Aristotelian Society* 100: 247–69.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford University Press.

Dowe, Phil. 2000. *Physical Causation*. Cambridge ; New York: Cambridge University Press.

Earman, John. 1986. *A Primer on Determinism*. Springer Science & Business Media.

———. 2011. "Sharpening the Electromagnetic Arrow(s) of Time." In *The Oxford Handbook of Philosophy of Time*, edited by Craig Callender. Oxford University Press. http://www.oxfordhandbooks.com.proxy-um.researchport.umd.edu/view/10.1093/oxfordhb/9780199298204.001.0001/oxfordhb-9780199298204-e-17.

Eberhardt, Frederick, and Richard Scheines. 2007. "Interventions and Causal Inference." *Philosophy of Science* 74 (5): 981–995.

Field, Hartry. 2003. "Causation in a Physical World." In *The Oxford Handbook of Metaphysics*, edited by Michael J. Loux and Dean W. Zimmerman, 435–60. Oxford University Press.

Fraassen, Bas C. van. 1980. *The Scientific Image*. Oxford : New York: Oxford University Press.

Frisch, Mathias. 2009a. "Causality and Dispersion: A Reply to John Norton." *The British Journal for the Philosophy of Science* 60 (3): 487–95.

———. 2009b. "'The Most Sacred Tenet'? Causal Reasoning in Physics." *The British Journal for the Philosophy of Science* 60 (3): 459–74. https://doi.org/10.1093/bjps/axp029.

———. 2016. *Causal Reasoning in Physics*. Cambridge University Press.

Hempel, Carl G. 1970. *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science*. New York; London: Free Press.

Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15 (2): 135–175.

Ismael, J. T. 2016. *How Physics Makes Us Free*. 1 edition. New York, NY, United States of America: Oxford University Press.

Lange, Marc. 2016. *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. 1 edition. New York, NY, United States of America: Oxford University Press.

Lewis, David. 1986. "Causal Explanation." In *Philosophical Papers Vol. Ii*, edited by David Lewis, 214–240. Oxford University Press.

Mach, Ernst. 1900. *Die principien der wärmelehre. Historisch-kritisch entwickelt,*. Leipzig: J.A. Barth.

———. 1905. *Erkenntnis und Irrtum. Skizzen zur Psychologie der Forschung.* Leipzig: Barth.

Maudlin, Tim. 2007. *The Metaphysics Within Physics*. Oxford ; New York: Oxford University Press.

Norton, John. 2009. "Is There an Independent Principle of Causality in Physics?" *British Journal for the Philosophy of Science* 60 (3): 475–486.

Norton, John D. 2003. "Causation as Folk Science." *Philosophers' Imprint* 3 (4): 1–22.

———. 2007. "Do the Causal Principles of Modern Physics Contradict Causal Anti-Fundamentalism?" In *Thinking about Causes: From Greek Philosophy to Modern Physics*, edited by P.K. Machamer and G. Wolters, 222–34. Pittsburgh: Pittsburgh University Press.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, U.K.; New York: Cambridge University Press.

———. 2009. *Causality : Models, Reasoning, and Inference*. 2nd. Cambridge U.K. ;;New York: Cambridge University Press.

Price, Huw. 1997. *Time's Arrow & Archimedes' Point : New Directions for the Physics of Time*. New York: Oxford University Press.

———. 2012. "Does Time-Symmetry Imply Retrocausality? How the Quantum World Says 'Maybe'?" *Studies in History and Philosophy of Science Part B* 43 (2): 75–83.

Price, Huw, Richard Corry, University of Sydney, and Centre for Time. 2007. *Causation, Physics, and the Constitution of Reality Russell's Republic Revisited*. Oxford; New York: Clarendon Press ; Oxford University Press.

Price, Huw, and Brad Weslake. 2009. "The Time-Asymmetry of Causation." In *The Oxford Handbook of Causation*. Oxford University Press.

Railton, Peter. 1981. "Probability, Explanation, and Information." *Synthese* 48 (2): 233–256.

Regt, Henk W. De, and Dennis Dieks. 2005. “A Contextual Approach to Scientific Understanding.” *Synthese* 144 (1): 137–70. https://doi.org/10.1007/s11229-005-5000-4.

Reichenbach, Hans. 1956. *The Direction of Time*. Berkeley: University of California Press.

Russell, Bertrand. 1912. “On the Notion of Cause.” *Proceedings of the Aristotelian Society* 7: 1–26.

———. 1992. *The Analysis of Matter*. Psychology Press.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. 1st US-1st Printing edition. Princeton, N.J: Princeton University Press.

———. 2006. *Four Decades of Scientific Explanation*. 1 edition. Pittsburgh: University of Pittsburgh Press.

Scriven, Michael. 1962. “Explanations, Predictions, and Laws.” In *Scientific Explanation, Space, and Time*, edited by Herbert Feigl and Grover Maxwell, 1st edition. Minneapolis: University of Minnesota Press.

Skow, Bradford. 2014. “Are There Non-Causal Explanations (of Particular Events)?” *The British Journal for the Philosophy of Science* 65 (3): 445–67. https://doi.org/10.1093/bjps/axs047.

Smith, Sheldon R. 2013. “Causation in Classical Mechanics.” In *The Oxford Handbook of Philosophy of Physics*, edited by Robert Batterman, 107–40. Oxford University Press.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines. 2000. *Causation Prediction & Search 2e*. MIT Press.

Van Fraassen, Bas C. 1989. *Laws and Symmetry*. Oxford; New York: Oxford University Press.

Watanabe, S. 1965. “Conditional Probability in Physics.” *Progress of Theoretical Physics Supplement* 65: 135–60.

Wiseman, Howard M., and Eric G. Cavalcanti. 2015. “Causarum Investigatio and the Two Bell’s Theorems of John Bell,” March. http://arxiv.org/abs/1503.06413.

Wood, Christopher J., and Robert W. Spekkens. 2015. “The Lesson of Causal Discovery Algorithms for Quantum Correlations: Causal Explanations of Bell-Inequality Violations Require Fine-Tuning.” *New Journal of Physics* 17 (3): 033002. https://doi.org/10.1088/1367-2630/17/3/033002.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

———. 2007. “Causation with a Human Face.” In *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, edited by Huw Price and Richard Corry. Oxford University Press.

———. 2017. “Scientific Explanation.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2017/entries/scientific-explanation/.