# AN EXPLANANS FOR TWO FACTS ABOUT CAUSATION

DOUGLAS KUTACH

ABSTRACT. An event $E_1$ never makes an event $E_2$ more likely by way of a nomic connection that goes from $E_1$ back in time to an event $C$ and then forward in time to $E_2$. Also, one can never exploit a nomic connection from an event $C$ to a desired prior event $E$ as a means of making $E$ more likely. I explain these two facts using facts about fundamental dynamical laws. These facts are relatively uncontroversial because they hold for all paradigm theories of fundamental physics. Also, no assumption of a primitive temporal or causal asymmetry is made. Thus, my explanation relies on less controversial hypotheses than alternative explanations.

## 1. INTRODUCTION

Here are two important facts about causation as we know it:

(1) An event $E_1$ never makes an event $E_2$ more likely by way of a nomic connection that goes from $E_1$ back in time to an event $C$ and then forward in time to $E_2$.
(2) One can never exploit a nomic connection from an event $C$ to a desired prior event $E$ as a means of making $E$ more likely.

Any explanation of them either employs a primitive asymmetry or doesn't. Theories using a primitive asymmetry, e.g., Hume (1739), Kant (1781), Broad (1923), Beauchamp and Rosenberg (1981), Ellis (2006), Maudlin (2007), can do so in various forms including assuming that causation is essentially asymmetric or that causation can only directed towards the future (where the future direction is an objective aspect of reality independent of the material layout of the universe) or that the counterfactual conditionals relevant to causal dependence presuppose the fixity of the past. Any of these primitive asymmetries allows one to provide simple explanations of both facts about causation, albeit at a theoretical cost. A minor drawback is ontological bloat, for one should not postulate primitive asymmetries beyond necessity. A more serious concern is the coordination problem: ensuring that any explanation in terms of the primitive asymmetry coordinate

*Date*: October 13, 2009.

properly with explanations given in terms of the microphysical evolution. A theory where a fundamental direction of causation can flip back and forth without consequence for the motions of particles and fields would undermine the explanatory value of the fundamental asymmetry. Although these problems are not insuperable, I am among the many who prefer a framework eschewing primitive asymmetries, e.g., Gasking (1955), Reichenbach (1956), von Wright (1971), Papineau (1985), Mackie (1973), Lewis (1979), Price (1992), Mellor (1995), Hausman (1998), Albert (2000), Dowe (2000), Kutach (2007), Loewer (2007).

My goal here is to provide a unified explanation of these two facts about causation without postulating a primitive asymmetry. The virtue of my explanation is that its explanans is significantly less controversial than those of other attempted explanations. Its premisses are hypothesized properties of the laws of fundamental physics and are not terribly controversial because they hold in all paradigm theories of fundamental physics. So long as the actual laws share these properties of paradigm laws, the two facts are provably true. The most controversial assumption is that influence and causation exist solely in virtue of such fundamental laws without there being additional special science laws that contravene the fundamental laws by permitting non-trivial zig-zagging of nomic connections. My explanation does not require any particular account of the matter asymmetries, e.g., no assumption of a fork asymmetry or entropy asymmetry or anything complicated like that, though such accounts can supplement my explanation of the second fact. The end result is a boon to anyone who wants to derive the direction of causation merely from fundamental laws plus some sort of asymmetry in the historical layout of the universe's contents because it automatically explains why it is useful for people to conceive of causation as essentially asymmetric even if the fundamental laws turn out not to possess the right kind of temporal asymmetry to ground a fundamental direction of causation.

With the leaner metaphysics that postulates only fundamental physics and refuses to cook the books in favor of unidirectional causation, it is prima facie mysterious why nomic connections cannot violate the two facts. This is clearest when the laws are deterministic in both temporal directions. Suppose event $C$ determines that $E_1$ and $E_2$ occur in the future, and that $E_1$ could fail to occur only in the absence of $C$, and that $E_2$ can occur only if $C$ occurs. By determinism, $E_1$ implies $C$, which in turn implies $E_2$. By the other conditions, a non-occurrence of $E_1$ implies a non-occurrence of $C$, which in turn implies a non-occurrence of $E_2$. Thus $E_1$ makes a difference as to whether $E_2$. A similar worry arises for past-directed causation. We take for granted our ability to

causally influence in at least some respects the state of the world in the near term future. But whatever differences we make on the near term future automatically determine differences in the distant past as well, and it may not be immediately clear what could forbid such influence from being exploitable for achieving desired effects.

## 2. The Explanans

In explaining the two facts, I assume causation holds in virtue of some fundamental structure of reality that resembles near enough the most famous theories proposed in physics. I only assume principles common to the following paradigm fundamental theories: Newtonian mechanics with gravitation, relativistic electromagnetism, general relativity, and non-relativistic quantum mechanics. These paradigm theories all describe reality using spacetime structure with material contents like particles and fields. In the case of quantum mechanics, a configuration space contains the universal wave packet in a way that resembles material contents in spacetime, and so can be subsumed under a more general structure that includes spacetime plus the configuration space. Let us say that the fundamental space where all fundamental stuff exists is the arena. The arena for non-quantum theories is just spacetime. The arena for quantum theories is spacetime plus configuration space.

The material contents inhabiting the arena can be described using the language of events. There are three kinds of events. Lowercase variables represent fine-grained events, i.e., maximally fine-grained specifications of fundamental stuff in some determinate region of the arena. Uppercase variables represent coarse-grained events, which are formally just a set of maximally fine-grained events occupying the same location. Coarse-grained events are really event types but we can speak of them as occurring because they are instantiated if and only if one of their realizers, i.e., elements, occurs. Uppercase variables with a line on top represent a contextualized event, which is simply a coarse-grained event with a probability measure over all its elements. Contextualized events allow us to represent coarse-grained events in a way where not all of their realizers are treated as equally likely, and it allows us to make sense of how a coarse-grained event can fix a unique probability for some other event even if its realizers, $c_i$, fix different probabilities for the same event $E$.

Every paradigm theory of physics postulates laws of temporal evolution that constrain which events at $t$ are compatible with which events at $t'$. Each theory's paradigm fundamental laws of temporal evolution provide rules whereby a large enough completely specified fine-grained

event at one time entails a probability distribution over events at other times. Whenever an event $c$ fixes (by way of the fundamental laws) a specific probability distribution that includes an event $e$, we say that $c$ termines $e$ and that $c$ is an terminant of $e$. If $c$ determines that $e$ occurs, that counts as a special case of termining. This definition easily extends to coarse-grained events because we can say that $c$ termines $E$ whenever it fixes a probability for $E$, again assuming only the fundamental laws. We cannot in general say that $C$ termines $E$ because different realizers of $C$ might fix different probabilities for $E$, but it is a well-formed claim that $\overline{C}$ termines $E$ because the probability that $\overline{C}$ fixes for $E$ is just the combined probability that each of $\overline{C}$'s realizers fix for $E$ weighted by the probability distribution that is part of the definition of $\overline{C}$.

The paradigm theories all obey the following four principles, which are the crucial premisses constituting the explanans. The principles might seem overly technical and thus opaque, but that is because I have tried to weaken them as much as possible. The basic idea behind them is very simple. In all the paradigm theories, the dynamical laws propagate states continuously through time without any rules that permit causal contribution to skip over intervening states. A preliminary definition helps to simplify the principles: An event $c$ is c-connected to an event $e$ iff every point of $e$ can be reached from some point of $c$ by way of a smooth path whose tangent is nowhere space-like.

- Transitivity: Any terminant of $e$ termines whatever $e$ termines. Transitivity does not imply, when $e_1$ termines $e_2$ which in turn termines $e_3$, that the probability $e_1$ fixes for $e_3$ equals the probability that $e_2$ fixes for $e_3$.
- Density: For any terminant $c$ of an event $e$, if there exists a location $x$, e.g., a subset of the arena, such that (1) every point of $x$ is c-connected between some point of $e$ and some point of $c$, and (2) no point of $c$ can be c-connected to $e$ without intersecting $x$, then there exists an event $i$ fully occupying $x$ such that $c$ termines $i$ and $i$ termines $e$. Density merely generalizes the claim that physical states evolve continuously through the arena without ever having any influence that hops over intervening states.
- Non-Spatiality: All terminants of $e$ are c-connected to $e$. Non-spatiality just expresses that events do not fix probabilities for what occurs elsewhere "at the same time." This principle is not violated by the non-locality in Newtonian gravitation or quantum mechanics (once the states are properly interpreted).

- Screening: The probability that a terminant $c$ fixes for $e$ is unaltered by augmenting $c$ with events that cannot be c-connected to $e$ without passing through $c$. Screening just generalizes the idea that any state of the world at a single time incorporates all the relevant information from its past (future) when it sets probabilities for what happens in its future (past).

## 3. Probability-Raising

Evaluating the consequences of these four principles requires making precise what it means for one event to make another more likely by way of a nomic connection. Because the ultimate purpose of the explanation is to help derive facts about causation using only fundamental laws and facts, it makes sense to identify the degree to which some event makes $E$ likely with the probability it fixes for $E$ using the fundamental laws alone, ignoring any accidental regularities and special science laws. (One could invoke other notions of law and probability, but my argument is not intended to apply to them necessarily.)

An absolutely crucial condition for applying the paradigm fundamental laws is this: None of the paradigm fundamental laws fix probabilities among the ordinary-sized events we usually care about. For example, we normally think that throwing an ordinary rock over the surface of a large calm liquid lake raises the probability that the lake will exhibit ripples ten seconds later. But the fundamental laws say nothing about the impact that the rock alone has on what happens ten seconds later. In order for the laws to be at all informative, they require the full specification of all the microscopic details of some vast event that includes the rock plus its larger environment. If the laws are relativistically local like the theory of relativistic electromagnetism, the event needs to span at least ten light-seconds in radius. If the laws are non-relativistic in the way that Newtonian gravitation or non-relativistic quantum mechanics is, the event needs to encompass the entire breadth of the universe. The upshot for understanding probability-raising is that a human-scale event $C$ cannot by itself raise the probability of anything beyond a few picoseconds later. So, to use the fundamental laws for $C$ at all, we need to construe the probability that $C$ fixes for $E$ as the probability that $\overline{C}$ fixes for $E$, where $\overline{C}$ is defined to be $C$ plus enough of $C$'s environment to constitute a terminant, plus a probability measure over such states so that $\overline{C}$ will fix a unique value for the probability of $E$.

We can interpret how much the localized event $C$ raises the probability of $E$ in the following way. First, we postulate a hypothetical state, $\overline{C}$, which is some chosen contextualized event that instantiates

$C$. Second, we postulate a hypothetical contrast state, $\overline{\neg C}$, which is some chosen contextualized event that instantiates a non-occurrence of $C$, typically a contextualized event that is exactly like $\overline{C}$ except that where $C$ occurs, we replace the physics with something that instantiates a non-occurrence of $C$ in whatever way is contextually appropriate. Third, we define the amount $C$ raises the probability of $E$ as the probability that the fundamental laws fix for $E$ using $\overline{C}$ minus the probability they fix for $E$ using $\overline{\neg C}$. The degree to which $C$ raises the probability of $E$ can then be symbolized as $p_{\overline{C}}(E) - p_{\overline{\neg C}}(E)$. For example, in assessing how much the throwing of the rock raises the probability of ripples, we (1) define $\overline{C}$ to be some coarse-graining of some actual state except that we instantiate a throw of a rock towards the lake, (2) define $\overline{\neg C}$ to be the exact same coarse-graining of the actual state except that we instantiate a person not throwing a rock, and (3) we infer using the fundamental laws that the $p_{\overline{C}}(E)$ is very nearly equal to one and that $p_{\overline{\neg C}}(E)$ is very nearly zero, thus verifying that the rock raises the probability of ripples.

Notice that any arbitrariness in how one characterizes $\overline{C}$ and $\overline{\neg C}$ and $E$ will affect the significance of the resulting value, $p_{\overline{C}}(E) - p_{\overline{\neg C}}(E)$, for any conclusions one wants to draw about influence. For example, suppose you are goofing off at time $t$ and on the other side of the planet at $t+1$, Queen Beatrix is lounging in peace in a low risk environment. When your boss catches you loafing, you insist that had you not been goofing off, the queen would have sustained a serious injury. How is that? In the actual world, assuming determinism just for the sake of simplifying the argument, your goofing off was an essential part of a complete physical state that fixed the probability of the queen's being injured at zero. If you had been not goofing off in the very particular way $\overline{\neg C}$—a state very carefully tailored to lead deterministically to the queen's being injured at $t+1$—then the objective probability of her being injured would have been one. Because the queen's lack of injury depended crucially on your having instantiated $\overline{C}$ rather than $\overline{\neg C}$, your boss should be celebrating your having saved the queen. Or, so you argue.

I think the best way to adjudicate such conflicting claims of probability raising (and associated claims of influence) is to permit you to pick whatever contrast state you want and to permit your boss to take into account that you did not pick a $\overline{\neg C}$ that was in any way natural but instead used the fact that you wanted to receive credit for saving the queen as the basis for your idiosyncratic choice of $\overline{\neg C}$. Genuine influence is what is left after accounting for any bias in one's choice of $\overline{C}$ and

$\overline{\neg C}$. This methodology parallels the conversational practice of granting the proponent of some argument leeway to define terminology however she likes but then holding her accountable for any resulting equivocation, circularity, or inapplicability to the intended subject matter. Clearly our interest in using the probabilities to inform us about causal influence requires that the parameters are not prejudiced in a way that presupposes what follows dynamically from what. However, rejecting bogus claims of influence does not require rules codifying what counts as circular or uninformative. We can rely on our instinctive ability to sniff out such cases just as we do when evaluating arguments.

## 4. The Unexploitability of Common Cause Linkages

In Fig. 1, three events are located in ways that exemplify a paradigmatic common cause, a situation where one event $C$ makes two distinct later events, $E_1$ and $E_2$, more likely than they would be in the absence of $C$. For concreteness, one can imagine that $C$ is the activation of a school's fire alarm, $E_1$ is some particular student's being frightened, and $E_2$ is a fire truck arriving at the school. The intuitive claim to be explained is that the student's being frightened does not make the appearance of a fire truck more likely in virtue of a connection between the alarm and the fright. This holds even if the fundamental laws ensure that the student's fright makes the sounding of the alarm more likely than it would have been if the student had been bored or sleepy. Of course, everyone agrees that the student's fright can raise the chance of a fire truck arriving in virtue of a direct nomic connection. For example, the fright might make the student more likely to phone the fire department to assure them that the alarm was not accidentally activated. I will assume $E_1$ temporally precedes $E_2$. By the end of the next section, it will be obvious what to say about cases where $E_2$ comes first or at the same time. (Fig. 1 depicts the terminants as if relativistic locality holds, but the argument nowhere depends on this. In non-relativistic theories, the terminants extend infinitely far out into space.)

The proof against the exploitability of the common cause connection is this: In order for $E_1$ to bear nomically on the probability of $E_2$ by way of $C$, it must do so in virtue of $E_1$'s being a part of some larger $\overline{E_1}$ that counts as a terminant of $\overline{C}$ which in turn is a terminant of $E_2$. If such terminants do not exist, there is no link between $E_1$ and $E_2$ to be exploited, so we only need to consider the case where they do exist. Let $\overline{N}$ be defined as the intersection of all the points c-connected to $E_2$ with some maximal space-like extension of $\overline{E_1}$. By density, $\overline{C}$ termines $\overline{N}$

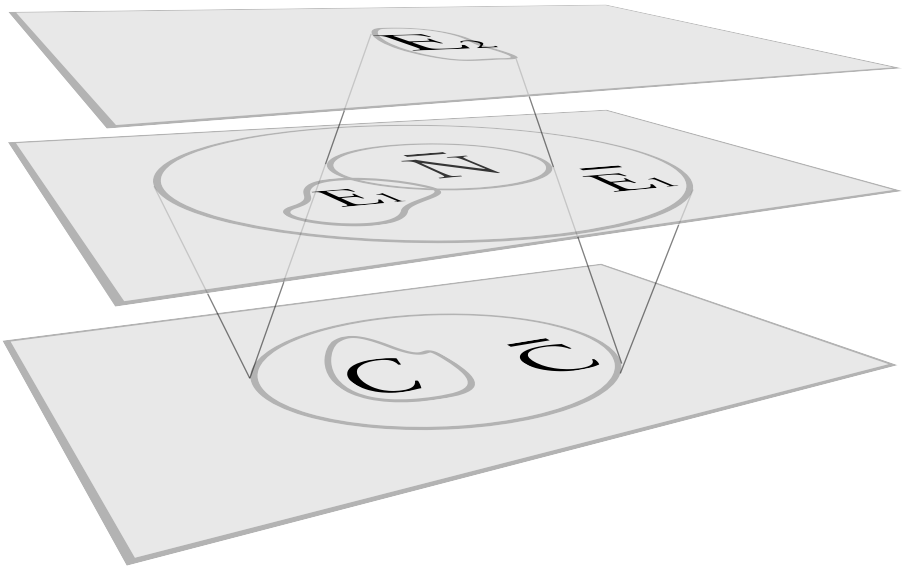


FIGURE 1. $\overline{E_1}$ screens $C$'s influence on $E_2$.

and $\overline{N}$ termines $E_2$. Due to screening, any probability that $\overline{C}$ sets for $E_2$ is screened by $\overline{N}$. Furthermore, $\overline{N}$ must be a subset of $\overline{E_1}$ because by transitivity $\overline{E_1}$ termines $\overline{N}$ and if $\overline{N}$ occupied any spacetime outside of $\overline{E_1}$, that would violate non-spatiality. Thus, any influence going by way of $C$ is screened by $\overline{E_1}$. Thus, the backtracking connection to $C$ adds nothing to the nomic connection going from $\overline{E_1}$ directly towards $E_2$. Q.E.D.

Because no reference was made to any definition of common cause, this explanation cannot be convicted of vacuously building the lack of effectiveness into the definition of 'common cause.' Common causes are inessential to the argument and are mentioned only because they superficially appear to allow non-trivial backtracking nomic connections.

Although that suffices for a complete explanation, it is pedagogically helpful to diagnose what is wrong with backtracking reasoning. Nothing is wrong with backtracking reasoning as a rule for handling hypothetical information. We can use counterfactual conditionals to express dependencies between $E_1$ and $E_2$ that arise because of reliable rules of thumb about how effects like $E_1$ probably arise from a previous $C$. This vindicates our saying (in simple cases at least), "If I had gotten evidence of $E_1$, $E_2$ would probably occur (given the fundamental laws and my evidence), and if I had gotten evidence that $E_1$ did not occur, $E_2$ would probably not occur (given the fundamental laws and my evidence)."

But regardless of the plausibility of such counterfactual claims, there are two reasons to reject their significance for establishing the existence of interesting causal backtracking. First, the rules of thumb only approximate the true fundamental laws of temporal evolution, and the paradigm fundamental laws entail absolutely nothing about what follows from states that are not fully filled out in microscopic detail. The only way the fundamental laws of temporal evolution can entail anything whatsoever about how $C$ failed to occur is to replace $\overline{E_1}$ with some fully specified instantiation of $\neg E_1$. But then, the backtracking reasoning is obviously redundant with regards to what (counterfactually) happens at that time. Second, it is important to remember that because there are no restrictions on the contrast state other than compatibility with the fundamental laws (and even that can be relaxed), one is always free to choose a contrast state constructed using rule of thumb inferences in backtracking fashion. But then, just like taking credit for the queen's lack of injury, such inferences presuppose the backtracking nomic connection and thus cannot count as evidence that one effect of a common cause genuinely raises the likelihood of the other.

## 5. The Unexploitability of Backwards Influence

The other fact to be explained is that one can never exploit a nomic connection from an event $C$ to a desired prior event $E$ as a means of making $E$ more likely. Actions that exploit $C$ do so through a process that is temporally asymmetric. One temporal end of an action is $M$, an event that instantiates something roughly describable as 'mulling over whether to do $C$' or 'waiting to do $C$' or some pre-decision condition like that. In our neighborhood of the universe, $M$ always evolves in the direction of time we call the future into a state instantiating a decision or an activation of a volition or something like that, which then evolves into a state instantiating $C$, which can then make other events more or less likely. The temporal orientation of actions does not need to be defined relative to a fundamental asymmetry. Because all actions in our part of the universe are aligned in the same temporal direction and because there are plenty of other reliable de facto asymmetries, we can interpret the explanandum as the claim that we cannot make desired events more likely when they are towards the past as defined by various de facto asymmetries. When $E$ is prior to $M$, the nomic connection—in order to be legitimately attributable to the action of intentionally bringing about $C$—must first go forward from $M$ through the various stages of the action until it gets to $C$ and only then backwards to reach

$E$. Thus, it is a backtracking nomic connection, which was previously proven to be superfluous. Because nothing in the previous argument depended on the direction of causation or of time or of any de facto asymmetries, that argument alone ensures that no actions to do $C$ can be effective in bringing about $E$ in virtue of the nomic connection going through $C$.

It might turn out, for all we know, that $M$ directly influences whether $E$ happens. Indeed, under all the paradigm theories, there exists at least some kind of nomic link from $C$ back to $M$ and then back to $E$. But even if the link were of a kind that permitted $M$ to make $E$ more likely, its extremely poor fit with what we regard as legitimate instances of 'exploiting $C$ as a means to make $E$ more likely' makes it not count as the kind of process needed to falsify the intended interpretation of the explanandum.

The only non-law fact that entered into the explanation in this section was that processes we describe as 'actions directed towards a goal' are all temporally aligned towards the future as defined by various de facto asymmetries. So, my explanation can be supplemented by a theory identifying the physical structures that best account for how actions (in our neighborhood of the universe at least) align temporally with each other and with the many de facto asymmetries. A virtue of my account is that it insulates the explanation of the unexploitability of the past from the details of what precisely accounts for the temporal alignment of actions.

## 6. Conclusion

The hypothesis that there are some true fundamental laws of temporal evolution exhibiting transitivity, density, non-spatiality, and screening, together with my background interpretation of what it means for one event to make another more likely by way of the fundamental laws, establishes both facts about causation without using any primitive asymmetry. The explanation also makes sense of why, in discussions of backwards causation, one often employs either a dynamical connection that by hypothesis skips back in time without proceeding by way of the intervening physical states. These take advantage of the falsity or inapplicability of the screening condition to permit dynamical connections directed towards the past that count not only as influence but influence that at least has a chance of being exploitable for advancing goals.

## References


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

[2] Beauchamp, T. and Rosenberg, R. (1981) *Hume and the Problem of Causation*. Oxford: Oxford University Press.

[3] Broad, C. D. (1923) *Scientific Thought*. New York: Harcourt, Brace and Co.

[4] Dowe, P. (2000) *Physical Causation*. Cambridge: Cambridge University Press.

[5] Ellis, G. (2006). "Physics in the Real Universe: Time and Spacetime," *General Relativity and Gravitation*. **38** (12), 1797–1824.

[6] Gasking, D. (1955) "Causation and Recipes," *Mind* 64, 479–487.

[7] Hausman, D. (1998) *Causal Asymmetries*, Cambridge: Cambridge University Press.

[8] Hume, D. (1739) *A Treatise of Human Nature*, London.

[9] Kutach, D. (2007) 'The Physical Foundations of Causation,' in Price, H. and Corry, R. (eds.), *Causation, Physics, and the Constitution of Reality*. Oxford: Oxford University Press.

[10] Lewis, D. (1979) 'Counterfactual Dependence and Time's Arrow,' *Noûs* **13**, 455–76, reprinted in *Philosophical Papers, Volume 2*, Oxford: Oxford University Press, 1986.

[11] Loewer, B. (2007) 'Counterfactuals and the Second Law,' in Price, H. and Corry, R. (eds.), *Causation, Physics, and the Constitution of Reality*. Oxford: Oxford University Press.

[12] Kant, I. (1781) *Critique of Pure Reason*.

[13] J. L. Mackie, (1973) *The Cement of the Universe*, Oxford: Oxford University Press.

[14] Maudlin, T. (2007) 'On the Passing of Time,' in *The Metaphysics in Physics*.

[15] Mellor, D. H. (1995) *The Facts of Causation*, New York: Routledge.

[16] Price, H. (1992) 'The Direction of Causation: Ramsey's Ultimate Contingency,' *PSA 1992, Volume 2*. East Lansing, MI: Philosophy of Science Association, pp. 253–67.

[17] Reichenbach, H. (1956) *The Direction of Time*, Berkeley: University of California Press.

[18] von Wright, G. (1971) *Explanation and Understanding*. Ithica, New York: Cornell University Press.

[19] von Wright, G. (1975) *Causality and Determinism*. New York: Columbia University Press.