Stanford Encyclopedia of Philosophy

Causal Approaches to Scientific Explanation

First published Fri Mar 17, 2023

This entry discusses some accounts of *causal* explanation developed after approximately 1990. For a discussion of earlier accounts of explanation including the deductive-nomological (DN) model, Wesley Salmon's statistical relevance and causal mechanical models, and unificationist models, see the general entry on scientific explanation. Recent accounts of non-causal explanation will be discussed in a separate entry. In addition, a substantial amount of recent discussion of causation and causal explanation has been conducted within the framework of causal models. To avoid overlap with the entry on causal models we do not discuss this literature here.

Our focus in this entry is on the following three accounts – Section 1 those that focus on mechanisms and mechanistic explanations, Section 2 the kairetic account of explanation, and Section 3 interventionist accounts of causal explanation. All of these have as their target explanations of *why* or perhaps *how* some phenomenon occurs (in contrast to, say, explanations of *what* something is, which is generally taken to be non-causal) and they attempt to capture causal explanations that aim at such explananda. Section 4 then takes up some recent proposals having to do with how causal explanations may differ in explanatory depth or goodness. Section 5 discusses some issues having to do with what is distinctive about causal (as opposed to non-causal) explanations.

We also make the following preliminary observation. An account of causal explanation in science may leave open the possibility that there are other sorts of explanations of a non-causal variety (it is just that the account does not claim to capture these, at least without substantial modifications) or it may, more ambitiously, claim that all explanations of the why/how variety are, at least in some extended sense, causal. The kairetic model makes this latter claim, as do many advocates of mechanistic models. By contrast, interventionist models, need not deny that there are non-causal explanations, although the version described below does not attempt to cover such explanations. Finally, we are very conscious that, for reasons of space, we omitted many recent discussions of causal explanation from this entry. We provide brief references to a number of these at the end of this article (Section 6).

---

1. Mechanisms and Mechanistic Explanations

Many accounts of causation and explanation assign a central importance to the notion of mechanism. While discussions of mechanism are present in the early modern period, with the work of Descartes and others, a distinct and very influential research program emerged with the "new mechanist" approaches of the late twentieth and early twenty-first century. This section focuses on work in this tradition.

Wesley Salmon's causal mechanical (CM) model of explanation (Salmon 1984) was an influential late twentieth century precursor to the work on mechanisms that followed. The CM model is described in the SEP entry on scientific explanation and readers are referred to this for details. For present purposes we just note the following. First, Salmon's model is proposed as an alternative to the deductive-nomological (DN) model and the "new mechanist" work that follows also rejects the DN model, although in some cases for reasons somewhat different from Salmon's. Like the CM model and in contrast to the DN model, the new mechanist tradition downplays the role of laws in explanation, in part because (it is thought) there are relatively few laws in the life sciences, which are the primary domain of application of recent work on mechanisms. Second, although Salmon provides an account of causal relationships that are in an obvious sense "mechanical", he focuses virtually entirely on physical examples (like billiard ball collisions) rather than examples from the life sciences. Third Salmon presents his model as an "ontic" account of explanation, according to which explanations are things or structures in the world and contrasts this with what he regarded as "epistemic" accounts of explanation (including in his view, the DN model) which instead conceive of explanations as representations of what is in the world (Salmon 1984). This "ontic" orientation has been important in the work of some of the new mechanists, such as Craver (2007a), but less so for others. Finally, Salmon's model introduces a distinction between the "etiological" aspects of explanation which have to do with tracing the causal history of some event *E* and the "constitutive" aspects which have to do with "the internal causal mechanisms that account for *E*'s nature" (Salmon 1984: 275). This focus on the role of "constitution" is retained by a number of the new mechanists.

We may think of the "new mechanism" research program properly speaking as initiated by writers like Bechtel and Richardson (1993 [2010]), Glennan (1996, 1997), and Machamer, Darden, and Craver (2000). Although these writers provide accounts that differ in detail,[1] they share common elements: mechanisms are understood as causal systems, exhibiting a characteristic organization, with multiple causal factors that work together in a coordinated manner to produce some effect of interest. Providing a mechanistic explanation involves explaining an outcome by appealing to the causal mechanism that produces it. The components of a mechanism stand in causal relationships but most accounts conceptualize the relationship between these components and the mechanism itself as a part-whole or "constitutive relationship" – e.g., a human cell is constituted by various molecules, compounds and organelles, the human visual system is constituted by various visual processing areas (including V1–V5) and an automobile engine may be constituted by pistons, cylinders, camshaft and carburetor, among other components. Such part/whole relations are generally conceptualized as non-causal – that is, constitution is seen as a non-causal relationship. Thus, on these accounts, mechanisms are composed of or constituted by lower-level causal parts that interact together to produce the higher-level behavior of the (whole) mechanism understood as some effect of interest. This part-whole picture gives mechanistic explanation a partially reductive character, in the sense that higher-level outcomes characterizing the whole mechanism are explained by the lower-level causes that produce them. In many accounts this is depicted in nested, hierarchical diagrams describing these relations between levels of mechanisms (Craver 2007a).

Although philosophical discussion has often focused on the role of constitutive relations in mechanisms and how best to understand these, it is, as noted above, also common to think of mechanism as consisting of factors or components that stand in causal ("etiological") relations to one another with accompanying characteristic spatial, temporal or geometrical organization. This feature of mechanism and mechanistic explanation is emphasized by Illari and Williamson (2010, 2012) and Woodward (2002, 2013). In particular, elucidating a mechanism is often understood as involving the identification of "mediating" factors that are

"between" the input to the mechanism and its eventual output – "between" both in the sense of causally between and in the sense that the operation of these mediating factors often can be thought of as spatially and temporally between the input to the mechanism and its output. (The causal structure and the spatiotemporal structure thus "mirror" or run parallel to each other.) Often this information about intermediates can be thought of as describing the "steps" by which the mechanism operates over time. For example, mechanistic explanations of the action potential will cite the (anatomical) structure of the neural cell membrane, the relative location and structure of ion channels (in this membrane), ion types on either side of this membrane, and the various temporal steps in the opening and closing of ion channels that generate the action potential. A step-by-step description of this mechanism cites all of these parts and their interactions from the beginning of the causal process to the end. In this respect a description of a mechanism will provide more detail than, say, directed acyclic graphs which describe causal relations among variables but do not provide spatio-temporal or geometrical information.

A hotly debated issue in the literature on mechanisms concerns the amount of detail descriptions of mechanisms or mechanistic explanations need to contain. While some mechanists suggest that mechanisms (or their descriptions) can be abstract or lacking in detail (Levy & Bechtel 2012), it is more commonly claimed that mechanistic explanations must contain significant detail – perhaps as much "relevant" detail as possible or at least that this should be so for an "ideally complete description" of a mechanism (see Craver 2006 and the discussion in Section 4). Thus, a mere description of an input-output causal relation, even if correct, lacks sufficient detail to count as a description of a mechanism. For example, a randomized control trial can support the claim that drug *X* causes recovery *Y*, but this alone doesn't elucidate the "mechanism of action" of the drug. Craver (2007a: 113–4) goes further, suggesting that even models that provide substantial information about anatomical structures and causal intermediaries are deficient *qua* mechanistic explanations if they omit detail thought to be relevant. For example, the original Hodgkin-Huxley (HH) model of the action potential identified a role for the opening and closing of membrane channels but did not specify the molecular mechanisms involved in the opening and closing of those channels. Craver (2006, 2007a, 2008) takes this to show that the HH model is explanatorily deficient – it is a "mechanism sketch" rather than a fully satisfactory mechanistic explanation. (This is echoed by Glennan who states that the monocausal model of disease – a one cause-one effect relationship – is "the sketchiest of mechanism sketches" [Glennan 2017: 226].) This "the more relevant detail the better" view has in turn been criticized by those who think that one can sometimes improve the quality of an explanation or at least make it no worse by omitting detail. For such criticism see, e.g., Batterman and Rice (2014), Levy (2014), Chirimuuta (2014), Ross (2015, 2020), etc. and for a response see by Craver and Kaplan (2020).[2]

The new mechanists differ among themselves in their views of causation and their attitudes toward general theories of causation found in the philosophical literature. Since a mechanism involves components standing in causal relations, one might think that a satisfactory treatment of mechanisms should include an account of what is meant by "causal relations". Some mechanists have attempted to provide such an account. For example, Craver (2007a) appeals to elements of Woodward's interventionist account of causation in this connection and for other purposes – e.g., to provide an account of constitutive relevance (Craver 2007b). By contrast, Glennan (1996, 2017) argues that the notion of mechanism is more fundamental than that of causation and that the former can be used to elucidate the latter – roughly, *X* causes *Y* when there is a mechanism connecting *X* to *Y*. Of course, for Glennan's project this requires that mechanism is elucidated in a way that doesn't appeal to the notion causation. Yet another view, inspired by Anscombe (1971) and advocated by Machamer, Darden, and Craver (MDC) (2000), Machamer (2004) and others, eschews any appeal to general theories of causation and instead describes the causal features of mechanism in terms of specific causal verbs. For example, according to MDC, mechanisms involve entities that engage in "activities", with examples of the latter including "attraction", "repulsion", "pushing" and so on (MDC 2000: 5). It is contended that no more general account according to which these are instances of some common genus (causation) is likely to be illuminating. A detailed evaluation of this claim is beyond the scope of this

entry, but we do wish to note that relatively general theories of causation that go beyond the cataloging of particular causal activities now flourish not just in philosophy but in disciplines like computer science and statistics (Pearl 2000 [2009]; Morgan & Winship, 2014) where they are often thought to provide scientific and mathematical illumination.

Another issue raised by mechanistic accounts concerns their scope. As we have seen these accounts were originally devised to capture a form of explanation thought to be widespread in the life sciences. This aspiration raises several questions. First, are all explanations in the life sciences "mechanistic" in the sense captured by some model of mechanistic explanation? Many new mechanists have answered this question in the affirmative but there has been considerable pushback to this claim, with other philosophers claiming that there are explanations in the life sciences that appeal to topological or network features (Lange 2013; Huneman 2010; Rathkopf 2018; Kostić 2020; Ross 2021b), to dynamical systems models (Ross 2015) and to other features deemed "non-mechanical" as with computational models in neuroscience (Chirimuuta 2014, 2018). This debate raises the question of how broadly it is appropriate to extend the notion of "mechanism" (Silberstein & Chemero 2013).

While the examples above are generally claimed to be non-causal and non-mechanistic, a further question is whether there are also types of *causal* explanation that are non-mechanistic. Answering this question depends, in part, on how "mechanism" is defined and what types of causal structures count as "mechanisms". If mechanisms have the particular features mentioned above – part-whole relationships, some significant detail, and mechanical interactions – it would seem clear that some causal explanations are non-mechanistic in the sense that they cite causal systems and information with different features. For example, causal systems including pathways, networks, and cascades have been advanced as important types of causal structures that do not meet standard mechanism characteristics (Ross 2018, 2021a, forthcoming). Other examples include complex causal processes that lack machine-like and fixed causal parts (Dupré 2013). This work often questions whether "mechanism" fruitfully captures the diversity of causal structures and causal explanations that are present in scientific contexts.

There is an understandable tendency among mechanists to attempt to extend the scope of their accounts as far as possible but presumably the point of the original project was that mechanistic explanations have some *distinctive* features. Extending the models too far may lead to loss of sight of these. The problem is compounded by the fact that "mechanism" is used in many areas of science as general term of valorization or approval, as is arguably the case for talk of the "mechanism" of natural selection or of "externalizing tendencies" as a "mechanism" leading to substance abuse. The question is whether these candidates for mechanisms have enough in common with, say, the mechanism by which the action potential is produced to warrant the treatment of both by some common model. Of course, this problem also arises when one considers the extent to which talk of mechanisms is appropriate outside of the life sciences. Chemists talk of mechanisms of reaction, physicists of the Higgs mechanism, and economists of mechanism design, but again this raises the question of whether an account of mechanistic explanation should aspire to cover all of these.

2. The Kairetic Account of Explanation

This account is developed by Michael Strevens in his *Depth* (2008) and in a number of papers (2004, 2013, 2018). Strevens describes his theory as a "two factor" account (Strevens 2008: 4). The first factor – Strevens' starting point – is the notion of causation or dependence (Strevens calls it "causal influence") that figures in fundamental physics. Strevens is ecumenical about what this involves. He holds that a number of different philosophical treatments of causal influence – conserved quantity, counterfactual or interventionist – will fit his purposes. This notion of causal influence is then used as input to an account of causal *explanation* – Strevens' second factor. A causal explanation of an individual event *e* (Strevens' starting point) assembles all

and only those causal influences that make a difference to (are explanatorily relevant to) e. A key idea here is the notion of *causal entailment* (Strevens 2008: 74).[3] A set of premises that causally entail that *e* occurs deductively entail this claim and do this in a way that "mirrors" the causal influences (ascertained from the first stage) leading to *e*. This notion of mirroring is largely left at an intuitive level but as an illustration a derivation of an effect from premises describing the cause mirrors the causal influences leading to the effect while the reverse derivation from effect to cause does not. However, more than mirroring is required for causal explanation: The premises in a causal entailment of the sort just described are subjected to a process (a kind of "abstraction") in which premises that are not necessary for the entailment of *e* are removed or replaced with weaker alternatives that are still sufficient to entail *e* – the result of this being to identify factors which are genuinely difference-makers or explanatorily relevant to *e*. The result is what Strevens calls a "stand-alone" explanation for *e* (Strevens 2008: 70). (Explanatory relevance or difference-making is thus understood in terms of what, so to speak, is minimally required for causal entailment, constrained by a cohesiveness requirement described below, rather than, as in some other models of explanation, in terms of counterfactuals or statistical relevance.) As an illustration, if the event *e* is the shattering of a window the causal influences on *e*, identified from fundamental physics, will be extremely detailed and will consist of influences that affect fine grained features of *e*'s occurrence, having to do, e.g., with exactly how the window shatters. But to the extent that the explanandum is just whether *e* occurs most of those details will be irrelevant in the sense that they will affect only the details of how the shattering occurs and not whether it occurs at all. Dropping these details will result in a derivation that still causally entails e. The causal explanation of *e* is what remains after all such details have been dropped and only what is necessary for the causal entailment of *e* is retained.

As Strevens is fully aware, this account faces the following apparent difficulty. There are a number of different causal scenarios that realize causes of bottle shatterings – the impact of rocks but also, say, sonic booms (cf. Hall 2012). In Strevens' view, we should not countenance causal explanations that disjoin causal models that describe such highly different realizers, even though weakening derivations via the inclusion of such disjunctions may preserve causal entailment. Strevens' solution appeals to the notion of *cohesion*; when different processes serve as "realizers" for the causes of *e*, these must be "cohesive" in the sense that they are "causally contiguous" from the point of view of the underlying physics. Roughly, contiguous causal processes are those that are nearby or neighbors to one another in a space provided by fundamental physics.[4] Sonic booms and rock impacts do not satisfy this cohesiveness requirement and hence models involving them as disjunctive premises are excluded. Fundamental physics is thus the arbitrator of whether upper-level properties with different realizers are sufficiently similar to satisfy the cohesion requirement. Or at least this is so for deep "stand alone" explanations in contrast to those explanations that are "framework" dependent (see below).

As Strevens sees it, a virtue of his account is that it separates difficult ("metaphysical") questions about the nature of the causal relationships (at least as these are found in physics which is Strevens' starting point) from issues about causal explanation, which are the main focus of the kairetic account. It also follows that most of the causal claims that we consider in common sense and in science (outside of fundamental physics) are in fact claims about causal explanation and explanatory relevance as determined by the kairetic abstraction procedure rather than claims about causation per se. In effect when one claims that "aspirin causes headache relief" one is making a rather complicated causal explanatory claim about the upshot of the application of the abstraction procedure to the causal claims that, properly speaking, are provided by physics. This contrasts with an account in which causal claims outside of physics are largely univocal with causal claims (assuming that there are such) within physics.

We noted above that Strevens imposes a cohesiveness requirement on his abstraction procedure. This seems to have the consequence that upper-level causal generalizations that have realizers that are rather disparate from the point of view of the underlying physics are defective *qua* explainers, even though there are many

examples of such generalizations that (rightly or wrongly) are regarded as explanatory. Strevens addresses this difficulty by introducing the notion of a *framework* – roughly a set of presuppositions for an explanation. When scientists "framework" some aspect of a causal story, they put that aspect aside (it is presupposed rather an explicit part of the explanation) and focus on getting the story right for the part that remains. A common example is to framework details of implementation, in effect black-boxing the low-level causal explanation of why certain parts of a system behave in the way they do. The resulting explanation simply presupposes that these parts do what they do, without attempting to explain why. Consequently, the black boxes in such explanations are not subject to the cohesion requirement, because they are not the locus of explanatory attention . Thus although explanations appealing to premises with disparate realizers are defective when considered by themselves as stand-alone explanations, we may regard such explanations as dependent on a framework with the framework incorporating information about a presupposed mechanism that satisfies the coherence constraint.[5] When this is the case, the explanation will be acceptable *qua* frameworked explanation. Nonetheless in such cases the explanation should in principle be deepened by making explicit the information presupposed in the framework.

Strevens describes his account as "descriptive" rather than "normative" in aspiration. Presumably, however, it is not intended as a description of the bases on which lay people or scientists come to accept causal explanations outside of fundamental physics – people don't actually go through the abstraction from fundamental physics process that Strevens describes when they arrive at or reason about upper-level causal explanations. Instead, as we understand his account, it is intended to characterize something like what must be the case from the point of view of fundamental physics for upper-level causal judgments to be explanatory – the explanatory upper-level claims must fit with physics in the right way as specified in Strevens' abstraction procedure and the accompanying cohesiveness constraint.[6] Perhaps then the account is intended to be descriptive in the sense that the upper-level causal explanations people regard as satisfactory do in fact satisfy the constraints he describes. In addition, the account is intended to be descriptive in the sense that it contends that as a matter of empirical fact people regard their explanations as committed to various claims about the underlying physics even if these claims are presently unknown – e.g., to claims about the cohesiveness of these realizers.[7] At the same time the kairetic account is also normative in the sense that it judges that explanations that fail to satisfy the constraints of the abstraction procedure are in some way unsatisfactory – thus people are correct to have the commitments described above.

Depth also contains an interesting treatment of the role of idealizations in explanation. It is often thought that idealizations involve the presence of "falsehoods", or "distortions". Strevens claims that these "false" features involve claims that do not have to do with difference-makers, in the sense captured by the abstraction procedure. Thus, according to the kairetic model, it does not matter if idealizations involve falsehoods or if they omit certain information since the falsehoods or omitted information do not concern difference-makers – their presence thus does not detract from the resulting explanation. Moreover, we can think of idealizations as conveying useful information about which factors are not difference-makers.

The kairetic account covers a great deal more that we lack the space to discuss including treatments of what Strevens calls "entanglement", equilibrium explanations, statistical explanation and much else.

As is always the case with ambitious theories in philosophy, there have been a number of criticisms of the kairetic model. Here we mention just two. First, the kairetic model assumes that all legitimate explanation is causal or at least that all explanation must in some way reference or connect with causal information. (A good deal of the discussion in *Depth* is concerned to show that explanations that might seem to be non-causal can nonetheless be regarded as working by conveying causal information.) This claim that all explanation is causal is by no means an implausible idea – until recently it was widely assumed in the literature on explanation (Skow 2014). Nonetheless this idea has recently been challenged by a number of philosophers (Baker 2005; Batterman 2000, 2002, 2010a; Lange 2013, 2016; Lyon 2012; Pincock 2007). Relatedly, the

kairetic account assumes that fundamental physics is "causal" – physics describes causal relations, and indeed lots of causal relations, enough to generate a large range of upper-level causal explanations when the abstraction procedure is applied. Some hold instead that the dependence relations described in physics are either not causal at all (causation being a notion that applies only to upper-level or macroscopic relationships) or else that these dependence relations lack certain important features (such as asymmetry) that are apparently present in causal explanatory claims outside of physics (Ney 2009, 2016). These claims about the absence of causation in physics are controversial but if correct, it follows that physics does not provide the input that Strevens' account needs.[8]

A second set of issues concern the kairetic abstraction process. Here there are several worries. First, the constraints on this process have struck some as vague since they involve judgments of cohesiveness of realizers from the point of view of underlying physics. Does physics or any other science really provide a principled, objective basis for such judgments? Second, it seems, as suggested above, that upper-level causal explanations often generalize over realizers that are very disparate from the point of view of the underlying physics. Potochnik (2011, 2017) focuses on the example, also discussed by Strevens, of the Lotka-Volterra (LV) equations which are applied to a large variety of different organisms that stand in predator/prey relations. Strevens uses his ideas about frameworks to argue that use of the LV equations is in some sense justifiable, but it also appears to be a consequence of his account (and Strevens seems to agree) that explanations appealing to the LV equations are not very deep, considered as standalone explanations. But, at least as a descriptive matter, Potochnik claims, this does not seem to correspond to the judgments or practices of the scientists using these equations, who seem happy to use the LV equations despite the fact that they fail to satisfy the causal contiguity requirement. Potochnik thus challenges this portion of the descriptive adequacy of Strevens' account. Of course, one might respond that these scientists *ought* to judge in accord with Strevens' account, but as noted above, this involves taking the account to have normative implications and not as merely descriptive.

A more general form of this issue arises in connection with "universal" behavior (Batterman 2002). There are a number of cases in which physical and biological systems that are very different from one another in terms of their low-level realizers exhibit similar or identical upper-level behavior (Batterman 2002; Batterman & Rice 2014; Ross 2015). As a well-known example, substances as diverse as ferromagnets and various liquid/gas systems exhibit similar behavior around their critical points (Batterman 2000, 2002). Renormalization techniques are often thought to explain this commonality in behavior, but they do so precisely by showing that the physical details of these systems do not matter for (are irrelevant to) the aspects of their upper-level behavior of interest. The features of these systems that are relevant to their behavior have to do with their dimensionality and symmetry properties among others and this is revealed by the renormalization group analysis (RGA) (Batterman 2010b). One interesting question is whether we can think of that analysis as an instance of Strevens' kairetic procedure. On the one hand the RGA can certainly be viewed as an abstraction procedure that discards non-difference-making factors. On the other hand, it is perhaps not so clear the RGA respects the cohesiveness requirements that Strevens proposes since the upshot is that systems that are very different at the level of fundamental physics are given a common explanation. That is, the RGA does not seem to work by showing (at least in any obvious way) that the systems to which it applies are contiguous with respect to the underlying physics.[9]

Another related issue is this: a number of philosophers claim that the RGA provides a non-causal explanation (Batterman 2002, 2010a; Reutlinger 2014). As we have seen, Strevens denies that there are non-causal explanations in his extended sense of "causal" but, in addition, if it is thought the RGA implements Strevens' abstraction procedure, this raises the question of whether (contrary to Strevens' expectations) this procedure can take causal information as input and yield a non-causal explanation as output. A contrary view, which may be Strevens', is that as long as the explanation is the result of applying the kairetic procedure to causal input, that result must be causal.

The issue that we have been addressing so far has to do with whether causal contiguity is a defensible requirement to impose on upper-level explanations. There is also a related question – assuming that the requirement is defensible, how can we tell whether it is satisfied? The contiguity requirement as well as the whole abstraction procedure with which it is associated is characterized with reference to fundamental physics but, as we have noted, users of upper-level explanations usually have little or no knowledge of how to connect these with the underlying physics. If Strevens' model is to be applicable to the assessment of upper-level explanations it must be possible to tell, from the vantage point of those explanations and the available information that surrounds their use, whether they satisfy the contiguity and other requirements but without knowing in detail how they connect to the underlying physics. Strevens clearly thinks this is possible (as he should, given his views) and in some cases this seems plausible. For example, it seems fairly plausible, as we take Strevens to assume, that predator/prey pairs consisting of lions and zebras are disparate from pairs consisting of spiders and house flies from the point of view of the underlying physics and thus constitute heterogeneous realizers of the LV equations.[10] On the other hand, in a case of pre-emption in which Billy's rock shatters a bottle very shortly before Suzy's rock arrives at the same space, Strevens seems committed to the claim that these two causal processes are non-contiguous – indeed he needs this result to avoid counting Suzy's throw as a cause of the shattering[11] (Non-contiguity must hold even if the throws involve rocks with the same mass and velocity following very similar trajectories, differing only slightly in their timing.) In other examples, Strevens claims that airfoils of different flexibility and different materials satisfy the contiguity constraint, as do different molecular scattering processes in gases – apparently this is so even if the latter are governed by rather different potential functions (as they sometimes are) (Strevens 2008: 165–6). The issue here is not that these judgments are obviously wrong but rather that one would like to have a more systematic and principled story about the basis on which they are to be made.

That said, we think that Strevens has put his finger on an important issue that deserves more philosophical attention. This is that there is something explanatorily puzzling or incomplete about a stable upper-level generalization that appears to have very disparate realizers: one naturally wants a further explanation of how this comes about – one that does not leave it as a kind of unexplained coincidence that this uniformity of behavior occurs.[12] The RGA purports to do this for certain aspects of behavior around critical points and it does not seem unreasonable to hope for accounts (perhaps involving some apparatus very different from the RGA) for other cases. What is less clear is whether such an explanation will always appeal to causal contiguity at the level of fundamental physics – for example in the case of the RGA the relevant factors (and where causal contiguity appears to obtain) are relatively abstract and high-level, although certainly "physical".

3. Interventionist Theories

Interventionist theories are intended both as theories of causation and of causal explanation. Here we provide only a very quick overview of the former, referring readers to the entry causation and manipulability for more detailed discussion of the former and instead focus on causal explanation. Consider a causal claim (generalization) of the form

(G) "*C* causes *E*"

where "*C*" and "*E*" are variables – that is, they refer to properties or quantities that can take at least two values. Examples are "forces cause accelerations" and "Smoking causes lung cancer". According to interventionist accounts (G) is true if and only if there is a possible intervention *I* such that if *I* were to change the value of *C*, the value of *E* or the probability distribution of *E* would change (Woodward 2003). The notion of an intervention is described in more detail in the causation and manipulability entry, but the basic idea is that this is an unconfounded experimental manipulation of *C* that changes *E*, if at all, via a route that goes

through C and not in any other way. Counterfactuals that describe would happen if an intervention were to be performed are called *interventionist counterfactuals*. A randomized experiment provides one paradigm of an intervention.

Causal explanations can take several different forms within an interventionist framework[13] – for instance, a causal explanation of some explanandum $E = e$ requires:

(3.1) Specification of a true causal generalization $E = f(C)$ – that is, some set of values of C is mapped into values of E by f – meeting the interventionist criterion for causation described above,

as well as

(3.2) One or more initial conditions $C = c$ (C being the independent variable or variables that figure in (3.1)) such that E is derivable from (3.1) and (3.2)

and also meeting the condition

(3.3) When combined with other initial conditions (3.1) tells us how the value e of E would change under changes in the initial conditions in specified in (3.2).

By meeting these conditions (and especially in virtue of satisfying (3.3)) an explanation answers what Woodward (2003) calls "what-if-things-had-been-different questions" (w-questions) about E – it tells us how E would have been different under changes in the values of the C variable from the value specified in (3.2).

As an example, consider an explanation of why the strength (E) of the electrical field created by a long straight wire along which the charge is uniformly distributed is described by $E = \lambda/2\pi r\epsilon_o$ where λ is the charge density and r is the distance from the wire. An explanation of this can be constructed by appealing to Coulomb's law (playing the role of (3.1) above) in conjunction with information about the shape of the wire and the charge distribution along it ((3.2) above). This information allows for the derivation of $E = \lambda/2\pi r\epsilon_o$ but it also can be used to provide answers to a number of other w-questions. For example, Coulomb's law and a similar modeling strategy can be used to answer questions about what the field would be if the wire had a different shape (e.g., if twisted to form a loop) or if it was somehow flattened into a plane or deformed into a sphere.

The condition that the explanans answer a range of w-questions is intended to capture the requirement that the explanans must be explanatorily relevant to the explanandum. That is, factors having to do with the charge density and the shape of the conductor are explanatorily relevant to the field intensity because changes in these factors would lead to changes in the field intensity. Other factors such as the color of the conductor are irrelevant and should be omitted from the explanation because changes in them will not lead to changes in the field intensity. As an additional illustration, consider Salmon's (1971a: 34) example of a purported explanation of (F) a male's failure to get pregnant that appeals to his taking birth control pills (B). Intuitively (B) is explanatorily irrelevant to (F). The interventionist model captures this by observing that B fails to satisfy the what-if-things-had-been-different requirement with respect to F: F would not change if B were to change. (Note the contrast with the rather different way in which the kairetic model captures explanatory relevance.)

Another key idea of the interventionist model is the notion of *invariance* of a causal generalization (Woodward & Hitchcock 2003). Consider again a generalization (G) relating C to E, $E = f(C)$. As we have seen, for (G) to describe a causal relationship at all it must at least be the case that (G) correctly tells how E would change under at least some interventions on C. However, causal generalizations can vary according to the range of interventions over which this is true. It might be that (G) correctly describes how E would

change under some substantial range *R* of interventions that set *C* to different values or this might instead be true only for some restricted range of interventions on *C*. The interventions on *C* over which (G) continues to hold are the interventions over which (G) is *invariant*. As an illustration consider a type of spring for which the restoring force *F* under extensions *X* is correctly described by Hooke's law:

(3.4) $F = -kX$

for some range *R* of interventions on *X*. Extending the spring too much will cause it to break so that its behavior will no longer be described by Hooke's law. (3.4) is invariant under interventions in *R* but not so for interventions outside of *R*. (3.4) is, intuitively, invariant only under a somewhat narrow range of interventions. Contrast (3.4) with the gravitational inverse square law:

(3.5) $F = Gm_1m_2/r^2.$

(3.5) is invariant under a rather wide range of interventions that set m_1, m_2, and r to different values but there are also values for these variables for which (3.5) fails to hold – e.g., values at which general relativistic effects become important. Moreover, invariance under interventions is just one variety of invariance. One may also talk about the invariance of a generalization under many other sorts of changes – for example, changes in background conditions, understood as conditions that are not explicitly included in the generalization itself. As an illustration, the causal connection between smoking and lung cancer holds for subjects with different diets, in different environmental conditions, with different demographic characteristics and so on.[14] However, as explained below, it is invariance under interventions that is most crucial to evaluating whether an explanation is good or deep within the interventionist framework.

Given the account of causal explanation above it follows that for a generalization to figure in a causal explanation it must be invariant under at least some interventions. As a general rule a generalization that is invariant under a wider range of interventions and other changes will be such that it can be used to answer a wider range of w-questions. (See section 4 below.) In this respect such a generalization might be regarded as having superior explanatory credentials – it at least explains more than generalizations with a narrower range of invariance. Generalizations that are invariant under a very wide range of interventions and that have the sort of mathematical formulation that allows for precise predictions are those that we tend to regard as laws of nature. Generalizations that have a narrower range of invariance like Hooke's "law" capture causal information but are not plausible candidates for laws of nature. An interventionist model of form (3.1–3.3) above thus requires generalizations with some degree of invariance or relationships that support interventionist counterfactuals, but it does not require laws. In this respect, like the other models considered in this entry, it departs from the DN model which does require laws for successful explanation (see the entry on scientific explanation).

Turning now to criticisms of the interventionist model, some of these are also criticisms of interventionist accounts of causation. Several of these (and particularly the delicate question of what it means for an intervention to be "possible") are addressed if not resolved in the causation and manipulability entry.

Another criticism, not addressed in the above entry, concerns the "truth makers" or "grounds" for the interventionist counterfactuals that figure in causal explanation. Many philosophers hold that it is necessary to provide a metaphysical account of some kind for these. There are a variety of different proposals – perhaps interventionist counterfactuals or causal claims more generally are made true by "powers" or "dispositions". Perhaps instead such counterfactuals are grounded in laws of nature, with the latter being understood in terms of some metaphysical framework, as in the Best Systems Analysis. For the most part interventionists, have declined to provide truth conditions of this sort and this has struck some metaphysically minded philosophers as a serious omission. One response is that while it certainly makes sense to ask for deeper explanations of

why various interventionist counterfactuals hold, the only explanation that is needed is an ordinary scientific explanation in terms of some deeper theory, rather than any kind of distinctively "metaphysical" explanation (Woodward 2017b). For example, one might explain why the interventionist counterfactual "if I were to drop this bottle it will fall to the ground" is true by appealing to Newtonian gravitational theory and "grounding" it in this way. (There is also the task of providing a semantics for interventionist counterfactuals and here there have been a variety of proposals – see, e.g., Briggs 2012. But again, this needn't take the form of providing metaphysical grounding.) This response raises the question of whether in addition to ordinary scientific explanations there are metaphysical explanations (of counterfactuals, laws and so on) that it is the task of philosophy to provide – a very large topic that is beyond the scope of this entry.

Yet another criticism (pressed by Franklin-Hall 2016 and Weslake 2010) is that the w-condition implies that explanations at the lowest level of detail are always superior to explanations employing upper-level variables – the argument being that lower-level explanations will always answer more w-questions than upper-level explanations. (But see Woodward (2021) for further discussion.)

4. Explanatory Depth

Presumably all models of causal explanation (and certainly all of the models considered above) agree that a causal explanation involves the assembly of causal information that is relevant to the explanandum of interest, although different models may disagree about how to understand causation, causal relevance, and exactly what causal information is needed for explanation. There is also widespread agreement (at least among the models considered above) that causal explanations can differ in how deep or good they are. Capturing what is involved in variations in depth is thus an important task for a theory of causal explanation (or for that matter, for any theory of explanation, causal or non-causal). Unsurprisingly different treatments of causal explanation provide different accounts of what explanatory depth consists in. One common idea is that explanations that drill down (provide information) about lower-level realizing detail are (to that extent) better – this is taken to be one dimension of depth even if not the only one.

This idea is discussed by Sober (1999) in the context of reduction, multiple realizability, and causal explanations in biology. Sober suggests that lower-level details provide objectively superior explanations compared to higher-level ones and he supports this in three main ways. First, he suggests that for any explanatory target, lower-level details can always be included without detracting from an explanation. The worst offense committed by this extra detail is that it "explains too much," while the same cannot be said for higher-level detail (Sober 1999: 547). Second, Sober claims that lower-level details do the "work" in producing higher-level phenomena and that this justifies their privilege or priority in explanations. A similar view is expressed by Waters, who claims that higher-level detail, while more general, provides "shallow explanations" compared to the "deeper accounts" provided by lower-level detail (1990: 131). A third reason is that physics has a kind of "causal completeness" that other sciences do not have. It is argued that this causal completeness provides an objective measure of explanatory strength, in contrast to the more "subjective" measures sometimes invoked in defenses of the explanatory credentials of upper level-sciences. As Sober (1999: 561) puts it,

> illumination is to some degree in the eye of the beholder; however, the sense in which physics can provide complete explanations is supposed to be perfectly objective.

Furthermore,

> if singular occurrences can be explained by citing their causes, then the causal completeness of physics [ensures] that physics has a variety of explanatory completeness that other sciences do not possess. (1999: 562)

Cases where some type-level effect (e.g., a disease) has a shared causal etiology at higher-levels, but where this etiology is multiply-realized at lower ones present challenges for such views (Ross 2020). In Sober's example, "smoking causes lung cancer" is a higher-level (macro) causal relationship. He suggests that lower-level realizers of smoking (distinct carcinogens) provide deeper explanations of this outcome. One problem with this claim is that any single lower-level carcinogen only "makes a difference to" and explains a narrow subset of all cases of the disease. By contrast, the higher-level causal factor "smoking" makes a difference to all (or most) cases of this disease. This is reflected in the fact that biomedical researchers and nonexperts appeal to smoking as the cause of lung cancer and explicitly target smoking cessation in efforts to control and prevent this disease. This suggests that there can be drawbacks to including too much lower-level detail.

The kairetic theory also incorporates, in some respects, the idea that explanatory depth is connected to tracking lower-level detail. This is reflected in the requirement that deeper explanations are those that are cohesive with respect to fundamental physics – at the very least we will be in a better position to see that this requirement is satisfied when there is supporting information about low-level realizers.[15] On the other hand, as we have seen, the kairetic abstraction procedure taken in itself pushes away from the inclusion of specific lower-level detail in the direction of greater generality which, in some form or other, is also regarded by most as a desirable feature in explanations, the result being a trade-off between these two desiderata. The role of lower-level detail is somewhat different in mechanistic models since in typical formulations generality per se is not given independent weight, and depth is associated with the provision of more rather than less relevant detail. Of course a great deal depends on what is meant by "relevant detail". As noted above, this issue is taken up by Craver in several papers, including most recently, Craver and Kaplan (2020) who discuss what they call "norms of completeness" for mechanistic explanations, the idea being that there needs to be some "stopping point" at which a mechanistic explanation is complete in the sense that no further detail needs to be provided. Clearly, whatever "relevant detail" in this connection means it cannot mean all factors any variation in which would make a difference to some feature of the phenomenon *P* which is the explanatory target. After all, in a molecular level explanation of some *P*, variations at the quantum mechanical level – say in the potential functions governing the behavior of individual atoms will often make some difference to *P*, thus requiring (on this understanding of relevance) the addition of this information. Typically, however, such an explanation is taken by mechanists to be complete just at the molecular level – no need to drill down further. Similarly, from a mechanistic point of view an explanation *T* of the behavior of a gas in terms of thermodynamic variables like pressure and temperature is presumably less than fully adequate since the gas laws are regarded by some if not most mechanists as merely "phenomenological" and not as describing a mechanism. A statistical mechanical explanation (SM) of the behavior of the gas is better *qua* mechanistic explanation but ordinarily such explanations don't advert to, say, the details of the potentials (DP) governing molecular interactions, even though variations in these would make some difference to some aspects of the behavior of the gas. The problem is thus to describe a norm of completeness that allows one to say that SM is superior to *T* without requiring DP rather than SM. Craver and Kaplan's discussion (2020) is complex and we will not try to summarize it further here except to say that it does try to find this happy medium of capturing how a norm of completeness can be met, despite its being legitimate to omit some detail.

A closely related issue is this: fine-grained details can be relevant to an explanandum in the sense that variations in those details may make a difference to the explanandum but it can also be the case that those details sometimes can be "screened off" from or rendered conditionally irrelevant to this explanandum (or approximately so) by other, more coarse grained variables that provide less detail, as described in Woodward 2021. For example, thermodynamic variables can approximately screen off statistical mechanical variables from one another. In such a case is it legitimate to omit (do norms about completeness permit omitting) the more fine-grained details as long as the more coarse-grained but screening off detail is included?

Interventionist accounts, at least in the form described by Woodward (2003), Hitchcock and Woodward (2003) offer a somewhat different treatment of explanatory depth. Some candidate explanations will answer

no w-questions and thus fail to be explanatory at all. Above this threshold explanations may differ in degree of goodness or depth, depending on the extent to which they provide more rather than less information relevant to answering w-questions about the explanandum – and thus more information about what the explanandum depends on. For example, an explanation of the behavior of a body falling near the earth's surface in terms of Galileo's law $v = gt$ is less deep than an explanation in terms of the Newtonian law of gravitation since the latter makes explicit how the rate of fall depends on the mass of the earth and the distance of the body above the earth's surface. That is, the Newtonian explanation provides answers to questions about how the velocity of the fall would have been different if the mass of the earth had been different, if the body was falling some substantial distance away from the earth's surface and so on, thus answering more w-questions than the explanation appealing to Galileo's law.

This account associates generality with explanatory depth but this connection holds only for a particular kind of generality. Consider the conjunction of Galileo's law and Boyle's law. In one obvious sense, this conjunction is more general than either Galileo's law or Boyle's law taken alone – more systems will satisfy either the antecedent of Galileo's law or the antecedent of Boyle's law than one of these generalizations alone. On the other hand, given an explanandum having to do with the pressure *P* exerted by a particular gas, the conjunctive law will tell us no more about what *P* depends on than Boyle's law by itself does. In other words, the addition of Galileo's law does not allow us to answer any additional w-questions about the pressure than are answered by Boyle's law alone. For this reason, this version of interventionism judges that the conjunctive law does not provide a deeper explanation of *P* than Boyle's law despite the conjunctive law being in one sense more general (Hitchcock & Woodward 2003).

To develop this idea in a bit more detail, let us say that the *scope* of a generalization has to do with the number of different systems or kinds of systems to which the generalization applies (in the sense that the systems satisfy the antecedent and consequents of the generalization). Then the interventionist analysis claims that greater scope per se does not contribute to explanatory depth. The conjunction of Galileo's and Boyle's law has greater scope than either law alone, but it does not provide deeper explanations.

As another, perhaps more controversial, illustration consider a set of generalizations N1 that successfully explain (by interventionist criteria) the behavior of a kind of neural circuit found only in a certain kind *K* of animal. Would the explanatory credentials of N1 or the depth of the explanations it provides be improved if this kind of neural circuit was instead found in many different kinds of animals or if N1 had many more instances? According to the interventionist treatment of depth under consideration, the answer to this question is "no" (Woodward 2003: 367). Such an extension of the application of N1 is a mere increase in scope. Learning that N1 applies to other kinds of animals does not tell us anything more about what the behavior of the original circuit depends on than if N1 applied just to a single kind of animal.

It is interesting that philosophical discussions of the explanatory credentials of various generalizations often assume (perhaps tacitly) that greater scope (or even greater potential scope in the sense that there are possible – perhaps merely metaphysically possible – but not actual systems to which the generalization would apply) per se contributes to explanatory goodness. For example, Fodor and many others argue for the explanatory value of folk psychology on the grounds that its generalizations apply not just to humans but would apply to other systems with the appropriate structure were these to exist (perhaps certain AI systems, Martians if appropriately similar to humans etc.) (Fodor 1981: 8–9). The interventionist treatment of depth denies there is any reason to think the explanatory value of folk psychology would be better in the circumstances imagined above than if it applied only to humans. As another illustration, Weslake (2010) argues that upper-level generalizations can provide better or deeper explanations of the same explananda than lower-level generalizations if there are physically impossible [but metaphysically possible] systems to which the upper-level explanation applies but to which the lower-level explanation does not (2010: 287), the reason being that in such cases the upper- level explanation is more general in the sense of applying to a wider variety of

systems. Suppose for example, that for some systems governed by the laws of thermodynamics, the underlying micro theory is Newtonian mechanics and for other "possible" or actual systems governed by the same thermodynamic laws, the correct underlying micro-theory is quite different. Then, according to Weslake, thermodynamics provides a deeper explanation than the either of the two micro-theories. This is also an argument that identifies greater depth with greater scope. The underlying intuition about depth here is, so to speak, the opposite of Strevens' since he would presumably draw the conclusion that in this scenario the generalizations of thermodynamics would lack causal cohesion if the different realizing microsystems were actual.

This section has focused on recent discussion of the roles played by the provision of more underlying detail, and generality (in several interpretations of that notion) in assessments of the depth of causal explanation. It is arguable that there are a number of other dimensions of depth that we do not discuss – readers are referred to Glymour (1980), Wigner (1967), Woodward (2010), Deutsch (2011) among many others.

5. Non-causal and Mathematical Explanation

We noted above that there has been considerable recent interest in the question of whether there are non-causal explanations (of the "why" variety) or whether instead all explanations are causal. Although this entry does not discuss non-causal explanations in detail, this issue raises the question of whether there is anything general that might be said about what makes an explanation "causal" as opposed to "non-causal". In what follows we review some proposals about the causal/non-causal contrast, including ideas that abstract somewhat from the details of the theories described in previous sections.

We will follow the philosophical literature on this topic by focusing on candidate explanations that target empirical explananda within empirical science but (it is claimed) explain these non-causally. These contrast with explanations within mathematics, as when some mathematical proofs are regarded as explanatory (of mathematical facts). Accounts of non-causal explanation in empirical science typically focus on explanatory factors that seem "mathematical", that abstract from lower-level causal details, and/or that are related to the explanatory target via dependency relations that are (in some sense) non-empirical, even though the explanatory target appears to be an empirical claim. A common suggestion is that explanations exhibiting one or more of these features, qualify as non-causal. Purported examples include appeals to mathematical facts to explain various traits in biological systems, such as the prime-number life cycles of cicadas, the hexagonal-shape of the bee's honeycomb, and the fact that seeds on a sunflower head are described by the golden angle (Baker 2005; Lyon & Colyvan 2008; Lyon 2012). An additional illustration is Lange's claim (e.g., 2013: 488) that one can explain why 23 strawberries cannot be evenly divided among three children by appealing to the mathematical fact that 23 is not evenly divisible by three. It is claimed that in these cases, explaining the outcome of interest requires appealing to mathematical relationships, which are distinct from causal relationships, in the sense that the former are non-contingent and part of some mathematical theory (e.g., arithmetic, geometry, graph theory, calculus) or a consequence of some mathematical axiom system.

A closely related idea is that in addition to appealing to mathematical relationships, non-causal explanations abstract from lower-level detail, with the implication that although these details may be causal, they are unnecessary for the explanation which is consequently taken to be non-causal. The question of whether it is possible to traverse each bridge in the city of Königsberg exactly once (hereafter just "traverse") is a much-discussed example. Euler provided a mathematical proof that whether such traversability is possible depends on higher-level topological or graph-theoretical properties concerning the connectivity of the bridges, as opposed to any lower-level causal details of the system (Euler 1736 [1956]; Pincock 2012). This explanatory pattern is similar to other topological or network explanations in the literature, which explain despite abstracting from lower-level causal detail (Huneman 2012; Kostic 2020; Ross 2021b). Other candidates for

non-causal explanations are minimal model explanations, in which the removal of at least some or perhaps all causal detail is used to explain why systems which differ microphysically all exhibit the same behavior in some respects (Batterman 2002; Chirimuuta 2014; Ross 2015; and the entry on models in science).

Still other accounts (not necessarily inconsistent with those described above) attempt to characterize some non-causal explanations in terms of the absence of other features (besides those described above). Woodward (2018) discusses two types of cases.

(5.1) Cases in which the factors cited in the explanans are not possible targets of physical interventions, although we can still ask what would be the case if those factors were different.
(5.2) Cases in which, although the factors in the explanans are possible targets of interventions, the dependency relation between explanans and explanandum is "mathematical" rather than contingent.

An example of (5.1) is a purported explanation relating the possibility of stable planetary orbits to the dimensionality of space – given natural assumptions, stable orbits are possible in a three-dimensional space but not possible in a space of dimensionality greater than three, so that the possibility of stable orbits in this sense seems to depend on the dimensionality of space. (For discussion see Ehrenfest 1917; Büchel 1963 [1969]; Callendar 2005). Assuming it is not possible to intervene to change the dimensionality of space, this explanation (if that is what it is) is treated as non-causal within an interventionist framework because of this impossibility. In other words, the distinction between explanations that appeal to factors that are targets of possible interventions and those that appeal to factors that are not targets of possible interventions is taken to mark one dividing line between causal and non-causal explanations.

In the second set of cases (5.2), there *are* factors cited in the explanans that can be changed under interventions but the relationship between this property and the explanandum is non-contingent and "mathematical". For example, it is certainly possible to intervene to change the configuration of bridges in Königsberg and in this way to change their traversability but the relation between the bridge configuration and their traversability is, as we have seen, non-contingent. Many of the examples mentioned earlier – the cicada, honeybee, and sunflower cases – are similar. In these cases, the non-contingent character of the dependency relation between explanans and explanandum is claimed to mark off these explanations as non-causal.

A feature of many of the candidates for non-causal explanation discussed above (and arguably another consideration that distinguishes causal from non-causal explanations) is that the non-causal explanations often seem to explain why some outcome is *possible* or *impossible* (e.g., why stable orbits are possible or impossible in spaces of different dimensions, why it is possible or not to traverse various configurations of bridges). By contrast it seems characteristic of causal explanations that they are concerned with a range of outcomes all of which are taken to be possible and instead explain why one such outcome in contrast to an alternative is realized (why an electric field has a certain strength rather than some alternative strength.)

While many have taken the above examples to represent clear cases of non-causal, mathematical explanation, others have argued that these explanations remain causal through-and-through. One example of this expansive position about causal explanation is Strevens (2018). According to Strevens, the Königsberg and other examples are cases in which mathematics plays a merely representational role, for example the role of representing difference-makers that dictate the movement of causal processes in the world. Strevens refers to these as "non-tracking" explanations, which identify limitations on causal processes that can explain their final outcome, but not the exact path taken to them (Strevens 2018: 112). For Strevens the topological structure represented in the Königsberg's case captures information about causal structure or the web of causal influence – in this way the information relevant to the explanation, although abstract, is claimed to be causal. While this argument is suggestive, one open question is how the kairetic account can capture the fact that some of these cases involve explanations of impossibilities, where the source of the impossibility is not

obviously "structural" (Lange 2013, 2016). For example, the impossibility of evenly dividing 23 by 3 does not appear to be a consequence of the way in which a structure influences some causal process.[16]

In addition to the examples and considerations just described, the philosophical literature contains many other proposed contrasts between causal and non-causal explanations, with accompanying claims about how to classify particular cases. For example, Sober (1983) claims that "equilibrium explanations" are non-causal. These are explanations in which an outcome is explained by showing that, because it is an equilibrium (or better, a unique equilibrium) , any one of a large number of different more specific processes would have led to that outcome. As an illustration, for sexually reproducing populations meeting certain additional conditions (see below), natural selection will produce an equilibrium in which there are equal numbers of males and females, although the detailed paths by which this outcome is produced (which conception events lead to males or females) will vary on different occasions. The underlying intuition here is that causal explanations are those that track specific trajectories or concrete processes, while equilibrium explanations do not do this. By contrast the kairetic theory treats at least some equilibrium explanations as causal in an extended sense (Strevens 2008: 267). Interventionist accounts at least in form described in Woodward (2003) also take equilibrium explanations to be causal to the extent that information is provided about what the equilibrium itself depends on. (That is, the interventionist framework takes the explanandum to be why this equilibrium rather than some alternative equilibrium obtains.) For example, the sex ratio equilibrium depends on such factors as the amount of parental investment required to produce each sex. Differences in required investment can lead to equilibria in which there are unequal numbers of males and females. On interventionist accounts, parental investment is thus among the causes of the sex ratio because it makes a difference for which equilibrium is realized. Interventionist accounts are able to reach this conclusion because they treat relatively "abstract" factors like parental investment as causes as long as interventions on these are systematically associated with associated with changes in outcomes. Thus, in contrast to some of the accounts described above, interventionism does not regard the abstractness per se of an explanatory factor as a bar to interpreting it as causal.

There has also been considerable discussion of whether computational explanations of the sort found in cognitive psychology and cognitive neuroscience that relate inputs to outputs via computations are causal or mechanistic. Many advocates (Piccinini 2006; Piccinini & Craver 2011) of mechanistic models of explanation have regarded such explanations as at best mechanism sketches, since they say little or nothing about realizing (e.g., neurobiological) detail. Since these writers tend to treat "mechanistic explanation", "causal explanation" and even "explanation" as co-extensional, at least in the biomedical sciences, they seem to leave no room for a notion of non-causal explanation. By contrast computational explanations count as causal by interventionist lights as long as they correctly describe how outputs vary under interventions on inputs (Rescorla 2014). But other analyses of computational models suggest that they are similar to non-causal forms of explanation (Chirimuuta 2014, 2018).

6. Additional Issues

Besides the authors discussed above, there is a great deal of additional recent work related to causal explanation that we lack the space to discuss. For additional work on the role of abstraction and idealization in causal explanation (and whether the presence of various sorts of abstraction and idealization in an explanation implies that it is non-causal) see Janssen and Saatsi (2019), Reutlinger and Andersen (2016), Blanchard (2020), Rice (2021), and Pincock (2022). Another set of issues that has received a great deal of recent attention concerns causal explanation in contexts in which different "levels" are present (Craver & Bechtel 2007; Baumgartner 2010; Woodward 2020) This literature addresses questions of the following sort. Can there be "upper-level" causation at all or does all causal action occur at some lower, microphysical level, with upper-level variables being casually inert? Can there be "cross-level" causation – e.g., "downward"

causation from upper to lower levels? Finally, in addition to the work on explanatory depth discussed in Section 4, there has been a substantial amount of recent work on distinctions among different sorts of causal claims (Woodward 2010; Ross 2021a; Ross & Woodward 2022) and on what makes some causes more explanatorily significant than others (e.g., Potochnik 2015).

Bibliography

Andersen, Holly, 2014a, "A Field Guide to Mechanisms: Part I: A Field Guide to Mechanisms I", *Philosophy Compass*, 9(4): 274–283. doi:10.1111/phc3.12119

——, 2014b, "A Field Guide to Mechanisms: Part II: A Field Guide to Mechanisms II", *Philosophy Compass*, 9(4): 284–293. doi:10.1111/phc3.12118

Anscombe, G. E. M., 1971, *Causality and Determination: An Inaugural Lecture*, Cambridge: Cambridge University Press. Reprinted in *Causation*, Ernest Sosa and Michael Tooley (eds.), Oxford/New York: Oxford University Press, 1993, 88–104.

Baker, Alan, 2005, "Are There Genuine Mathematical Explanations of Physical Phenomena?", *Mind*, 114(454): 223–238. doi:10.1093/mind/fzi223

Batterman, Robert W., 2000, "Multiple Realizability and Universality", *The British Journal for the Philosophy of Science*, 51(1): 115–145. doi:10.1093/bjps/51.1.115

——, 2002, *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, (Oxford Studies in Philosophy of Science), Oxford/New York: Oxford University Press. doi:10.1093/0195146476.001.0001

——, 2010a, "On the Explanatory Role of Mathematics in Empirical Science", *The British Journal for the Philosophy of Science*, 61(1): 1–25. doi:10.1093/bjps/axp018

——, 2010b, "Reduction and Renormalization", in *Time, Chance, and Reduction*, Gerhard Ernst and Andreas Hüttemann (eds.), Cambridge/New York: Cambridge University Press, 159–179. doi:10.1017/CBO9780511770777.009

——, 2021, *The Middle Way: A Non-Fundamental Approach to Many-Body Physics*, New York: Oxford University Press. doi:10.1093/oso/9780197568613.001.0001

Batterman, Robert W. and Collin C. Rice, 2014, "Minimal Model Explanations", *Philosophy of Science*, 81(3): 349–376. doi:10.1086/676677

Baumgartner, Michael, 2010, "Interventionism and Epiphenomenalism", *Canadian Journal of Philosophy*, 40(3): 359–383. doi:10.1080/00455091.2010.10716727

Bechtel, William and Robert C. Richardson, 1993 [2010], *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton, NJ: Princeton University Press. Second edition, Cambridge, MA: The MIT Press, 2010.

Blanchard, Thomas, 2020, "Explanatory Abstraction and the Goldilocks Problem: Interventionism Gets Things Just Right", *The British Journal for the Philosophy of Science*, 71(2): 633–663. doi:10.1093/bjps/axy030

Briggs, Rachael, 2012, "Interventionist Counterfactuals", *Philosophical Studies*, 160(1): 139–166. doi:10.1007/s11098-012-9908-5

Büchel, W., 1963 [1969], "Warum hat unser Raum gerade drei Dimensionen?", *Physik Journal*, 19(12): 547–549. Translated and adapted as "Why Is Space Three-Dimensional?", Ira. M. Freeman (trans./adapter), *American Journal of Physics*, 37(12): 1222–1224. doi:10.1002/phbl.19630191204 (de) doi:10.1119/1.1975283 (en)

Callender, Craig, 2005, "Answers in Search of a Question: 'Proofs' of the Tri-Dimensionality of Space", *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 36(1): 113–136. doi:10.1016/j.shpsb.2004.09.002

Chirimuuta, M., 2014, "Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience", *Synthese*, 191(2): 127–153.

doi:10.1007/s11229-013-0369-y
——, 2018, "Explanation in Computational Neuroscience: Causal and Non-Causal", *The British Journal for the Philosophy of Science*, 69(3): 849–880. doi:10.1093/bjps/axw034
Craver, Carl F., 2006, "When Mechanistic Models Explain", *Synthese*, 153(3): 355–376. doi:10.1007/s11229-006-9097-x
——, 2007a, *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Clarendon Press. doi:10.1093/acprof:oso/9780199299317.001.0001
——, 2007b, "Constitutive Explanatory Relevance":, *Journal of Philosophical Research*, 32: 3–20. doi:10.5840/jpr20073241
——, 2008, "Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential", *Philosophy of Science*, 75(5): 1022–1033. doi:10.1086/594543
Craver, Carl F., and Bechtel, William, 2007, "Top-down Causation Without Top-down Causes" *Biology & Philosophy*, 22: 547–563. doi:10.1007/s10539-006-9028-8
Craver, Carl F. and David M. Kaplan, 2020, "Are More Details Better? On the Norms of Completeness for Mechanistic Explanations", *The British Journal for the Philosophy of Science*, 71(1): 287–319. doi:10.1093/bjps/axy015
Deutsch, David, 2011, *The Beginning of Infinity: Explanations That Transform the World*, New York: Viking.
Dupré, John, 2013, "Living Causes", *Aristotelian Society Supplementary Volume*, 87: 19–37. doi:10.1111/j.1467-8349.2013.00218.x
Ehrenfest, Paul, 1917, "In What Way Does It Become Manifest in the Fundamental Laws of Physics that Space Has Three Dimensions?", *KNAW, Proceedings*, 20(2): 200–209. [Ehrenfest 1917 available online]
Euler, Leonhard, 1736 [1956], "Solutio problematis ad geometriam situs pertinentis", *Commentarii Academiae scientiarum imperialis Petropolitanae*, 8: 128–140. Translated as "The Seven Bridges of Königsberg", in *The World of Mathematics: A Small Library of the Literature of Mathematics from A'h-Mosé the Scribe to Albert Einstein*, 4 volumes, by James R. Newman, New York: Simon and Schuster, 1:573–580.
Fodor, Jerry A., 1981, *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, MA: MIT Press.
Franklin-Hall, L. R., 2016, "High-Level Explanation and the Interventionist's 'Variables Problem'", *The British Journal for the Philosophy of Science*, 67(2): 553–577. doi:10.1093/bjps/axu040
Jansson, Lina, & Saatsi, Juha, 2017, "Explanatory abstractions", *The British Journal for the Philosophy of Science*, 70(3): 817–844. doi:10.1093/bjps/axx016
Glennan, Stuart S., 1996, "Mechanisms and the Nature of Causation", *Erkenntnis*, 44(1): 49–71. doi:10.1007/BF00172853
——, 1997, "Capacities, Universality, and Singularity", *Philosophy of Science*, 64(4): 605–626. doi:10.1086/392574
——, 2017, *The New Mechanical Philosophy*, Oxford: Oxford University Press. doi:10.1093/oso/9780198779711.001.0001
Glymour, Clark, 1980, "Explanations, Tests, Unity and Necessity", *Noûs*, 14(1): 31–50. doi:10.2307/2214888
Halina, Marta, 2018, "Mechanistic Explanation and Its Limits", in *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, Stuart Glennan and Phyllis Illari (eds.), New York: Routledge, 213–224.
Hall, Ned, 2012, "Comments on Michael Strevens's *Depth*", *Philosophy and Phenomenological Research*, 84(2): 474–482. doi:10.1111/j.1933-1592.2011.00575.x
[EG2] Hitchcock, Christopher and James Woodward, 2003, "Explanatory Generalizations, Part II: Plumbing Explanatory Depth", *Noûs*, 37(2): 181–199. [For EG1, see Woodward & Hitchcock 2003.] doi:10.1111/1468-0068.00435
Huneman, Philippe, 2010, "Topological Explanations and Robustness in Biological Sciences", *Synthese*, 177(2): 213–245. doi:10.1007/s11229-010-9842-z
Illari, Phyllis McKay and Jon Williamson, 2010, "Function and Organization: Comparing the Mechanisms of Protein Synthesis and Natural Selection", *Studies in History and Philosophy of Science Part C: Studies*

in History and Philosophy of Biological and Biomedical Sciences, 41(3): 279–291. doi:10.1016/j.shpsc.2010.07.001

——, 2012, "What Is a Mechanism? Thinking about Mechanisms across the Sciences", *European Journal for Philosophy of Science*, 2(1): 119–135. doi:10.1007/s13194-011-0038-2

Kostić, Daniel, 2020, "General Theory of Topological Explanations and Explanatory Asymmetry", *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1796): 20190321. doi:10.1098/rstb.2019.0321

Kaplan, David Michael and Carl F. Craver, 2011, "The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective", *Philosophy of Science*, 78(4): 601–627. doi:10.1086/661755

Lange, Marc, 2013, "What Makes a Scientific Explanation Distinctively Mathematical?", *The British Journal for the Philosophy of Science*, 64(3): 485–511. doi:10.1093/bjps/axs012

——, 2016, *Because without Cause: Non-Causal Explanations in Science and Mathematics*, (Oxford Studies in Philosophy of Science), New York: Oxford University Press. doi:10.1093/acprof:oso/9780190269487.001.0001

Levy, Arnon, 2014, "What Was Hodgkin and Huxley's Achievement?", *The British Journal for the Philosophy of Science*, 65(3): 469–492. doi:10.1093/bjps/axs043

Levy, Arnon and William Bechtel, 2013, "Abstraction and the Organization of Mechanisms", *Philosophy of Science*, 80(2): 241–261. doi:10.1086/670300

Lyon, Aidan, 2012, "Mathematical Explanations Of Empirical Facts, And Mathematical Realism", *Australasian Journal of Philosophy*, 90(3): 559–578. doi:10.1080/00048402.2011.596216

Lyon, Aidan and Mark Colyvan, 2008, "The Explanatory Power of Phase Spaces", *Philosophia Mathematica*, 16(2): 227–243. doi:10.1093/philmat/nkm025

Machamer, Peter, 2004, "Activities and Causation: The Metaphysics and Epistemology of Mechanisms", *International Studies in the Philosophy of Science*, 18(1): 27–39. doi:10.1080/02698590412331289242

[MDC] Machamer, Peter, Lindley Darden, and Carl F. Craver, 2000, "Thinking about Mechanisms", *Philosophy of Science*, 67(1): 1–25. doi:10.1086/392759

Mackie, J. L., 1974, *The Cement of the Universe: A Study of Causation*, (The Clarendon Library of Logic and Philosophy), Oxford: Clarendon Press. doi:10.1093/0198246420.001.0001

Morgan, Stephen L. and Christopher Winship, 2014, *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, second edition, (Analytical Methods for Social Research), New York, NY: Cambridge University Press. doi:10.1017/CBO9781107587991

Ney, Alyssa, 2009, "Physical Causation and Difference-Making", *The British Journal for the Philosophy of Science*, 60(4): 737–764. doi:10.1093/bjps/axp037

——, 2016, "Microphysical Causation and the Case for Physicalism", *Analytic Philosophy*, 57(2): 141–164. doi:10.1111/phib.12082

Pearl, Judea, 2000 [2009], *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press. Second edition 2009. doi:10.1017/CBO9780511803161

Piccinini, Gualtiero, 2006, "Computational Explanation in Neuroscience", *Synthese*, 153(3): 343–353. doi:10.1007/s11229-006-9096-y

Piccinini, Gualtiero and Carl Craver, 2011, "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches", *Synthese*, 183(3): 283–311. doi:10.1007/s11229-011-9898-4

Potochnik, Angela, 2011, "Explanation and Understanding: An Alternative to Strevens' Depth", *European Journal for Philosophy of Science*, 1(1): 29–38. doi:10.1007/s13194-010-0002-6

——, 2015, "Causal patterns and adequate explanations", *Philosophical Studies*, 172: 1163–1182. doi:10.1007/s11098-014-0342-8

——, 2017, *Idealization and the Aims of Science*, Chicago, IL: University of Chicago Press.

Pincock, Christopher, 2007, "A Role for Mathematics in the Physical Sciences", *Noûs*, 41(2): 253–275. doi:10.1111/j.1468-0068.2007.00646.x

——, 2012, *Mathematics and Scientific Representation*, (Oxford Studies in Philosophy of Science),

Oxford/New York: Oxford University Press. doi:10.1093/acprof:oso/9780199757107.001.0001
——, 2022, "Concrete Scale Models, Essential Idealization, and Causal Explanation", *The British Journal for the Philosophy of Science*, 73(2): 299–323. doi:10.1093/bjps/axz019
Rathkopf, Charles, 2018, "Network Representation and Complex Systems", *Synthese*, 195(1): 55–78. doi:10.1007/s11229-015-0726-0
Rescorla, Michael, 2014, "The Causal Relevance of Content to Computation", *Philosophy and Phenomenological Research*, 88(1): 173–208. doi:10.1111/j.1933-1592.2012.00619.x
Reutlinger, Alexander, 2014, "Why Is There Universal Macrobehavior? Renormalization Group Explanation as Noncausal Explanation", *Philosophy of Science*, 81(5): 1157–1170. doi:10.1086/677887
Reutlinger, Alexander and Andersen, Holly, 2016, "Abstract versus Causal Explanations?", *International Studies in the Philosophy of Science*, 30(2): 129–146. doi:10.1080/02698595.2016.1265867
Reutlinger, Alexander and Saatsi, Juha (eds.), 2018, *Explanation beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, Oxford: Oxford University Press. doi:10.1093/oso/9780198777946.001.0001
Rice, Collin, 2021, *Leveraging Distortions: Explanation, Idealization, and Universality in Science*, Cambridge, MA: The MIT Press.
Ross, Lauren N., 2015, "Dynamical Models and Explanation in Neuroscience", *Philosophy of Science*, 82(1): 32–54. doi:10.1086/679038
——, 2018, "Causal Selection and the Pathway Concept", *Philosophy of Science*, 85(4): 551–572. doi:10.1086/699022
——, 2020, "Multiple Realizability from a Causal Perspective", *Philosophy of Science*, 87(4): 640–662. doi:10.1086/709732
——, 2021a, "Causal Concepts in Biology: How Pathways Differ from Mechanisms and Why It Matters", *The British Journal for the Philosophy of Science*, 72(1): 131–158. doi:10.1093/bjps/axy078
——, 2021b, "Distinguishing Topological and Causal Explanation", *Synthese*, 198(10): 9803–9820. doi:10.1007/s11229-020-02685-1
——, forthcoming, "Cascade versus Mechanism: The Diversity of Causal Structure in Science", *The British Journal for the Philosophy of Science*, first online: 5 December 2022. doi:10.1086/723623
Ross, Lauren N. and James F. Woodward, 2022, "Irreversible (One-Hit) and Reversible (Sustaining) Causation", *Philosophy of Science*, 89(5): 889–898. doi:10.1017/psa.2022.70
Salmon, Wesley C., 1971a, "Statistical Explanation", in Salmon 1971b: 29–87.
—— (ed.), 1971b, *Statistical Explanation and Statistical Relevance*, Pittsburgh, PA: University of Pittsburgh Press.
——, 1984, *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ: Princeton University Press.
Silberstein, Michael and Anthony Chemero, 2013, "Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences", *Philosophy of Science*, 80(5): 958–970. doi:10.1086/674533
Skow, Bradford, 2014, "Are There Non-Causal Explanations (of Particular Events)?", *The British Journal for the Philosophy of Science*, 65(3): 445–467. doi:10.1093/bjps/axs047
Strevens, Michael, 2008, *Depth: An Account of Scientific Explanation*, Cambridge, MA: Harvard University Press.
——, 2004, "The Causal and Unification Approaches to Explanation Unified: Causally", *Noûs*, 38(1): 154–176. doi:10.1111/j.1468-0068.2004.00466.x
——, 2013, "Causality Reunified", *Erkenntnis*, 78(S2): 299–320. doi:10.1007/s10670-013-9514-8
——, 2018, "The Mathematical Route to Causal Understanding", in Reutlinger and Saatsi 2018: 117–140 (ch. 5).
Sober, Elliott, 1983, "Equilibrium Explanation", *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 43(2): 201–10.
——, 1999, "The Multiple Realizability Argument against Reductionism", *Philosophy of Science*, 66(4):

542–564. doi:10.1086/392754
Waters, C. Kenneth, 1990, "Why the Anti-Reductionist Consensus Won't Survive the Case of Classical Mendelian Genetics", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1990(1): 125–139. doi:10.1086/psaprocbienmeetp.1990.1.192698
Weslake, Brad, 2010, "Explanatory Depth", *Philosophy of Science*, 77(2): 273–294. doi:10.1086/651316
Wigner, Eugene Paul, 1967, *Symmetries and Reflections: Scientific Essays of Eugene P. Wigner*, Bloomington, IN: Indiana University Press.
Woodward, James, 2002, "What Is a Mechanism? A Counterfactual Account", *Philosophy of Science*, 69(S3): S366–S377. doi:10.1086/341859
——, 2003, *Making Things Happen: A Theory of Causal Explanation*, Oxford/New York: Oxford University Press. doi:10.1093/0195155270.001.0001
——, 2006, "Sensitive and Insensitive Causation", *The Philosophical Review*, 115(1): 1–50. doi:10.1215/00318108-2005-001.
——, 2010, "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation", *Biology & Philosophy*, 25(3): 287–318. doi:10.1007/s10539-010-9200-z
——, 2013, "Mechanistic Explanation: Its Scope and Limits", *Aristotelian Society Supplementary Volume*, 87: 39–65. doi:10.1111/j.1467-8349.2013.00219.x
——, 2017a, "Explanation in Neurobiology: An Interventionist Perspective", in *Explanation and Integration in Mind and Brain Science*, David M. Kaplan (ed.), Oxford: Oxford University Press, ch. 5.
——, 2017b, "Interventionism and the Missing Metaphysics: A Dialogue", in *Metaphysics and the Philosophy of Science: New Essays*, Matthew Slater and Zanja Yudell (eds.), New York: Oxford University Press, 193–228. doi:10.1093/acprof:oso/9780199363209.003.0010
——, 2018, "Some Varieties of Non-Causal Explanation", in Reutlinger and Saatsi 2018: 117–140.
——, 2020, "Causal Complexity, Conditional Independence, and Downward Causation", *Philosophy of Science*, 87(5): 857–867. doi:10.1086/710631
——, 2021, "Explanatory Autonomy: The Role of Proportionality, Stability, and Conditional Irrelevance", *Synthese*, 198(1): 237–265. doi:10.1007/s11229-018-01998-6
[EG1] Woodward, James and Christopher Hitchcock, 2003, "Explanatory Generalizations, Part I: A Counterfactual Account", *Noûs*, 37(1): 1–24. [For EG2, see Hitchcock & Woodward 2003.] doi:10.1111/1468-0068.00426

Academic Tools

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up topics and thinkers related to this entry at the Internet Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

[Please contact the author with suggestions.]

Related Entries

causal models | causation: and manipulability | causation: regularity and inferential theories of | mathematics:

explanation in | models in science | scientific explanation

Acknowledgments

Thanks to Carl Craver, Michael Strevens and an anonymous referee for helpful comments on a draft of this entry.

Library of Congress Catalog Data: ISSN 1095-5054