Special-science counterfactuals

Christian List*
LMU Munich

Published in *The Monist* 105(2), 2022, pp. 194–213

Abstract: On the standard analysis, a counterfactual conditional such as "If *P* had been the case, then *Q* would have been the case" is true in the actual world if, in all nearest possible worlds in which its antecedent (*P*) is true, its consequent (*Q*) is also true. Despite its elegance, this analysis faces a difficulty if the laws of nature are deterministic. Then the antecedent could not have been true, given prior conditions. So, it is unclear what the relevant "nearest possible worlds" are. David Lewis suggested that they are ones in which a local breach of the laws occurred: a "small miracle". Others have suggested that they are ones in which the initial conditions were different ("backtracking"). I propose another response. It builds on the idea that the special sciences, where counterfactual reasoning is most common, operate at a higher level of description from fundamental physics, and that the world may behave indeterministically at higher levels even if it behaves deterministically at the fundamental physical one. The challenge from determinism can then be bypassed for many special-science counterfactuals.

1. Introduction

Counterfactual reasoning is extremely common in the special sciences. In disciplines ranging from biology, ecology, and medicine to history, law, and the social sciences, counterfactuals play an important role. Think of statements such as the following:

- If atmospheric CO_2 had remained at pre-industrial levels, there would be less bleaching of the world's coral reefs.
- If the patient had been treated earlier, he would have lived longer.
- If the central bank hadn't raised the interest rate, inflation would have gone up.
- If Jones hadn't administered the poison, Smith wouldn't have died.

We consider such counterfactuals not only intelligible, but also informative. They tell us something about the causes of certain events: which factors made a difference to which outcomes. Indeed, a counterfactual analysis of causation is very common in the special sciences. The significance and robustness of various causal factors is often also identified by considering counterfactual variations of certain antecedent conditions.

* I thank Sander Beckers, Helen Beebee, Michael Esfeld, Barry Loewer, the participants of two MCMP seminars at LMU Munich in March and October 2021, and an anonymous reviewer for helpful comments and suggestions. I also gratefully acknowledge my continuing affiliation with the LSE's Philosophy Department as a Visiting Professor.

The prominence of counterfactual reasoning, however, raises the question of how we should interpret counterfactual conditionals. What are their truth-conditions? The standard approach is to invoke "possible worlds" to evaluate them (Lewis 1973, Stalnaker 1968). In brief, a counterfactual conditional is true on such an analysis if, in all nearest possible worlds in which the antecedent, the "if" clause, is true (e.g., atmospheric CO_2 had remained at pre-industrial levels), the consequent, the "then" clause, is also true (e.g., less coral bleaching would have occurred). This analysis is elegant and plausible, but faces a difficulty if the laws of nature are deterministic. The truth of a conditional's antecedent will then have been impossible in the actual world at the relevant time, given everything that happened before: the antecedent was "historically inaccessible". So, it is unclear what the "nearest possible worlds" are in which the antecedent is true. In a deterministic world, atmospheric CO_2 couldn't have remained at pre-industrial levels, given the actual prior conditions.

My aim is to present a new response to this problem. It builds on the idea that the special sciences, where we typically engage in counterfactual reasoning, operate at a higher level of description from fundamental physics, where we find the basic laws of nature, and that the world need not behave deterministically at the relevant higher levels even if it behaves deterministically at the fundamental physical one. The challenge from determinism can be bypassed if our counterfactual reasoning takes place at those special-science levels at which determinism doesn't hold.

In Section 2, I explain the challenge in more detail. In Section 3, I review some standard responses. In Section 4, I give an informal sketch of my own response, followed by a more formal development in Section 5. In Section 6, I conclude with some further discussion.

2. The challenge

Consider a counterfactual conditional such as the following:

(C) If John F. Kennedy had not been assassinated, Lyndon B. Johnson would not have become President of the United States in 1963.

This conditional seems true. According to the standard analysis, we can vindicate this verdict by

- considering the nearest possible worlds in which the antecedent (the "if" clause) is true, for instance worlds in which the bullet missed Kennedy or in which Lee Harvey Oswald didn't fire his shot, and
- asking whether the consequent (the "then" clause) is also true in such worlds.

If Johnson wouldn't have become President in 1963 in those worlds, then conditional (C) is indeed true. Generally, a conditional of the form "If *P* had been the case, then *Q* would have been the case" is true in the actual world if and only if *Q* is true in all nearest possible worlds in which *P* is true.[1]

However, this analysis runs into a difficulty if the world is deterministic. Determinism is the thesis that, given the world's initial conditions, only one sequence of events is nomologically possible, i.e., permitted by the laws of nature. The laws of nature, thus, constrain the possible sequences of events such that, once the world's initial conditions are fixed, everything that happens thereafter is fixed too. Suppose this is true. Then the scenario in which Kennedy wasn't killed in 1963 wasn't possible at that time. Holding fixed everything that happened before, the only possible sequence of events under the laws of nature was one in which Kennedy was assassinated. The antecedent of the conditional was – as we may put it – *historically inaccessible* in the actual world. Technically, a proposition is *historically inaccessible* in a world if there is no initial segment of that world's history that admits a nomologically possible continuation in which that proposition is true; I will formalize this later.

If we still wish to claim that conditional (C) is true, we must say that if Oswald hadn't killed Kennedy,

(1) the laws of nature would have been breached (or would at least have been different), or
(2) the distant past, say at the time of the Big Bang, would have been different.

So, counterfactual reasoning would force us to suspend the conventional assumptions of the fixity of the laws or the fixity of the past, insofar as the nearest worlds in which a conditional's antecedent is true would have to satisfy (1) or (2). Neither option seems particularly plausible. The only alternative would be to say that

(3) every counterfactual conditional is true by default.

If the laws and the past are fixed and determinism is true, there simply are no nearest possible worlds in which Kennedy wasn't killed, and then it is vacuously true that in *every* such world Johnson wouldn't have become President in 1963. But that would trivialize our analysis of counterfactuals. We would have to agree not only that if Kennedy had survived, Johnson wouldn't have become President in 1963, but also that if Kennedy had survived, the Earth would have been invaded by extra-terrestrials in 1964, which is preposterous. Thus option (3) is unattractive.

---

[1] For simplicity, this analysis relies on the *limit assumption*, made by Stalnaker but not by Lewis, which guarantees that *if* there are any possible worlds in which *P* is true, *then* some of them count as *nearest* to the actual world. Stalnaker further assumes that there is a *unique* nearest possible world in which *P* is true. Lewis does not make either assumption. Without the limit assumption, the truth-condition for "If *P* had been the case, then *Q* would have been the case" becomes this: *if* there are any possible worlds in which *P* is true, then, among them, some in which *Q* is true are nearer to the actual world that any in which *Q* is not true.

In sum, we are faced with a trilemma (see Dorr 2016). To uphold the judgment that conditionals such as (C) can be true in a deterministic world, we must choose between three seemingly unattractive options:

(1) giving up the fixity of the laws of nature,
(2) giving up the fixity of the past, and
(3) trivializing all counterfactuals.

What can we say in response?

3. Some responses

Assuming we don't wish to trivialize or abandon standard counterfactual reasoning, we must choose between options (1) and (2). Alternatively, we must argue that the world isn't deterministic, so that the challenge doesn't arise.

Let us briefly run through these responses. Some scholars choose option (1) and lift the assumption that the laws of nature are fixed when it comes to counterfactuals. One way of developing this idea, suggested by Lewis (1973), is to postulate "small miracles" to make sense of conditionals such as (C) against the background of determinism. A "small miracle" is a localized breach of the laws of nature making the antecedent true at the relevant time, where this breach doesn't affect any prior events and lasts only very briefly, so that the laws come into effect again once the "miracle" has happened. The nearest possible worlds in which Kennedy wasn't assassinated could then have been ones in which everything was equal until the crucial morning, but the bullet took a different path, missing Kennedy. If determinism is true, such worlds would have been nomologically impossible, given prior conditions. But Lewis finds this acceptable, because the postulated breach of the laws would be minor and would save us from having to say that if Oswald had decided not to pull the trigger, the world's initial conditions would have been different. In this way, Lewis is able to avoid any apparent implications of backwards causation or counterfactual dependence of the distant past on the present (for discussion, see Tomkow and Vihvelin 2017).

A more radical way of giving up the fixity of the laws, inspired by David Hume's ideas, would be to argue that, as the world's history is unfolding, the laws of nature are not fixed in a metaphysically binding, necessitarian way at all, and that those laws descriptively emerge only in the limit or at the hypothetical end of time. On such a Humean, "non-governing" conception, the "laws" are just a description of the patterns and regularities that best summarize the "mosaic" of actual contingent facts that hold in a world across time (Beebee 2000; see also Braddon-Mitchell 2001). Before the end of time, it thus remains open what the descriptively adequate system of laws will be. If the antecedent of some counterfactual had been true, say, if Oswald had decided not to pull the trigger, a different chain of events would have unfolded, thereby rendering a slightly different system of laws descriptively adequate. A downside of this response is that it abandons the conventional "governing conception" of the laws of nature, under which these impose strictly binding modal constraints on what can or cannot happen in any world at any time.[2]

To retain the conventional assumption of fixed laws, other scholars choose option (2) and give up the fixity of the past in counterfactual reasoning, especially the fixity of the world's initial conditions. The result is a so-called "backtracking" interpretation of counterfactuals. According to it, if Oswald had missed his target or decided to act otherwise, the world's initial conditions would have been slightly different from what they actually were. Dorr (2016), Esfeld (2021), and Loewer (2020) have recently defended versions of this idea, the latter two in relation to human action and free will (for analyses supportive of backtracking, see also Loewer 2007 and Tomkow and Vihvelin 2017). By allowing backtracking, they need not postulate any miracles and can also retain a governing conception of the laws. However, they must confront the worry that they are committed to some form of backwards causation. If I had chosen to have tea instead of coffee this morning, for instance, the world's initial conditions would have been slightly different on the given proposal, and if we understand causation as counterfactual difference-making, it looks (at least on some versions of such an understanding) as if I could have causally affected what happened before my birth. I will come back to this problem in Section 6.

The remaining alternative is to argue that the world isn't deterministic, so that the identified challenge for counterfactual reasoning doesn't arise. If the laws of nature are indeterministic, they admit "forks in the road", i.e., turning points at which the course of events could have gone one way or another. In the case of counterfactual (C), one might say that there was a fork in the road on that fateful November day. While in the actual world the assassination happened, a random gust of wind could have deflected the bullet or Oswald could have freely chosen not to fire. So, there was a nomologically possible course of events with the same past in which Kennedy survived. The nearest possible worlds in which Kennedy wasn't assassinated could then have coincided with the actual world in both the initial conditions and the laws.

---

[2] On the relative merits of regularity-based versus necessity-based accounts of laws in relation to the analysis of counterfactuals, see also Noordhof (2020, section 14.3.5). For a discussion of the modal resilience of the laws on a Humean conception, see further Loew and Jaag (2020). One might wonder whether laws that descriptively emerge only in the limit or at the hypothetical end of time could be consistent with determinism at all. However, on one standard definition of determinism, even laws of a non-governing sort could have a functional form that is deterministic, in the sense that, given a particular set of initial conditions, only one sequence of events is consistent with the relevant functional form. Of course, on a more demanding definition of determinism, which refers to some kind of necessitation, it is less clear whether non-governing laws could count as deterministic. For a discussion of different ways of defining determinism, contrasting "entailment" and "necessitation" definitions, see Steward (2021).

It is unclear, however, whether nature is indeterministic in this way. In fundamental physics, the jury is still out on this matter. Special and general relativity theory, two leading theories of macroscopic phenomena, seem to support determinism, as does classical Newtonian mechanics. Quantum mechanics, a leading theory of microscopic phenomena, is often thought to support indeterminism (especially on its traditional Copenhagen interpretation), but it also admits deterministic reinterpretations (e.g., the Bohmian one). Moreover, the competing physical theories have not yet been integrated into a "grand unified theory", and it is open whether such a theory would support determinism or not. It therefore seems a gamble to make our analysis of conditionals contingent on how certain debates in physics play out. Nevertheless, I will argue that determinism can be denied in a principled way.

4. A sketch of my response

My response builds on the observation that, outside ordinary discourse, counterfactual reasoning (especially of the sort that is most naturally captured by the standard analysis) is more common in the special sciences than in fundamental physics, i.e., in disciplines ranging from biology to the social sciences.[3] Crucially, these sciences represent the world at a different level of description from fundamental physics. The significance of this, I will explain, is that the world may behave indeterministically at the relevant special-science levels, even if it behaves deterministically at the level of physics. So, there may be "forks in the road" at the level of description that matters for the analysis of counterfactuals, even if there are no such forks at the fundamental physical level. This implies that many special-science counterfactuals have historically accessible antecedents in a relevant sense, and the choice between the unpalatable options (1), (2), and (3) does not arise.

I develop this response in three steps. First, I draw a distinction between the world as described at the fundamental physical level and the world at a higher level, associated with some special science, and suggest that the latter, not the former, matters for the analysis of special-science counterfactuals. Then I show that indeterminism at a higher level can go along with determinism at a lower one. Finally, I explain what this implies for counterfactuals.

[3] Fundamental physics sometimes uses thought experiments, but these are arguably quite different from paradigmatic counterfactuals such as conditional (C) above. The historical inaccessibility of certain hypothetical conditions in a thought experiment poses no problem akin to the problems it poses for special-science counterfactuals such as (C). The observation that, outside the rather special context of thought experiments, counterfactual reasoning is more common in the special sciences than in fundamental physics is also consonant with the often-made suggestion that fundamental physics is not the most hospitable context for the notion of causation under a counterfactual interpretation, and that causation is more of a special-science notion than a fundamental physical one. Woodward (2009, p. 257), for instance, writes: "the view that fundamental physics is not a hospitable context for causation and that attempts to interpret fundamental physical theories in causal terms are unmotivated, misguided, and likely to breed confusion is probably the dominant, although by no means universal, view among contemporary philosophers of physics."

4.1 The distinction between the physical level and higher levels

A salient feature of the practice of science is its organization into different domains of enquiry. When we describe and explain phenomena in different domains, we use different concepts and categories: we operate at different *levels of description* (see List 2019b for an account of levels and further literature review). For example, in fundamental physics, we refer to particles, fields, and forces; in biology, we refer to cells, organisms, and ecosystems; and in psychology, we refer to agents and their mental states. We follow this practice for good reasons. Different explanatory tasks require the use of a different conceptual repertoire, which, in turn, allows us to identify different features and patterns in the world. As a result, each science focuses on a somewhat different class of facts and properties, namely those that can be described using the domain-specific concepts and categories. Fundamental physics focuses on *physical-level facts and properties*, while the special sciences, such as biology, psychology, and sociology, focus on certain *higher-level facts and properties*. Crucially, higher-level facts may be hard or impossible to describe using only lower-level language. Facts about agents and their actions, for instance, may be hard or impossible to describe in purely physical terms; and even facts about cells or organisms may be hard to capture using the language of fundamental physics alone.[4]

If we define the *world at the physical level* as the totality of physical-level facts, and we define the *world at a given higher level*, associated with some special science, as the totality of corresponding higher-level facts, we can see that the two are distinct. Of course, the world at a higher level is not unrelated to the world at the physical level but depends on it. Plausibly, higher-level facts supervene on physical-level ones. But the converse dependence does not hold. For instance, the biological-level facts are insufficient to settle the physical-level ones, and so the world at the biological level is more coarse-grained than its physical-level counterpart. Biology abstracts away from extraneous physical details, and it does so for good explanatory reasons.

For each level, there is thus a level-specific set of nomologically possible worlds. At the physical level, it consists of all physical-level worlds that satisfy the fundamental physical laws. At the biological level, it consists of all biological-level worlds that satisfy the relevant biological laws. At the psychological level, it consists of all psychological-level worlds that satisfy the relevant psychological laws. And so on. Importantly, while some readers might be inclined to view such level-specific worlds as nothing more than epistemic constructs, I prefer an ontic interpretation. I

[4] The difficulty of reducing higher-level descriptions to lower-level ones arguably already occurs when we move from the level of physics to that of chemistry. Despite the close connections between physics and chemistry, philosophers of chemistry have argued that even seemingly straightforward chemical notions, such as acidity, need not necessarily have an exactly matching physical counterpart. See, e.g., Manafu (2015). For a book-length study of the question of whether chemistry is reducible to physics, see Hettema (2012).

suggest that the levelled nature of our explanatory practices lends support to a levelled ontology, in which each level is associated with its own level-specific set of possible worlds.

Now let us return to the subject of counterfactuals. The idea that, when we answer questions in a particular domain, we must do so at the relevant level of description applies here, too. Counterfactuals in fundamental physics must obviously be evaluated at the fundamental physical level, by asking what would be the case in the nearest possible physical-level worlds. However, counterfactuals in any special science, such as biology, psychology, or sociology, must be evaluated at the appropriate higher level, i.e., by asking what would be the case in the nearest possible higher-level worlds.

Standard practice corroborates this approach. Consider the earlier example: "If the central bank hadn't raised the interest rate, inflation would have gone up." On a natural reading, this conditional is true if, and only if, in all nearest economically possible worlds in which there was no interest-rate increase, inflation would have gone up. A nearest such economically possible world can be understood as a hypothetical complete specification of the economy in which there was no interest-rate increase, with all other features of the economy held equal. We are not asking what the world would have been like at the level of particle physics if the antecedent had been true, or at the level of molecular biology. That would be the wrong level of analysis. Rather, we consider what the world would have been like at the level of macro- or micro-economics. Thus, we are focusing on the nearest possible worlds at an appropriate higher level, not at the physical one.

4.2 Why higher-level indeterminism can coexist with lower-level determinism

One might think that if the fundamental physical laws are deterministic, then the world is deterministic *simpliciter*, regardless of the level of description. But this is incorrect. The distinction between determinism and indeterminism is level-specific: the question of whether the world is deterministic or indeterministic is ill-posed until we specify the level at which we are asking this question. The world at a physical level could well be deterministic, while the world at some higher level could be indeterministic.

To make this precise, I draw on a simple dynamical-systems model (List 2014, List and Pivato 2015). At each level, we think of the world as a sequence of states across time, called a *history*, where the states in the sequence are specified at the relevant level. At the physical level, the world is a sequence of physical-level states: a *physical-level history*. At the biological level, the world is a sequence of biological-level states: a *biological-level history*. And so on.

For each level, we can think of the laws as specifying which histories are possible at that level, and which not. The fundamental physical laws, for instance, specify which sequences of physical states constitute nomologically possible physical-level histories. Figure 1 (from List 2019a) gives a toy example. Dots represent physical-level states, and lines from the bottom to the top represent nomologically possible physical-level histories, across six time periods, from $t = 1$ to $t = 6$.

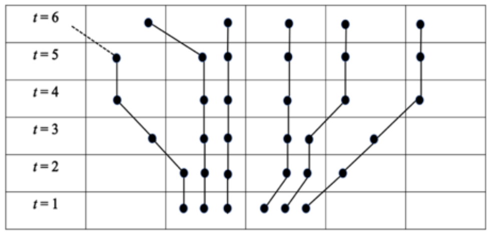

Figure 1: Physical-level histories

In this example, the physical-level histories are clearly deterministic: any initial segment of any one of them admits only one nomologically possible continuation. There is never any branching.

Now, if we turn to a higher level, say that of biology, psychology, or sociology, we must consider the appropriate higher-level histories. As noted, a higher-level history is a sequence of higher-level states across time, rather than a sequence of physical-level states. I have mentioned the assumption that higher-level states *supervene* on physical-level states: if we fix the detailed physical-level state of the universe, we thereby also fix the resulting higher-level state, such as the biological-level state. If we also make the standard assumption that higher-level properties are *multiply realizable* at the lower level, then it follows that the higher-level states are more coarse-grained than the physical-level states. That is, a given higher-level state – a macro-state, say a biological, psychological, or sociological one – may be realized by a variety of different physical micro-states. A higher-level state thus corresponds to an equivalence class of physical-level states, consisting of its different possible physical realizers. These realizing states are then indistinguishable at the higher level.

In line with this, suppose that whenever two or more distinct physical-level states fall into the same cell within the rectangular grid in Figure 1, they give rise to the same higher-level state, say, the same biological, psychological, or sociological state. The higher-level states and resulting histories will then be as shown in Figure 2 (also from List 2019a).

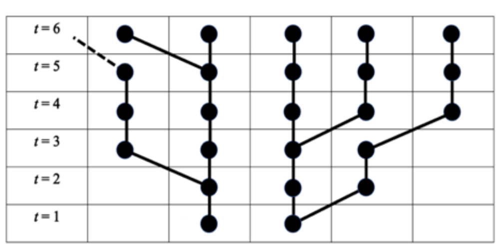

Figure 2: Higher-level histories

Evidently, these higher-level histories – sequences of higher-level states – are indeterministic. Given the same initial segment of a history, two or more distinct continuations are possible. If the higher-level state at time $t = 1$ is the state on right-hand side in the bottom row, for instance, then the history could unfold in three ways.

This example illustrates that even when there is determinism in physical-level histories, there could be indeterminism in higher-level histories. The converse scenario is also coherent: indeterminism in lower-level histories can co-exist with determinism in higher-level histories. I will not discuss that scenario here, but note that physical-level indeterminism is neither necessary nor even sufficient for higher-level indeterminism (for earlier similar or related results, see Loewer 2007, Werndl 2009, Butterfield 2012, and Yoshimi 2012).

One might ask whether higher-level indeterminism in the presence of physical-level determinism is just a contrived theoretical possibility or whether it is realistic. My answer is that our best theories in the special sciences represent many higher-level phenomena as indeterministic, from random mutations in genetics and the weather in meteorology to human choices in psychology and the financial markets in economics. Indeed, the human and social sciences offer explanations of human behaviour in which the idea that people face choices between different courses of action is indispensable. The kinds of intentional explanations which economists, sociologists, political scientists, and historians give rely centrally on the assumption that people are choice-making agents: they might make these choices rationally or not, and they may be influenced by background conditions, but crucially, different options are open to them – an idea I have called "agential-level indeterminism" (List 2014).

Furthermore, the use of indeterministic models in the special sciences is independent of any assumptions about fundamental physics. Some scientists explain the apparent higher-level indeterminism away by suggesting that it should be viewed as merely epistemic: as stemming from our ignorance of the micro-details underpinning our macro-models or from our computational limitations. The reason why we postulate either randomness or free choices in the special sciences, they say, is that we lack the ability to compute the full deterministic state evolution of the underlying physical system. In response, however, we can say this: since the distinction between determinism and indeterminism is level-specific, it would be unsystematic to treat only physical-level indeterminism – should it obtain – as *ontic* (a feature of reality) while treating any higher-level indeterminism as *merely epistemic* (List and Pivato 2015). A naturalistic approach to ontology would support a realist attitude towards the commitments of our best scientific theories *at all levels*. If our best theories represent the phenomena in their domains as indeterministic, then so be it: we should take this at face value, as something real (ibid., List 2019a).

4.3 Lessons for the analysis of counterfactual conditionals

I have argued that the world is plausibly indeterministic at the levels that matter for many of the special sciences. In consequence, many counterfactuals in the relevant domains have historically accessible antecedents, after all. If, from the perspective of history or psychology, Oswald faced a real choice on that November morning, then there is a possible world *at the relevant level* in which Kennedy wasn't assassinated and where the level-specific laws and the past coincide with those of the actual world. Such a world would have been nomologically possible at the time, given the level-specific prior conditions, and could be viewed as a nearest possible world in which the antecedent of conditional (C) is true. This would non-trivially support the truth of (C) while respecting fixity of the laws and fixity of the past.

I suggest that this kind of case is not the exception but the rule when it comes to special-science counterfactuals. Recall the earlier examples:

- If atmospheric CO_2 had remained at pre-industrial levels, there would be less bleaching of the world's coral reefs.
- If the patient had been treated earlier, he would have lived longer.
- If the central bank hadn't raised the interest rate, inflation would have gone up.
- If Jones hadn't administered the poison, Smith wouldn't have died.

Each of these is best analysed at a level well above fundamental physics and one where the relevant antecedent was in fact historically accessible. In climate science and ecology, for instance, postulating choice nodes at which human societies could embark upon different pollution paths is not just appropriate but standard practice. Equally, in medicine and macroeconomics, it is standard to postulate choice nodes at which different medical or economic interventions are possible. Finally, in the law, we routinely assume that a person who caused some harm could have acted otherwise (List 2019a, ch. 4).

In each example, the question of whether the physical state evolution of the universe would have permitted the antecedent of the relevant conditional to be true is neither here nor there. What matters is that, at the appropriate higher level, the antecedent could have been true, given the more coarse-grained higher-level state at the relevant time and the nomological constraints that apply at that level.

My proposal, then, is that we evaluate any special-science conditional strictly at the appropriate higher level. For instance, take a conditional of the form "If $\mathbb{P}$ had been the case, then $\mathbb{Q}$ would have been the case", where $\mathbb{P}$ and $\mathbb{Q}$ denote higher-level expressions. My convention, for the rest of this paper, will be to use the outline font for higher-level expressions. To evaluate the given conditional, we consider the nomologically possible worlds at the appropriate higher level and ask whether $\mathbb{Q}$ (the consequent) is true in all nearest higher-level worlds in which $\mathbb{P}$ (the antecedent) is true: the *nearest* $\mathbb{P}$*-worlds*. Specifically, we take a higher-level world $\boldsymbol{w}$ to qualify as a nearest $\mathbb{P}$-world to the actual world *only if*

(1) $\mathbb{P}$ is true in $\boldsymbol{w}$; and
(2) $\boldsymbol{w}$ shares some initial segment (up to some time *t*) with the actual world.

If there is only one world $\boldsymbol{w}$ that satisfies (1) and (2), then that qualifies as the unique nearest possible world in which $\mathbb{P}$ is true. If, on the other hand, there is more than one such world, as we may often expect, then we either treat all of them as nearest or, more plausibly, we employ some tie-breaking criterion to pick out a subset, or a single one, among them. This criterion could, but need not, include the constraint that, for a $\mathbb{P}$-world $\boldsymbol{w}$ to count as "nearest" to the actual world, its shared initial segment with the actual world must be maximally long, i.e.,

(3) there are no worlds $\boldsymbol{w}'$ distinct from $\boldsymbol{w}$ which also satisfy (1) and (2) and share a longer initial segment with the actual world.[5]

However, if there are several branching points at which some $\mathbb{P}$-world would have been accessible from the actual world, worlds that branch off later need not always qualify as "nearer" to the actual world than worlds that branch off earlier. For instance, the nearest worlds in which my bag was stolen yesterday were presumably not ones in which it was stolen just before midnight (even though such worlds might have been accessible), but rather ones in which it was stolen during a cramped bus ride in the afternoon (a point recognized, e.g., by Bennett 2003, p. 220).

In any case, for my purposes, it suffices to note that, given higher-level indeterminism, constraints (1) and (2) are in principle satisfiable, and different case-specific tie-breaking criteria may be available to narrow down the worlds satisfying these constraints. Crucially, the higher-level $\mathbb{P}$-worlds on which we focus in this analysis of a conditional not only match the actual world at the higher level in both the laws and the past, but the match is perfect up to the crucial "fork in the road" at which the truth-value of $\mathbb{P}$ was settled in one way rather than another. We have thereby bypassed the trilemma that arose under determinism.

[5] For (3) to be satisfiable along with (1) and (2), we must make the following *limit assumption*: if $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, $\boldsymbol{w}_3$, … is a sequence of nomologically possible $\mathbb{P}$-worlds which all share an initial segment with the actual world and where those initial segments are weakly increasing in length, then there exists a nomologically possible $\mathbb{P}$-world $\boldsymbol{w}^*$ which shares an initial segment with the actual world that is at least as long as all those the $\boldsymbol{w}_i$s share with the actual world. This limit assumption is guaranteed to hold if time is discrete. Without the limit assumption, we may adopt Lewis's analysis of conditionals mentioned in footnote 1, and replace (3) with the stipulation that, whenever a world $\boldsymbol{w}$ shares a longer initial segment with the actual world than another world $\boldsymbol{w}'$, then $\boldsymbol{w}$ is nearer to the actual world than $\boldsymbol{w}'$. A further tie-breaking criterion could be used in case $\boldsymbol{w}$, $\boldsymbol{w}'$ share an equally long initial segment with the actual world.

In sum, I have suggested that we should truth-evaluate counterfactuals at the level at which they are expressed. If a counterfactual is expressed in the language of fundamental physics, then the appropriate level of truth-evaluation will be the fundamental physical one. But that's not the typical case. As many counterfactuals are expressed in the language of some special science, the level of truth-evaluation should be a higher one. And since higher-level worlds can be indeterministic even if lower-level worlds are deterministic, the analysis of special-science counterfactuals is then immune to the difficulty that arose at a deterministic physical level.

5. A simple formalization

To formalize my proposal (building on List 2014, List and Pivato 2015), I begin with the notion of a "physical-level world". This is a fully specified history of the world at the physical level. Let $T = \{0,1,2,\ldots\}$ be the set of all points in time, where, for simplicity, time is discrete and has a beginning.[6] Let S be the set of all possible physical-level states in which the world could be at some time. Call S the *physical-level state space*. We can then define a *physical-level history* as a function h from T into S, which assigns to each time $t \in T$ the state $h(t) \in S$ in which the world is at that time. In our example of Figure 1 above, the little dots represent physical-level states in S, and the lines from bottom to top represent physical-level histories. Let Ω_{nom} denote the set of all physical-level histories that are *nomologically possible*, i.e., permitted by the physical laws of nature. Unless the laws of nature are totally unconstraining, Ω_{nom} will be a proper subset of the set of all *logically possible* histories (functions from T into S), which we may denote Ω_{log}.[7]

To distinguish between deterministic and indeterministic histories, let me introduce the notion of an *initial segment* of a history h up to time t, denoted h_t. This is the restriction of the function h to all points in time up to and including t. A history $h \in \Omega_{nom}$ is *deterministic* if its initial segment up to any time $t \in T$ admits only one continuation in Ω_{nom}; formally, for any $h' \in \Omega_{nom}$ and any $t \in T$, $h'_t = h_t$ implies $h' = h$. History $h \in \Omega_{nom}$ is *indeterministic* if its initial segment up to some time $t \in T$ admits two or more distinct continuations in Ω_{nom}; formally, there exists some $h' \in \Omega_{nom}$ and some $t \in T$ such that $h'_t = h_t$ but $h' \neq h$. Deterministic laws of nature imply that all histories in Ω_{nom} are deterministic. Indeterministic laws imply that some histories in Ω_{nom} are indeterministic.

[6] Alternatively, one could take T to be the set of non-negative real numbers, but this would give rise to some technical complications that I wish to avoid. These are due to the fact that bounded sets of real numbers, unlike bounded sets of natural numbers, do not generally have a minimum or maximum but only an infimum or supremum. In particular, the limit assumption, discussed in previous footnotes, could be violated.

[7] In this way, Ω_{nom} encodes (at least extensionally) the laws of nature. Different specifications of the laws of nature would correspond to different specifications of Ω_{nom}. More restrictive laws would pick out fewer histories as nomologically possible, more permissive laws would pick out more histories.

We next turn to the truth-evaluation of counterfactuals. By hypothesis, physical-level histories are specified sufficiently richly to settle the truth-value of every ordinary indicative sentence expressed in the language of fundamental physics. By an "ordinary indicative sentence", I mean a sentence that uses no modal operators or conditionals and that is declarative rather than interrogative, expressing an ordinary proposition. At every physical-level history h, such a sentence, say P, has a truth-value: true or false. The *propositional content* of P, denoted $[P]$, will be the set of all those physical-level histories (a subset of Ω_{log}) at which P is true.

Consider a conditional of the form "If P had been the case, then Q would have been the case", where P and Q are ordinary indicative sentences. Following the standard analysis, this sentence is true at history h in Ω_{nom} if and only if Q is true at all nearest histories from h at which P is true. To select those nearest "P-histories", we need to introduce a *selection function* f_P which assigns to each history $h \in \Omega_{\text{nom}}$ some set of histories $f_P(h)$ at which P is true, to be interpreted as the nearest P-histories from h. Then the conditional is true at h if and only if $f_P(h) \subseteq [Q]$.[8]

Our problem is this. Let h be some history in Ω_{nom}, say the actual physical-level history, and let P be the antecedent of our counterfactual, where P isn't true at h. Then the following three constraints on $f_P(h)$ are mutually incompatible if h is deterministic:

Fixity of the past: $f_P(h)$ contains only histories that share some initial segment with h, i.e., for every $h' \in f_P(h)$, $h'_t = h_t$ for some $t \in T$.

Fixity of the laws: $f_P(h)$ contains only histories that are nomologically possible, i.e., $f_P(h) \subseteq \Omega_{\text{nom}}$.

Non-triviality: $f_P(h)$ is non-empty.

Thus, we are faced with the trilemma discussed earlier. We can also characterize the standard responses. Denying fixity of the past means allowing $f_P(h)$ to contain histories that share no initial segment with h. That's the "backtracking" response. Denying fixity of the laws means allowing $f_P(h)$ to contain nomologically impossible histories; $f_P(h)$ will then be a subset of Ω_{log} but no longer of Ω_{nom}. That's the "local miracles" response. Alternatively, one might deny that a set Ω_{nom} of nomologically possible histories can be antecedently specified at all, as in some versions of the Humean response. Finally, denying non-triviality means setting $f_P(h) = \varnothing$, thereby rendering the relevant counterfactuals vacuously true. That's the "trivializing" response.

If the physical-level history h is deterministic, then these are the only ways in which we might go about truth-evaluating a conditional of the sort discussed, consistently with the standard analysis.

[8] As noted, my analysis uses the limit assumption. Without it, the analysis must be modified as indicated in earlier notes.

But now consider a conditional whose antecedent and consequent are expressed not in the language of fundamental physics, but in that of some special science. As I have argued, the appropriate level of description is now a higher level, not the physical one.

Let me formalize the notion of a "higher-level world". This is a fully specified history of the world at the relevant higher level, associated with the special science in question. Higher-level worlds have to be specified as richly as needed to settle the truth-values of all ordinary indicative sentences in the language of the given special science, but no more richly than that.

As before, let time be represented by $T = \{0,1,2,\ldots\}$. Now let $\mathbb{S}$ (in outline font) be the set of all possible *higher-level* states in which the world could be at any particular point in time. Call this a *higher-level state space*. As noted, I assume that higher-level states *supervene* on physical states, but are *multiply realizable* by them. That is, each higher-level state may admit different physical realizers and thus corresponds to an equivalence class of physical-level states. We can capture this by introducing a *supervenience mapping* σ from the physical-level state space S to the higher-level state space $\mathbb{S}$, which assigns to each physical-level state $s \in S$ the higher-level state $\mathbb{s} = \sigma(s) \in \mathbb{S}$ that is realized by it. Multiple realizability means that σ is a many-to-one mapping. Each physical-level history h now gives rise to a resulting higher-level history $\mathbb{h} = \sigma(h)$, which is a function from T into $\mathbb{S}$, defined as follows:

$$\text{for each time } t \in T, \mathbb{h}(t) = \sigma(h(t)).$$

The set of all nomologically possible higher-level histories, denoted $\mathbb{\Omega}_{\text{nom}}$, can simply be defined as the image of Ω_{nom} under the supervenience mapping σ. Formally,

$$\mathbb{\Omega}_{\text{nom}} = \{\mathbb{h} : \mathbb{h} = \sigma(h) \text{ for some } h \in \Omega_{\text{nom}}\}.$$

The set of all logically possible higher-level histories, in turn, is the set of all logically possible functions from T into $\mathbb{S}$ and will be denoted $\mathbb{\Omega}_{\text{log}}$. In the example of Figure 2, the thick dots represent higher-level states in $\mathbb{S}$, and the lines from bottom to top represent nomologically possible higher-level histories.

As illustrated by Figures 1 and 2, determinism in physical-level histories does not imply determinism in higher-level histories. Formally, determinism in h is compatible with indeterminism in $\mathbb{h} = \sigma(h)$. Similarly, indeterminism in h is compatible with determinism in $\mathbb{h} = \sigma(h)$. Here, determinism and indeterminism at the higher level are defined in perfect analogy to the earlier definitions for the physical level: a higher-level history $\mathbb{h} \in \mathbb{\Omega}_{\text{nom}}$ is *deterministic* if its initial segment up to any time $t \in T$ admits only one continuation in $\mathbb{\Omega}_{\text{nom}}$, and *indeterministic* if its initial segment up to some time $t \in T$ admits two or more distinct such continuations. The *initial segment* of $\mathscr{h}$ up to t is also defined as before, as the restriction of the function $\mathscr{h}$ to all points in time up to and including t.

Now consider a special-science counterfactual of the form "If $\mathscr{P}$ had been the case, then $\mathscr{Q}$ would have been the case", where $\mathscr{P}$ and $\mathscr{Q}$ are ordinary indicative sentences expressed in the appropriate higher-level language. Each of $\mathscr{P}$ and $\mathscr{Q}$ expresses a higher-level proposition, namely a subset of Ω_{log}. Specifically, $[\mathscr{P}]$ and $[\mathscr{Q}]$ are the sets of all those higher-level histories at which $\mathscr{P}$ and $\mathscr{Q}$ are true, respectively. To evaluate our conditional, we must introduce a selection function over higher-level histories, denoted $\mathscr{f}_{\mathscr{P}}$, which assigns to each higher-level history $\mathscr{h} \in \Omega_{nom}$ the set $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$ of nearest higher-level histories at which $\mathscr{P}$ is true. Our conditional is true at history $\mathscr{h}$ if and only if $\mathscr{Q}$ is true at every history in $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$, formally $\mathscr{f}_{\mathscr{P}}(\mathscr{h}) \subseteq [\mathscr{Q}]$.

If history $\mathscr{h}$ is indeterministic, then some higher-level antecedents $\mathscr{P}$ can be historically accessible even if they are not true at $\mathscr{h}$ itself. In this case, it becomes possible to satisfy the following three constraints simultaneously:

Fixity of the past: $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$ contains only histories that share some initial segment with $\mathscr{h}$, i.e., for every $\mathscr{h}' \in \mathscr{f}_{\mathscr{P}}(\mathscr{h})$, $\mathscr{h}_t = \mathscr{h}'_t$ for some $t \in T$.

Fixity of the laws: $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$ contains only higher-level histories that are nomologically possible, i.e., $\mathscr{f}_{\mathscr{P}}(\mathscr{h}) \subseteq \Omega_{nom}$.

Non-triviality: $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$ is non-empty.

A concrete definition of $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$ satisfying these constraints would take $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$ to be (or to be a non-empty subset of) the set of all those $\mathscr{P}$-histories $\mathscr{h}' \in \Omega_{nom}$ for which the shared initial segment with $\mathscr{h}$ is maximally long. Formally, the initial segment that $\mathscr{h}'$ shares with $\mathscr{h}$ is *maximally long* if, for any other $\mathscr{P}$-history $\mathscr{h}'' \in \Omega_{nom}$ and any time $t \in T$, if $\mathscr{h}''_t = \mathscr{h}_t$ then $\mathscr{h}'_t = \mathscr{h}_t$. Alternatively, one might take $\mathscr{f}_{\mathscr{P}}(\mathscr{h})$ to be some other set of $\mathscr{P}$-histories $\mathscr{h}' \in \Omega_{nom}$ that share *some* initial segment with $\mathscr{h}$, not insisting on maximal length.

This formalizes the proposal introduced in the last section. In short, we evaluate a counterfactual conditional at history $\mathscr{h}$ by rewinding $\mathscr{h}$ until we reach a fork in the road at which the conditional's antecedent, say $\mathscr{P}$, could become true. This could be either the most recent such fork or instead the most plausible fork, by some standard. We then ask whether the conditional's consequent, $\mathscr{Q}$, is true at all the $\mathscr{P}$-histories that are "reachable" from that fork (or at those histories among them that are deemed "nearest" to $\mathscr{h}$ by some further tie-breaking criterion). If the answer is "yes", the conditional is true at history $\mathscr{h}$. If it is "no", then not.

6. Concluding discussion

In conclusion, I address some further questions about my proposal.

6.1 What about mixed-level conditionals?

On my proposal, any counterfactual conditional is truth-evaluated at the level at which its antecedent and consequent are expressed. This is unambiguous when both the antecedent and the consequent are expressed at the same level of description, for instance using the same special-science language. Things become less clear, however, when a conditional mixes levels of description, for instance by using a special-science language to express the antecedent and the language of fundamental physics to express the consequent.

An example would be "If Kennedy had not been assassinated, the physical microstate of the universe would have been such-and-such", where "such-and-such" stands for some fine-grained microphysical property. If we tried to truth-evaluate this conditional at the higher level, we might be able to identify some nearest higher-level worlds in which the antecedent is true, but these would be too coarse-grained to settle the truth-value of the consequent. So, the truth-value of the conditional would be undefined. By contrast, if we tried to truth-evaluate the conditional at the physical level, then – under physical-level determinism – the conditional's antecedent would cease to be historically accessible and my proposal would no longer apply. Moreover, we would be truth-evaluating the antecedent at a level that is arguably too fine-grained for it. Recall that claims about Kennedy are not best analysed at the level of fundamental physics.

The converse constellation, i.e., a conditional with physical-level antecedent and higher-level consequent, is somewhat less problematic, since it could be truth-evaluated at the physical level without the same risk of ill-definedness. If higher-level facts supervene on physical-level facts, the higher-level consequent could be assigned a truth-value at any physical-level world, via supervenience. However, if physical-level determinism holds, we would still face our original trilemma, due to the historical inaccessibility of any counterfactual antecedent.

On my view, we should avoid such mixing of levels, and strictly speaking, mixed-level conditionals have no well-defined truth-values.[9] Indeed, if languages are built up using proper formation rules, as in formal logic, then no conditionals are ever formed in which the antecedent

---

[9] I have previously argued against conflating levels in my work on free will (e.g., List 2014, 2019a,c). Others who have cautioned against level-mixing include the philosopher Siderits (2008), who argues against the mistake of "illegitimately mix[ing] two distinct vocabularies" in our analysis of free will, and the physicist Carroll (2016), who similarly argues that it would be a mistake, in our explanations of various phenomena, to mix vocabularies from different levels of description.

and the consequent are expressed in different languages, associated with different levels of description.

However, if one still wanted to formulate mixed-level conditionals, I would suggest treating them as if they were expressed at the lower one of the two levels, though with the implication that we would be truth-evaluating either the consequent or the antecedent at a level that is too fine-grained relative to its content. Moreover, we cannot generally assume that, for any higher-level sentence, there exists a corresponding lower-level sentence whose propositional content perfectly matches that of the given higher-level sentence, modulo supervenience (List 2019b, section 4.3). It would therefore be non-ideal to treat the mixed-level conditional as if it were a lower-level one: there might be no well-formed sentence in the lower-level language that expresses it.

An advantage of not assigning any truth-values to mixed-level conditionals is that this allows us to refrain from taking a stand on statements such as "if Kennedy had not been assassinated, the micro-state of the universe at the time of the Big Bang would have been different". Yet, we would still be able to say that if Kennedy had not been assassinated, the prior macro-states of the universe (at the level of description at which we are talking about Kennedy) would have been the same, and no macro-level laws would have been breached either.

6.2 Backtracking through the backdoor?

When we truth-evaluate a special-science counterfactual with antecedent $\mathbb{P}$ at history $\mathbb{h}$, on my proposal, we go back in time as far as needed to reach a (plausible) branching point at which the initial segment of $\mathbb{h}$ could have been continued so as to make $\mathbb{P}$ true. The nearest $\mathbb{P}$-histories that we pick out are then nomologically possible (i.e., in $\mathbb{\Omega}_{\text{nom}}$) and share $\mathbb{h}$'s initial segment up to the relevant time. We can thus retain fixity of the laws and fixity of the past at the appropriate level, without trivializing counterfactuals.

A critic might object that I have introduced backtracking through the backdoor. Consider any higher-level $\mathbb{P}$-history $\mathbb{h}'$ which matches history $\mathbb{h}$ up to time t (i.e., $\mathbb{h}'_t = \mathbb{h}_t$), while diverging thereafter (i.e., $\mathbb{h}' \neq \mathbb{h}$). Assuming that higher-level histories supervene on physical-level ones and that there is determinism at the physical level, it will follow that while $\mathbb{h}$ and $\mathbb{h}'$ coincide up to time t, none of their physical-level realizers will share any initial segment. Take any physical-level history h in the inverse image of $\mathbb{h}$ under the supervenience mapping σ (i.e., any $h \in \Omega_{\text{nom}}$ such that $\sigma(h) = \mathbb{h}$), and take any h' in the inverse image of $\mathbb{h}'$ under σ. Then, given physical-level determinism, the initial segments of h and h' up to time t (indeed, up to *any* time) will be distinct, i.e., $h'_t \neq h_t$.

In other words, if we focus not on the nearest *higher-level* histories at which $\mathbb{P}$ is true, but on the physical-level histories realizing them, then my proposal seems very similar to the backtracking proposal. The "fork in the road" at which the higher-level history $\mathbb{h}'$ branches off from history $\mathbb{h}$ at time t so as to render $\mathbb{P}$ true is not a real fork in the road, the critic might say, but only an artefact of the coarse-graining at the higher level. Once we consider the underlying physical-level histories, we must go back all the way to the beginning of time and set different initial conditions to put the world on a trajectory in which $\mathbb{P}$ comes out as true. Thus, my proposal looks like backtracking through the backdoor.

My response is this. First, my proposal is still structurally very different from the standard backtracking proposal. It does not involve any backtracking at the level that matters, namely the higher level at which we truth-evaluate the counterfactual in question. The nearest $\mathbb{P}$-histories that we are considering are higher-level histories, not physical-level ones, and *qua* higher-level histories they do share an initial segment with history $\mathbb{h}$ up to time t. Moreover, from the higher-level perspective, it is irrelevant what goes on at the subvenient physical level.

Second, if we take the levelled picture of the world seriously, we must not regard a "fork in the road" in history $\mathbb{h}$ as unreal merely because it occurs at the higher level. Just as the distinction between determinism and indeterminism is level-specific, so the notion of "forks in the road" is level-specific too. And just as I have suggested that higher-level indeterminism should be considered a real phenomenon, so higher-level "forks in the road" should be considered real too.

Third, on the backtracking proposal, a statement such as "if Oswald had decided not to pull the trigger, the initial conditions of the universe would have been different" comes out as true, and this might then seem to imply that there can be backwards causation. This implication arises if, as is common in technical theories of causation, counterfactual difference-making is taken to be either constitutive or indicative of causation. By contrast, on my proposal, there is no such implication. The problematic conditionals, which usually involve higher-level antecedents and physical-level consequents, are assigned no truth-values at all, and any conditionals whose antecedents and consequents are expressed at the higher level are assigned the intuitively correct truth-values, without any implications of backwards causation.

Dorr, Esfeld, and Loewer have tried to make the notion of backtracking more palatable by pointing out that, although the initial conditions of the universe would have been different if Oswald had decided not to pull the trigger, the difference would have been so minor as to be macroscopically irrelevant. As Loewer notes in his own discussion of the Oswald case, the relevant alternative possible worlds can "keep the macro state fixed while altering the microstate so as to realize the alternative decision", and so, "these alternatives don't 'back track' macroscopically" (2007, p.

319). In this way, we would have to give up only what one might call *fixity of the micro-past*, while we could retain *fixity of the macro-past*.[10] I agree that the standard backtracking proposal seems less unpalatable once we recognize this point. After all, the proposal would entail backwards causation *at most* with respect to macroscopically irrelevant aspects of the world. Indeed, for Loewer, the counterfactual dependence of the world's initial conditions on present decisions wouldn't qualify as "causation" at all, since we wouldn't be able to use that dependence to control the past.

Still, by separating the micro- and macro-levels more explicitly, my proposal allows us to explain systematically why fixity of the macro-past is the requirement that matters – namely because, at the higher level, that's simply what fixity of the past amounts to – and further why, at the higher level, there is no backwards causation at all.

6.3 What about conditionals with historically inaccessible antecedents?

Even if higher-level histories may permit branching when physical-level histories don't, this does not guarantee that *all* special-science conditionals will have historically accessible antecedents. It is a further question how much branching there is in any given higher-level history, and whether, in the case of a particular conditional, its antecedent could have become true at some point in the actual past. The scenario in which Kennedy wasn't assassinated might count as historically accessible, while the scenario in which extra-terrestrials landed on Earth in 1964 might not.

One is then faced with the question of how to handle special-science conditionals whose antecedents are historically inaccessible. As we have seen, one would have to choose between three responses: deem such conditionals trivially true, postulate local miracles or non-fixed laws, or invoke backtracking. A similar choice would arise for physical-level conditionals in a deterministic physical world. So, which response is the right one?

---

[10] Esfeld (2021) has proposed a particularly striking version of this idea. In characterizing the physical state of the world, he emphasizes the distinction between the *position properties* of all physical particles on the one hand and their *dynamic properties* on the other. He suggests that if Oswald had decided not to pull the trigger, the initial state of the world would have been different only in the dynamic properties but not in the position properties. Moreover, he suggests that if Oswald had decided not to pull the trigger, no position properties of any particles would have been different at all *at any time prior to Oswald's decision*. Only dynamic properties would have been different. Thus, he proposes a backtracking interpretation of the relevant conditionals that allows us to retain the fixity of the past with respect to all position properties, while only giving up the fixity of the past with respect to dynamic properties. The challenge for Esfeld is to show that relaxing fixity of the past with respect to dynamic properties does not also commit one to relaxing fixity of the past with respect to position properties. In particular, he needs to show that there are plausible deterministic physical theories that would allow two physical-level histories of the world to coincide perfectly (not just approximately) in all their position properties up to time *t* while only differing in dynamic properties up to *t*, where the two histories diverge after *t* in both position and dynamic properties. If this challenge could be met, then this would strengthen the claim that backwards causation could be minimal even on a backtracking proposal (by not affecting any position properties).

I don't need to take a stand on this question here, and I can see the merits of each of the two non-trivializing responses. A non-*ad-hoc* version of the "local miracles" response can be found in technical work on interventionism, where the notion of an "intervention" in a causal system has been given a precise meaning (e.g., Pearl 2000, Woodward 2003). In particular, the "submodel" of a causal model induced by an intervention represents the world with the Lewisian "miracle". And so, the "local miracles" response can be spelt out systematically, though metaphysical questions about it remain.

The "backtracking" response has been defended by Dorr, Esfeld, and Loewer, who have argued that it would commit us *at most* to a minor and macroscopically irrelevant form of backwards causation. Furthermore, as discussed above, my own proposal looks similar to those scholars' backtracking proposals when it is viewed (albeit contrary to my preferred interpretation) from a subvenient lower-level perspective. A further consideration that might speak for the backtracking response from the perspective of my proposal is that it is the limiting case when we have to go back all the way to the beginning of time and set different initial conditions to reach a world in which the conditional's antecedent is true. If, in my definitions, we allow the maximally long shared initial segment between a $\mathbb{P}$-world and the actual world to be empty (of length zero), then the backtracking proposal emerges as an instance of my own proposal applied to a historically inaccessible antecedent.

Still, if my analysis is right, then there is only a limited need for traditional backtracking, since far more conditionals have historically accessible antecedents than conventionally thought.

References

Beebee, H. 2000. "The Non-Governing Conception of Laws of Nature," *Philosophy and Phenomenological Research* 61(3): 571–594.

Bennett, J. 2003. *A Philosophical Guide to Conditionals*, Oxford: Oxford University Press.

Braddon-Mitchell, D. 2001. "Lossy Laws," *Noûs* 35(2): 260–277.

Butterfield, J. 2012. "Laws, Causation and Dynamics at Different Levels," *Interface Focus* 2(1): 101–114.

Carroll, S. 2016. *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*, New York: Dutton.

Dorr, C. 2016. "Against Counterfactual Miracles," *Philosophical Review* 125(2): 241–286.

Esfeld, M. 2021. "Super-Humeanism and free will," *Synthese* 198(7): 6245–6258.

Hettema, H. 2012. *Reducing Chemistry to Physics: Limits, Models, Consequences*, PhD thesis, University of Groningen.

Lewis, D. K. 1973. *Counterfactuals*, Cambridge, MA: Harvard University Press.

List, C. 2014. "Free will, Determinism, and the Possibility of Doing Otherwise," *Noûs* 48(1): 156–178.

List, C. 2019a. *Why Free Will is Real*, Cambridge, MA: Harvard University Press.

List, C. 2019b. "Levels: Descriptive, Explanatory, and Ontological," *Noûs* 53(4): 852–883.

List, C. 2019c. "What's Wrong with the Consequence Argument: A Compatibilist Libertarian Response," *Proceedings of the Aristotelian Society* 119(3): 253–274.

List, C., and M. Pivato 2015. "Emergent Chance," *Philosophical Review* 124(1): 119–152.

Loew, C., and S. Jaag 2020. "Humean Laws and (Nested) Counterfactuals," *Philosophical Quarterly* 70(278): 93–113.

Loewer, B. 2007. "Counterfactuals and the Second Law," pp. 293–326 in H. Price and R. Corry, eds., *Causation, Physics, and the Constitution of Reality*, Oxford: Oxford University Press.

Loewer, B. 2020. "The Consequence Argument Meets the Mentaculus," working paper, Rutgers University.

Manafu, A. 2015. "A Novel Approach to Emergence in Chemistry," pp. 39–55 in E. Scerri and L. McIntyre, eds., *Philosophy of Chemistry: Growth of a New Discipline*, Heidelberg: Springer.

Noordhof, P. 2020. *A Variety of Causes*, Oxford: Oxford University Press.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*, New York: Cambridge University Press.

Stalnaker, R. C. 1968. "A Theory of Conditionals," pp. 98–112 in N. Rescher, ed., *Studies in Logical Theory*, Oxford: Blackwell.

Siderits, M. 2008. "Paleo-Compatibilism and Buddhist Reductionism," *Sophia* 47(1): 29–42.

Steward, H. 2021. "What Is Determinism? Why We Should Ditch the Entailment Definition," pp. 17–43 in M. Hausmann and J. Noller, eds., *Free Will: Historical and Analytic Perspectives*, Cham: Palgrave Macmillan.

Tomkow, T., and K. Vihvelin 2017. "The Temporal Asymmetry of Counterfactuals", working paper, <http://philarchive.org/archive/TOMTTA-5>.

Werndl, C. 2009. "Are Deterministic Descriptions and Indeterministic Descriptions Observationally Equivalent?" *Studies in History and Philosophy of Science Part B* 40(3): 232– 242.

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

Woodward, J. 2009. "Agency and Interventionist Theories," pp. 234–262 in C. Hitchcock and P. Menzies, eds., *Oxford Handbook of Causation*, Oxford: Oxford University Press.

Yoshimi, J. 2012. "Supervenience, Dynamical Systems Theory, and Non-reductive Physicalism," *British Journal for the Philosophy of Science* 63(2): 373–398.