Causation, physics, and fit*

Christian Loew†

Abstract

Our ordinary causal concept seems to fit poorly with how our best physics describes the world. We think of causation as a time-asymmetric dependence relation between relatively local events. Yet fundamental physics describes the world in terms of dynamical laws that are, possible small exceptions aside, time symmetric and that relate global time slices. My goal in this paper is to show why we are successful at using local, time-asymmetric models in causal explanations despite this apparent mismatch with fundamental physics. In particular, I will argue that there is an important connection between time asymmetry and locality, namely: understanding the locality of our causal models is the key to understanding why the physical time asymmetries in our universe give rise to time asymmetry in causal explanation. My theory thus provides a unified account of why causation is local and time asymmetric and thereby enables a reply to Russell's famous attack on causation.

1 Introduction

Our ordinary causal concept seems to fit poorly with how our best physics describes the world.[1] We think of causation as a *time-asymmetric* dependence relation between relatively *local* events. Yet fundamental physics describes the world in terms of dynamical laws that are, possible small exceptions aside, time symmetric and that relate global time slices. According to these laws, events depend for their occurrence on everything that

---

*To appear in *Synthese*. The final publication is available at Springer via `http://dx.doi.org/10.1007/s11229-016-1029-9`.

†Universität zu Köln, Philosophisches Seminar, Albertus-Magnus-Platz 1, 50923 Köln, Germany.

[1] See, e.g., Field (2003), Price and Weslake (2009), Russell (1913), and many of the papers in Price and Corry (2007). I am taking the useful notions of "fit" and "poor fit" from Schaffer (2010).

happens in a large spatial region and, moreover, the past depends on the future in the exact same way as vice versa.

An extreme reaction to this seemingly poor fit is Russell's causal eliminativism. Russell (1913, p. 1) famously argued that causation is "a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm." His advice to get rid of causation, however, is not feasible. Even philosophers who agree with Russell about the limited role of causation in fundamental physics maintain that causation is central to understanding the physical world (see, e.g., Hitchcock 2007a; Woodward 2007). Science would be crippled and our survival chances diminished without a causal concept (Cartwright 1979, pp. 419–420; Field 2003, p. 336).

A different response involves explaining why we are successful at using local, time-asymmetric causal models in our explanations despite the global and time-symmetric character of the fundamental physical laws. Such an account has to appeal both to the goals of causal explanation and to the physical circumstances in our universe to show why it is advantageous to use time-asymmetric and local models in causal explanation. This account would then show how causation fits into the physical world and also how the special sciences, which use causal models in their explanations, relate to fundamental physics. In this paper, I will defend such an account.

Most theories in the literature give separate accounts of the locality and the time asymmetry of causation.[2] Time asymmetry is usually regarded as the more serious problem. A long-standing conjecture is that the time asymmetry of causation is grounded in the special initial boundary conditions of our actual universe (see, e.g., Field 2003 and Loewer 2012). However, exactly how time asymmetries from the boundary conditions give rise to a time asymmetry of causation remains puzzling—for example, how exactly are these time asymmetries connected to the fact that we causally explain outcomes in terms

---

[2] Eagle (2007), Elga (2007), and Woodward (2007) address why it makes sense for us to conceive of causation as a relation between relatively localized events or variables. Albert (2000), Eckhardt (2006), Field (2003), Hausman (1998), Loewer (2007, 2012), Price (2007), and Price and Weslake (2009) address why it makes sense for us to have a time-asymmetric causal concept. Frisch (2014) and Kutach (2013) discuss both issues but not in a unified fashion.

of earlier rather than later events?[3] I will argue that there is an important connection between time asymmetry and locality, namely: understanding the locality of our causal models is the key to understanding why the physical time asymmetries in our universe give rise to time asymmetry in causal explanation. My theory thus provides a unified account of why causation is local and time asymmetric despite its apparent poor fit with physics.

In §2, I spell out in more detail the challenge of reconciling the time asymmetry and locality of our causal concept with the fundamental physical laws. In §3, I adopt Woodward's interventionist theory as a plausible background assumption about what constitutes a causal explanation. In §4, I use this account of causal explanation to show why local causal models in the forward direction are explanatory despite the global character of the laws. In §5, I then use this response to the locality worry to show why causal explanation is time asymmetric: I argue that our local causal models have no analogs in the backward direction. Specifically, the physical circumstances that allow us to build local causal models that are explanatory in the forward direction are absent in the backward direction. In §6, I use this result to show why there are no causal explanations in the backward direction: locality is an essential feature of causal explanation, and so the unavailability of local models in the backward direction shows why later events do not causally explain earlier events. In §7, I sketch how my account provides a reply to Russell's attack on causation as well as how it complements existing accounts of the temporal direction of causation.

2 Locality and Time Asymmetry

The causal models we use in everyday life and in scientific explanations are local and time asymmetric in that they depict effects as dependent on a *small number* of *earlier* conditions. For example, we might explain the shattering of a window by citing as its causes the earlier existence of the intact window and the stone flying toward it. But how can this information about a small number of localized events constitute an explanation, and why do causal explanations cite events at earlier rather than later times?

---

[3] Hausman (1998, chapter 4 and 12) provides the most comprehensive treatment of this question, but he does not address the locality of our causal models. Most of the other authors focus on our agency and give no explicit account of the time asymmetry of causal explanation. Yet the time asymmetry of explanation, at least prima facie, seems more general than just a time asymmetry in our abilities as agents (cf. Hausman 1982, p. 47).

Most philosophers think that causal relations reduce to patterns of lawful dependence, where this dependence is specified in terms of minimal sufficiency, probability raising, counterfactuals, or lawful regularities.[4] So it is natural to think that causal models are explanatory because they latch on to the world's lawful structure. The fundamental physical laws tell us how the complete state of the world at any one time depends on its states at other times. Causal models are then explanatory because they represent, in a piece-wise fashion, how particular localized events at one time lawfully depend on particular localized events at other times (cf. Hall 2004). Yet the apparent lesson from Russell (1913) is that due to their locality and time asymmetry, our causal models do an exceedingly poor job of latching on to the world's lawful structure.

This mismatch is most striking with regard to time asymmetry. Most plausible candidates for the fundamental dynamical laws of physics are deterministic in both temporal directions. According to these laws, the state of the universe at any one time uniquely fixes its state at all later and earlier times. Hence, otherwise identical systems that differ in their final condition also differ in their initial condition, just as otherwise identical systems that differ in their initial condition also differ in their final condition.[5] So lawful dependence goes in both temporal directions: the initial state of a system lawfully depends on its final state just as much as vice versa.

As an example, consider a ball that is moving across a billiard table without any significant outside influences. The fundamental physical laws allow us to predict the final condition of the ball from its initial condition; but they equally allow us to retrodict the initial condition of the ball from its final condition. Moreover, we can assume that the ball's momentum was different at some earlier time and use the physical laws to calculate its new momentum at a later time—or we can assume the ball's momentum to be different at a later time, thus creating a new final condition, and use the physical laws to calculate its new momentum at an earlier time.

Due to this lawful time symmetry, fundamental physics not only allows us to build forward-looking causal models that depict how later events lawfully depend on earlier

---

[4] See Hall (2005) and the references therein.

[5] I assume that the laws are deterministic, but the problem does not significantly change if we move to probabilistic laws, as long as these laws have the same probabilistic character in either temporal direction. This is the case if the state of the world at any one time fixes a unique probability distribution over all earlier and later states (Field 2003, p. 437). In this case, systems that differ only in their final condition also differ (at least typically) in what probability distribution the laws assign over the system's initial condition.

events; it also allows us to build backward-looking models that depict how earlier events lawfully depend on later events (I will say more about these backward-looking models in §5). Given the availability of backward-looking models, it seems at least arbitrary and possibly misleading for us to single out forward-looking models in our explanations. Why would backward-looking models not be equally as explanatory as the forward-looking models that we ordinarily use in causal explanations?

One might object that backward-looking models would not be causal models. After all, we ordinarily think that causation goes only in the forward direction. This objection, however, misses the point. The point is not that backward-looking models would be causal models. The point is that these models seem to latch on to the lawful structure of our universe just as well as the forward-looking models that we regard as "causal" and that we use in our explanations. It thus needs to be explained why it is a good idea for us to use causal, forward-looking models in our explanations rather than backward-looking models.

The second apparent mismatch between causation and fundamental physics concerns locality. Typical causal models depict outcomes as dependent on a small number of relatively localized events. For example, a causal explanation of a window shattering might only mention the earlier existence of the intact window and a stone flying toward it. Yet, as Field points out, fundamental physics lacks this extreme locality because "no reasonable laws of physics, whether deterministic or indeterministic, will make the probability of what happens at a time depend on only finitely many localized antecedent states." (Field 2003, p. 439)

By explaining outcomes in terms of a small number of relatively localized earlier events, our causal models leave out numerous other events on which the outcome lawfully depends. For instance, the shattering of a window lawfully depends on many events not mentioned in a typical causal model. If billions of air molecules all bumped into the stone from the same direction, if the earth's gravitational force were significantly smaller, or if a burst of energy from outer space hit the stone in midair, for example, then the window would not shatter even if a stone flew toward it. Thus, just as there appears to be a mismatch between the time asymmetry of causation and the time symmetry of the physical laws, there also appears to be a mismatch between the local character of causation and the global character of the physical laws. How we can understand why an outcome occurred based on information about only a small subset of the events that physically determine its probability needs to be explained.

A response to both of these challenges must show why our actual practice of building local and time-asymmetric causal models is useful for explaining the physical world despite the apparent mismatch with fundamental physics. It must show why backward-looking models would not be as successful as forward-looking models, and it must also show why are our local models are successful despite leaving out numerous factors that also determine features of the modeled system.

3 The Goals of Causal Explanation

To justify why local, time-asymmetric causal models can be explanatory, I need to make some assumptions about what constitutes a causal explanation. In the following, I will adopt Woodward's interventionist theory as a background assumption.[6] Woodward's theory of causal explanation has been very influential, and many alternative accounts of causal explanation share the features of Woodward's theory that will play a role in my argument.

Woodward thinks of a causal model as a set of variables and a set of equations. Variables represent possible states of a system. We can think of their values as events. For example, the extension of a spring corresponds to a variable whose values are events of the spring having some particular extension. A causal model's equations specify how variables depend on interventions on other variables. For example, an equation might tell us how the force with which a spring resists its displacement depends on interventions on the spring's extension.

Interventions are "idealized experimental manipulations" (Woodward 2003, p. 94), though we need not be able to carry them out in practice. Woodward's definition of an intervention is complex, but the details will not matter for present purposes. It will suffice to say that a manipulation of a variable X counts as an intervention with respect to another variable Y just in case the manipulation sets X to a new value and is not correlated with the value of Y via any other causal route except through X.

Woodward characterizes causal explanations in terms of interventions. According to Woodward (2003, p. 201), "the underlying or unifying idea in the notion of causal explanation is the idea that an explanation must answer a what-if-things-had-been-different question, or exhibit information about a pattern of dependency." More specifically, causal explanations describe how, for a range of counterfactual circumstances, the to-be-explained

---

[6] See Woodward (2003, 2007) and Hitchcock and Woodward (2003).

variable depends on interventions on the explaining variables. One of Woodward's examples is Hooke's law, which describes the behavior of springs:

(H) $F = -k X$

where X is the amount by which the spring is displaced from its relaxed position, F is the force with which the spring resists its displacement, and k characterizes the 'stiffness' of the spring.

According to Woodward, Hooke's law explains why a spring resists its displacement with a particular force because the law specifies for a wide range of possible circumstances how this force depends on interventions on the spring's displacement and stiffness. It is essential that the explanation not only allows us to derive the actual value of F from the actual values of k and X, but also that it correctly predicts the value of F for a range of counterfactual circumstances. Woodward distinguishes two aspects of this modal robustness: *invariance* and *stability*. Invariance concerns interventions on variables that are part of the model. Stability concerns interventions on variables that are part of the background conditions but are not represented in the model. For example, the spring's displacement, X, is part of the model, whereas the spring's temperature and location are part of the background conditions.

Hooke's law is invariant because it correctly predicts how the value of F would be different for a wide range of interventions on k and X. For example, (H) tells us what value F would take if we were to extend the spring by five inches or by ten inches. Moreover, Hooke's law is stable because (H) still holds for various ways in which the background conditions could have been different. For example, if the spring were moved or if its temperature were a bit higher, then the equation, at least typically, would still allow us to correctly predict the value of F from a wide range of interventions on the values of X and k.

Woodward emphasizes that explanations need not be invariant and robust in all possible circumstances. For example, if an intervention increases the value of X by too much, then the spring becomes overextended and the restoring force is significantly less than (H) predicts, and if the temperature of the spring rises to an extreme degree, then (H) also no longer correctly predicts how the extension of the spring depends on interventions on the other variables. Nonetheless, Hooke's law is explanatory because it is invariant and robust for interventions that create circumstances that count as likely or typical. As Woodward (2007, p. 79) puts it: "[W]e are particularly interested in invariance and stability under changes that are not too infrequent or unlikely to occur, around here, right now, and less interested in what would happen under changes that are extremely unlikely or which seem 'farfetched'." Thus, models are explanatory if they supply dependencies that are invariant and stable in a range of typical circumstances.

Woodward's interventionist theory is very plausible because it accounts in a natural way for which relations are explanatory and which are not. For example, though we can derive the later occurrence of a storm from the barometer reading, there is no explanatory relationship between the two variables. Woodward's theory accounts for this fact because the storm does not depend on interventions on the barometer reading. If we were to intervene on the barometer reading (for instance, by manipulating its display), this would not make a difference to the storm (Woodward 2003, p. 15). So the relation is not invariant.

However, even invariant relationships may fail to be explanatory if they are not stable. Imagine a generalization like the one in Hooke's law, except that it turns out that the generalization correctly predicts the behavior of a spring only when the spring has a very specific temperature T. For a spring that happens to be at temperature T, this generalization will allow us to derive the force with which the spring resists displacement for a wide range of interventions on the spring's displacement and stiffness. Nonetheless, this generalization would not be an adequate explanation because it would look like an accident that the spring resists displacement with the force in question—since in many typical circumstances (where the spring has a different temperature), the spring will respond differently to the same interventions. By contrast, Hooke's law continues to hold in these circumstances. Woodward's account captures this additional demand by requiring that causal explanations be stable.

Moreover, it is plausible that causal explanations only need to be stable and invariant in typical or not too far-fetched circumstances. Causal explanations in the special sciences and in everyday life aim to account for why systems behave the way they do in the circumstances we encounter around here, around now. So it makes sense that causal explanations only need to be stable and invariant in these kinds of circumstances, not in ones that we typically do not encounter.[7]

---

[7] Which circumstances count as likely or typical is a difficult question. There are two ways to go about answering it. First, some philosophers argue that we get a global probability distribution over all macrostates from the foundation of statistical mechanics (Albert 2000, chapter 5). We could then use this probability distribution to determine which changes in variables are likely or to be expected. Second, some philosophers argue that we should invoke pragmatic factors to determine which values of

Hooke's law nicely illustrates Woodward's account, but the account also applies to the kinds of qualitative explanations that are common in everyday life. For example, we might explain why a window shattered in terms of a stone earlier flying toward the intact window. This model is invariant because it not only tells us that the window does shatter given that the stone actually flies toward it, but also that the window would not have shattered if an intervention had prevented the stone from flying toward it. Moreover, the model is stable because whether the window shatters depends on interventions on whether a stone flies toward it not just in the actual circumstances but also in other typical circumstances. For example, in counterfactual circumstances where the temperature is different, there is a bit more wind, or the stone is a bit heavier, etc., it is still true that if an intervention had prevented the stone from flying toward the window, then the window would not have shattered.

The guiding idea of Woodward's account is that what distinguishes explanatory from non-explanatory relationships is a certain kind of modal robustness (Woodward 2007, p. 76). Woodward tries to capture this robustness in terms of invariance and stability, given interventions. Many other accounts agree that modal robustness is central to causal explanations, though they either try to capture this robustness with tools other than interventions or have somewhat different standards of robustness.[8] I am adopting Woodward's account because it has been very influential and will be familiar to many readers. In the following, I will use it to justify the locality and time asymmetry of our causal models. My argument will also work given other accounts that emphasize modal robustness as a central virtue of causal explanations, though I do not have the space to show this here.

4 Defending the Locality of Causal Models

With Woodward's account of causal explanation in hand, we can justify the locality in our causal models by showing that these models meet the goals of causal explanation. Due to the global character of the fundamental dynamical laws, the behavior of systems in our universe lawfully depends on numerous distant influences. However, Elga (2007) argues that we can nonetheless treat systems, at least for the purpose of explaining their

---

a variable count as the default and therefore as to be expected (Hitchcock 2007b, pp. 506–507). Either route could be adopted for my present purposes.

[8] See, e.g., Franklin-Hall (2015), Lange (2009), Strevens (2004), and Weslake (2010).

macroscopic behavior, as if they were isolated from most of these distant influences. In this section, I will flesh out Elga's argument and use it to show that our local causal models are invariant and robust in typical circumstances, despite leaving out many variables on which the outcome depends.[9] Moreover, this section provides the set-up for my novel account of why causal models are time asymmetric. In §5 and §6, I will argue that the time asymmetry of our causal models falls directly out of Elga's account of locality because the physical conditions that allow for local causal models in the forward direction are absent in the backward direction.

Typical causal models involve "coarse-grained" variables that characterize macroscopic properties of systems (Field 2003, p. 445; Woodward 2007, §4.5). For example, we might want to explain why my office window is *broken* rather than *unbroken.* This variable is coarse-grained because its values can be realized in multiple ways by precise microstates. There are numerous different ways for the molecules making up my window to constitute a broken or an unbroken window.

If we wanted to explain the precise microstate of a shattered window, we would not be able to find a local model that is invariant and stable. After all, many forces in our universe, in particular gravitational and electromagnetic forces, have non-zero values over large distances. So any change in the surroundings of the window would make a difference to its exact microstate. Hence, there are many typical circumstances where the precise microstate of my shattered office window would not have obtained even if the stone had flown toward it.

Coarse-grained variables allow us to ignore features of the background conditions that do not make a difference regarding whether the window is broken or unbroken. Nonetheless, it is far from trivial that we can explain events such as the window shattering with a local causal model. Numerous outside influences could interfere even with dependencies among coarse-grained variables. For example, if a high-energy beam pulverized the stone in midair or if billions of air molecules bumped into the stone at the same time, the window would not shatter even if a stone did fly toward it. So to show that in typical circumstances, the window's shattering invariantly and robustly depends on the stone flying toward it, we must show that such interferences do not happen in typical circumstances.

Elga (2007) provides just such an argument. He argues that the kinds of inter-

---

[9] To avoid cumbersome phrases, I will sometimes say of causal models that they are "robust and invariant" when I mean that they specify dependencies that are robust and invariant. I will also sometimes drop the qualification "in typical circumstances."

ferences from afar that could make a difference to the macroscopic behavior of systems are extremely rare in typical circumstances. This rarity of interferences is due to three conditions that hold in virtue of the nature of the fundamental dynamical laws and the actual boundary conditions. Elga states two of these conditions explicitly, while the third one is implicit in his discussion. I will illustrate and motivate the three conditions with the example of a stone shattering my office window.

The first condition is what I will call "Insensitivity":

(I) Insensitivity. The evolution of most macroscopic systems toward the future is relatively insensitive to small changes in their initial conditions (see Elga 2007, p. 110).

Consider the stone and the window as a single system. This system lawfully evolves from an initial condition where the stone flies toward the intact window into a final condition where the stone and glass shards lie under a shattered window. This evolution is relatively insensitive because it is extremely unlikely for a small difference in the initial state of the stone or the window to make a difference to the system's future macroscopic behavior. Given most microscopic changes to the initial state of the stone or the window, the system still would evolve into a future macroscopic state where the window is shattered and glass pieces and a stone lie on the floor.

Elga motivates Insensitivity with statistical mechanics. Statistical mechanics provides a probability distribution over the possible microstates of systems such as a stone flying toward a window, and this probability distribution entails that small microscopic differences are extremely unlikely to make a difference to the future macroscopic behavior of a system (Elga 2007, p. 110). Moreover, Insensitivity can be motivated through ordinary experience. We observe on numerous occasions that systems that are similar in their macroscopic properties also display similar future macroscopic behaviors. For example, we observe that windows typically shatter when a stone flies toward them. This observation supports Insensitivity because these systems plausibly differ in their exact microscopic properties, and so the observation suggests that small microscopic differences typically do not matter to the macroscopic behavior of these systems.

Insensitivity does not hold for all macroscopic systems. For some systems, small differences in their initial conditions will almost certainly make a difference to their future macroscopic behavior. For example, consider a rock that is delicately balanced on a mountain top. The rock eventually falls in one direction, but given any small difference it would have fallen in a different direction. (I will argue in §6 that we can still causally explain systems like the delicately balanced rock by attributing chancy dispositions to them. See also Elga 2007, pp. 111–117.)

It is a remarkable fact about our universe that many of the systems that we try to explain causally obey Insensitivity. For example, numerous outside influences affect my office window, such as gravitational forces from nearby bodies or air molecules bumping into it. However, due to Insensitivity, we can omit from our causal models all variables that in typical circumstances only change in ways that make at most a small difference to the window, because these changes will not affect its macroscopic behavior. Frisch (2014, p. 67) makes a similar observation, pointing out that causal models need not include variables representing spatial regions "if whatever is physically going [on] in these regions can be taken to be irrelevant to the phenomenon at issue." So to show that local causal models can be stable and invariant in typical circumstances, we need to show that most variables in typical circumstances only make very small differences to a system. The remaining two conditions guarantee this fact.

The second condition states that in typical circumstances, most outside forces that act on physical systems are either extremely small or almost constant:

(II) Stability of Forces. Most outside forces that act on systems in our universe are either extremely small or almost constant (Elga 2007, p. 109).

Elga defends this condition by considering the four kinds of forces that are found in our universe: strong, weak, electromagnetic, and gravitational. Strong and weak forces are negligible except over very small distances. Gravitational and electromagnetic forces can be significant even across large distances if they are exerted by massive or heavily charged bodies. However, the gravitational forces that are strong around here, such as the force exerted by the earth, tend to be almost constant. Moreover, there are not very many heavily charged bodies around here. As a consequence, the outside forces that act, for example, on my office window and the stone flying toward it are almost constant in typical circumstances. Hence any intervention on the background conditions that changes these variables to values that would count as typical only make a small difference to the forces acting on these systems. This is not to say that we cannot imagine circumstances where the external forces acting on objects like stones can vary strongly. For example, we can imagine conditions where stones are frequently pulverized by high-energy beams from outer space or where the earth's gravitational force sometimes drastically changes. But in typical circumstances, such large changes in the forces acting on objects from afar are very uncommon, and so we can leave the relevant variables out of our models.

However, even if small changes in forces do not make a difference to the macroscopic behavior of systems and large changes are rare, we still need to rule out the possibility that numerous small forces are coordinated in such a way that they jointly make a difference to the macroscopic behavior of a system. For example, a tiny difference in the movements of some air molecules would not make a difference to whether or not a stone flying toward a window shatters the window. However, if billions of air molecules all bumped into the stone from the same direction at the same time, they might divert its trajectory such that it would no longer shatter the window. Therefore, a third condition is that there are no such correlations in typical circumstances. Elga does not explicitly mention this third condition, but it is implicit in his argument.

Randomness rules out extreme correlations among distinct variables occurring in typical circumstances:

(III) Randomness. Typically, there are no extreme correlations between the values of distinct variables.

Randomness is supported by our experience that genuine coincidences are rare. For example, we can typically treat the effects that different air molecules have on the trajectory of a stone as random, and so leave them out from our models. Randomness again holds due to the particular boundary conditions in our actual universe.[10] We can imagine boundary conditions that would give rise to coordinated behavior between distinct systems that would strike us as miraculous (Penrose 1989, chapter 7). In fact, we will see that such miraculous-looking correlations are common in the backward direction.

The three conditions jointly explain why, despite the global character of the physical dynamical laws, we can build local causal models with coarse-grained variables that are explanatory. This serves to fend off the worry that there is something deficient about local causal models. Rather than misrepresenting physical reality, local models represent

---

[10] Many philosophers have defended Randomness, as well as the fact that it has no analog in the backward direction, as central to the time asymmetry of causation. Randomness is entailed by the "statistical postulate" (Albert 2000, p. 96 and Loewer 2012, p. 124), the "asymmetry of randomness" (Frisch 2014, p. 243), the "asymmetry of independence" (Hausman 1998, chapter 4), the "initial micro-chaos condition" (Horwich 1987, p. 72), the "asymmetry of bizarre coincidences" (Kutach 2013, p. 175), and the "Principle of the Independence of Incoming Influences" (Price 1996, p. 26). I will say more below about how my account relates to these other accounts.

invariant and stable dependencies between local, coarse-grained variables.

5 The Time Asymmetry of Local Models

I have shown that local causal models in the forward direction can meet the goals of causal explanation. In this section, I will argue that these local models have no analogs in the backward direction because the physical conditions that allow us to find invariant and stable local models in the forward direction do not all obtain in the backward direction. In §6, I will then argue that this time asymmetry among local models accounts for the time asymmetry of causal explanation.

To show that there are no local models in the backward direction that are stable and invariant in typical circumstances, we need a way of assessing how earlier variables depend on later variables. As mentioned above, Woodward (2003, 2007) models interventions after ideal experimental manipulations. Ideal experimental manipulations, however, already presuppose a time asymmetry because in these manipulations we change variables 'from the past' (cf. Price 2007, p. 268 and Reichenbach 1956, p. 45). For example, to prepare a billiard ball to have a certain momentum at a time t, we would interact with the ball before t (by pushing it or making it roll down a slope) in order to create the desired condition at t.

It would be illegitimate to rely on these time-asymmetric interventions in the current context. We are trying to show that there is a physical time asymmetry between how earlier and later variables lawfully depend on each other that can ground a time asymmetry of causal explanation. In making this argument, we have to take care to not smuggle in a time asymmetry through the tool we use to test for lawful dependence, namely, the notion of an intervention. We therefore need a time-neutral notion of an intervention that guarantees that any time asymmetry of dependence we find is part of the physical world rather than a projection of ours.

The simplest time-neutral notion of an intervention is one where we suppose the value of a given variable to be changed without assuming any process that changes this value. Given this change in the intervened-on variable, we then 'update' variables at earlier and later times in accordance with the physical laws. Such interventions are time neutral because we are not assuming any changes (in the past or in the future) that make it the case that the variable in question has a new value. We can imagine such intervention in terms of the 'hand of God' reaching down and changing the value of the variable (cf.

Weslake 2006, p. 139).

In practice, we can typically change a given variable only by virtue of also changing other variables. However, interventions need only be conceptually possible (Woodward 2003, p. 132), and there is no conceptual difficulty with regard to time-neutral interventions. For example, in a time-neutral intervention on the momentum of a billiard ball at a time t, we change its momentum at t but hold all other variables fixed at t, and then use the laws of nature to propagate the changed state forward and backward to see how variables at other times change in response to this change at t. This procedure then tells us how both the earlier and the later momentum of the ball lawfully depend on its momentum at t.

In contrast to other ways of evaluating counterfactuals, such as the one proposed by Lewis (1986), counterfactuals based on time-neutral interventions are not meant to capture our ordinary judgments. They are a heuristic for illustrating how variables at one time depend on variables at other times in accordance with the physical laws. Having a clear picture of what lawful dependence is then allows us to justify why only lawful dependence in the forward direction underwrites causal explanations.[11]

Time-neutral interventions qualify as interventions in Woodward's sense. In Woodward's own words, "the intuitive idea is that an intervention on X with respect to Y changes the value of X in such a way that if any change occurs in Y, it occurs only as a result of the change in the value of X and not from some other source" (2003, p. 14). Time-neutral interventions on a variable X trivially guarantee that any associated change in Y is due to the change in X and not due to another source, because X is the only variable we change. With the notion of a time-neutral intervention in hand, we can show that local causal models in the forward direction have no analogs in the backward direction.

Physical processes in our universe are thermodynamically either *irreversible* or *re-*

---

[11] We sometimes assert "backtracking" counterfactuals, in which we reason from the non-occurrence of an event to the non-occurrence of its past causes and then forward again (Lewis 1986, p. 33). For example, we might reason that if the barometer reading had been different, then the earlier air pressure would have had to be different, and so the storm would also not have occurred. These counterfactuals are different from ones based on time-neutral interventions because they involve changing present facts other than the intervened-on variable. For example, when we backtrack from a difference in the barometer reading to a difference in the earlier air pressure and then reason forward again, we also assume that the air pressure is different at the time of the barometer reading. A time-neutral intervention on the barometer reading, in contrast, changes only the barometer reading and holds all other variables at the time, such as the air pressure, fixed. So time-neutral interventions, but not backtracking, tell us how variables at other times would depend on a difference in just the barometer reading.

versible. Irreversible processes go along with an increase in entropy. Examples include the melting of an ice cube, the shattering of a window, and the burning of wood. These processes are called “irreversible” because we never observe them in reverse temporal order. For example, we never observe an ice cube forming in a glass of water at room temperature or glass shards assembling into a window. Reversible processes, by contrast, do not involve an increase in entropy, and their time reversal is common and unremarkable. For example, consider a ball moving from one end of a horizontal billiard table to the other. The time reversal of this process, where the ball moves in the opposite direction, would be unsurprising. I will argue that neither type of process can be represented by local backward-looking models that are invariant and stable. More precisely, for each type of process, at least one of the conditions that allow us to build local models in the forward direction (namely, Insensitivity, Stability of Forces, and Randomness) has no analog if we try to model the process in the backward direction.

Consider an irreversible process such as a stone shattering my office window. As shown earlier, we can build invariant and stable models that depict how later stages of the process depend on earlier stages. For example, we can explain the window shattering as dependent on a stone flying earlier toward the intact window. However, we cannot build invariant and stable models that depict how earlier stages of the process depend on later stages, because the macroscopic past of systems undergoing thermodynamically irreversible processes is extremely sensitive to small changes to the final state of these processes. If the final state were just slightly different, it would almost certainly lawfully entail a different initial macroscopic state of the system (see, e.g., Dummett 1964; Elga 2000; Horwich 1987).

This sensitivity in the backward direction is easiest to see if we visualize the window shattering in reverse, that is, what we would see if we videotaped the process and played the movie backward. Viewed in reverse, we would see glass shards and a stone on the floor vibrate and eventually jump upward, with the glass pieces colliding and forming a smooth surface inside the wall opening just as the stone passes through. It is easy to see that this process is extremely sensitive to small changes in its initial condition. The initial state requires an extremely delicate coordination among its components (the particles in the floor, the glass pieces, and the stone) in order to lawfully evolve into a final macroscopic state in which a stone flies away from an intact window.

Numerous small changes could disrupt this coordination such that the initial state no longer lawfully evolves into this kind of macroscopic outcome. During the window shattering, the kinetic energy from the glass shards and stone that fall to the ground disperses into heat energy in the form of vibrations of the molecules in the floor. Hence, if we view the process in reverse, billions of particles in the floor bump into the glass pieces and the stone in parallel movements to set them in motion. In addition, the movements of each glass piece and the stone are coordinated with each other such that the glass pieces form an intact window just as the stone passes through. Finally, the glass pieces have to be in the right molecular state to form chemical bonds upon collision. Any small change in the floor, the glass pieces, the stone, or the outside forces would disrupt this delicate coordination. For example, if there were some added vibrations in the floor, if the gravitational forces were slightly different, or if some glass pieces were moved even a tiny bit, then the initial state where the glass pieces and the stone lie on the floor would almost certainly no longer lawfully evolve into a final state where a stone flies away from an intact window.

This lesson from visualizing the process in reverse shows that in the actual window shattering, numerous small differences in its final state would make it be the case that it no longer lawfully entails the initial macrostate. Visualizing the process in reverse illustrates the delicately coordinated behavior the particles need to undergo in order to get from the final state to the initial one. However, the particle motions we visualize in the forward direction when we consider the process in reverse are the exact same motions the particles undergo in the backward direction in the actual window shattering (cf. Elga 2000). So it is equally true that if the final state of the window shattering were different in some small respect, it would almost certainly no longer lawfully entail an earlier macrostate where a stone flies toward an intact window. Moreover, this sensitivity in the backward direction is a feature of all irreversible processes. Small changes to the final state of an irreversible process almost certainly make it the case that the final state no longer lawfully entails the initial macrostate (Albert 2000; Penrose 1989; Sklar 1993).

This sensitivity rules out local backward-looking models that are invariant and stable. We could try to explain why the window is intact in terms of what happened later. The intact state of the window does depend on time-neutral interventions on the later state in which the glass pieces and the stone lie on the floor, and such interventions create a new final state that is exactly like the actual state, except that neither glass pieces nor a stone lie on the floor. The simplest such state that can be conceived is one where the stone and the glass pieces are absent and instead replaced by some additional air molecules. If we take this new final state and run the laws backward, there is no intact window at the earlier time. After all, the intervention creates a state where the window is broken but where no glass pieces lie on the floor. Hence, if we evolve this state backward in accordance with the physical laws, the window is also broken earlier because, with the glass pieces absent, there are no materials that could fill the gap in the glass. The intact state of the window therefore depends on the intervention on the later state.[12]

This dependence, however, cannot support causal explanations because it is not stable. Stability would require that the dependence continues to hold in typical circumstances. However, we have just seen that if the final state of an irreversible process were just slightly different, it would almost certainly no longer lawfully entail the earlier macrostate. So if some of the background circumstances at the time when the glass pieces and the stone lay on the floor were slightly different, the state would almost certainly no longer entail the earlier existence of the intact window. In such circumstances, there is no intact window in the past regardless of whether a stone and glass pieces are on the floor, and so whether there is an intact window no longer depends on interventions on whether a stone and glass pieces later lie on the floor. The relevant model, therefore, is not stable in typical circumstances and hence not explanatory.

Remember that a necessary condition for our being able to find robust and invariant models in the forward direction is Insensitivity: the evolution of macroscopic systems toward the future is relatively insensitive to changes in their initial state. Insensitivity, however, has no analog in the backward direction for irreversible processes. The past evolution of systems undergoing irreversible processes is extremely sensitive to small differences in their final state. Hence, to build stable and invariant backward-looking models

---

[12] Albert (2000) and Loewer (2007, 2012) argue that the macroscopic past, at least in typical circumstances, never counterfactually depends on small macroscopic changes to the future. According to the Albert-Loewer account, a time-neutral intervention on the state of the glass shards and the stone on the floor almost certainly would not make a difference to the earlier macroscopic state of the window, though it would lead to small microscopic differences in the past. This restriction against backward dependence at the macroscopic level is supposed to follow from a lawful constraint on the initial macrostate of the universe, the so-called "past-hypothesis," and a probability distribution over possible microstates that could realize this initial macrostate. However, it is extremely contentious whether the past-hypothesis can be justified on physical grounds (Earman 2006) as well as whether it really would rule out macroscopic dependence toward the past (see, e.g., Frisch 2007, 2014; Price and Weslake 2009). My account also appeals to the special initial conditions of our universe (viz., that Randomness and Insensitivity have no equivalents in the backward direction) to supply a time asymmetry of causal explanation. However, it puts much less stress on the physics because it grounds a time asymmetry of causal explanation even if our best physics tells us (as I think it does) that the macroscopic past does depend on the macroscopic future.

of irreversible processes we would have to include all the numerous later variables that can make a difference to the past macroscopic behavior of the system. Hence, there are no local, backward-looking models of irreversible processes that meet the goals of causal explanation.

What about thermodynamically reversible processes? Imagine a stone sitting on my desk. This process is reversible. If we watched it on video, we could not even tell whether the tape is played forward or backward. In contrast to irreversible processes, many reversible processes are not sensitive in the backward direction. For example, small changes to the final state of a stone sitting on my desk are unlikely to make a difference to whether it was already sitting there earlier. Nonetheless, there is an important time asymmetry with respect to modeling reversible processes.

In the forward direction, we would explain the current location of the stone as dependent on its previous location. If nothing interferes, then the stone just remains where it was. However, there are many typical circumstances where something interferes with the stone sitting on my desk, such as someone picking up or pushing the stone. So an explanation also needs to mention at least those interferences that count as typical in the circumstances at hand. We can accommodate such interferences by mentioning a small number of macroscopic conditions (cf. Elga 2007, p. 106). For instance, we can provide an invariant and stable model that depicts the location of the stone as dependent on its earlier location plus the non-interference of eager cleaning personnel.

In the backward direction, by contrast, we cannot account for likely interferences by citing only a small number of later macroscopic conditions, because many typical interferences are due to irreversible processes that are extremely sensitive in the backward direction. Consider a scenario where a cleaner accidentally knocks the stone to a different part of my desk. If we visualize this process in reverse, we see how the stone suddenly begins to vibrate, because billions of air molecules and molecules in the desk bump into it, and moves toward the cleaner's arm. The interfering factor in the backward direction, therefore, is the coordinated behavior of the surrounding air molecules and the molecules in my desk. This process is irreversible and hence extremely sensitive in the backward direction. If any of these molecules moved just slightly differently, this difference would 'infect' the movements of other molecules, and hence, the molecules as a whole would no longer bump into the stone in the coordinated pattern necessary to knock the stone toward the cleaner's arm. So to model how the molecules in the desk would interfere with the past position of the stone, we have to mention even small external forces acting on the molecules—for if these forces were just slightly different, the molecules would no longer interfere with the past macroscopic behavior of the rock. Hence, a backward-looking model that represents possible interferences would have to be extremely global.

Moreover, we cannot just set aside these interferences as occurring only in atypical circumstances. Randomness has no equivalent in the backward direction, which means that delicate correlations among how forces act toward the past are common. Such correlations happen, for example, every time some object breaks or is dropped onto a surface.[13] So even when we model reversible processes, we need to take into account possible interferences by irreversible processes, which makes it impossible to model the processes locally. To sum up, the time asymmetry of local models arises because some of the physical conditions, namely, Insensitivity and Randomness, that allow us to build local models in the forward direction are absent in the backward direction.

6 The Time Asymmetry of Causal Explanation

Why does the fact that our local causal models in the forward direction have no analogs in the backward direction give rise to a time asymmetry in causal explanation? After all, even if there are no invariant and stable backward-looking local models, we can (at least in principle) build global models that are invariant and stable. The fundamental dynamical laws allow us, for example, to build a model of the breaking of a window that tells us how it depends on any intervention on past and future events. So why would a backward-looking model that tells us how the shattering depends on all variables at some future time not be a causal explanation? After all, such a model would be invariant and stable to the highest degree, namely, in all possible circumstances.

I will argue that the time asymmetry at the level of local models gives rise to a time asymmetry of causal explanation simpliciter because global models cannot support causal explanations. One reason is practical. Due to our epistemic limitations, we only can know about a relatively small, macroscopic portion of the world's state at any time. So it is essential for our practice of causal explanation that models depict outcomes as dependent on a small number of macroscopic conditions. In fact, even when we model systems in fundamental physics, we leave out information to make our models computationally more

---

[13] Martin Amis's (1991) novel Time's Arrow beautifully illustrates that numerous correlations between distinct systems that we would regard as miraculous were they to occur in the forward direction are common in the backward direction.

tractable (Frisch 2014, pp. 67–68). So global backward-looking models, even if they were in other respects analogous to forward-looking causal models, would not be of any practical use (cf. Eagle 2007, p. 184). This observation by itself is enough to justify, on practical grounds, why we only use forward-looking models in our explanations despite the time symmetry of the physical laws.

However, we can say more. Even if they were practically feasible, global models would lack an essential feature of causal models. An essential feature of causation is the distinction between actual and merely potential causes. Consider an example from Field (2003, p. 439): Sara is aiming the water hose at the fire while Sam is praying that the fire will go out. The fire's going out depends on both Sara's and Sam's actions: If Sara had not aimed the water hose at the fire, it would not have gone out. But if Sam had shot a hole in Sara's water hose instead of idly praying, the fire also would not have gone out.

A model of the fire going out that includes both Sara's and Sam's actions achieves greater stability than a model that includes only Sara's action. After all, the former model also correctly predicts what would have happened to the fire if Sam had shot a hole in Sara's water hose. Such an inclusive model, however, would not be a causal model because it does not distinguish between variables whose values bring about the outcome (Sara's action) from variables that are merely potential causes (Sam's action). As Woodward (2007, p. 84) puts it, a global model that includes any variable that potentially could make a difference to an outcome "threatens to collapse the distinction between causal and temporal priority and with it the whole point of the former notion."

This restriction against global models explains why global backward-looking models are not causal despite their invariance and stability. As an example, go back to the model that purports to explain the window's intactness as dependent on conditions at the later time when glass pieces and a stone lie on the floor. Any change to a variable at this later time that results in even a small change in the forces acting on the glass pieces or the stone would make a difference to the earlier intact window. So, since numerous variables that are part of the background conditions would exert such a force, an invariant and stable backward-looking model would have to include all of these variables. Such a model could then not distinguish between actual and potential causes and so would not count as a causal model. Hence, we have more than just practical reasons for disregarding backward-looking models as underwriting causal explanations.

A possible objection to this argument denies that the distinction between actual and potential causes is essential to causal models. Frisch (2014, p. 68) argues, against

Field (2003) and Woodward (2007), that the distinction is merely "an important pragmatic component of our concept of cause, which reflects the close connection between the concepts of cause and of explanation." According to Frisch, a model that cites both Sam's and Sara's actions as causes of the fire going out still counts as causal. It is merely that we typically would not cite Sam's action in a causal explanation because we regard changes to it that would make a difference to the fire going out, such as his shooting a hole in Sara's water hose, as too far-fetched.

I have two replies to Frisch's argument. First, the fact that the distinction is pragmatic does not entail that the distinction is not essential to causation. Most philosophers who endorse the difference-making conception of causation that I have adopted in this paper (and that Field, Frisch, and Woodward share) hold that all causal discourse is essentially pragmatic. On this conception, causation is closely connected to counterfactuals, specifically, to what would have happened if the cause had not occurred. Evaluating any causal claim therefore requires an antecedently understood or contextually supplied specification of how to conceptualize a situation where the cause is absent or different (Lewis 2004, p. 90 and Paul and Hall 2014, p. 51–52), and this specification, it seems, can only be pragmatic (see, e.g., Hitchcock 2007b). So pragmatic considerations are required for the notion of causation to even make sense. Hence, the fact that the distinction between potential and actual causes may rely on pragmatic factors does not show that it is not essential to causation.

Second, even if Frisch is right that the distinction is not essential to causation, it is still central to our practice of causal explanation. So the fact that backward-looking models do not allow us to distinguish actual from potential causes strengthens my earlier point that these models cannot support our ordinary practice of causal explanation. Someone might object to the idea that backward-looking models are merely pragmatically unsuited for causal explanations. After all, we ordinarily believe that backward-looking models are not causal at all. However, we also do not ordinarily think that Sam's praying is a cause of the fire's going out. So if Frisch is right that the distinction between causes and potential causes is pragmatic, we should not be surprised if the temporal direction of causation also turns out to be pragmatic.

Another possible objection to my argument is that if locality is an essential feature of causal explanations, then very sensitive or unstable outcomes in the future would not be causally explicable. For example, consider a pencil that is delicately balanced on my fingertip and that eventually falls. If we want to explain why the pencil fell in the direction it did, it seems we need to include every variable whose value could potentially make a difference to the pencil, because the smallest difference in the forces acting on the pencil would make a difference to how it fell. So it might seem that any invariant and stable model of the pencil's falling would have to be global and therefore would not be causal on my account. However, the objection concludes, it is implausible that outcomes like the pencil's falling should have no causal explanation.

I respond that a global model of the pencil's falling would indeed not be a causal model. In fact, such a global model could not underwrite a causal explanation precisely because it does not allow for a meaningful distinction between potential and actual causes. For example, consider the air molecules in various regions around the pencil. Regarding the air molecules in any such region we can ask: Did the motions of these molecules causally act on the pencil, or does the pencil's falling merely depend on them because their motions could have affected it although they actually did not? There is no meaningful answer to this question within the global model, because to achieve invariance and stability the model has to indiscriminately include the motions of all of these molecules.

Nonetheless, there is a causal explanation of the pencil's falling because there is a local, chancy model. This model attributes to the pencil an equal chance of falling in either direction. It thus allows us to single out the pencil's equal chances of falling in either direction while omitting numerous other variables on which its falling depends, such as the motions of air molecules. Attributing such chancy dispositions allows us to explain the behavior of the pencil within a local model that is stable and invariant. Due to Randomness, we can assume that, in typical circumstances, the pencil has an equal chance of falling to any side despite the numerous small forces acting on it (cf. Elga 2007, §5.3). So I can allow that sensitive or unstable phenomena have causal explanations without agreeing that global models would be causal. At the same time, we cannot build local backward-looking models by invoking chances. Since Randomness has no equivalent in the backward direction, we cannot attribute stable dispositions to behave in a certain way toward the past to systems (cf. Arntzenius 1997).

7 Conclusion

I have shown that the same physical conditions that account for why causal models can be local also account for the time asymmetry of causal models. My theory thus provides a satisfying reply to Russell's (1913) attack on causation. Russell argues that we cannot make sense of causation as a relation between localized events due to the global character of the physical laws. Moreover, he argues that once we think of causation as a relation between more global states, there is no time asymmetry of causation because of the time symmetry of the fundamental dynamical laws. My theory shows how we can make sense of causation as a relation between localized events (or variables), and it also shows that we find the time asymmetry of causal explanation at this local level.

The time asymmetry of causal explanation is, according to my account, ultimately due to the fact that Randomness and Insensitivity have no equivalents in the backward direction. This time asymmetry is plausibly closely related to the thermodynamic asymmetry, because Insensitivity and Randomness are part of many accounts of why our universe manifests a thermodynamic asymmetry.[14] Thus, my account is, in this respect, similar to many other accounts that hold that the time asymmetry of causation reduces to these same physical conditions.[15]

My account is distinctive in that it provides an explicit story of the time asymmetry of causal explanation. A successful account of the time asymmetry of causation needs to explain why the physical conditions in our universe make it a good idea for us to have a time-asymmetric causal concept. Most theories try to meet this challenge by arguing that a time-asymmetric causal concept is useful because we can control the future but not the past.[16] However, that does not seem to be the full story. Another important dimension of causation is its connection to explanation. We use our causal concept to track which events explain which other events, and it is not clear that our interest in explanations only concerns which events we can control. My account thus addresses an important piece of the puzzle of how causation fits into the physical world by showing why, given our interest in explanation, it is a good idea for us to have a time-asymmetric causal concept. Insofar as other accounts argue that the very same physical conditions also ground a time asymmetry regarding control, my account and these others are complementary and go

---

[14] See Albert (2000) and Horwich (1987). It is plausible that these physical asymmetries are non-causal and that we have therefore reduced the causal asymmetry to a non-causal physical asymmetry (but see Frisch 2014).

[15] See, e.g., Kutach (2013), Loewer (2012), and Price and Weslake (2009). Frisch (2014) also holds that the time asymmetry of causation is closely related to these physical time asymmetries, though he is skeptical that the former can be reduced to the latter.

[16] Hausman (1998) is an exception in that he offers a worked-out account of the time asymmetry of causal explanation. My account, however, is more general than Hausman's, which applies only to open systems, even though there may be no truly open systems in our universe (Hausman 1998, p. 67).

hand in hand.

Acknowledgements

Thanks to Bob Adams, Siegfried Jaag, Matthew Kotzen, Marc Lange, L.A. Paul, and John Roberts for comments on earlier drafts of this paper. Thanks also to two anonymous referees for this journal. Work on this paper was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG), (Grant: FOR 1063).

References

[1] Albert, D. (2000). Time and chance. Cambridge, MA: Harvard University Press.

[2] Amis, M. (1991). Time's arrow. New York: Vintage.

[3] Arntzenius, F. (1997). Transition chances and causation. Pacific Philosophical Quarterly, 78, 149–168.

[4] Cartwright, N. (1979). Causal laws and effective strategies. Noûs, 13, 419–437.

[5] Dummett, M. (1964). Bringing about the past. Philosophical Review, 73, 338–359.

[6] Eagle, A. (2007). Pragmatist causation. In Price & Corry (2007), pp. 156–190.

[7] Earman, J. (2006). The "past hypothesis": Not even false. Studies in History and Philosophy of Modern Physics, 37, 399–430.

[8] Eckhardt, W. (2006). Causal time asymmetry. Studies in History and Philosophy of Modern Physics, 37, 439–466.

[9] Elga, A. (2000). Statistical mechanics and the asymmetry of counterfactual dependence. Philosophy of Science, 68(Supplement), 313–324.

[10] Elga, A. (2007). Isolation and folk physics. In Price & Corry (2007), pp. 106–119.

[11] Field, H. (2003.) Causation in a physical world. In M. Loux & D. Zimmerman (Eds.), Oxford handbook of metaphysics (pp. 435–460). Oxford: Oxford University Press.

[12] Franklin-Hall, L. (2015). Explaining causal selection with explanatory causal economy: Biology and beyond. In C. Malaterre & P.-A. Braillard (Eds.), Explanation in biology (pp. 413–438). Dordrecht: Springer.

[13] Frisch, M. (2007). Causation, counterfactuals, and entropy. In Price & Corry (2007), pp. 351–396.

[14] Frisch, M. (2014). Causal reasoning in physics. Cambridge: Cambridge University Press.

[15] Hall, N. (2004). Rescued from the rubbish bin: Lewis on causation. Philosophy of Science, 71, 1107–1114.

[16] Hall, N. (2005). Causation. In F. Jackson & M. Smith (Eds.), The Oxford handbook of contemporary philosophy (pp. 505–533). Oxford: Oxford University Press.

[17] Hausman, D. (1982). Causal and explanatory asymmetry. In PSA: Proceedings of the biennial meeting of the Philosophy of Science Association, vol. 1 (pp. 43–54).

[18] Hausman, D. (1998). Causal asymmetries. Cambridge: Cambridge University Press.

[19] Hitchcock, C., and Woodward, J. (2003) Explanatory generalizations, part II: Plumbing explanatory depth. Noûs, 37, 181–199.

[20] Hitchcock, C. (2007a). What Russell got right. In Price & Corry (2007), pp. 45–65.

[21] Hitchcock, C. (2007b). Prevention, preemption, and the principle of sufficient reason. Philosophical Review, 116, 495–532.

[22] Horwich, P. (1987). Asymmetries in time. Cambridge, MA: MIT Press.

[23] Kutach, D. (2013). Causation and its basis in fundamental physics. New York: Oxford University Press.

[24] Lange, M. (2009). Laws and lawmakers: Science, metaphysics and the laws of nature. Oxford: Oxford University Press.

[25] Lewis, D. (1986). Counterfactual dependence and time's arrow. In Philosophical papers, vol. II (pp. 32–52). Oxford: Oxford University Press.

[26] Lewis, D. (2004). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), Causation and counterfactuals (pp. 75–106). Cambridge, MA: MIT Press.

[27] Loewer, B. (2007). Counterfactuals and the second law. In Price & Corry (2007), pp. 293–326.

[28] Loewer, B. (2012). Two accounts of laws and time. Philosophical Studies, 160, 115–137.

[29] Paul, L. A., & Hall, N. (2014). Causation: A user's guide. Oxford: Oxford University Press.

[30] Penrose, R. (1989). The emperor's new mind. Oxford: Oxford University Press.

[31] Price, H. (1996). Time's arrow and Archimedes' point. Oxford: Oxford University Press.

[32] Price, H. (2007). Causal perspectivalism. In Price & Corry (2007), pp. 250–292.

[33] Price, H., & Corry, R. (2007). Causation, physics and the constitution of reality. Oxford: Oxford University Press.

[34] Price, H. & Weslake, B. (2009). The time-asymmetry of causation. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), The Oxford handbook of causation (pp. 414–446). Oxford: Oxford University Press.

[35] Reichenbach, H. (1956). The direction of time. Berkley: University of California Press.

[36] Russell, B. (1913). On the notion of cause. Proceedings of the Aristotelian Society, 13, 1–26.

[37] Schaffer, J. (2010). Review of Price and Corry's Causation, physics, and the constitution of reality. Mind, 119, 844–848.

[38] Sklar, L. (1993). Physics and chance: Philosophical issues in the foundations of statistical mechanics. New York: Cambridge University Press.

[39] Strevens, M. (2004). The causal and unification approaches to explanation unified—causally. Noûs, 38, 154–179.

[40] Weslake, B. (2006). Review of Making things happen. Australasian Journal of Philosophy, 84, 136–140.

[41] Weslake, B. (2010). Explanatory depth. Philosophy of Science, 77, 273–294.

[42] Woodward, J. (2003). Making things happen. Oxford: Oxford University Press.

[43] Woodward, J. (2007). Causation with a human face. In Price & Corry (2007), pp. 66–105.