Dependencies Across Spacetime:
The Geometry of Chance and Causation

by

Jens Jaeger

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Philosophy

New York University

August, 2025

Cian Dorr

For all farm animals.

Acknowledgements

I used to think one's intellectual development would slow after the early twenties. Looking back at my years at NYU shows me how very wrong I was.

Cian Dorr has shaped my intellectual development like nobody else. His unmatched sharpness and intellectual openness make him not only a great philosopher but an equally great advisor. A two-hour meeting with Cian is one of the philosophically most productive experiences you can have. Everything I'll ever write will bear his influence.

Throughout my time at NYU, Michael Strevens was an unending source of invaluable philosophical guidance. His detailed feedback at once helped me solidify my arguments and never lose sight of the bigger picture. I am thankful for all he has taught me over the years. David Chalmers also had a big impact on my trajectory; my first ever publication was written under his expert guidance. A true generalist, he always saw connections between my work and other areas of philosophy which I did not. Having Dave just down the hall for all these years was a true blessing. I vividly remember meeting Dmitri Gallow on a sunny afternoon in Melbourne two years ago. After two years of mostly solitary work on my dissertation, the meeting gave me renewed conviction that my work mattered. I'm incredibly fortunate that he agreed to serve as an external committee member. His generous and extremely incisive feedback led to significant improvements.

During my time at NYU I have learned lots from Tim Maudlin, about physics, philosophy, and teaching. His clarity of explanation is the standard I shall always strive for. I'm deeply grateful for his introductions to people in the foundations of physics community around New York, and for his and Vishnya's unending hospitality and friendship. Brad Weslake has always been extremely kind with his time. His competent counsel on several pieces of my work, including chapter 4, proved vital for improving them. Like many others in the discipline, I have also been fortunate to get to know Al Hájek. His suppport greatly contributed to my feeling welcome in the philosophical community at large.

I was blessed with a special NYU cohort: Bar Luzon, Cristina Ballarini, Evan Behrle, Jorge Ferreira, Ariel Melamedoff, and Justin Zacek. From an impromptu first-year metaphysics reading group to countless hangouts, my cohort was a constant source of intellectual and social support. Strong personalities met an equal willingness to compromise. My friendships with Richard Roth, whose generosity matches his smarts, Bar Luzon, with her rare blend of high agency and social instinct, and Chris Sun, whose physics expertise is only matched by his compassion, have been invaluable. My time at NYU was further enhanced by knowing the wonderful Kevin Coffey, Sophie Côte, Jane Friedman, Hartry Field and Julia Wrigley, Rob Hopkins, Anja Jauernig, Matt Mandelkern, Lara Maszynski, Jessica Moss, Aidan Penn, Ali Rezaei, Ethan Russo, Lauren Somers, Sharon Street, Trevor Teitel, and Patrick Wu. Thank you all.

Outside of philosophy, I am lucky to count several individuals my life-long friends: Alessandra Velez, Dominik Peters, Kolja Sand, and Marco Mühlenjost. There is no greater blessing than knowing that I can always count on you. A few non-human friends—Ceci, Goose, Milo, and Zigo—reminded me that life exists outside footnotes.

I am fortunate to have the best nuclear family in the world: my parents Gitti and Peter, who did everything right, my brother Lutz, the best nocturnal fraction tutor in the world, and his fiancée Rebecca, who makes him very happy. Finally, for her affection, patience, and support, I owe more than I can express here to Susanna Romanini.

I really couldn't have asked for a more stimulating sequel to my early twenties. Stuck on a time loop, I'd be happy to relive this period over and over again.

Abstract

Philosophical investigations into the nature of chance and causation have typically assumed classical spacetime backgrounds. My dissertation argues that this practice misses important lessons. By tying chance functions to a global time parameter—a structure of classical spacetimes—orthodox theories of chance cannot state important Markov-style principles and fail outright in spacetimes that lack a linear time structure, e.g. due to closed time-like curves. By contrast, the *urchance* approach, developed in the dissertation's first half, assumes no such parameter, works in any spacetime geometry, captures the needed independence principles, and resolves the puzzle of chances on closed time-like curves. Two general morals follow: chance isn't as intimately tied to time or causation as is usually thought, and chances can vary even across intrinsically duplicate trials.

On the causation side, counterfactualist reductions are unable to deal with *synchronic* nomic constraints—law-based, yet non-causal, links between simultaneous events. These constraints occur not only in temporal-loop geometries but also in common-or-garden theories like Maxwellian electrodynamics. To save the counterfactualist approach to causation, its defenders must separate genuinely dynamical counterfactual influences from those produced solely by synchronic constraints.

Contents

List of Figures

List of Tables

List of Appendices

Introduction

Chance and causation need a spacetime upgrade. By tying chance functions to times, orthodox accounts of chance are unable to capture genuinely *spatiotemporal* patterns of nomic dependence, even in ordinary spacetimes. In worlds with spacetime loops, orthodox accounts malfunction altogether. Counterfactual analyses of causation suffer similar defects: they treat all law-supported subjunctive conditionals alike, unable to distinguish those that reflect genuine causal influence from those that arise from non-causal nomic constraints. This dissertation addresses both problems. Part I (Chapters 1–3) develops an account of chance that's sensitive to spatiotemporal structure and applies it to the problem of spacetime loops. Part II (Chapters 4–5) addresses the problem of causation under synchronic, non-causal nomic constraints.

Part I: Chance

Part I develops an *urchance* formalism. It assigns to each possible world a single mathematical object, an *urchance function*, that fully encodes the content of the laws at that world. The framework applies equally to deterministic and indeterministic theories and it requires no temporal machinery beyond a well-defined spacetime background—in particular, times needn't be strictly ordered. Since urchance functions can be conditioned on the physical state of *any* spacetime region, not just times, they can capture nomic relationships between states of arbitrary regions.

An example of such a relationship is the *Parental Markov* condition: a region's immediate past screens it off from any other past or present event. As Part I will make clear, Parental Markov generally only holds for local laws in *acyclic* spacetimes. To cover the general case, I propose *Strong Boundary Markov*: take any spacetime region R and wrap it in a thin buffer zone—a *thick boundary*; the state of this boundary alone then screens off everything inside R from everything outside. Strong Boundary Markov is naturally formulable in the urchance framework and it is indifferent to whether spacetime is acyclic or riddled with loops: provided the laws are local, it holds everywhere, even at worlds where Parental Markov fails.

Chapter 1 ("Loops and the Geometry of Chance") uses the condition to develop an account of chances on loops. It shows how Strong Boundary Markov reduces problems of cyclic chances to acyclic problems, with the latter handled by a further principle expressible within the urchance framework. The resulting chance assignments in cyclic worlds are principled, fully classical, and under certain conditions approach the familiar acyclic values for long-duration loops.

Chapter 2 ("Urchance Ideology") asks what urchance functions have to look like to meet their stated goal, of capturing the entire content of a set of laws. To encode not only all of a set of laws' precise chance predictions, but also its nomic dependence relations in the absence of precise chances, an urchance function must consist of a *set* of primitively conditional probability functions. (It's still convenient to model it as a function, namely one which sends each ordered pair of propositions (A, B) to a projection function, evaluating each member of said set at the pair. This also allows us to stick to the phrase "urchance *function*".) Primitive conditionality allows conditioning on probability-0 bits of history, ubiquitous in any continuous physics, and *sets* of probability functions can capture our Markov conditions in the absence of precise chances. The chapter argues that rival approaches and weaker imprecision formalisms fail to meet one or another of these requirements.

Chapter 3 ("Classical Probability vs. Chance Invariance") turns to the literature's main rival to Chapter 1's position. The rival jettisons classical probability and allows outright contradictions to have positive chance in worlds with loops. This, I argue, generates a raft of problems. Among other things, counterlogical chances would spread to ordinary time-travel-free spacetimes—a disastrous result—and the proposal offers no workable replacement for standard probability theory. The solution is to keep classical probability theory intact, while eschewing the link between chances and times which tempted my predecessors into simplistic chance-invariance principles and, ultimately, into the non-classical view.

Part II: Causation

While every law of nature supports counterfactuals, not every law supports *causal* counterfactuals. *Synchronic laws*—so called because they relate events that are simultaneous or, in relativistic terms, not causally separated—fail the causal test. Synchronic laws have two hallmark features: (1) they relate simultaneous (or not causally separated) events and (2) each event is already fully explained by its local dynamical past (if any). This suggests that these laws are non-causal: if the synchronic connections they induce *were* causal, they would (by (2)) be mere overdetermination and yet (by (1)) be particularly theoretically costly, since non-local.

Chapter 4 ("Counterfactual Dependence Is Not Sufficient for Causation") provides two examples of synchronic laws: Gauss's Law in Maxwellian electrodynamics,[1] and consistency constraints along spacetime loops (closed causal curves). I then argue that these laws undermine every extant counterfactual analysis of causation. Counterfactual analyses reduce causal claims to claims about subjunctive conditionals evaluated in the

[1]A law linking a physical entity to the *spatial divergence* of a distinct field.

“standard” context.[2] Yet, as I argue in Chapter 4, the counterfactual subjunctives supported by synchronic laws are *true* in standard contexts, but also non-causal. As a result, Lewis’s original theory, its immediate successors, and modern structural-equation models all misfire when synchronic laws are in play.

Any adequate theory must be able to exclude counterfactual dependencies partially underwritten by synchronic laws, in favor of those purely underwritten by *diachronic* (or dynamical) laws. In **Chapter 5** (“Causation and Diachronic Counterfactuals”), I sketch what the first steps toward such a theory might look like.

Guide for the Reader

While the two parts of the dissertation are self-contained, the chapters *within* each part build on each other. Readers mainly interested in chance can follow the progression from application (Chapter 1) through formal foundations (Chapter 2) to critique of alternatives (Chapter 3). Those concerned with causation may start directly with Chapter 4’s challenge to counterfactualist analyses and then move to the outlook in Chapter 5.

[2] The context is also sometimes labeled “non-backtracking”. The label is misleading: on Lewis’s (1979) own account of the standard context, counterfactual worlds involve a “ramping” period (Bennett’s (2003) words). As a result, for example, “if I now raised my hand, the immediate past would be different than it actually is” resolves as true whenever I don’t raise my hand.

Part I

Chance

Chapter 1

Loops and the Geometry of Chance[1]

1.1 Introduction

You're a star pharmacist, and you've invented a universal antidote, able to cure any poisoning. Unfortunately, the antidote isn't perfectly reliable: normally, given any poisoning, there's a 50% chance that it'll cure it. One day, your evil sibling travels back in time, intending to lethally poison your grandfather, back when he was still an infant. Determined to save grandpa, you grab two antidotes and follow your sibling into the wormhole. ("Better to bring more than one!", you think.) Upon finding infant grandpa, poisoned, you administer the first antidote. Alas, it doesn't work. The second antidote is your last hope. You administer it—and success: the paleness vanishes from grandpa's face, he is cured.

As you administered the first antidote, what was the chance that it would be effective? Perhaps 0? After all, it already failed: its failure is what causes the second antidote's success, which causes grandpa's survival, which causes your being born... On the other hand, the antidote's failure is also still future—some time will pass until it occurs—and the present leaves it open which of the two antidotes is ultimately effective. So perhaps the chance is 0.5, because that's what it normally is? No. I'll argue that, on a salient

[1] A version of this chapter has been published: Jäger (forthcoming).

interpretation of "as you administered the first antidote", the answer is 2/3.

The essay's broader question is this: *Given* a time travel structure, what are the chances? This is distinct from asking what the chances are of wormholes and other time-travel structures arising in the first place—a question reserved for physicists studying the mechanisms behind such structures.

Why care about *our* question? Where time travel involves *spacetime loops*—informally, trajectories which travel back in time to their origins—it has been of continued interest to physicists.[2] Part of a philosopher of science's responsibility is to interpret what physicists study.

But more importantly, our inquiry yields significant philosophical insights. Previously, philosophical investigations into chance have largely assumed standard spacetime backgrounds.[3] As we'll see, this practice misses important lessons. Assuming that spacetime loops are metaphysically possible, our account challenges two orthodoxies about chance. First, it'll show that chances aren't as intimately tied to time or causation as is usually thought; and second, that chances aren't necessarily constant across intrinsically duplicate trials. I'll replace these orthodoxies with a view on which chances are tied, not to temporal histories, but to *chance setups*, and on which chances on loops differ from the ordinary chances in systematic, scrutable ways. Our final account provides a complete theory for chances on loops, for any setup.[4]

The account is founded on two principles about the relationship between chance and spacetime structure. One key principle concerns the familiar idea of "screening off". In assessing the probability of some proposition C, we say that A screens off B from C if A renders any information provided by C about B irrelevant. For example, suppose I'd like

[2] E.g., Gödel (1949); Carter (1968); Echeverria *et al.* (1991a); Earman (1995).

[3] *Viz.*, globally hyperbolic Lorentzian backgrounds, or classical Newtonian or Galilean backgrounds.

[4] Provided the background geometry is static, as e.g. in special relativity. A generalization to theories with dynamical spacetime structure (such as general relativity) is a topic for future work. Still, I hope the present proposal marks significant progress in that direction.

to know if I carry a certain genetic marker. Given complete information about my *parents'* genes (A), no information about my *grandparents'* genes (C) should affect my confidence of my carrying the gene (B): complete information about A *screens off* any information C would provide about B. With respect to *chance*, what screens off what is partially determined by spacetime structure. Specifically, I defend the idea that, given a local dynamics, what's happening at a spacetime region's *boundary* screens off what happens on the region's inside from what happens on its outside. This provides a systematic connection between chances on loops and the ordinary, time-travel-free chances.

The essay is structured as follows. To streamline the discussion, section 1.2 introduces a simple, stripped-down spacetime loop scenario—grandpa will reappear later. In Sections 1.3–1.5, I survey two accounts of this case. One is based on the orthodox "temporalist" framework of chance, promulgated by Lewis (1987); the other is based on the idea that chances are invariant across intrinsically duplicate trials. The first account often trivializes chances on spacetime loops, and the second account leads to inconsistency. Both should thus be rejected. This sets the stage for my positive proposal. In Sections 1.6 and 1.7, I explain the proposal's two core principles, *Acyclic Chance Invariance* and *Strong Boundary Markov*. Section 1.8 applies these two principles to the simple loop scenario. Section 1.9 revisits the stochastic grandfather paradox and develops a general recipe for deriving chances on loops. Section 1.10 corrects a misconception regarding chances on loops. It shows that, in stark contrast to the acyclic case, in *cyclic* spacetimes the dynamics alone manages to fix *unconditional* chances over the states of the universe, thus generating a "complete probability map of the world". Similarly, the dynamics generically fixes precise expectations about what will emerge from future wormholes. Section 1.11 concludes.

1.2 A Simple Case

The previous scenario involves a spacetime loop: your saving your infant grandfather leads to him growing up and having children, one of whom bears you and your sibling, which leads you to eventually enter the wormhole, ending up in front of infant grandfather. For the purpose of discovering general principles about cases like this, let's start with a particularly simple spacetime loop world, called CIRCLE.

Imagine a single stationary particle, occupying a single point of space, persisting in circular time. For concreteness, let the cycle have a period of 100 billion years—that is, persisting for 100 billion years from now gets the particle back to the present. Heuristically, you may picture the spacetime as an infinitely thin strip of paper, with an arrow drawn parallel to the strip, and whose ends you've glued together. This circular strip is a crude representation of a universe consisting of a single point of space, persisting in circular time, with the arrow indicating the time's direction. Now picture a small marble sitting at every point along the strip. The marbles represent the particle's different time slices, successively occupied by the particle as it persists through time. Together, this yields a crude representation of a single particle in a circular, one-dimensional spacetime.[5]

Let's stipulate a simple dynamics for the particle. Let the particle have two intrinsic magnitudes, *color* and *clock*. The particle's clock grows in proportion to the time passed, until it reaches 24 hours, at which point it restarts from 0. The particle sometimes changes colors, exactly at clock restart points. You only ever observe three possible colors and three possible transitions: *red to green*, *green to blue*, and *blue to red*. (See fig. 1.1.) Moreover, you observe a color change at about 2 out of 10 reset points.

Suppose the particle is currently red, and you'd like to calculate the chance that it'll still

[5]Mathematically, we can represent CIRCLE by a one-dimensional, oriented, closed Lorentzian (or, equivalently in the one-dimensional case, Riemannian) manifold—that is, a circle equipped with a metric. Note that the one-dimensional is illustrative not only because it's particularly simple, but also because one-dimensional oriented spaces commonly appear as base spaces in fibre bundle constructions of other spaces, e.g. of Galilean spacetime or of the configuration-space-in-time relevant to some interpretations of quantum mechanics.

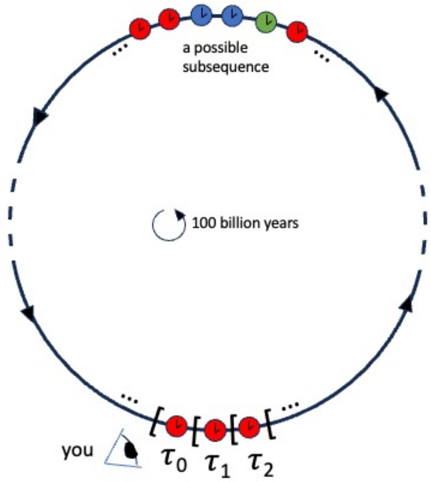


Figure 1.1: A sketch of CIRCLE.

be red tomorrow. Two hypotheses may jump out, paralleling those about the stochastic grandfather paradox. One might think that the chance that the particle changes color at the next reset point is 1 if it actually changes color, and 0 if it doesn't. After all, whatever the particle does has *already* happened. But then again, whatever happens is *still* to happen, and the particle's current color leaves tomorrow's color open. So perhaps the chance of a color switch is just what it ordinarily is. The following sections survey two proposals, capturing the two hypotheses. I argue that both should be rejected.

1.3 Against Temporalism

Chances vary in time: Wilbur and Orville Wright flipped a coin to settle who would fly first. As the coin was flipped, both had a positive chance of being the first to fly. But because Orville lost the flip, *today* there's chance 0 that he flew first. Lewis's (1987) framework therefore relativizes chances to times:

Temporalism about chance:

1. Necessarily, chance is a function of two arguments, a proposition and a

time.[6]

2. Necessarily, if chance is the function ch, then for any time t, $ch(\cdot, t)$ assigns chance 1 to t's temporal history, i.e. to the strongest truth "entirely about matters of particular fact" (Lewis, 1987) at times at or before t.[7]

The second clause causes trouble. In CIRCLE, every time precedes every other time. Given this, clause 2 trivializes all chances in CIRCLE: it implies that every chance function assigns chance 1 to the particle's actual complete color history, and 0 to every non-actual history. As a valid argument:

(i) For all times t and t', t' is at or before t.

(ii) For all times t and t', if t' is at or before t, then the temporal history of t entails what color the particle is at t'.

(iii) For any time t, $ch(\cdot, t)$ assigns chance 1 to t's temporal history.

(iv) For all propositions A and B and times t, if A entails B and $ch(A, t) = 1$, then $ch(B, t) = 1$.

$\therefore$ For all times t and t', $ch(\cdot, t)$ assigns chance 1 to what color the particle is at t'.

[6]Officially, Lewis adds a *third* argument, a world, where "$ch(A, t, w)$" (or "$P_{tw}(A)$" in his notation) refers to "the chance, at time t and world w, of A's holding" (87). By contrast, we're setting up chance as a *contingent* relation here (between a proposition, time, and real number). I find this latter setup more perspicuous, because it's neutral on the underlying account of contingency (e.g., if it's worlds-based, or something else). But nothing of significance will hang on this here.

[7]Actually, the original quote says "entirely about matters of particular fact at times *no later than* t" (Lewis, 1987). This, incidentally, doesn't fall prey to the objection below: since in CIRCLE every time is later than any other time, this account simply places no constraints at all on $ch(\cdot, t)$ in CIRCLE. Now, it's clear that Lewis himself doesn't use "no later than t" to distinguish it from "at or before t". (As evidenced a few paragraphs later, where he writes, of a proposition A about states of affairs at time t_A, that "[i]f t is later than t_A, then A is admissible at t" (Lewis, 1987). Temporalism entails that A is admissible, and hence receives chance 1, at t only if "t is later than t_A" entails "t_A is *no later* than t". But "t is later than t_A" is equivalent to "t_A is before t", and so Lewis must assume that "t_A is before t" entails "t_A is no later than t". Since Lewis also thinks that A is admissible at t if t_A is *simultaneous* with t, it follows that he must assume that "t_A is at or before t" entails "t_A is no later than t".) In any case, an account which places no constraints at all on $ch(\cdot, t)$ in CIRCLE is seriously incomplete. For example, surely $ch(\cdot, t)$ will at least assign chance 1 to the state of the world at t. Temporalism, as formulated in the main text, is thus a proposal for filling out the account.

Premise (ii) follows from the definition of temporal history; (iii) follows from temporalism's clause 2; and (iv) follows from the fact that $ch(\cdot, t)$ is a probability function. Premise (i), meanwhile, is supported by two thoughts: (a) there is some small duration ε such that, for any t, all times no more than ε into t's past are before t, and (b) "before" is transitive. While it's logically possible to either deny (a) or deny (b), neither possibility seems attractive. Regarding (a): surely, if *any* times are before t in CIRCLE, it includes those in t's most recent past. Meanwhile, holding that *no* times are before t falsifies evident truths: for example, despite the particle's changing color from red to green, it wouldn't be the case that the particle has *previously* been red. Moreover, the reply would at best secure silence about CIRCLE. Yet our aim is a positive theory about chances on loops. Regarding (b): denying transitivity burdens us with arbitrary cutoffs—when is t' just far enough in t's past that it's no longer "before" t? I see no principled way to draw this distinction.

The trivialization of chances in CIRCLE is problematic for two connected reasons.[8] First, recall the regularity you observe: whenever the particle has a given color, in about 2 out of 10 cases it'll have a different color the next day. It would seem extremely natural—and useful, and informative—to try to describe this behavior in non-trivial chance-theoretic terms. Indeed, it seems just as natural to do so in CIRCLE as it does in any *linear* world. But that's incompatible with Clause 2.

Secondly, a universe with circular time is still compatible with our empirical evidence.[9]

[8] Lewis (1987) is aware that temporalism has issues with time travel. He notes that the existence of time travelers may make some past information inadmissible: "That is why I qualified my claim that historical information is admissible, saying only that it is so 'as a rule'." (ibid., 274) But Lewis mentions this problem only to discard it: he states that he merely wants to argue that "the Principal Principle captures our common opinions about chance" and those common opinions, he says, "may rest on a naive faith that past and future cannot possibly get mixed up". (ibid., 274) I find it doubtful that our common opinions include any clear judgment about the possibility of time travel. In any case, Lewis admits that "[a]ny serious physicist, if he remains at least open-minded both about the shape of the cosmos and about the existence of chance processes, ought to do better" (ibid., 274). Philosophers should, too.

[9] For example, it is compatible with our empirical evidence that our universe is (representable by) a four-dimensional time-like closed Lorentzian manifold. Such a manifold is permitted by Einstein's field equations, and if spatially flat, accords with general astronomical observations about the shape of our universe. (The role of the Past Hypothesis in such a world would be played by the posit that there is a low-entropy macrostate *at some time*, with entropy increasing bidirectionally from there.)

If this speculative scenario was true, would it falsify all scientific theories involving non-trivial chances? Would it mean, for example, that radioactive decay wasn't stochastic after all? The answer is clearly no. But this contradicts temporalism's clause 2. So we should reject temporalism.[10]

1.4 Urchance

The problem with temporalism is that its chance functions must be certain of complete temporal histories. To deliver non-trivial chances in cyclic spacetimes, the correct chance theory has to be more flexible.

As Hall (2004b) observes, temporalism can be reformulated by stipulating the existence of an "urchance" function, given to us by the fundamental physical laws. The chance function at a time t, ch_t, is then the result of conditioning said urchance function on the temporal history up to t, H_t, i.e., $ch_t(\cdot) = \text{urchance}(\cdot|H_t)$. One advantage of the urchance approach is that it automatically ensures that chance functions at different times "cohere" with one another, *viz.* that chance functions at later times result from conditioning those at earlier times on intervening history. Coherence is mandated by canonical chance deference principles like the Principal Principle.[11]

Now, the temporalist considers *only* the results of conditioning the urchance function on complete temporal histories. But sometimes the result of conditioning the urchance function on *non-history* propositions is also well-defined. This much follows already from

[10]Cusbert (2018; 2022) has recently suggested replacing temporalism with a "causal history view" of chance: instead of having $ch(\cdot, t)$ assign chance 1 to t's temporal history, have it assign chance 1 to t's *causal* history instead. (Cusbert's formalism replaces "times" with "globally connected sets" of events—this difference doesn't matter here.) But this fares no better than temporalism in CIRCLE: the particle's color at each day is caused by the color the previous day. Moreover, *causal histories* are transitively closed (even if causation isn't). So Cusbert's view also trivializes chance in CIRCLE.

[11]To see this: the Principal Principle (in one of its canonical formulations, Lewis (1987)) says that, where Cr_0 is any rational initial credence function, t any time, H_t the world's actual history up to t, and T the true chance theory, $Cr_0(A|H_tT) = ch_t(A)$. Where $H_{[t,t^+]}$ is the intervening history from t to t^+, it follows that $ch_{t^+}(A) = Cr_0(A|H_{t^+}T) = Cr_0(A|H_{[t,t^+]}H_tT) = ch_t(A|H_{[t,t^+]})$, provided $H_{[t,t^+]}$ has positive chance at t.

the probability laws alone: where urch is the urchance function and H_t the world's history up to t, whenever $\text{urch}(\cdot|H_t)$ assigns positive probability to some A, $\text{urch}(\cdot|H_tA) = \text{urch}(\cdot \wedge A|H_t)/\text{urch}(A|H_t)$ is a well-defined probability function even if H_tA isn't a complete temporal history. The fundamental dynamical laws go yet beyond this. Consider any bounded history segment. Since it is bounded, it doesn't entail any complete temporal history; yet, together with the fundamental dynamical laws, the segment's state generically entails a well-defined probability distribution over the segment's possible futures. (In the deterministic case, this probability distribution is trivial.) So, insofar as urchance encodes exactly the content of the fundamental dynamical laws, the result of conditioning it on non-history propositions is often well-defined too.

In my view, urchance *is* naturally thought of as encoding (exactly) the content of the fundamental dynamical laws. The results of conditioning it on non-history propositions are then objective chance functions: since they follow from the fundamental dynamical laws alone, they are objective; they are "single-case" probabilities—*viz.*, $\text{urch}(\cdot|B)$ exists and is generally non-trivial, even if an event described by B occurs only once; they respect dynamical symmetries; and they go with straightforward deference principles.[12]

Now, the fundamental dynamical laws aren't generally so powerful that $\text{urch}(A|B)$ is precisely defined for *all* physically specifiable A and B. Some B are too weak for some A: for example, a contingent proposition generally won't have a precise probability conditional on a logical tautology. In the following paragraphs, I outline a framework for capturing this predictive weakness using imprecise probability. While the framework underpins the remainder of the paper, it is designed so that the paper's main philosophical insights remain accessible without it. (Specifically, the formalism is designed so that all main text equations are interpretable under the simplifying, albeit false, assumption that urch is a

[12] One such principle, compatible with the Principal Principle: where Cr_0 is any rational prior credence function, A and E are any propositions, and $\langle \text{urchance}(A|E) = x\rangle$ is the proposition that the urchance of A, conditional on E, is x: $Cr_0(A|E \wedge \langle \text{urchance}(A|E) = x\rangle) = x$.

precise probability function.) Readers who prefer to skip formalism may proceed to the final paragraph of this section.

To capture urchance's predictive weaknesses, we express it not in terms of a single probability function, but in terms of a *set* of probability functions. Intuitively, these functions are all the *precisifications* of the fundamental laws' probability judgments.[13] Accordingly, A has a precise chance x conditional on B iff all functions in the set assign A probability x conditional on B. Where the functions disagree, the best we can do is assign A a set of chance values conditional on B. Officially, we let these functions be two-place, total, *primitively conditional* probability functions (cf. Hájek (2003)).[14]

Now, one way to proceed would be to identify urchance directly with the set of probability functions induced by the true fundamental dynamical laws. More conveniently, however, we define it as follows. Where $\mathbf{u}$ is the set of all precisifications of the dynamical laws' probability judgments, urchance $\mathrm{urch}(\cdot|\cdot)$ is a function from pairs of propositions into functions on $\mathbf{u}$: specifically, for any pair (X, Y) of propositions, $\mathrm{urch}(X|Y)$ is the function from $\mathbf{u}$ into $[0, 1]$ mapping $u \in \mathbf{u}$ to $u(X|Y)$—that is, $\mathrm{urch}(X|Y)(u) = u(X|Y)$. Intuitively, $\mathrm{urch}(X|Y)$ isn't merely a set of values, but (additionally) keeps track of which member of $\mathbf{u}$ assumes which value.

This way of defining urchance has the advantage that arithmetic operations on $\mathrm{urch}(\cdot|\cdot)$

[13]Cf. van Fraassen (1984), who presents a set-based (or "representor"-based) formalism for credences. There are significant philosophical advantages to a set-based formalism over less expressive alternatives, such as one using partial urchance functions. For example, a set-based formalism can express probabilistic independencies even where no precise unconditional probabilities exist (cf. Joyce (2010))—this enables our definitions of Markov properties in Section 1.7. A set-based formalism also streamlines the proofs in Appendix C.

[14] Primitive conditionality is needed because sometimes A has precise chance conditional on B even if B has (precisely) zero prior chance. Popper (1968, App. *IV and *V) offers a convenient axiomatization of total primitively conditional probability. His axioms—specifically, the version with mere finitely additivity—are what I'll mean throughout by "the laws of probability". The axioms include the standard multiplicative axiom, $p(A \wedge B, C) = p(A, B \wedge C) \cdot p(B, C)$, finite additivity, and the probabilistic analogue of explosion, $p(A, B \wedge \neg B) = 1$. We'll also assume that any member of the set induced by some fundamental dynamical laws is "maximally sure" exactly of anything metaphysically entailed by those laws. That is, for all members p of the set, $p(A|\neg A) = 1$ iff A is metaphysically entailed by L. In particular, no member assigns maximal prior chance to any nomically contingent matter of fact.

can be defined in the standard functional way, i.e., they apply point-wise. For example, $\text{urch}(A|B) \cdot \text{urch}(C|D)$ denotes the function which maps each $u \in \mathbf{u}$ to the value $u(A|B) \cdot u(C|D)$. Similarly, equality is point-wise: $\text{urch}(A|B) = \text{urch}(C|D)$ says that the both sides agree everywhere on their domain, i.e. for all $u \in \mathbf{u}$, $u(A|B) = u(C|D)$, while $\text{urch}(A|B) = x$ says that the left-hand side has constant value x, i.e. for all $u \in \mathbf{u}$, $u(A|B) = x$. Intuitively, whenever you see an expression like $\text{urch}(A|B)$ in an equation, read it as the equation's holding *determinately*, i.e. for all members of $\mathbf{u}$. (The same goes for inequalities, $>$, and approximate equalities, $\approx$.) Conditioning is equally straightforward: for any proposition B, $\text{urch}(\cdot|\cdot \wedge B)$ maps any pair (X, Y) of propositions to the function on $\mathbf{u}$ mapping each $u \in \mathbf{u}$ to $u(X|Y \wedge B)$.

Our definition entails that, whenever all members of $\mathbf{u}$ satisfy an equation involving only arithmetic operations and conditioning, urch satisfies an equation of the exact same form. These equations include the probability axioms. For example, since all members of $\mathbf{u}$ satisfy the multiplicative axiom, urch satisfies an expression of the same form, $\text{urch}(AB|C) = \text{urch}(A|BC) \cdot \text{urch}(B|C)$. Defining urchance as we have allows us to use syntactically familiar equations.

Where urch is the urchance function, I'll call B the "background proposition" for the function $\text{urch}(\cdot|\cdot \wedge B)$.[15] The urchance formalism enables theories of non-trivial chances in cyclic worlds: chance functions with background propositions weaker than any temporal history can assign non-trivial probabilities to contingent matters of fact. The second account attempts to build on this flexibility to construct a theory of non-trivial chances—unsuccessfully so.

[15] The terminology of "background proposition" is also used in Nelson (2009) and Cusbert (2018). Other authors with flexible chance formalisms include Meacham (2005); Nelson (2009); Briggs (2010); Handfield & Wilson (2014); Cusbert (2018).

1.5 Against Chance Invariance

Consider again an infinitely thin strip of paper, but this time leave its ends unjoined. Call the represented world LINE (or L for short).[16] This world is inhabited by the same sort of particle as CIRCLE, subject to the same laws. Now suppose you are told that, in LINE, upon clock reset there's (precisely) a *0.2 chance* that the particle changes color (from red to green, green to blue, or blue to red). What can we infer from this about the chances in CIRCLE? What would the transition chances be if the spacetime was cyclic rather than linear?

A natural idea is that the chances would be unchanged—that transition chances in circular spacetimes are just what they are in identical "linear" situations. This is suggested by the popular idea that objective chances are, as Schaffer (2003) puts it, "stable". Arntzenius & Hall (2003, p. 178) express the idea as follows: "if … two processes going on in different regions of spacetime are exactly alike, your [theory should assign] to their outcomes the same single-case chances". Or, as Schaffer (2007, p. 125) puts it, "chance values should remain constant across intrinsically duplicate trials" within the same world. While these principles strictly speaking only concern chance assignments within the *same* world, they have obvious and natural generalizations that also cover chance assignments *across* worlds with the same laws. These generalizations dictate the same transition chances for CIRCLE as for LINE.[17]

Arntzenius and Hall don't provide a formally precise version of their principle, and Schaffer's (2007) presentation assumes temporalism.[18] So let's formulate a principle ourselves, for the case of CIRCLE and LINE, using urchance. (Instead of "stability" I choose

[16] Mathematically, we can represent the world by a one-dimensional, oriented, open Lorentzian (or, equivalently in the 1D case, Riemannian) manifold—that is, a line equipped with a metric.

[17] We could also make CIRCLE and LINE world-mates, by connecting them (say) by a space-like line. Schaffer's and Arntzenius and Hall's principles then apply directly.

[18] Horacek (2005, p.428) and Effingham (2020, p.152) each propose similar principles which also presuppose temporalism. Cusbert (2022, p. 617,625-6) meanwhile formulates a version of stability weak enough to count even the trivialized chances on CIRCLE as "stable" relative to the chances on LINE. This doesn't strike me as a promising formulation of stability, going against the intuitions the principle is meant to capture. Moreover, Cusbert's principle doesn't get us out of the trivialization trap.

the label “invariance”, which I find more fitting.) Informally, the idea is that the chance of an event conditional on the state of an interval in LINE equals the chance of a duplicate event, conditional on the state of a duplicate interval in CIRCLE, provided the temporal distances between the event and the interval are the same in both worlds.

More carefully, where t and s are times, let the *forward distance from t to s* be the smallest duration from t to s, i.e., the smallest amount of time yo have to persist to get from t to s. In CIRCLE, there's a forward distance from any time to any other time. In LINE, a forward distance from t to s exists iff t occurs earlier than s. Where I and J are *intervals*, let the *forward distance from I to J* be the forward distance from the starting point of I's closure to the starting point of J's closure.[19] For example, in CIRCLE (fig. 1.1), the forward distance from τ_0 to τ_2 is 2 days, and the forward distance from τ_2 to τ_0 is 100 billion years minus 2 days. Moreover, say that two pairs of intervals, (I, J) and (I^*, J^*), are *temporally congruous* iff I and I^* have equal duration, J and J^* have equal duration, and the forward distance from I to J equals the forward distance from I^* to J^*.

Since CIRCLE and LINE have the same fundamental dynamical laws, they share the same urchance function. Denote by $\text{urch}_C(\cdot|\cdot)$ and $\text{urch}_L(\cdot|\cdot)$ the results of conditioning that function on a complete description of CIRCLE's and LINE's spacetime geometry, respectively.[20]

> ***Hypothesis.*** **Chance Invariance:** Let (I, J) and (I^*, J^*) be temporally congruous pairs of intervals in CIRCLE and LINE, respectively. Then
>
> $$\text{urch}_C(P(J)|Q(I)) = \text{urch}_L(P(J^*)|Q(I^*)),$$
>
> for any qualitative intrinsic properties P and Q.

[19] If one of the two closures doesn't have a starting point or if there is no forward distance from I's starting point to J's starting point, the forward distance from I to J is ill-defined.

[20] That is, where $\mathcal{C}$ is a complete description of CIRCLE's geometry and $\mathcal{L}$ is a complete description of LINE's geometry, $\text{urch}_C(\cdot|\cdot) :=\text{urch}(\cdot| \cdot \wedge\mathcal{C})$ and $\text{urch}_L(\cdot|\cdot) :=\text{urch}(\cdot| \cdot \wedge\mathcal{L})$. We'll always understand complete geometric descriptions to include a “that's all” clause—i.e., they say that they describe all geometrical relationships between spacetime regions.

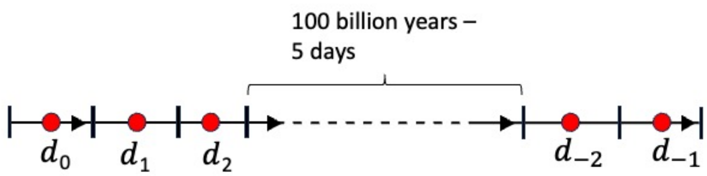


Figure 1.2: A sketch of LINE.

Chance Invariance captures the thought that locally duplicate situations, like $Q(I)$ and $Q(I^*)$, generate the same chances for locally duplicate outcomes, like $P(J)$ and $P(J^*)$, provided the relevant temporal distances are the same.

To illustrate the principle, let RED(τ) be the proposition that τ is a 24-hour interval and that a single particle exists throughout τ, is red throughout τ, and has a clock reading of 0 at the start of τ. Chance Invariance then requires that $\text{urch}_C(\text{RED}(\tau_1)|\text{RED}(\tau_0)) = 0.8$. For consider any two successive days d_0 and d_1 in LINE. The pairs (d_0, d_1) and (τ_0, τ_1) are temporally congruous. Moreover, from the dynamics in LINE, we have $\text{urch}_L(\text{RED}(d_1)|\text{RED}(d_0)) = 0.8$. So, by Chance Invariance,

$$\text{urch}_C(\text{RED}(\tau_1)|\text{RED}(\tau_0)) = \text{urch}_L(\text{RED}(d_1)|\text{RED}(d_0)) = 0.8.$$

Alas, Chance Invariance is inconsistent: it defines conditional chances in CIRCLE twice over, with conflicting results. Essentially, what goes wrong is that, except for antipodes, any two days in CIRCLE bear two distinct forward distances to each other: depending on which day comes first, the forward distance is either the short way or the long way around CIRCLE. For example, the pair (τ_0, τ_1) has a forward distance of one day, whereas the pair (τ_1, τ_0) has a forward distance of 100 billion years minus one day. As a result, the two pairs are temporally congruous to very different pairs in LINE. This makes Chance Invariance yield conflicting results.

For illustration, let's calculate $\text{urch}_C(\text{GREEN}(\tau_1)|\text{RED}(\tau_2) \wedge \text{RED}(\tau_0))$—the chance, in CIRCLE, of the particle's being green at τ_1, conditional on its being red on the two adjacent days. First consider the pairs (τ_0, τ_1) and (τ_1, τ_2) in CIRCLE and the temporally congruous days (d_0, d_1) and (d_1, d_2) in LINE—see figure 1.2. Chance Invariance yields the following:[21]

$$\mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) = \mathrm{urch}_L(\mathrm{GREEN}(d_1)|\mathrm{RED}(d_2) \wedge \mathrm{RED}(d_0)). \tag{1.5}$$

But the pair (τ_0, τ_1) is also temporally congruous with (d_{-2}, d_{-1}) while the "inverse" pair

[21] *Proof:* By the multiplicative axiom,

$$\begin{aligned}\mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) \cdot \mathrm{urch}_C(\mathrm{RED}(\tau_2)|\mathrm{RED}(\tau_0)) = \\ = \mathrm{urch}_C(\mathrm{GREEN}(\tau_1) \wedge \mathrm{RED}(\tau_2)|\mathrm{RED}(\tau_0)).\end{aligned} \tag{1.1}$$

From the congruity of (τ_0, τ_1) and (d_0, d_1) and of (τ_1, τ_2) and (d_1, d_2), we also have that (τ_0, τ_2) and (d_0, d_2) are congruous and that $(\tau_0, \tau_1 \cup \tau_2)$ and $(d_0, d_1 \cup d_2)$ are congruous. So, by Chance Invariance,

$$\mathrm{urch}_C(\mathrm{RED}(\tau_2)|\mathrm{RED}(\tau_0)) = \mathrm{urch}_L(\mathrm{RED}(d_2)|\mathrm{RED}(d_0)), \tag{1.2}$$

and

$$\mathrm{urch}_C(\mathrm{GREEN}(\tau_1) \wedge \mathrm{RED}(\tau_2)|\mathrm{RED}(\tau_0)) = \mathrm{urch}_L(\mathrm{GREEN}(d_1) \wedge \mathrm{RED}(d_2)|\mathrm{RED}(d_0)). \tag{1.3}$$

From eqs. 1.1–1.3,

$$\begin{aligned}\mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) \cdot \mathrm{urch}_L(\mathrm{RED}(d_2)|\mathrm{RED}(d_0)) = \\ = \mathrm{urch}_L(\mathrm{GREEN}(d_1) \wedge \mathrm{RED}(d_2)|\mathrm{RED}(d_0)).\end{aligned} \tag{1.4}$$

But, from the given dynamics in LINE, $\mathrm{urch}_L(\mathrm{RED}(d_2)|\mathrm{RED}(d_0)) = 0.8^2 > 0$, and so, from eq. 1.4,

$$\begin{aligned}\mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) &= \frac{\mathrm{urch}_L(\mathrm{GREEN}(d_1) \wedge \mathrm{RED}(d_2)|\mathrm{RED}(d_0))}{\mathrm{urch}_L(\mathrm{RED}(d_2)|\mathrm{RED}(d_0))} \\ &= \mathrm{urch}_L(\mathrm{GREEN}(d_1)|\mathrm{RED}(d_2) \wedge \mathrm{RED}(d_0)),\end{aligned}$$

where the second line follows by the multiplicative axiom. ■

(τ_2, τ_1) is temporally congruous with (d_0, d_{-1}). This yields:[22]

$$\mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) = \mathrm{urch}_L(\mathrm{GREEN}(d_{-1})|\mathrm{RED}(d_0) \wedge \mathrm{RED}(d_{-2})). \tag{1.10}$$

But eqs. 1.5 and 1.10 conflict. Because the dynamics in LINE disallows RED-GREEN-RED transitions,

$$\mathrm{urch}_L(\mathrm{GREEN}(d_1)|\mathrm{RED}(d_2) \wedge \mathrm{RED}(d_0)) = 0.$$

[22]*Proof:* By the multiplicative axiom,

$$\begin{aligned} \mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) \cdot \mathrm{urch}_C(\mathrm{RED}(\tau_0)|\mathrm{RED}(\tau_2)) = \\ = \mathrm{urch}_C(\mathrm{GREEN}(\tau_1) \wedge \mathrm{RED}(\tau_0)|\mathrm{RED}(\tau_2)). \end{aligned} \tag{1.6}$$

From the congruities of (τ_2, τ_1) and (d_0, d_{-1}) and of (τ_0, τ_1) and (d_{-2}, d_{-1}), we obtain that (τ_2, τ_0) is congruous with (d_0, d_{-2}) and $(\tau_2, \tau_0 \cup \tau_1)$ is congruous with $(d_0, d_{-2} \cup d_{-1})$. So, by Chance Invariance,

$$\mathrm{urch}_C(\mathrm{RED}(\tau_0)|\mathrm{RED}(\tau_2)) = \mathrm{urch}_L(\mathrm{RED}(d_{-2})|\mathrm{RED}(d_0)), \tag{1.7}$$

and

$$\mathrm{urch}_C(\mathrm{GREEN}(\tau_1) \wedge \mathrm{RED}(\tau_0)|\mathrm{RED}(\tau_2)) = \mathrm{urch}_L(\mathrm{GREEN}(d_{-1}) \wedge \mathrm{RED}(d_{-2})|\mathrm{RED}(d_0)). \tag{1.8}$$

Plugging eqs. 1.7 and 1.8 into eq. 1.6 yields

$$\begin{aligned} \mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) \cdot \mathrm{urch}_L(\mathrm{RED}(d_{-2})|\mathrm{RED}(d_0)) = \\ = \mathrm{urch}_L(\mathrm{GREEN}(d_{-1}) \wedge \mathrm{RED}(d_{-2})|\mathrm{RED}(d_0)). \end{aligned} \tag{1.9}$$

But $\mathrm{urch}_L(\mathrm{RED}(d_{-2})|\mathrm{RED}(d_0)) \approx 1/3 > 0$, for the particle's current color provides essentially no evidence about its far-future color. (To derive this formally, divide the probability-weighted sum of color trajectories compatible with $\mathrm{RED}(d_{-2}) \wedge \mathrm{RED}(d_0)$ by the probability-weighted sum of color trajectories compatible with $\mathrm{RED}(d_0)$, leading to a formula similar to the expression in fn. 45.) So, from eq. 1.9,

$$\begin{aligned} \mathrm{urch}_C(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_2) \wedge \mathrm{RED}(\tau_0)) &= \frac{\mathrm{urch}_L(\mathrm{GREEN}(d_{-1}) \wedge \mathrm{RED}(d_{-2})|\mathrm{RED}(d_0))}{\mathrm{urch}_L(\mathrm{RED}(d_{-2})|\mathrm{RED}(d_0))} \\ &= \mathrm{urch}_L(\mathrm{GREEN}(d_{-1})|\mathrm{RED}(d_0) \wedge \mathrm{RED}(d_{-2})), \end{aligned}$$

where the second line follows from the multiplicative axiom. ■

But the same dynamics also entails that[23]

$$\mathrm{urch}_L(\mathrm{GREEN}(d_{-1})|\mathrm{RED}(d_0) \wedge \mathrm{RED}(d_{-2})) = 0.2 > 0.$$

Contradiction. So Chance Invariance is inconsistent.

1.6 Two Vestiges of Invariance

Despite its inconsistency, Chance Invariance has intuitive appeal. Moreover, I think we can salvage its appealing parts in the form of weaker, and jointly consistent, principles.

First, its restriction to *loop-free* worlds remains consistent: two intrinsically duplicate situations in two nomically compatible *loop-free* worlds generate the same chance distributions. Let's call this weaker principle *Acyclic Chance Invariance*.

Below I'll spell out a precise version of this principle. To do this, a few concepts are needed, which will reappear throughout the rest of the paper (and thus shouldn't be skipped). Informally, a *curve* is any directed line (straight or curved) through spacetime.[24] A *causal curve* is any curve through spacetime which could be the trajectory of a material particle.[25,26] Causal curves can be either *future-directed* or *past-directed*. A *closed causal curve*

[23] *Proof:* Because d_{-2} is located in between d_2 and d_{-1}, $\mathrm{RED}(d_{-2})$ "screens off" $\mathrm{RED}(d_2)$ from $\mathrm{GREEN}(d_{-1})$:

$$\mathrm{urch}_L(\mathrm{GREEN}(d_{-1})|\mathrm{RED}(d_0) \wedge \mathrm{RED}(d_{-2})) = \mathrm{urch}_L(\mathrm{GREEN}(d_{-1})|\mathrm{RED}(d_{-2})). \tag{1.11}$$

(In section 1.7 I'll discuss "screening off" much more—the relevant principle, Parental Markov, is true in acyclic worlds and yields eq. 1.11.) From the dynamics for LINE we moreover have

$$\mathrm{urch}_L(\mathrm{GREEN}(d_{-1})|\mathrm{RED}(d_{-2})) = 0.2. \blacksquare$$

[24] Formally, a *curve* is a function $c : I \to \mathcal{M}$ from an interval $I \subseteq \mathbb{R}$ into the spacetime $\mathcal{M}$. A *line* is the image of a curve. For bounded curves, we'll set $I = [0, 1]$ without loss of generality.

[25] Formally, we define a *causal curve* to be any differentiable curve with everywhere light-like (null) or time-like tangent vector.

[26] "Causal" thus has its technical meaning from physics, in terms of spacetime structure. I don't presuppose that this exactly—or even approximately—tracks the philosopher's various notions of causality. The fundamental dynamical laws speak the language of spacetime structure, not causation—so we're interested in the former, not the latter.

(or "spacetime loop", as I called it earlier) is any causal curve which loops back in on itself, i.e. ends at its starting point. Informally, a region R's *causal past*, $J^-(R)$, is that part of spacetime from which a material particle can eventually reach R—we include in this the entirety of R itself. Meanwhile, R's *causal future*, $J^+(R)$, is that part of spacetime which a material particle can eventually reach *from* the region—again, including R itself.[27] R's *proper causal future* $K^+(R)$ is the difference between the causal future and R, i.e. $K^+(R) := J^+(R) \setminus R$; *$R$'s proper causal past* $K^-(R)$ is defined analogously, $K^-(R) := J^-(R) \setminus R$. Finally, throughout, where urch is a possible urchance function and $\mathcal{M}$ a possible spacetime, $\text{urch}_{\mathcal{M}}$ denotes the result of conditioning urch on a complete description of $\mathcal{M}$'s geometry.

Now for the precise version of Acyclic Chance Invariance (those who'd like to avoid formalism may skip this paragraph). We'll formulate the principle for pairs (R_1, R_2) of disjoint regions where R_2 is "strictly to the future" of R_1, i.e., $R_2 \subseteq K^+(R_1)$ and $R_1 \cap K^+(R_2) = \emptyset$, which I'll also write as $R_1 < R_2$. Intuitively, whenever $R_1 < R_2$, $J^+(R_1) \cap J^-(R_2)$ comprises the region "between" R_1 and R_2.

> ***Thesis.* Acyclic Chance Invariance:** Let $\mathcal{M}$ and $\mathcal{M}'$ be spacetimes with no closed causal curves. Let R_1 and R_2 be regions in $\mathcal{M}$ with $R_1 < R_2$. Finally, let R_1' and R_2' be regions in $\mathcal{M}'$ with $R_1' < R_2'$, such that there is an isometry $\Phi : J^+(R_1) \cap J^-(R_2) \to J^+(R_1') \cap J^-(R_2')$ with $\Phi(R_1) = R_1'$ and $\Phi(R_2) = R_2'$.[28] Then,
>
> $$\text{urch}_{\mathcal{M}}(P(R_2)|Q(R_1)) = \text{urch}_{\mathcal{M}'}(P(R_2')|Q(R_1')),$$
>
> for any qualitative intrinsic properties Q and P.

Acyclic Chance Invariance is our first vestige of Chance Invariance.

[27] More exactly, I define R's causal future [causal past], $J^+(R)$ [$J^-(R)$], as the union of R with all points p such that there is a future-directed [past-directed] causal curve starting in R and ending in p.

[28] If Φ exists, you might call (R_1, R_2) and (R_1', R_2') *spatiotemporally congruous*, generalizing the concept of *temporally congruous* from Section 1.5.

The second vestige of Chance Invariance concerns the *cyclic* chances. As we've learned in the previous section, they can't be strictly identical to the acyclic chances. Still, they shouldn't just be arbitrary either. Instead, they should be derivable from the acyclic chances in a principled way.

To sharpen this up, suppose you're given some possible fundamental dynamical laws, defining an urchance function urch. The laws' *acyclic chances* is simply the collection of all functions $\text{urch}_{\mathcal{K}}(\cdot|\cdot)$ such that $\mathcal{K}$ is a spacetime without closed causal curves. The laws' *cyclic chances* are defined in exactly the same way, except that $\mathcal{K}$ now ranges over all spacetimes *with* closed causal curves. Say that a theory of cyclic chances is *dynamically scrutable* iff the cyclic chances can be inferred from the acyclic chances in a principled way. A demand for dynamic scrutability is our second vestige of Chance Invariance. It and Acyclic Chance Invariance are, I say, what remains of Chance Invariance when stripped of its inconsistency.

Our work is now cut out for us: we must find a (1) *consistent* and (2) *dynamically scrutable* theory of cyclic chances that (3) *avoids trivialization*. Chance Invariance satisfies (2) and (3), but not (1). A consistent but arbitrary way of assigning cyclic chances satisfies (1) and (3), but not (2). And Temporalism satisfies (1) and (2) but not (3)—it makes chances conditional on CIRCLE's geometry dynamically scrutable, but trivially so: the only allowed background proposition is the entire history. We are looking for a theory which checks all three boxes.

1.7 Spacetime Markov: What's Inside Doesn't Matter

As we saw in the introduction, when it comes to assessing the probability of a proposition, some information can override, or *screen off*, other information. When it comes to chance, what screens off what is partially determined by spacetime structure. For example, once we know the current state of a radioactive atom and its local environment, any additional information about the atom's causal past, e.g. how long it has been in its excited state, has no additional impact on the chance of its decaying within the next 10 seconds. Similarly, provided the dynamics are local[29] what happens at space-like separation also doesn't matter. In short, the atom's current state screens off everything outside of its own causal future.

Generalizing this yields the following proposal: a region's immediate proper causal past screens it off from everything outside of the region's causal future. This is the spacetime-theoretic analogue of the well-known "Causal Markov Condition". In its generic form, the Condition states that an event's immediate causes screen it off from any of its non-effects (cf. Hitchcock & Rédei (2021)). The Causal Markov Condition and its predecessor, Reichenbach's "Common Cause Principle", occupy important roles in debates about the metaphysics of time (Reichenbach, 1956) and the nature of causation (Hitchcock & Rédei, 2021). Now, how might its spacetime-theoretic analogue bear on what the chances on loops are? The idea is this: for some spacetime regions we can screen off *whether or not they're part of a world with closed causal curves*. As we shall see, this allows us to derive the cyclic chances from the acyclic chances.

1.7.1 Parental Markov

Let's start by stating the spacetime analogue of the Causal Markov Condition more precisely.

Informally, we say that a spacetime region S "screens off" region R from T iff, conditional on the complete state of S, urchance judges information about T as irrelevant to the

[29] Throughout this essay, I restrict myself to local dynamics. However, at least some non-local theories can be handled by my framework with mild modifications. For example, in the case of non-relativistic GRW—a stochastic theory of non-relativistic quantum mechanics, due to Ghirardi *et al.* (1986)—one can maintain Parental Markov (see below) relative to Galilean spacetime structure (where any differentiable curve intersecting each time slice at most once counts as "causal"), provided we restrict urchance to the algebra generated by the maximal intrinsic states of time slices. Similarly, a theory involving superluminal particles might posit a non-Minkowskian spacetime structure with an alternative notion of "causal", relative to which Parental Markov (see below) could still hold.

state of R. More formally:

***Def.* Screening Off:** Where urch is the world's urchance function, $\mathcal{M}$ its spacetime, and R, S, T any spacetime regions: S *screens off* R *from* T iff

$$\text{urch}_{\mathcal{M}}(Q_1(R)|Q_2(S) \wedge Q_3(T)) = \text{urch}_{\mathcal{M}}(Q_1(R)|Q_2(S)),$$

for any maximal intrinsic property Q_2 and any intrinsic properties Q_1 and Q_3 such that $Q_2(S) \wedge Q_3(T)$ is possible according to $\text{urch}_{\mathcal{M}}$.[30]

Next we'll define what it means to contain a region's "immediate proper causal past". For any region A, denote A's complement by $A^{\perp}$. Let a *thick parent* of R be any (possibly empty) region P, disjoint from R, which is such that every future-directed causal curve which starts in $(P \cup R)^{\perp}$ and ends in R has a non-trivial subcurve[31] in P before ever intersecting R.[32] Intuitively, this says that approaching any region from the past, you'll have to spend at least some time in its thick parents. Any thick parent contains a region's "immediate proper causal past".

Why "thick" parent? One reason is that dynamical laws often require velocities or other time derivatives as inputs for generating chance distributions.[33] But many, including myself, are inclined to endorse reductive views about time derivatives, such as the "at-at"

[30]On "maximal" and "possible according to $\text{urch}_{\mathcal{M}}$": an intrinsic property of a spacetime region is *maximal* iff it includes a "that's all" clause. That is, where Q is a maximal intrinsic property, $Q(R)$ says that every particular matter of fact intrinsic to R is entailed by it. "Possible": in our Popperian formalism, where u is a primitively conditional probability function, a proposition A is *impossible according to* u iff $u(\neg A|A) = 1$. Where $\mathbf{u}$ is the set of primitively conditional probability functions induced by the fundamental dynamical laws, call any function from pairs of propositions into functions on $\mathbf{u}$ an *urchance candidate*. Then, where $\mathfrak{u}$ is an urchance candidate, say that A is *impossible according to* $\mathfrak{u}$ iff $\mathfrak{u}(\neg A|A) = 1$ (where "$=$" is point-wise). This explains the proviso that $Q_2(S) \wedge Q_3(T)$ be *possible* according to $\text{urch}_{\mathcal{M}}$. For example, where $S' \subseteq S$, we can generically choose a possible intrinsic property $Q_{2'}$ of S' such that $Q_{2'}(S') \wedge Q_2(S)$ is impossible according to $\text{urch}_{\mathcal{M}}$. This would have the consequence that S generically wouldn't even screen off R from any of S's own subsets—that's not a promising notion of "screening off".

[31]A *subcurve* of $c : I \to \mathcal{M}$ is any restriction of c to a subinterval of I. A continuous curve is *non-trivial* iff its image consists of at least two points (which, in a continuous spacetime, is equivalent to its consisting of continuum-many points).

[32]That is, where $c : I \to \mathcal{M}$, there are three disjoint subintervals, $I_1, I_2, I_3 \subseteq I$ with $I_1 < I_2 < I_3$, such that $c[I_1] \subseteq (R \cup P)^{\perp}$, $c[I_2] \subseteq P$ is non-trivial, and $c[I_3]$ intersects R.

[33]Specifically, this is the case if the dynamical law involves second- or higher-order differential equations.

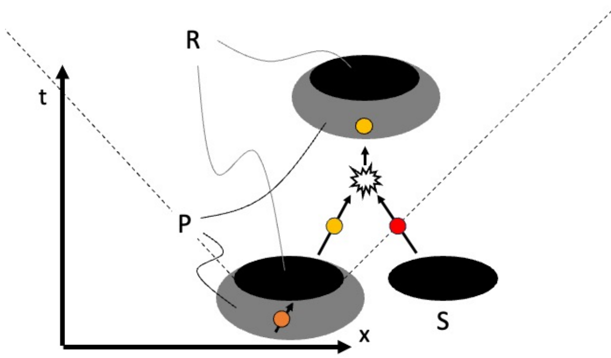


Figure 1.3: A counterexample to Unrestricted Parental Markov.

theory of velocities (Russell, 1903, Ch. 54). In this case, the maximal intrinsic state of a "thin" region doesn't specify particle velocities in that regions. A parent's *thickness* makes it so that at least one-sided derivatives are generally reductively definable.

In keeping with standard physics terminology, say that a spacetime region A is a *cause* of spacetime region B iff there is a future-directed causal curve starting in A and ending in B; *effect* is the converse of *cause*. One possible idea for the spacetime analogue of the Causal Markov Condition is this:

> ***Hypothesis.*** **Unrestricted Parental Markov:** Any thick parent of a spacetime region screens it off from any region not caused by it.

But Unrestricted Parental Markov is generically false. The problem is that, even in worlds with no closed causal curves, some regions generically cause their own thick parents: consider the disconnected region R (a union of two black ovals) in fig. 1.3. Conditioning on its thick parent P (the union of the two grey regions, disjoint from R) doesn't generally screen off R from S, despite S's being disjoint from R's causal future.[34]

[34]For a concrete example, consider the world PLANE, inhabited by red and yellow particles, with the following dynamics: whenever two particles of the *same* color collide, they fuse into a red particle; and whenever two particles of *different* colors collide, they fuse into a yellow particle. Additionally, there are orange particles, which cannot collide with anything. The orange particles have a short lifespan, at whose end they decay, with equal chance, either into a yellow or a red particle. P's lower part contains such an orange particle (and nothing else), and P's upper part a yellow particle (and nothing else). Suppose that the spacetime distances are such that the orange particle is guaranteed to decay inside R. P's state together with

This suggests a weaker principle. Let a *pure* thick parent of R be any thick parent of R not caused by R.

> ***Thesis.*** **Parental Markov:** Any *pure* thick parent of a spacetime region screens it off from any region not caused by it.

This principle is plausibly true in any spacetime without closed causal curves and with a local dynamics.

Alas, we are interested in spacetimes *with* closed causal curves. Here Parental Markov is less useful: any region which intersects some closed causal curve without containing it whole causes all of its own parents, and so lacks *any* pure thick parents. In a world like CIRCLE, Parental Markov is therefore entirely vacuous. Even worse, in some loop worlds Parental Markov is outright *false*—Appendix A.1 provides an example. We can't use a principle that's at best vacuous and at worst false. So: back to the drawing board. Are there other properties about "screening off" we can exploit?

1.7.2 Boundary Markov

Yes. A cognate principle is that a region is screened off by its *thick boundaries.* In Minkowski spacetime, Parental Markov already entails that this is true for a fairly general class of regions (as I prove in Appendix C). But in contrast to Parental Markov, the generalized principle—which will be called *Boundary Markov*—remains non-trivial and plausibly true for *all* regions, in worlds with loops and in worlds without, provided that the dynamics is local. (In fact, satisfying Boundary Markov is plausibly part of what it *is* for a dynamics to be local.)

Before stating the principle, let's build an intuition for it. Consider again LINE (cf. section 1.5). Suppose that you know the world's urchance function. You then observe the

S's containing a *red* particle entails that R's lower part contains a yellow particle. Yet P's state together with S's containing a *yellow* particle entails that R's lower part contains a red particle. So P's state doesn't screen off R from S.

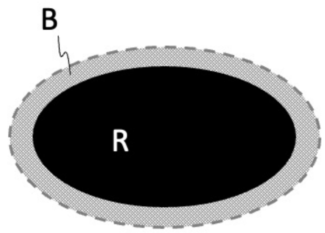


Figure 1.4: B is a thick boundary of R.

particle on two days, d_9 and d_{11}, learning BLUE(d_9) and RED(d_{11}), respectively. Suppose you'd now like to calculate the chances of events outside of d_{10}. For this, is it worth examining d_{10}, to establish if BLUE(d_{10}) or RED(d_{10})? No: as far as the chances of events *outside* of d_{10} are concerned, d_{10}'s state doesn't matter given d_9's and d_{11}'s states. Any information d_{10} may carry about events outside of it is *screened off* by their union.

The union $d_9 \cup d_{10} \cup d_{11}$ contains what I'll call a *thick neighborhood* of d_{10}, and $d_9 \cup d_{11}$ is a *thick boundary* of d_{10}. More generally, a *thick neighborhood* of a region R is any open superset N of R such that every continuous curve starting in $N^\perp$ and ending in R has a non-trivial subcurve in $N \backslash R$ before ever intersecting R.[35] Intuitively, a thick neighborhood is like a city plus its suburbs: coming from the outside, you have to spend some time in the 'burbs to get to the city. Meanwhile, a *thick boundary* of R is any region B disjoint from R such that $R \cup B$ contains a thick neighborhood of R. In our analogy, the suburbs themselves are a thick boundary of the city, as is any region that *contains* the suburbs but no part of the city. See e.g. fig. 1.4, where B is a thick boundary of R.[36] As I prove in Appendix B, in any spacetime with the same topology as $\mathbb{R}^n$, N is a thick neighborhood of R iff N is an open superset of R's closure.

[35]Formally: where $c : [0,1] \to \mathcal{M}$, the interval $[0,1]$ can be partitioned into subintervals I_0, I_1, I_2 with $I_0 < I_1 < I_2$, such that $c[I_0] \subseteq N^\perp$, $c[I_1] \subseteq N \backslash R$ is non-trivial, and $c[I_2]$ intersects R.

[36]It's worth noting some edge cases: the entire spacetime has the empty set as a thick boundary ("the universe doesn't have suburbs"), and the empty set has every region as a thick boundary ("everything is a suburb of the empty set").

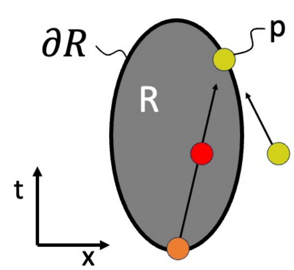
We are ready to state Boundary Markov:[37]

> ***Thesis.*** **Boundary Markov:** Any thick boundary of a spacetime region screens it off from any region disjoint from it.

Appendix C proves that, in Minkowski spacetime, given plausible conglomerability and locality assumptions for urchance, Parental Markov already entails a certain restriction of Boundary Markov.

Every region has a thick boundary (indeed often infinitely many), even in worlds with closed causal curves. Moreover, there are plausibly no locality-respecting counterexample to Boundary Markov. (As I said, Boundary Markov may indeed be partially constitutive of locality.) Appendix A.2 explicitly demonstrates that in the cyclic case in which Parental Markov is false, Boundary Markov is still true. Henceforth I'll assume that, necessarily, if the dynamics is local, Boundary Markov is true.

[37]Besides accommodating reductive accounts of derivatives, a boundary's thickness plays a second role. Even for non-reductivists, a *thin* boundary generally wouldn't screen off its inside from its outside. (This contrasts with parental thickness: to my knowledge, given instantaneous derivatives, thin parents do screen off a region from its non-effects.) To see this, consider again PLANE (cf. fn. 34), now with the following setup:



where ∂R is just the ordinary topological ("thin") boundary. Conditioning on the complete state of ∂R doesn't screen off R's inside from the outside. Let the distances be such that the orange particle on ∂R is guaranteed to decay inside R. Conditional on the state of ∂R, with its yellow particle at point p, the probability that R contains a red particle is then equal to the probability, conditional on the state of ∂R, that there is a yellow particle on course to collide with it at p. (Recall that the only way for a red particle to transform into a yellow one is to collide with another yellow particle.) Since ∂R contains little information about the universe's initial conditions, this probability is highly imprecise. (And if it was precise, it would be near zero—it would be more likely that the orange particle decayed into a yellow particle to begin with.) But conditional on the existence of a yellow particle on collision course with p, the chance rises to 1. So ∂R doesn't screen off R from the outside. Meanwhile, any *thick* boundary would automatically include either the information about the inside particle's color or the information about the outside particle's color (or both), and hence screen them off each other.

1.7.3 Strong Boundary Markov

We are almost there. So far we've understood screening off a region as screening off its *matter content*. But plausibly not only matter content can be screened off, but also a region's *internal geometry.* This insight holds the key to our theory of chances on loops. For whenever a region intersects *all* spacetime loops, the existence of spacetime loops partially depends on the region's internal geometry. But then screening off the region's internal geometry also screens off whether there are any spacetime loops. This gives us the desired connection between cyclic and acyclic chances.

To first get a better feel for the strenghtened notion of *screening off*, consider LINE. Suppose you've studied the world's geometry everywhere outside of d_{10}, but haven't examined d_{10} at all; you're even unsure about aspects of d_{10}'s internal geometry—say its length, or whether it is connected, etc. When calculating the chance of events in d_{10}'s complement, does your ignorance matter? No, I say, because d_9 and d_{11} screen off d_{10}'s internal geometry. Where LINE* (or L^* for short) is a world just like LINE except that (say) d_{10} isn't connected, and where A is a proposition about the complement of d_{10}, we get the following identity:

$$\text{urch}_L(A|\text{BLUE}(d_9) \wedge \text{RED}(d_{11})) = \text{urch}_{L^*}(A|\text{BLUE}(d_9) \wedge \text{RED}(d_{11})).$$

(Where, as always, urch_{L^*} denotes the result of conditioning urch on a complete description of L^*'s geometry.)

The previous definition of "screening off" compares urchance functions on the same spacetime geometries. One way to define the stronger notion instead compares the original spacetime with the result of *deleting* the screened-off region. For any spacetime $\mathcal{M}$, let $\mathcal{M}\backslash X$ denote the result of deleting region X from $\mathcal{M}$.[38] Moreover, let $\text{urch}_{\mathcal{M}\backslash X}$ be the

[38]More precisely, we'll first define the deletion operation for manifolds. Let $M = (\Omega, \mathcal{A}, g)$ be a Lorentzian manifold with topological space Ω, atlas $\mathcal{A}$, and (pseudo-)metric field g. For $X \subseteq \Omega$, define $M\backslash X := (\Omega\backslash X, \mathcal{A}|_{\Omega\backslash X}, g|_{\Omega\backslash X})$, where $\mathcal{A}|_{\Omega\backslash X} := \{\phi|_{U\backslash X} | (\phi : U \to \mathbb{R}^n) \in \mathcal{A}\}$ is the set of all restrictions of coordinate

result of conditioning urch on a complete geometrical description of $\mathcal{M}\backslash X$. (This description includes a "that's all" clause—cf. fn. 20—i.e. a proviso that $\mathcal{M}\backslash X$ is *all of spacetime*. Accordingly, $\text{urch}_{\mathcal{M}\backslash X}$ is not the result of conditioning urch on an *incomplete* description that's *true* at $\mathcal{M}$, but instead the result of conditioning urch on a *complete* description that's *false* at $\mathcal{M}$.)

> ***Def.* Strong Screening Off:** Where urch is the world's urchance function, $\mathcal{M}$ the spacetime, and R, S, T are any spacetime regions: *S strongly screens off R from T* iff
>
> $$\text{urch}_{\mathcal{M}}(Q_1(R)|Q_2(S) \wedge Q_3(T)) = \text{urch}_{\mathcal{M}\backslash T}(Q_1(R)|Q_2(S)),$$
>
> for any maximal intrinsic property Q_2 and any intrinsic properties Q_1 and Q_3 such that $Q_2(S) \wedge Q_3(T)$ is possible according to $\text{urch}_{\mathcal{M}}$.[39]

---

charts to their domains minus X. (When X has non-differentiable boundary, $(\Omega\backslash X, \mathcal{A}|_{\Omega\backslash X})$ will generally be neither a manifold nor manifold-with-boundary. But no matter: $g|_{\Omega\backslash X}$ retains all the metric structure we need to make sense of the dynamics.) Let now $\mathcal{M}$ be a spacetime represented by M. Then we define $\mathcal{M}\backslash X$ to be the part of $\mathcal{M}$ represented (under the same representation) by $M\backslash X$.

[39] To see how Strong Boundary Markov entails that thick boundaries screen off a region's internal geometry, consider the operation of *adding* a region to a spacetime. I'll define this precisely at the end of the footnote; for now, simply note that if $\mathcal{M} + X$ is a result of adding region X to $\mathcal{M}$, the operation $\cdot\backslash X$ (fn. 38) reverses that addition: $(\mathcal{M} + X)\backslash X = \mathcal{M}$. Let now B be a thick boundary of R in $\mathcal{M}$, and let A be a proposition purely about $(R \cup B)^{\perp}$. Let $\mathcal{M}^* := (\mathcal{M}\backslash R) + R^*$ be the result of adding some region R^* to $\mathcal{M}\backslash R$ such that B is also a thick boundary of R^* in $\mathcal{M}^*$. (Think of $\mathcal{M}^*$ as resulting from "replacing" R by R^* in $\mathcal{M}$ in some way.) By the above, we have $\mathcal{M}^*\backslash R^* = ((\mathcal{M}\backslash R) + R^*)\backslash R^* = \mathcal{M}\backslash R$. Moreover, by Strong Boundary Markov,

$$\text{urch}_{\mathcal{M}^*}(A|Q_1(B) \wedge Q_2(R^*)) = \text{urch}_{\mathcal{M}^*\backslash R^*}(A|Q_1(B)), \text{ and}$$
$$\text{urch}_{\mathcal{M}}(A|Q_1(B) \wedge Q_3(R)) = \text{urch}_{\mathcal{M}\backslash R}(A|Q_1(B))$$

(where Q_1 is a maximal intrinsic property and Q_2 and Q_3 are intrinsic properties such that $Q_1(B) \wedge Q_2(R)$ is possible according to $\text{urch}_{\mathcal{M}}$ and $Q_1(B) \wedge Q_3(R^*)$ is possible according to $\text{urch}_{\mathcal{M}^*}$). So,

$$\text{urch}_{\mathcal{M}}(A|Q_1(B) \wedge Q_2(R)) = \text{urch}_{\mathcal{M}^*}(A|Q_1(B) \wedge Q_3(R^*)).$$

In other words, B screens off A from anything in B, including B's internal geometry.

Now, to define *adding*, first define it for manifold-like structures. Specifically, let $M = (\Omega, \mathcal{A}, g)$ be the result of deleting a (possibly empty) region from a Lorentzian manifold. Then, for any X disjoint from Ω, $M + X = (\Omega \cup X, \mathcal{A}^{\Omega\cup X}, g^{\Omega\cup X})$ is *a result of adding X to M* iff

(i) $\mathcal{A}^{\Omega\cup X}$ is a set of charts whose domains form an open (relative to some chosen topology on $\Omega \cup X$, inducing Ω's topology) cover of $\Omega \cup X$,

(ii) $\mathcal{A}^{\Omega\cup X}|_{\Omega} = \mathcal{A}$ (see fn. 38 for the definition of $|.$ on sets of charts), Replacing "screens off" in Boundary Markov by "strongly screens off" yields

> ***Thesis.*** **Strong Boundary Markov:** Any thick boundary of a spacetime region strongly screens it off from any region disjoint from it.

I think that, accepting Boundary Markov, you should also accept Strong Boundary Markov. Intrinsic geometrical information isn't privileged over information about matter content: both are screened off by thick boundaries.

1.8 Cutting Loops

Our theory of chances on loops is now complete: it consists of Strong Boundary Markov and Acyclic Chance Invariance. Let us put them to work.

To start simply, consider SMALL CIRCLE (or SC for short), which is just like CIRCLE except only three days long. We'd like to calculate $\text{urch}_{SC}(\text{RED}(\tau_1)|\text{RED}(\tau_0))$—the chance of the particle's being red at τ_0 conditional on its being red the day prior. To do so, partition τ_0 into any three non-trivial intervals $\{\tau_0^0, \tau_0^1, \tau_0^2\}$ with $\tau_0^0 < \tau_0^1 < \tau_0^2$—see fig. 1.5. Note that $\tau_0^0 \cup \tau_0^2$ is a thick boundary of τ_0^1, and hence, given Strong Boundary Markov, strongly screens it off from τ_1.

Now, whether the spacetime is a cycle partially depends on τ_0^1's internal geometry. In particular, if τ_0^1 has a "hole"—i.e., if it's empty or its ends otherwise don't connect—the

(iii) all charts in $\mathcal{A}^{\Omega\cup X}$ satisfy the usual smoothness condition (i.e., all transition maps are smooth),

(iv) $g^{\Omega\cup X}$ is an extension of g to $\Omega \cup X$ smooth relative to $\mathcal{A}^{\Omega\cup X}$.

The non-uniqueness of $M + X$ is reflected in (i), specifically in the choice of a topology on $\Omega \cup X$ and the charts (if any) whose domains overlap both Ω and X—intuitively, these two choices determine *how* X is added to M. Extending the deletion operation (fn. 38) to $M + X$ in the obvious way, we have, by conditions (ii) and (iv), $(M + X)\backslash X = ((\Omega \cup X)\backslash X, (\mathcal{A}^{\Omega\cup X})|_{(\Omega\cup X)\backslash X}, g^{\Omega\cup X}|_{(\Omega\cup X)\backslash X}) = (\Omega, \mathcal{A}, g) = M$, as desired. Finally, if $\mathcal{M}$ is a *spacetime* represented by M, *a result of adding X to $\mathcal{M}$* is any mereological fusion of $\mathcal{M}$ with a spacetime region such that the result is represented (under the same representation relation) by some $M + X$.

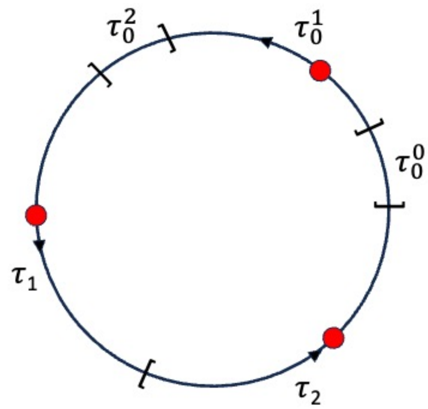


Figure 1.5: A sketch of SMALL CIRCLE. Each τ_i is a half-open interval (closed toward the past, open toward the future).

spacetime isn't a cycle. For concreteness, consider the possibility where τ_0^1 is empty. The resulting spacetime is SMALL LINE (or SL for short), as follows:[40]

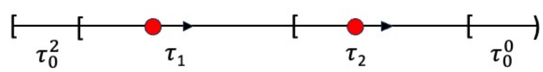


Figure 1.6: A sketch of SMALL LINE.

Strong Boundary Markov now lets us express the chances on SMALL CIRCLE in terms of the chances on SMALL LINE. But given Acyclic Chance Invariance, we already know the chances on SMALL LINE: they are based on the same transition chances as those on LINE (cf. Section 1.5). And now we are done: we've derived the cyclic chances from the acyclic chances.

Let's do this slowly. Let $[\text{RED}_i/\text{RED}_m/\text{RED}_f]$ be the property of containing a red particle whose clock initially reads $[0:00/t_0^1/t_0^2]$ and grows in proportion to the time passed—where t_0^1 is τ_0^1 's starting time and t_0^2 is τ_0^2's starting time. According to Acyclic Chance Invariance, the transition chances in SMALL LINE directly follow from those in

[40]Mathematically, this is a *manifold with boundary.*

LINE:[41]

$$\begin{aligned}
\mathrm{urch}_{SL}(\mathrm{RED}(\tau_1)|\mathrm{RED}_f(\tau_0^2)) &= 0.8,\\
\mathrm{urch}_{SL}(\mathrm{GREEN}(\tau_1)|\mathrm{RED}_f(\tau_0^2)) &= 0.2,\\
&\dots\\
\mathrm{urch}_{SL}(\mathrm{RED}_i(\tau_0^0)|\mathrm{RED}(\tau_2)) &= 0.8,\\
\mathrm{urch}_{SL}(\mathrm{GREEN}_i(\tau_0^0)|\mathrm{RED}(\tau_2)) &= 0.2,\\
&\text{etc.}
\end{aligned} \tag{1.12}$$

Strong Boundary Markov relates the urchance at SMALL CIRCLE to the urchance at SMALL LINE, as follows:

$$\begin{aligned}
\mathrm{urch}_{SC}(\mathrm{RED}(\tau_1)|\mathrm{RED}(\tau_0)) &= \mathrm{urch}_{SC}(\mathrm{RED}(\tau_1)|\mathrm{RED}_i(\tau_0^0) \wedge \mathrm{RED}_f(\tau_0^2) \wedge \mathrm{RED}_m(\tau_0^1))\\
&= \mathrm{urch}_{SL}(\mathrm{RED}(\tau_1)|\mathrm{RED}_i(\tau_0^0) \wedge \mathrm{RED}_f(\tau_0^2)).
\end{aligned} \tag{1.13}$$

where the first line follows because, given a complete description of SMALL CIRCLE's geometry, $\mathrm{RED}(\tau_0)$ is logically equivalent to $\mathrm{RED}_i(\tau_0^0) \wedge \mathrm{RED}_f(\tau_0^2) \wedge \mathrm{RED}_m(\tau_0^1)$, and the second line follows by Strong Boundary Markov.

Eq. 1.13 gives the desired connection between cyclic and acyclic chances. We can easily derive its right-hand side, $\mathrm{urch}_{SL}(A|\mathrm{RED}_i(\tau_0^0) \wedge \mathrm{RED}_f(\tau_0^2))$, from eqs. 1.12; the result is $64/65 \approx 0.98$.[42] So, by eq. 1.13,

$$\mathrm{urch}_{SC}(\mathrm{RED}(\tau_1)|\mathrm{RED}(\tau_0)) = \frac{64}{65} \approx 0.98. \tag{1.15}$$

[41] Using the official definition of Acyclic Chance Invariance, the relevant isometries are obvious: for example, for the first two lines in eqs. 1.12, any isometry from $J^+(\tau_0^2) \cap J^-(\tau_1) = \tau_0^2 \cup \tau_1$ into a subset of LINE will do; *mutatis mutandis* for the other lines.

[42] *Proof*: Using the multiplicative axiom twice,

$$\begin{aligned}
&\mathrm{urch}_{SL}(\mathrm{RED}(\tau_1)|\mathrm{RED}_i(\tau_0^0) \wedge \mathrm{RED}_f(\tau_0^2)) \cdot \mathrm{urch}_{SL}(\mathrm{RED}_i(\tau_0^0)|\mathrm{RED}_f(\tau_0^2)) =\\
&\quad = \mathrm{urch}_{SL}(\mathrm{RED}(\tau_1) \wedge \mathrm{RED}_i(\tau_0^0)|\mathrm{RED}_f(\tau_0^2))\\
&\quad = \mathrm{urch}_{SL}(\mathrm{RED}_i(\tau_0^0)|\mathrm{RED}(\tau_1) \wedge \mathrm{RED}_f(\tau_0^2)) \cdot \mathrm{urch}_{SL}(\mathrm{RED}(\tau_1)|\mathrm{RED}_f(\tau_0^2)).
\end{aligned}$$

This is the cyclic transition chance we were looking for. *Contra* Chance Invariance, the chance for the particle to remain red is *higher* in SMALL CIRCLE than the usual 0.8.

We've derived this result from entirely general principles about chance. It's also worth noting that it makes sense pretheoretically. *Contra* Chance Invariance, we should really have expected that

$$\text{urch}_{SC}(\text{RED}(\tau_1)|\text{RED}(\tau_0)) > 0.8. \tag{1.16}$$

Provided that $\text{urch}_{SL}(\text{RED}_i(\tau_0^0)|\text{RED}_f(\tau_0^2)) > 0$—which we'll show below—we can rewrite this:

$$\begin{aligned}\text{urch}_{SL}&(\text{RED}(\tau_1)|\text{RED}_i(\tau_0^0) \wedge \text{RED}_f(\tau_0^2)) = \\ &= \frac{\text{urch}_{SL}(\text{RED}_i(\tau_0^0)|\text{RED}(\tau_1) \wedge \text{RED}_f(\tau_0^2)) \cdot \text{urch}_{SL}(\text{RED}(\tau_1)|\text{RED}_f(\tau_0^2))}{\text{urch}_{SL}(\text{RED}_i(\tau_0^0)|\text{RED}_f(\tau_0^2))}.\end{aligned} \tag{1.14}$$

Let's evaluate each part of the quotient separately. Note that, since SMALL LINE is acyclic, Parental Markov is in good standing there. The first factor in the numerator:

$$\begin{aligned}\text{urch}_{SL}&(\text{RED}_i(\tau_0^0)|\text{RED}(\tau_1) \wedge \text{RED}_f(\tau_0^2)) = \\ &= \text{urch}_{SL}(\text{RED}_i(\tau_0^0)|\text{RED}(\tau_1)) = \\ &= \text{urch}_{SL}(\text{RED}_i(\tau_0^0)|\text{RED}(\tau_2) \wedge \text{RED}(\tau_1)) \cdot \text{urch}_{SL}(\text{RED}(\tau_2)|\text{RED}(\tau_1)) = \\ &= \text{urch}_{SL}(\text{RED}_i(\tau_0^0)|\text{RED}(\tau_2)) \cdot \text{urch}_{SL}(\text{RED}(\tau_2)|\text{RED}(\tau_1)) = \\ &= 0.8^2,\end{aligned}$$

where the first and third equalities follow from Parental Markov for SMALL LINE, the second equality follows from the probability laws plus the fact that only $\text{RED}(\tau_2)$ is nomically compatible with $\text{RED}(\tau_1) \wedge \text{RED}_i(\tau_0^0)$, and the final equality follows from the transition chances (eqs. 1.12) for SMALL LINE. The numerator's second factor follows immediately from the transition chances:

$$\text{urch}_{SL}(\text{RED}(\tau_1)|\text{RED}_f(\tau_0^2)) = 0.8.$$

Finally, to calculate the denominator, observe that

$$\begin{aligned}\text{urch}_{SL}&(\text{RED}_i(\tau_0^0)|\text{RED}_f(\tau_0^2)) = \\ &= \sum_{\pi \in \{\text{RED,GREEN}\}} \text{urch}_{SL}(\text{RED}_i(\tau_0^0)|\pi(\tau_1) \wedge \text{RED}_f(\tau_0^2)) \cdot \text{urch}_{SL}(\pi(\tau_1)|\text{RED}_f(\tau_0^2)),\end{aligned}$$

where the second line follows because LINE's dynamics disallow immediate RED-to-BLUE transitions. For $\pi = \text{RED}$, the right-hand side's summand is just the numerator, whose value we've just calculated: 0.8^3. (As promised, this also proves that the denominator is positive.) For $\pi = \text{GREEN}$, we perform exactly analogous calculations. Noting that $\text{BLUE}(\tau_2)$ is the only option for τ_2 compatible with $\text{GREEN}(\tau_1)$ and $\text{RED}_i(\tau_0^0)$, the right-hand side's summand then comes out to $0.2^2 \cdot 0.2 = 0.2^3$. Plugging everything into eq. 1.14,

$$\text{urch}_{SL}(\text{RED}(\tau_1)|\text{RED}_i(\tau_0^0) \wedge \text{RED}_f(\tau_0^2)) = \frac{0.8^3}{0.8^3 + 0.2^3} = \frac{64}{65} \approx 0.98. \ \blacksquare$$

For, intuitively, $\text{RED}(\tau_0)$ *doubly* supports $\text{RED}(\tau_1)$: the latter is not only a likely *effect* of $\text{RED}(\tau_0)$, but also a likely *cause*. Going from $\text{RED}(\tau_1)$ to $\text{RED}(\tau_0)$ involves transitioning from $\text{RED}(\tau_1)$ to $\text{RED}(\tau_2)$ to $\text{RED}(\tau_0)$—two "likely" transitions. By contrast, going from $\text{GREEN}(\tau_1)$ to $\text{RED}(\tau_0)$ involves two "unlikely" transitions: $\text{GREEN}(\tau_1)$ to $\text{BLUE}(\tau_2)$ to $\text{RED}(\tau_0)$. So, pretheoretically, $\text{RED}(\tau_0)$ should favor $\text{RED}(\tau_1)$ both because it preferentially causes it *and* because it's preferentially caused *by* it. Eq. 1.13 captures this "double support" intuition. It says that the result of conditioning the urchance on today's particle being red equals (as far as events outside of τ_0 are concerned) the result of conditioning the acyclic urchance function on today's particle *and* the particle three days from now being red—thus *doubly supporting* tomorrow's redness.

We can also understand eq. 1.13 in terms of "unraveling" the cyclic spacetime. Consider SL^+, an extension of SMALL LINE obtained by adding[43] a copy of $\tau_0^0 \cup \tau_0^1$ to the beginning of τ_0^2 and a copy of $\tau_0^1 \cup \tau_0^2 \cup \tau_1 \cup \tau_2$ to the end of τ_0^0 in the obvious ways. The result is a concatenation of two copies of SMALL LINE. Denoting parts of the later copy with $+$ superscripts, we can sketch SL^+ as follows (with the original SL highlighted):

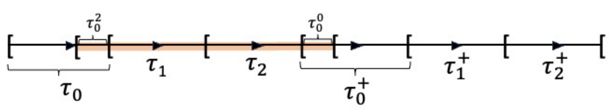


Figure 1.7: A sketch of SL^+.

Now, SL^+ is a result of "unraveling" the SMALL CIRCLE twice—more formally, it is a *double cover* of SMALL CIRCLE. Since $\tau_0^2 \cup \tau_0^0$ is a thick boundary of $\tau_1 \cup \tau_2$ in SL^+, by Strong Boundary Markov, it follows that[44]

$$\text{urch}_{SL}(\text{RED}(\tau_1)|\text{RED}_f(\tau_0^2) \wedge \text{RED}_i(\tau_0^0)) = \text{urch}_{SL^+}(\text{RED}(\tau_1)|\text{RED}(\tau_0) \wedge \text{RED}(\tau_0^+)).$$

[43] See fn. 39 for a rigorous definition of "adding".

[44] To see this rigorously: let RED^n be the property of being a concatenation of n 24h intervals the first of which satisfies RED. Given complete geometric descriptions of SL or of SL^+ (each entailing that τ_2 is a 24h interval following τ_1), $\text{RED}(\tau_1)$ is equivalent to $\text{RED}^2(\tau_1 \cup \tau_2)$. Similarly, given a complete geometric

So, by eq. 1.13,

$$\text{urch}_{SC}(\text{RED}(\tau_1)|\text{RED}(\tau_0)) = \text{urch}_{SL^+}(\text{RED}(\tau_1)|\text{RED}(\tau_0) \wedge \text{RED}(\tau_0^+)).$$

Thus, our account implies that the transition chances for $\text{RED}(\tau_0)$ on CIRCLE can also be obtained by unraveling the spacetime twice, into a double cover, and then conditioning on the proposition that both copies of τ_0 in the cover are red.

Our theory of chances on loops ticks all three boxes: since they're fully derived from the acyclic chances, the conditional chances in SMALL CIRCLE are dynamically scrutable. As eq. 1.15 shows, the approach also avoids trivialization. And if (as I claim) locality entails Strong Boundary Markov, the conjunction of Strong Boundary Markov and Acyclic Chance Invariance is (*ipso facto*) consistent with any local dynamics whose chance prescriptions are invariant across acyclic worlds—which arguably includes any *plausible* local dynamics.

There is a hidden fourth benefit to our account. Sometimes it salvages Chance Invariance in the short term, as it were, *asymptotically*. In the original CIRCLE case, with a period of 100 billion years, the short-term cyclic transition chances are essentially indistinguishable from the acyclic transition chances (the difference between $\text{urch}_C(\text{GREEN}(\tau_1)|\text{RED}(\tau_0))$ and 0.2 is smaller than $0.8^{10^{12}}$). More generally, unless the chance to switch color in the *acyclic* case is *extremely* close to 0 or 1, the short-term cyclic transition chances converge extremely quickly to the acyclic transition chances as the loop length increases. This makes sense intuitively: knowing the particle's color in the far future provides little evidence

description of SL^+, $\text{RED}(\tau_0^+)$ is equivalent to $\text{RED}^3(\tau_0^+ \cup \tau_1^+ \cup \tau_2^+)$. We thus have

$$\begin{aligned}\text{urch}_{SL^+}(\text{RED}(\tau_1)|\text{RED}(\tau_0) \wedge \text{RED}(\tau_0^+)) &= \text{urch}_{SL^+}(\text{RED}^2(\tau_1 \cup \tau_2)|\text{RED}(\tau_0) \wedge \text{RED}^3(\tau_0^+ \cup \tau_1^+ \cup \tau_2^+)) \\ &\overset{SBM}{=} \text{urch}_{SL}(\text{RED}^2(\tau_1 \cup \tau_2)|\text{RED}_f(\tau_0^2) \wedge \text{RED}_i(\tau_0^0)) \\ &= \text{urch}_{SL}(\text{RED}\tau_1|\text{RED}_f(\tau_0^2) \wedge \text{RED}_i(\tau_0^0)),\end{aligned}$$

where the second line follows by Strong Boundary Markov because RED^2 and RED^3 are (non-maximal) intrinsic properties of $\tau_1 \cup \tau_2$ and $\tau_0^+ \cup \tau_1^+ \cup \tau_2^+$, respectively, $\text{RED}(\tau_0)$ is equivalent to a conjunction $\phi(\tau_0 \backslash \tau_0^2) \wedge \text{RED}_f(\tau_0^2)$ with ϕ intrinsic to $\tau_0 \backslash \tau_0^2$, and $\text{RED}^3(\tau_0^+ \cup \tau_1^+ \cup \tau_2^+)$ is equivalent to a conjunction $\psi((\tau_0^+ \cup \tau_1^+ \cup \tau_2^+) \backslash \tau_0^0) \wedge \text{RED}_i(\tau_0^0)$ with ψ intrinsic to $\tau_0^+ \cup \tau_1^+ \cup \tau_2^+) \backslash \tau_0^0$. ■

about the near-term colors.[45] More generally, in a cyclic spacetime, our approach asymptotically ensures short-term Chance Invariance whenever, according to the *acyclic* dynamics, far-future states are increasingly probabilistically independent of near-future states.[46]

1.9 Saving Grandpa

Let's apply our theory to the stochastic "grandfather paradox" from the introduction. For concreteness, let the spacetime $\mathcal{M}$ be Minkowskian except for one topological quirk, a *wormhole*. We can represent the wormhole with two duplicate (three-dimensional) bounded space-like surfaces, w_1 and w_2: every future-directed causal curve intersecting w_1 immediately exits at w_2 (without intersecting w_2), from where it continues future-ward, and every future-directed causal curve intersecting w_2 immediately exits at w_1 (without intersecting w_1), from where it continues future-ward.[47] See figure 1.8, which includes an

[45] More formally, where $q \in [0,1)$ is the acyclic chance for a particle to switch color during the next transition and l is the loop length, we obtain (if $\lfloor l/3 \rfloor = 0$, the numerator sum is set to 0)

$$\mathrm{urch}_L(\mathrm{GREEN}(\tau_1)|\mathrm{RED}(\tau_0) \wedge \mathrm{RED}(\tau_{-1})) = \frac{\sum_{n=1}^{\lfloor l/3 \rfloor} (1-q)^{l-3n} q^{3n} \binom{l-1}{3n-1}}{\sum_{n=0}^{\lfloor l/3 \rfloor} (1-q)^{l-3n} q^{3n} \binom{l}{3n}} \tag{1.17}$$

For increasing l, this converges to q extremely quickly.

[46] To see this: suppose we're given a cyclic spacetime $\mathcal{C}$. Let τ be a region in $\mathcal{C}$ bounded by two time-slices. As before, we partition τ into three parts τ^0, τ^1, τ^2. (In higher-dimensional cyclic spacetimes, $\tau, \tau^0, \tau^1, \tau^2$ are all hypercuboids.) Let $\mathcal{L} := \mathcal{C} \backslash \tau^1$. The longer the return time in $\mathcal{C}$, the greater the forward distance from τ^2 to τ^0 in $\mathcal{L}$. Let τ_+ be a time in τ^2's near-term future. Given a long return time and approximate probabilistic independence of far-future from near-future states, we have

$$\mathrm{urch}_{\mathcal{L}}(Q(\tau_+)|P(\tau^2) \wedge P(\tau^0)) \approx \mathrm{urch}_{\mathcal{L}}(Q(\tau_+)|P(\tau^2))$$

for any possible maximal intrinsic property P of τ—where "$P(\tau^i)$" denotes the strongest proposition entirely about τ^2 entailed by $P(\tau)$—and any intrinsic property Q. Hence, by Strong Boundary Markov,

$$\begin{aligned} \mathrm{urch}_{\mathcal{C}}(Q(\tau_+)|P(\tau)) &= \mathrm{urch}_{\mathcal{C}}(Q(\tau_+)|P(\tau^2) \wedge P(\tau^1) \wedge P(\tau^0)) \\ &\overset{SBM}{=} \mathrm{urch}_{\mathcal{L}}(Q(\tau_+)|P(\tau^2) \wedge P(\tau^0)) \\ &\approx \mathrm{urch}_{\mathcal{L}}(Q(\tau_+)|P(\tau^2)). \end{aligned}$$

But $\mathrm{urch}_{\mathcal{L}}(Q(\tau_+)|P(\tau^2))$ is just the ordinary acyclic transition chance from $P(\tau)$ to $Q(\tau_+)$. ■

[47] To keep derivatives everywhere well-defined, w_1 and w_2's (two-dimensional) boundaries are deleted.

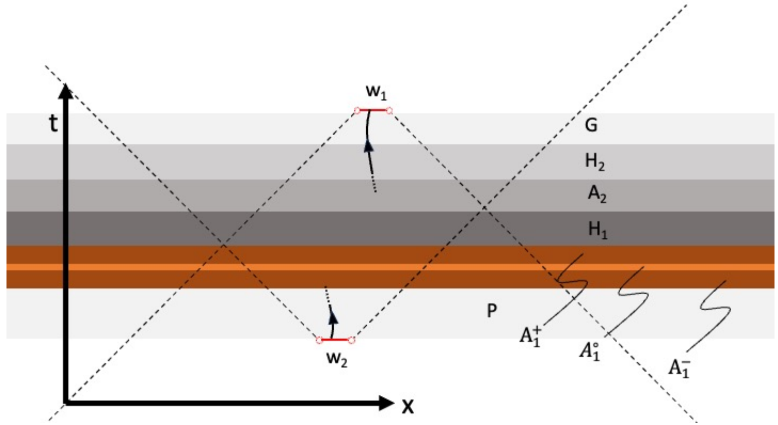


Figure 1.8: A sketch of $\mathcal{M}$, the spacetime in the grandfather case.

example trajectory through the wormhole.[48]

Figure 1.8 also indicates six other spacetime regions: P is the region where the poisoning occurs; A_1 (A_2) is the region where you (would) administer the first (second) antidote, with A_1 additionally tri-partitioned into A_1^-, A_1°, and A_1^+; H_1 (H_2) is the region where the first (second) antidote takes (would take) effect; G is the region where grandpa has children, one of whom bears you and your sibling, who then travel through the wormhole at the end of G. (The regions aren't drawn to scale.)

To keep things simple, consider only finitely many possible maximal intrinsic states for each region.[49] Associating each state with a natural number, we can write property assignments to a region R conveniently as $R = n$. In the following all properties are to be read as intrinsic to the respective region.[50] Let a bracketed expression $(\phi)_i$ indicate that "ϕ" is to be added for the values i.

[48]The choice of w_1 and w_2 is non-unique: any other pair of duplicate bounded space-like surfaces with the same boundaries as w_1 and w_2 generates the same wormhole. Arguably, "wormhole" is therefore most naturally identified with the union of w_1's and w_2's domains of dependence. But to keep things simple, I'll keep framing things in terms of w_1 and w_2 specifically.

[49]We *could* consider infinitely many possible states for each region, and then condition on their disjunctions, using Regional Conglomerability (see Appendix C).

[50]So, for example: "infant" means something like *having the physiology typical of a neonate*, rather than *having been born some time ago*; "antidote about to be administered" means something like *your standing ready with the antidote, with the intention to administer it*, etc; and "antidote just administered" means *the bottle's being lifted from the infant's mouth, while the antidote is entering the infant's body*, etc. For readability I'll use the shorter expressions.

- $P = 0, 1$: infant (not)$_0$ poisoned in P
- $A_1^- = 0, 1$: in A_1^-, infant (healthy)$_0$ (sick)$_1$ and antidote 1 (not)$_0$ about to be administered
- $A_1^\circ = 0, 1$: in A_1°, infant (healthy)$_0$ (sick)$_1$and antidote 1 (not)$_0$ being administered
- $A_1^+ = 0, 1$: in A_1^+, infant (healthy)$_0$ (sick)$_1$ and antidote 1 (not)$_0$ just administered
- $A_1 = 0, 1 : (A_1^- = A_1^\circ = A_1^+ = 0)_0\ (A_1^- = A_1^\circ = A_1^+ = 1)_1$
- $H_1 = 0, 1, 2$: antidote 1 (not)$_{0,1}$ taking effect on (healthy)$_0$ (sick)$_{1,2}$ infant in H_1[51]
- $A_2 = 0, 1$: in A_2, infant (healthy)$_0$ (sick)$_1$ and antidote 2 (not)$_0$ administered
- $H_2 = 0, 1, 2$: antidote 2 (not)$_{0,1}$ taking effect on (healthy)$_0$ (sick)$_{1,2}$ infant in H_2
- $G = 0, 1$: in G, infant (not)$_0$ alive, and (not)$_0$ eventually growing up to have two grandchildren

One salient interpretation of "chance, upon administration, of the first antidote's working" is $\text{urch}_{\mathcal{M}}(H_1 = 2|A_1 = 1)$—the urchance, in $\mathcal{M}$, of the first antidote's working conditional on administration in A_1. To calculate this, first note that $A_1^+ \cup A_1^-$ is a thick boundary of A_1° (cf. fig. 1.8). Thus consider $\mathcal{M}' := \mathcal{M} \backslash A_1^\circ$, the result of deleting A_1° from $\mathcal{M}$, sketched in figure 1.9. Since, given a complete geometric description of $\mathcal{M}$, $A_1 = 1$ is necessarily equivalent to $A_1^- = A_1^\circ = A_1^+ = 1$, we obtain, by Strong Boundary Markov:

$$\begin{aligned} \text{urch}_{\mathcal{M}}(H_1 = 2|A_1 = 1) &= \text{urch}_{\mathcal{M}}(H_1 = 2|A_1^- = A_1^\circ = A_1^+ = 1) \\ &\overset{SBM}{=} \text{urch}_{\mathcal{M}'}(H_1 = 2|A_1^- = A_1^+ = 1). \end{aligned} \tag{1.18}$$

Since A_1° intersects all closed causal curves in $\mathcal{M}$, $\mathcal{M}'$ contains no closed causal curves at all. So we can derive the right-hand side of eq. 1.18 from the acyclic chances.

[51] An antidote can, of course, only take effect on a sick person.

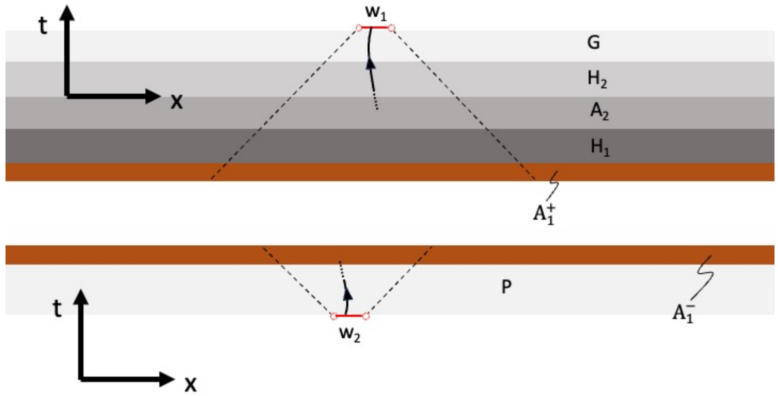


Figure 1.9: The result of deleting A_1° from $\mathcal{M}$.

To do this, let's assume for simplicity that the antidote's actions are the only indeterministic processes in $\mathcal{M}'$. In worlds without closed causal curves, there's a 50% chance that a given antidote works. Hence:

$$\text{urch}_{\mathcal{M}'}(H_1 = i | A_1^+ = 1) = \text{urch}_{\mathcal{M}'}(H_2 = i | A_2 = 1) = 0.5. \tag{1.19}$$

To pin down the rest of the acyclic dynamics, note that $\mathcal{M}'$ contains two disconnected infant spacetime worms: one starting in P (at birth) and ending at A_1^-'s future border, the other starting at A_1^+'s past border and ending somewhere in G (as the infant grows into an adolescent). These spacetime worms are disconnected because grandfather himself never travels through the wormhole. Call the spacetime worm from P to A_1^-'s future border *young infant*, and the spacetime worm from A_1^+'s past border to G *older infant*. For any propositions A, B, let $A \Rightarrow B$ denote that, deterministically, if A, then B.[52] (For convenience, I'll often suppress explicit mention of "deterministically" in the following, writing "...iff..." instead of "deterministically, ...iff...".) $A \Leftrightarrow B$ denotes $A \Rightarrow B \wedge B \Rightarrow A$.

(a) $H_1 = 0 \Leftrightarrow A_1^+ = 0$: older infant is healthy coming into H_1 iff he is healthy in A_1^+

[52]In our urchance formalism, ⌜deterministically, A given B⌝ is equivalent to ⌜$\text{urch}(A|B \wedge Q) \equiv 1$ for all propositions Q⌝.

(b) $A_2 = 1 \Leftrightarrow H_1 = 1$: an antidote is administered in A_2 iff older infant is still sick at the end of H_1

(c) $H_2 = 0 \Leftrightarrow A_2 = 0$: older infant is healthy coming into H_2 iff he is already healthy in A_2

(d) $G = 1 \Leftrightarrow H_2 = 0 \vee H_2 = 2$: older infant grows up to become a grandfather in G iff either he is healthy going into H_2, or he is sick going into H_2 but antidote 2 works

(e) $P = 1 \Leftrightarrow G = 1$: young infant is poisoned in P iff older infant in G grows up to be a grandfather[53]

(f) $A_1^- = 1 \Leftrightarrow P = 1$: antidote about to be administered in A_1^- iff young infant is poisoned in P

By eqs. (d), (e) and (f) we have $A_1^- = 1 \Leftrightarrow H_2 = 0 \vee H_2 = 2$—an antidote is about to be administered to the young infant in A_1^- iff the older infant survives in H_2 (either by being already healthy at the start of H_2 or by being healed in H_2). Meanwhile, from (a), (b), and (c) we have $A_1^+ = 1 \wedge H_2 = 0 \Leftrightarrow A_1^+ = 1 \wedge H_1 = 2$—if an antidote has entered the infant's stomach in A_1^+, then he is healthy at the start of H_2 iff the first antidote works in H_1. Those two equivalences jointly entail the following:

$$A_1^- = A_1^+ = 1 \Leftrightarrow (H_1 = 2 \vee H_2 = 2) \wedge A_1^+ = 1,$$

i.e., if an antidote has just been administered to the older infant in A_1^+, then [an antidote is *about* to be administered to the young infant in A_1^- iff at least one antidote works]. Plugging this into the right-hand side of eq. 1.18, we obtain the following:

$$\mathrm{urch}_{\mathcal{M}}(H_1 = 2 | A_1 = 1) = \mathrm{urch}_{\mathcal{M}'}(H_1 = 2 | A_1^+ = 1 \wedge (H_1 = 2 \vee H_2 = 2)). \tag{1.20}$$

[53]This holds because one of the older infant's grandchildren poisons the young infant and, by assumption, nobody else possibly does.

That is, the probability, in $\mathcal{M}$, that the first antidote is effective conditional on its administration equals the probability, *in* $\mathcal{M}'$, that the first antidote is effective conditional on its administration *and* at least one of the two antidotes' working.

It's intuitively clear that the latter probability is greater than $1/2$. That's because the guarantee that one of the two antidotes works raises the chance of each one's working—it excludes the possibility that both fail. More precisely, we find that[54]

$$\mathrm{urch}_{\mathcal{M}'}(H_1 = 2|A_1^+ = 1 \wedge (H_1 = 2 \vee H_2 = 2)) = \frac{1/2}{3/4} = \frac{2}{3}. \tag{1.22}$$

So, by eq. 1.20,

$$\mathrm{urch}_{\mathcal{M}}(H_1 = 2|A_1 = 1) = \frac{2}{3} > \frac{1}{2}.$$

Equally, the fact that one of the two antidotes works raises the chance of the second one's working: an analogous calculation yields $\mathrm{urch}_{\mathcal{M}}(H_2 = 2|A_1 = 1) = 1/3$, higher than the usual $1/4$ (note that the second antidote is administered only if the first fails). We also obtain $\mathrm{urch}_{\mathcal{M}}(H_2 = 2|H_2 = 1) = 1$—given that antidote 1 fails, antidote 2 *must* work. In the presence of spacetime loops, chances differ in non-trivial but scrutable ways.

Our two scenarios provide us with a general recipe for deriving chances on loops.

General Recipe:

1. Given a spacetime $\mathcal{M}$ with closed causal curves, identify a region R intersecting all closed causal curves. Let Q_R a qualitative intrinsic state of R.

[54] *Proof:* By the multiplicative axiom,

$$\mathrm{urch}_{\mathcal{M}'}(H_1 = 2|A_1^+ = 1 \wedge (H_1 = 2 \vee H_2 = 2)) = \frac{\mathrm{urch}_{\mathcal{M}'}(H_1 = 2|A_1^+ = 1)}{\mathrm{urch}_{\mathcal{M}'}(H_1 = 2 \vee H_2 = 2|A_1^+ = 1)}, \tag{1.21}$$

provided the denominator isn't zero. Eq. 1.19 gives us $\mathrm{urch}_{\mathcal{M}'}(H_1 = 2|A_1^+ = 1) = 1/2$. Moreover, we obtain $\mathrm{urch}_{\mathcal{M}'}(H_2 = 2|A_1^+ = 1) = 1/4$ via the usual methods (the second antidote is administered only if the first one fails, and, if it is administered, has a $1/2$ chance of healing the poisoning—all of this is encoded in eq. 1.19, (a), (b), and (c)). Finally, $H_1 = 2$ and $H_2 = 2$ are mutually exclusive (because, again, the second antidote is administered only if the first antidote fails). Plugging this all into eq. 1.21, we obtain

$$\mathrm{urch}_{\mathcal{M}'}(H_1 = 2|A_1^+ = 1 \wedge (H_1 = 2 \vee H_2 = 2)) = (1/2)/(3/4) = 2/3. \blacksquare$$

2. Identify a thick boundary B of R. Let Q_B be its complete intrinsic state.

3. Using Acyclic Chance Invariance, calculate, for any propositions X and Y entirely about $\mathcal{M} \backslash R$,

$$\text{urch}_{\mathcal{M} \backslash R}(X|Y \wedge Q_B(B)).$$

4. By Strong Boundary Markov, this equals

$$\text{urch}_{\mathcal{M}}(X|Y \wedge Q_R(R) \wedge Q_B(B)).$$

This way, Strong Boundary Markov and Acyclic Chance Invariance determine certain *conditional* chances. A natural follow-up question is whether those conditional chances determine other chances too. The answer is a resounding Yes. In cyclic spacetimes, they determine precise *unconditional* chances for virtually all propositions: we generically obtain a full "probability map of the universe".[55] Meanwhile, where a wormhole is embedded into a larger spacetime, we obtain precise chance distributions over states of the loop region—including over what emerges from the wormhole—conditional on the state of the world *prior* to the loop region. The next section explains.

1.10 Marginal Chances

Several people have suggested to me in conversation that, generically, there are no well-defined chances of what comes out of a future wormhole, conditional only on the state of the world prior to it. In the same vein, one might think that there is no privileged way of assigning marginal (i.e., unconditional) chances over the states of a cyclic universe.

I once believed both things too. Certainly in *acyclic* worlds, like Minkowski spacetime, the transition chances alone generically *don't* fix marginal chance distributions over the

[55]Cf. Loewer (2019). Of course, in contrast to the Mentaculus, our probability maps for cyclic spacetimes don't require anything like a Past Hypothesis—everything is fixed by the *dynamics* alone.

states of the world; additionally, one requires a marginal chance distribution over the universe's possible initial conditions.[56]

But things are different with cyclic worlds. A specification of all transition chances generically fixes even a *marginal* chance distribution over the possible states of a cyclic world. Intuitively, the reason for this is that cyclic worlds have one "extra" transition compared to their linear counterparts: their "ends" also connect. This extra transition generically imposes additional constraints on the marginal distribution over the loop, enough to fix it uniquely.[57]

To see this, consider a generalized version of CIRCLE, where the loop is n days round-trip and there are k possible particle colors, represented by natural numbers. (In the original CIRCLE case, $n \approx 5.5 \cdot 10^{12}$, and $k = 3$.) The probability axioms imply that, for every $i = 1, ..., n$ and $j = 1, ..., k$ (where we identify a 0 in an index with n):[58]

$$\text{urch}_C(\tau_i = j) = \sum_{l=1}^{k} \text{urch}_C(\tau_i = j | \tau_{i-1} = l) \cdot \text{urch}_C(\tau_{i-1} = l). \tag{1.23}$$

[56]Some derive such a distribution from statistical mechanical considerations (e.g. Albert, 2000; Loewer, 2019). This abandons the idea that the urchance function is determined by the dynamical laws alone. In any case, the considerations cannot apply to closed causal curves, which lack unidirectional entropic arrows.

[57] Mellor (1995, Sec. 17.3) once tried to leverage a mathematically similar fact into an argument *against* the possibility of spacetime loops, by claiming that it's impossible for transition chances to constrain marginal chances (or, in his framework, marginal *limiting frequencies*) in this way. See Berkovitz (2001, pp.14-5) for a cogent rebuttal.

Berkovitz (2001, pp.15-20) then develops an alternative argument on Mellor's behalf, here presented in a slightly simplified form. Consider a large set of causal loops, each intrinsically like CIRCLE, connected by space-like lines so as to be world-mates. As a purely mathematical fact, the *transition* frequencies within this set (generically) depend on each other. For example, in the case of SSC—see Appendix 4—the frequencies of $\tau_1 = 1$ given $\tau_0 = 0$, of $\tau_1 = 1$ given $\tau_0 = 1$, and of $\tau_0 = 0$ given $\tau_1 = 0$, together logically fix the frequency of $\tau_0 = 0$ given $\tau_1 = 1$. But, Berkovitz says, not so for the transition *chances* (or, following Mellor, the "chances of effects with and without their causes"): those *are* logically independent of each other. So, in the world with many causal loops, the cyclic transition chances will generically differ radically from the cyclic limiting frequencies; in particular, they plausibly won't equal them with chance 1, violating a "law of large numbers" constraint for chance. Berkovitz responds that this violation is acceptable, because the relevant reference classes on causal loops are "biased". But in fact the response is much simpler: cyclic transition chances *aren't* logically independent of each other. What's logically independent—and what Berkovitz must therefore have in mind with his "chances of effects with and without their causes"—are the *acyclic* transition chances. But *of course* nobody expects *them* to satisfy a law of large number constraint with respect to the *cyclic* transitions frequencies. So the argument's conclusion is entirely unproblematic—indeed, desired.

[58]Marginal urchances are defined in the obvious way: $\text{urch}_C(A) := \text{urch}_C(A|\top)$, where $\top$ is any tautology.

From the acyclic dynamics, our account obtains all cyclic transition probabilities, $\text{urch}_C(\tau_i = j|\tau_{i-1} = l)$. Hence eq. 1.23 yields $n \cdot k$ equations in $n \cdot k$ unknowns. Since we also know that, for every $i = 1, ..., n$,

$$\sum_{j=1}^{k} \text{urch}_C(\tau_i = j) = 1, \tag{1.24}$$

we can eliminate n equations and n unknowns from this system, leaving us with $n \cdot (k-1)$ equations in $n \cdot (k-1)$ unknowns. In Appendix D, we see that these equations generically are linearly independent and have a unique solution.

Solving the system (see eq. D.1 in the Appendix—I'll skip the calculation here) for SMALL CIRCLE (i.e., $n = k = 3$, and every color i has only itself and $i+1$ as a possible successor, with $1/65 \approx 0.015$ chance to transition to $i+1$), we get the following result, for all $i = 1, 2, 3$:

$$\text{urch}_{SC}(\text{RED}(\tau_i)) = \text{urch}_{SC}(\text{GREEN}(\tau_i)) = \text{urch}_{SC}(\text{BLUE}(\tau_i)) = 1/3.$$

A sensible result, given the symmetry in transition probabilities between the colors: every color has, besides itself, a unique permissible successor, and each color has the same chance of switching to its respective successor. Breaking this symmetry in the transition probabilities also breaks the symmetry in the marginals. For example, if $\text{urch}_{SC}(\text{RED}(\tau_i)|\text{RED}(\tau_{i-1})) = 0.5$ for all $i = 1, 2, 3$ and the remaining transition chances are unchanged, we get $\text{urch}_{SC}(\text{RED}(\tau_i)) = 1/66 \approx 0.015$ and $\text{urch}_{SC}(\text{GREEN}(\tau_i)) = \text{urch}_{SC}(\text{BLUE}(\tau_i)) = 65/132 \approx 0.492$ for all $i = 1, 2, 3$. (The reader may verify this by plugging the given transition probabilities into eq. D.1 in Appendix D.) We can of course also break the symmetry between the *times*, i.e., impose time-dependent transition probabilities, which then also makes the marginal probabilities time-dependent.

Once we have marginal chance distributions over states of the loop, we can derive many other conditional chance distributions via the ratio formula. For example, consider a two-dimensional version of CIRCLE, i.e. a flat 2D spacetime rolled up along the time-like direction into a cylinder.[59] For any subset of the cylinder, our account determines the chance of any proposition conditional on any state of the subset, provided only that the latter has positive marginal chance.

I said that the linear equations are "generically" independent, because in special circumstances they aren't, allowing for multiple solutions (i.e., in our framework, a non-constant $\text{urch}_{\mathcal{M}}$). Roughly, this happens when there are *too many* deterministic transitions. For illustration, consider a variant of SMALL CIRCLE, where the particle is guaranteed to remain at its current color, i.e., for every $\text{COL} \in \{\text{RED}, \text{GREEN}, \text{BLUE}\}$,

$$\text{urch}_{SC}(\text{COL}(\tau_i)|\text{COL}(\tau_{i-1})) = 1.$$

Given these transition chances, the only constraint the system imposes on the marginals is $\text{urch}_{SC}(\text{COL}(\tau_{i-1})) = \text{urch}_{SC}(\text{COL}(\tau_i))$. Any probabilistically coherent[60] urchance function which satisfies this constraint is nomically allowed. (In Appendix D, I show this explicitly for the simplest non-trivial case, with $n = k = 2$. The case of SC, i.e. $n = k = 3$, is computationally more complex, but doesn't offer additional insight.)

So much for the case of a cyclic spacetime. The case of a spacetime with a wormhole is mathematically similar. Generically, our account yields well-defined transition chances on the region between the wormhole mouths, *conditional* on the state of the world prior to the wormhole. Once we have those transition chances, the remaining calculation is exactly the same. It follows that, conditional on the state of the world prior to the wormhole, there is (generically) a precise chance distribution over states of the loop region, including a precise chance distribution over what emerges from the wormhole. Only when enough transition probabilities are trivial—e.g. if the dynamics is deterministic—is there no such precise chance distribution.

[59] Mathematically, we can represent this by a two-dimensional, oriented, closed Lorentzian manifold.

[60] I.e., satisfying, for $i = 1, 2, 3$, $\sum_{\substack{\text{COL}\in \\ \{\text{RED},\text{GREEN},\text{BLUE}\}}} \text{urch}_{SC}(\text{COL}(\tau_i)) = 1$ (where addition is, as always, point-wise).

1.11 Conclusion

My theory of chances on loops consists of Strong Boundary Markov and Acyclic Chance Invariance. Strong Boundary Markov is an *a priori* plausible constraint on local laws, following from the idea that there shouldn't be a difference between *geometric* information about a region and other kinds of information about it—if the dynamics is local, both should be screened off by a thick boundary. Acyclic Chance Invariance, meanwhile, is a consistent weakening of an initially attractive, yet inconsistent, Chance Invariance principle. The weakening says that chances are invariant among *loop-free* worlds. Given the acyclic dynamics, these two general principles fix everything there is about chance in cyclic worlds.

Our theory satisfies all theoretical criteria we've set out. It avoids temporalism's triviality problem, and it avoids the inconsistency plaguing general Chance Invariance. Still, it manages to preserve the two next best things to Chance Invariance. The first is Acyclic Chance Invariance. The second is the idea that chances are "dynamically scrutable": while not identical to them, the cyclic chances should be *derivable from* the acyclic chances in a principled way. In our theory, Strong Boundary Markov and Acyclic Chance Invariance provide this principled connection. Finally, we also saw how under certain conditions—namely when, according to the acyclic chances, far-future events are increasingly probabilistically independent of near-future events—chances are *practically* invariant in the near-term.

While the essay's focus is objective chance, it naturally has implications for *rational credence*. Objective chance constrains credence via plausible deference principles.[61] For example, following the previous section, if you know the true (precise) transition chances governing a cyclic world, generically you should have precise prior credences over the world's possible states. Likewise, if you know the true chance laws, and are well-informed

[61]See fn. 12 for a deference principle for urchance.

about the *current* state of the world, you should, in general, have precise expectations about what will emerge from a future wormhole.

Programmatically, this essay supports flexible chance formalisms, by demonstrating how they solve problems eluding other approaches. On our urchance formalism, *all* propositions—not just temporal or causal histories—are eligible background propositions. Of course, some propositions will be more informative than others. But we've seen that even regions much smaller than entire temporal or causal history regions can be highly informative.[62] Some of these regions cover exactly the local environments of coin flips, of roulette wheels, or of decks of cards. The concept of a background proposition thus comes to subsume the concept of a *chance setup*. Indeed, conceptual economy suggests identifying the two: every background proposition is a chance setup, every chance setup a background proposition. With this identification, our framework then enshrines a view Popper (1959) embraced long ago: that chances are intimately tied, not to time or causation, but to chance setups ("*arrangements*", as he called them).

[62]That is, assuming we additionally supply information about the world's background geometry—something which the temporalist also has to do.

Chapter 2

Urchance Ideology

2.1 Introduction

Given the state of the world at one time, the laws of physics predict how things are at other times. If the laws are *deterministic*, the prediction singles out a unique possibility for each time. If the laws are *indeterministic*, generally multiple possibilities are permitted, and those possibilities tend to get probabilities assigned.

A paradigm case of a set of deterministic physical laws are Maxwell's laws of electrodynamics (with Lorentz's force law). A paradigm case of a set of indeterministic physical laws are the various GRW versions of quantum mechanics (Ghirardi *et al.*, 1986; Tumulka, 2006; Maudlin, 2019). These versions are all highly non-local laws;[1] but hypothetical examples of local stochastic theories are easy to construct too—for example, consider the result of supplementing Maxwell's laws with a (local) toy theory of stochastic decay.

The classification of laws into deterministic and indeterministic is a matter of their *diachronic* predictive profiles. But physical laws also routinely impose *synchronic* constraints:

[1] There are two aspects to the non-locality: the fundamental physical states are non-separable, and the probability distributions over future evolutions, generated from the states at a time, violate Parental Markov and Boundary Markov. (Relativistic GRW, as developed by Tumulka (2006), even involves *temporal* action at a distance: a region's pure thick parent generally doesn't screen it off even from its *ancestors*.)

how things could be at some time, given how some other things are *at that same time*. (In Chapter 5, we'll see Gauss's Law—one of Maxwell's laws of electrodynamics—as one example of this.) A more general job description for physical laws should thus be couched in *spatiotemporal* language: given how things are in one *spacetime region*, the laws of physics predict how other things are in the same or other spacetime regions.

Urchance entities are meant to capture exactly this more general, spatiotemporal predictive profile. Given the complete state of a sufficiently large and suitably shaped spacetime region, the laws can generate chance values for many propositions. For instance, given the exact state of a Cauchy surface, typical laws may determine chance values for all propositions about the surface's future. But if the region is too small or oddly shaped, even its complete state may fail to generate precise chance assignments for other regions. Similar remarks apply to *incomplete* specifications of a region's state.

This suggests two desiderata for our urchance formalism: (1) it should be permissive about the states and spacetime regions it relates, and (2) accommodate imprecise chance assignments when needed. Now, even when precise chances are unavailable, the laws may still impose substantive structural constraints, such as independence and invariance conditions. We've already seen examples in Chapter 1: Parental Markov, Boundary Markov, and Chance Invariance. A third desideratum, then, is that (3) the formalism allow the formulation of these principles even in the absence of precise chances. A fourth desideratum is a more familiar demand: (4) the formalism should balance expressive power with simplicity—minimizing theoretical posits, facilitating simple theories, and being easy to work with.

(Now, while I'm personally inclined to think that all of these desiderata also speak in favor of the formalism's being *metaphysically fundamental*—or a "joint-carving" piece of ideology (Sider, 2011)—I won't press that claim here. Those skeptical of metaphysical fundamentality and its epistemic basis can also treat this chapter simply as a case for the formalism's utility.)

This chapter defends Chapter 1's urchance formalism as meeting these desiderata best. According to the formalism, urchance entities are (equivalent to) sets of complete, primitively conditional, finitely additive probability functions, defined on the Cartesian square of the Boolean algebra of all propositions. The proposal has six parts: "probability functions," "primitively conditional," "set," "complete," "the Boolean algebra of all propositions," and "finitely additive." I'll take them in order.

2.2 Six Choice Points

2.2.1 "Probability Functions"

Evidently, the urchance formalism must make room for probability functions because physical laws possibly make probabilistic predictions. In fact, nothing more is needed, for probability functions also capture deterministic dynamics: on my preferred probability calculus, A deterministically entails B according to a probability function p iff $p(B|A \wedge Q) = 1$ for all propositions Q (including $Q = \neg B$).

2.2.2 "Primitively Conditional"

A probability function p, defined over a Boolean algebra with meet $\wedge$ and complement $\neg$, is *primitively conditional* iff it is a two-place function—written $p(\cdot|\cdot)$—which satisfies the ratio formula $p(A|B \wedge C) = p(A \wedge B|C)/p(B|C)$ whenever the right-hand side is defined *and* where $p(A|B \wedge C)$ is sometimes well-defined even when said right-hand side is undefined. In our formalism, both $p(A \wedge B|C)$ and $p(B|C)$ are always well-defined (see subsection 2.2.4); and so the right-hand side is undefined *iff* $p(B|C) = 0$.

Why think that chance functions are primitively conditional probability functions? As Hájek (2003, sec. 4.2) emphasizes, the chance of A conditional on $B \wedge C$ *does* appear to be sometimes well-defined even when the chance of B conditional on C is zero. He illustrates this with the case of a chance-zero event conditional on itself: "the conditional probability: $P_{12:01}$(an eclipse will occur at 15:52, given an eclipse will occur at 15:52) ... should equal 1" (287). Now, if this were the only kind of case where the ratio formula fails, the proponent of single-place probability might still be able to save herself. She would have to give up the attractive idea that conditional probability generalizes propositional logic—which includes that $p(A|B) = 1$ if $B \vDash A$—in its full generality. But she could still retain a restriction, according to which conditional probability generalizes propositional reasoning in the positive-probability sector.

However, another attractive idea is on the chopping block: that *chance functions at later times are the result of conditioning those at earlier times on intervening history* (cf. Hall, 2004b). For, in a continuous probabilistic physics, intervening histories generally have chance 0. To illustrate, there is now, pre-t, a well-defined chance of some radioactive atom's decaying between t and $t + \Delta t$, conditional on some given atom decaying into it at *exactly* instant t—even though the latter is a chance-0 event. Thus, if the first decay *does* happen at exactly t, the single-placer can't say that the chance function at t is the result of conditioning our pre-t chance function on intervening history—for, pre-t, that history has chance 0. In general, we need primitive conditionality to derive later chance functions from earlier ones by conditionalization.

Now, there is another alternative to single-place probability besides primitive conditionality. Some make conditional probability a *three-place* relation, between two propositions and a *partition* of the underlying sample space (Kolmogorov, 1950, V.§2). Easwaran (2008) endorses this as the correct account of *subjective probabilities*, or *credences*. However, I don't think partition-relativity is the right tool for conditional chance. Partitions are convenient for encoding symmetry considerations favoring one conditional-on-0 probability over another; subsection 2.2.6 supplies an example. But they don't obviously help encode *nomic* considerations favoring one conditional-on-0 probability over another. The laws themselves select the admissible conditional chance functions, and they don't do so by privileging partitions in any obvious way. Which partition of logical space underwrites the law-given chances conditional on our chance-0 decay above? Do any symmetry principles motivate that partition? I wouldn't know how to answer these questions.

Now, one might worry that, when several symmetries *are* relevant and the laws fail to break between them, we need partition-relativity to avoid making an arbitrary choice. But that's not so. Our urchance formalism avoids symmetry breaking without partition-relativity. If the laws fail to break between multiple symmetry considerations, the urchance entity simply incorporates *all* chance functions left open by the laws (see next section); that is, it'll contain equal representatives for each symmetry consideration and thus won't collapse to a single function tied to an arbitrary partition.

As far as I can see, then, a partition-relative framework has no utility for modeling chance.

2.2.3 "Set"

As we mentioned in the introduction, a world's fundamental physical laws generally won't fix precise chance values for all pairs of propositions. Indeed, pairs of propositions with precise probability assignments tend to be an elite bunch. For example, conditional on the state of some part P of a Cauchy surface, both Maxwellian electrodynamics and its "stochastic enhancement" assign precise probabilities to states of P's *future domain of dependence*, $D^+(P)$,[2] but generally not to possible states of any spacetime region disjoint from $D^+(P)$.

Chapter 1 proposes to capture this predictive weakness in terms of *sets* of (precise) probability functions (following van Fraassen (1984) also called "representors"). The

[2]That is, the largest region such that all inextendible past-directed causal curves starting in it intersect P.

resulting formalism is highly expressive, able to discriminate strictly more probability assignments than popular alternatives. Consider the following list of competitors.[3] The urchance entity may be either...

(i) a coherent (see below) partial probability function,

(ii) a coherent assignment of lower probabilities (e.g., Borel, 1924)[4]

(iii) a coherent comparative probability ranking (possibly incomplete) (Koopman, 1940; Hawthorne, 2016),

(iv) a coherent lower prevision, i.e. a coherent assignment of lower expectation values to all real-valued random variables (Walley, 1991),[5]

(v) a convex set of probability functions,

(vi) a coherent set-valued function (White, 2009),

(vii) a coherent general prevision, i.e. a coherent assignment of *sets* of expectation values to all real-valued random variables.

"Coherence" in each case means that there exists a set of complete probability functions which exactly reproduces the given imprecise assignment.[6] Every coherent imprecise

[3] The list is partially inspired by Joyce (2010), who considers (ii), (iii), (vi), and (vii).

[4] This is equivalent to Kyburg's (1983) formalism of coherent closed-interval-valued functions. (A generalization of Kyburg's formalism, in terms of coherent *interval*-valued functions, would be strictly stronger than (ii), but strictly weaker than (vi).)

[5] Technically, all *bounded* and *measurable* (relative to a given probability space) real-valued random variables. This ensures that the expectation functional, at any given such random variable, is continuous on the space of probability functions even for relatively small topologies. (This is relevant to the proofs in Appendix E.)

[6] In detail:

- A partial probability function is coherent iff: there exists a non-empty set of (complete) probability functions which all agree with the partial function whenever the latter is defined *and* for which the partial function is defined whenever the set agrees on a single value.
  - ▷ The last condition ensures that coherent partial probability functions are "inferentially closed": they must assign a value to every event whose probability is already fixed by the values it gives elsewhere. In my view, it would be hard to justify a partial function p as a probability function,

assignment determines a unique *maximal* such set. Call this the imprecise assignment's *corresponding representor*, written $\mathbf{P}_\xi$ with ξ the given imprecise assignment.

The concept of a corresponding representor produces a relation of relative expressiveness. Let T and T' be two of the modeling frameworks (i)–(vii). We say that T is *at least as expressive as* T' iff every corresponding representor of a member of T' is a corresponding representor of a member of T:

T is *at least as expressive as* T' iff: $\{\mathbf{P}_{\xi'} \mid \xi' \in T'\} \subseteq \{\mathbf{P}_\xi \mid \xi \in T\}$.

With this on the table, we can start our investigation into relative expressiveness. Let's start with (iv), coherent lower previsions. Walley's *Lower Envelope Theorem* (1991, §3.3) shows that the corresponding representor of a lower prevision $\underline{P}$ is a *closed convex* set of probability functions—namely the set

$$\mathbf{P}_{\underline{P}} = \bigcap_{A \in \mathcal{A}} \pi_A^{-1}([\underline{P}(\mathbf{1}_A), \overline{P}(\mathbf{1}_A)]), \text{ with } \pi_A(p) = p(A)$$

(where as usual $\overline{P}(X) := -\underline{P}(-X)$, and $\mathbf{1}_A$ is the indicator function of A on the sample space Ω). Convexity is immediate by the convexity of intervals: whenever $p(A)$,$q(A) \in [a,b]$, then $(\lambda p + (1-\lambda)q)(A) = \lambda p(A) + (1-\lambda)q(A) \in [a,b]$. And closedness follows on

if, say, $p(A)$, $p(B)$, and $p(A \wedge B)$ are all well-defined but $p(A \vee B)$ ill-defined—additivity is part of what it is to be a probability function.

- An assignment of lower probabilities is coherent iff it is the lower envelope of a set of probability functions.
- A comparative probability ranking is coherent iff there exists a non-empty set of probability functions which all agree with the ranking on its comparisons and for which the ranking is defined whenever the set agrees on a comparison.
- A lower prevision is coherent iff it is the lower envelope of a set of linear previsions (assignments of expectation values generated by probability functions.)
- A convex set of probability functions is coherent iff it is non-empty.
- A set-valued function is coherent iff there exists a set of probability functions whose range of probability values at each point is the value of the set-valued function at that point.
- A general prevision is coherent iff there exists a set of probability functions whose range of expected values at each random variable is the value of the general prevision at that variable.

any standard topology on $[0,1]^{\mathcal{A}}$ (where $\mathcal{A}$ is the given algebra over Ω).[7] But a restriction to closed sets and a restriction to convex sets each prevent us from capturing arguably relevant nomic distinctions.

Start with closedness. Closed sets of probability functions can't tell the difference between laws of nature saying nothing about the chance of an event E and the laws merely insisting that its chance be positive.[8] In both cases, the only closed set that satisfies the constraint is the "maximally dilated" one, with $\mathbf{P}(E) = [0,1]$.[9] Hence the lower-prevision formalism cannot distinguish between two distinct possibilities about physical laws—unless, implausibly, we assume that one of them is metaphysically impossible.

The situation is even worse with convexity. Sometimes we need to say that two propositions are independent even though none has precise chance. Crucially, Parental Markov and Boundary Markov do this. Think, for instance, of a Minkowski spacetime sliced by a time-like "thick" boundary: once the exact state of that boundary is fixed, events in the two spatial halves should be independent, despite the fact that each conditional chance may still be interval-valued. A natural proposal is:

> A and B are independent given C according to a set S of probability functions iff they are independent given C according to every member of S.

But this proposal fails dramatically for convex sets of probability functions. Whenever a set is dilated both on A and B (so it contains two probability functions that disagree on

[7] More precisely, on any topology at least as fine-grained on $[0,1]^{\mathcal{A}}$ as the product topology. Intuitively, the product topology is the coarsest (i.e., smallest) topology on $[0,1]^{\mathcal{A}}$ containing all finite intersections and arbitrary unions of sets of the form $\underbrace{[0,1] \times ... \times [0,1] \times I \times [0,1] \times ...}_{|\mathcal{A}| \text{ times}}$, where $I \subseteq [0,1]$ is open in $[0,1]$. The product topology is exactly the *coarsest* topology on $[0,1]^{\mathcal{A}}$ on which all evaluation maps π_A are continuous. A much more fine-grained topology on $[0,1]^{\mathcal{A}}$ is the *box topology*, consisting of arbitrary unions of sets of the form $\{ \prod_{A \in \mathcal{A}} U_A \mid U_A \text{ open in } [0,1]\}$. In either case, the evaluation maps are continuous on $[0,1]^{\mathcal{A}}$, hence continuous on $\Delta(\mathcal{A})$, and hence the pre-image of closed intervals under them closed in $\Delta(\mathcal{A})$. So $\mathbf{P}_{\underline{P}}$ is closed *qua* arbitrary intersection of closed sets.

[8] This is true whether we equip $\Delta(\mathcal{A})$ with the subspace topology induced by the product topology on $[0,1]^{\mathcal{A}}$ or with the subspace topology induced by the box topology.

[9] I'm using the obvious shorthand here: $\mathbf{P}(E) := \pi_E(\mathbf{P}) = \{P(E) \mid P \in \mathbf{P}\}$.

each marginal), some member of its convex hull must judge A and B to be dependent. So, under a convex representor two propositions can be independent (conditional on a third) only if at least one of them has a precise probability (conditional on the third). The reason is the following elementary fact:[10]

> **Theorem 1. Convex closure generically destroys independence.** If A and B are independent according to both p and q, then for any $0 < \lambda < 1$, they are *dependent* according to $\lambda p + (1-\lambda)q$ unless either $p(A) = q(A)$ or $p(B) = q(B)$.

A more sophisticated idea, available for *closed* convex sets, is to appeal to the set's extreme points. Any closed convex set of probability functions has a non-empty set of extrema—the minimal family that generates the set by convex closure.[11] One might therefore propose:

> A and B are independent given C according to closed convex set S iff A and B are independent given C according to all of S's extrema.

This criterion passes *some* basic tests. For example, it preserves independence: whenever all extrema of a closed convex set agree that two propositions are independent conditional on a third then, after conditioning on the third proposition, the extrema agree that the two propositions are (now unconditionally) independent. Conditional independence is preserved.

[10] *Proof of Theorem* 1: Let $p(A|B) = p(A)$ and $q(A|B) = q(A)$, and define $r := \lambda p + (1-\lambda)q$ for $0 < \lambda < 1$. Suppose, for contradiction, that $p(A) \neq q(A)$, $p(B) \neq q(B)$, and $r(A|B) = r(A)$. Because $p(B) \neq q(B)$, $r(B) > 0$, and so $r(A|B) = \frac{r(A \cap B)}{r(B)}$. Hence $\frac{r(A \cap B)}{r(B)} = r(A)$; writing this out:

$$\frac{\lambda \cdot p(A) \cdot p(B) + (1-\lambda) \cdot q(A) \cdot q(B)}{\lambda \cdot p(B) + (1-\lambda) \cdot q(B)} = \lambda p(A) + (1-\lambda) q(A).$$

Multiplying this out, collecting terms, and dividing by $\lambda(1-\lambda) \neq 0$, we obtain

$$\big(p(A) - q(A)\big) \cdot \big(q(B) - p(B)\big) = 0.$$

So either $p(A) = q(A)$ or $p(B) = q(B)$. Contradiction. ■

[11] This is a special case of the Krein-Milman theorem (e.g. Voigt, 2020, p. 132).

However, the rule *fails* to preserve conditional *dependence*. It can happen that A and B are dependent given C (because at least one extreme point witnesses the dependence), yet once we condition on C, the posterior extrema unanimously judge A and B as independent. The reason is that conditioning can push the witnessing extreme point into the interior of the convex set, thereby silencing it on the proposed rule.[12] So the proposed sophistication

[12] *Proof*: Let A, B, and C be propositions, and

$$\Omega := \{ABC, AB\overline{C}, A\overline{B}C, A\overline{B}\overline{C}, \overline{A}BC, \overline{A}B\overline{C}, \overline{A}\overline{B}C, \overline{A}\overline{B}\overline{C}\}$$

(where for space reasons I'm suppressing "$\wedge$" and write $\neg X$ as $\overline{X}$). Let $S := \overline{\mathrm{conv}\{P_1, P_2, P_3\}}$ be the closed convex hull of the following three probability functions over $\mathcal{P}(\Omega)$:

Event	P_1	P_2	P_3
ABC	3/32	9/64	3/16
$AB\overline{C}$	0	0	0
$A\overline{B}C$	5/32	27/64	1/16
$A\overline{B}\overline{C}$	0	0	0
$\overline{A}BC$	5/32	3/64	9/16
$\overline{A}B\overline{C}$	0	0	0
$\overline{A}\overline{B}C$	3/32	9/64	3/16
$\overline{A}\overline{B}\overline{C}$	1/2	1/4	0

Table 2.1: Extrema of a convex hull of probability functions

P_1 is an extremum of S, for

$$P_1(C) = \frac{1}{2} < P_2(C) = \frac{3}{4} < P_3(C) = 1.$$

(Indeed, P_2 and P_3 are also extrema.) Moreover, according to P_1, A and B are dependent conditional on C:

$$\begin{aligned} P_1(A|BC) = \frac{P_1(ABC)}{P_1(BC)} &= \frac{3/32}{3/32 + 5/32} = \frac{3}{8} \\ \neq \frac{1}{2} &= \frac{1/4}{1/4 + 1/4} = \frac{P_1(AC)}{P_1(C)} = P_1(A|C). \end{aligned}$$

Thus, according to the current proposal, A and B aren't screened off by C, according to S. But now, where $P_i^C \equiv P_i(\cdot|C)$ denotes the result of conditioning P_i on C :

Event inside C	P_1^C	P_2^C	P_3^C
AB	3/16	3/16	3/16
$A\overline{B}$	5/16	9/16	1/16
$\overline{A}B$	5/16	1/16	9/16
$\overline{A}\overline{B}$	3/16	3/16	3/16

Table 2.2: After conditioning the extrema on C

Notice that we have

$$P_1^C = 0.5P_2^C + 0.5P_3^C.$$

doesn't produce a sensible notion of *independence according to a closed convex set* after all.

In summary, Walley's account (iv), according to which urchance entities are (equivalent to) *closed convex* sets, fails on two fronts. First, the closedness requirement blurs genuine nomic differences (between "no constraints on $p(E)$" and "$p(E) > 0$"). Second, and arguably even more devastatingly, closed convex sets cannot represent the imprecise conditional independences required by Parental or Boundary Markov.

What about the other frameworks? Framework (v) (closed and non-closed convex sets) still struggles with the second problem, and so is unsatisfactory too. Walley (1991, §4.5.3(d), §2.7.3) also proves that framework (iv) is at least as expressive as framework (iii) (comparative probability) and strictly more expressive than (ii) (coherent lower probabilities). Because framework (iv) already fails to resolve either of our two central difficulties, neither of the weaker formalisms (ii) or (iii) can succeed. Finally, coherent lower probabilities are strictly more expressive than coherent partial probability functions (framework (i)).[13] (See fig. 2.1 for a summary of these expressiveness relations.) In short, none of the formalisms (i) through (v) is expressive enough.

Turn now to (viii). Coherent *general* previsions—assignments of *sets* of expectation values to gambles—are obviously strictly more expressive than coherent lower previsions, assignments of closed *intervals* of expectation values to gambles. Indeed, general previsions *can* sometimes represent independence between two propositions even when neither has

Hence $\text{ext}(S^C) \subseteq \{P_2^C, P_3^C\}$. But it's easy to check that

$$P_2^C(A|B) = \frac{3}{4} = P_2^C(A), \text{ and}$$
$$P_3^C(A|B) = \frac{1}{4} = P_3^C(A).$$

So, according to the present proposal, A and B are unconditionally screened off according to S^C. ■

[13]To see this: any coherent partial probability function has a unique coherent lower probability function, namely the lower envelope of the set of all probability functions agreeing with p whenever the latter is defined. Moreover, two different coherent lower probability functions may generate the same coherent partial probability function; e.g. any lower probability function dilated on every non-trivial proposition generates the empty partial function. So (ii) is strictly more expressive than (i).

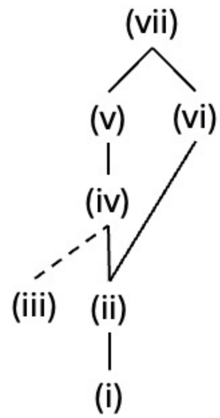


Figure 2.1: Relative expressive power: a solid (dashed) line means that the upper theory is strictly more expressive than (at least as expressive as) the lower theory.

precise chance.[14] However, this only happens under certain conditions. Specifically, to be able to represent A and B as independent, the prevision's corresponding representor must be disconnected.

> **Theorem 2.** ***Connectedness entails convexity.*** Let $\mathfrak{P}$ be a coherent general prevision on an algebra $\mathcal{A}$ over sample space Ω, with $A, B \in \mathcal{A}$. If the set $\mathbf{P}_{\mathfrak{P}}$ is connected (in the product topology on $\Delta(\mathcal{A})$), then it is convex.

Theorem 2 is proved in Appendix E. The following is an easy corollary of Theorems 1 and 2:[15]

[14]An elementary example: consider $\Omega = \{AB, A\overline{B}, \overline{A}B, \overline{AB}\}$ (where I suppress the conjunction symbol and write $\bar{\cdot}$ instead of $\neg\cdot$). We can write probability functions p on $\mathcal{P}(\Omega)$ as four-vectors $(p(AB), p(A\overline{B}), p(\overline{A}B), p(\overline{AB}))$. Now consider the representor $S := \{p, q\}$, consisting of the two probability functions

$$p := (1/4, 1/4, 1/4, 1/4),$$
$$q := (1/3, 1/3, 1/6, 1/6).$$

There are only four other probability functions which, for each indicator gamble on Ω's elements, agree with either p or q: $r_1 = (1/4, 1/3, 1/6, 1/4)$, $r_1 = (1/4, 1/3, 1/4, 1/6)$, $r_1 = (1/3, 1/4, 1/4, 1/6)$, $r_1 = (1/3, 1/4, 1/6, 1/4)$. But now we immediately see that, for every $i = 1, ..., 4$, $\mathbb{E}_{r_i}(\mathbf{1}_A) = 7/12$, which differs both from $\mathbb{E}_p(\mathbf{1}_A) = 1/2$ and $\mathbb{E}_q(\mathbf{1}_A) = 2/3$. So, the only probability functions which, for every random variable X on Ω, generate an expectation value in $\{\mathbb{E}_p(X), \mathbb{E}_q(X)\}$ are p and q themselves. ■

[15]*Proof*: If $\mathbf{P}_{\mathfrak{P}}$ is connected then, by Theorem 2, $\mathbf{P}_{\mathfrak{P}}$ is convex. Now suppose that $\mathbf{P}_{\mathfrak{P}}(A)$ and $\mathbf{P}_{\mathfrak{P}}(B)$ are

Corollary. ***Independence requires precision or disconnectedness.*** Let $\mathfrak{P}$ be a coherent general prevision on an algebra $\mathcal{A}$ over sample space Ω, with $A, B \in \mathcal{A}$. If $\mathbf{P}_{\mathfrak{P}}$ is connected (in the product topology), then A and B are dependent according to some member of $\mathbf{P}_{\mathfrak{P}}$ unless either $\mathbf{P}_{\mathfrak{P}}(A)$ or $\mathbf{P}_{\mathfrak{P}}(B)$ are singleton.

The following is a concrete instance of this corollary: given the sample space $\Omega := \{AB, A\overline{B}, \overline{A}B, \overline{AB}\}$, general previsions can't represent A and B as independent if both have *interval* imprecision.[16] An example involving an infinite sample space would be one where $\Omega = \{ABC_i, A\overline{B}C_i, \overline{A}BC_i, \overline{AB}C_i\}_{i\in\mathcal{I}}$, A and B are each independent of the C_i, and the probability on the coarse-graining $\Omega^* = \{C_i\}_{i\in\mathcal{I}}$ is connected (in the product topology).

But coherent general previsions are strictly more expressive than White's (2009) coherent set-valued functions (framework (vii)). If two set-valued function f and g disagree on some A (i.e., $f(A) \neq g(A)$), they produce different expectation sets for the indicator gamble on A. Hence any two distinct coherent set-valued functions correspond to different

both non-singleton. It follows that there are $p, q \in \mathbf{P}_{\mathfrak{P}}$ such that $p(A) \neq q(A)$ and $p(B) \neq q(B)$. If A and B are dependent according to either p or q, we are done. So assume that A and B are independent according to both p and q. Since $\mathbf{P}_{\mathfrak{P}}$ is convex, there is a $\lambda \in (0,1)$ such that $r := \lambda p + (1-\lambda)q \in \mathbf{P}_{\mathfrak{P}}$. By Theorem 1, A and B are dependent according to r. ■

[16] That is: there is no coherent general prevision $\mathfrak{P}$ on Ω such that:

1. $\mathfrak{P}(\mathbf{1}_A) = [a_0, a_1]$ and $\mathfrak{P}(\mathbf{1}_B) = [b_0, b_1]$ for some $a_0 < a_1$ and $b_0 < b_1$.
2. All members of $\mathbf{P}_{\mathfrak{P}}$ judge A and B as independent.

Proof: Suppose for contradiction that $\mathbf{P}_{\mathfrak{P}}$ satisfies both 1 and 2. Defining $p_{a,b}$ such that

$$\begin{aligned}
p_{a,b}(AB) &= ab, \\
p_{a,b}(A\overline{B}) &= a(1-b), \\
p_{a,b}(\overline{A}B) &= (1-a)b, \\
p_{a,b}(\overline{AB}) &= (1-a)(1-b),
\end{aligned}$$

we have

$$\mathbf{P}_{\mathfrak{P}} = \{p_{a,b} \mid (a,b) \in [a_0, a_1] \times [b_0, b_1]\}.$$

Continuous variation of a and b (in $[0,1]^2$) leads to continuous variation in $\Delta(\mathcal{P}(\Omega)) \subseteq [0,1]^{16}$ (for finite n, the product topology on $[0,1]^n$ is just the Euclidean topology). Thus, since $[a_0, a_1] \times [b_0, b_1]$ is connected, $\mathbf{P}_{\mathfrak{P}}$ is connected. By 1. again, neither $\mathbf{P}_{\mathfrak{P}}(A)$ or $\mathbf{P}_{\mathfrak{P}}(B)$ are singleton. So, by the previous corollary, some member of $\mathbf{P}_{\mathfrak{P}}$ judges A and B to be dependent. ■

coherent general previsions.[17] Given that we reject coherent general previsions for lack of expressive power, we must reject (vii) as well.

Of the formalisms we've seen, only the unrestricted set-based formalism offers enough expressive resources to formulate our Markovian principles while avoiding running together what seem like genuine nomic differences.

2.2.4 "Complete"

Now, our preferred formalism is arguably not *maximally* expressive either: we require the urchance entity to be (equivalent to) a set of *complete*—rather than partial—primitively conditional probability functions. This excludes a few genuine logical possibilities. Yet (1) the case that these possibilities correspond to genuine nomic differences strikes me as much less convincing, and (2) unlike the shift from formalisms (i) through (vii)) to our set-based formalism, the extra expressive power would come at a substantial theoretical cost. That being said, if one disagreed with me on (1), one *could* swallow those costs and reformulate the Chapter 1 theory in the broader—but markedly more cumbersome—formalism that allows partial functions in representors.

The probability gaps of a partial function indicate which propositions don't receive precise chances from the laws of nature, with the rest of the function imposing bounds on admissible chance values. But sets of complete functions already let us do exactly that. Admitting partial functions into a representor merely adds representational redundancy.

Suppose, for instance, that the laws do not fix a precise chance for a particle to occupy one energy state rather than another, instead allowing any value in the range $[a, b]$. Our formalism models this by including every complete function whose value for that proposi-

[17] To see the "strictly": consider a situation where ω_1, ω_2, and ω_3 have imprecise probabilities $[0, 1]$, $[0, 1/2]$, and $[0, 1/2]$, respectively. Here a set-valued function can't distinguish between a representor according to which ω_2 and ω_3 have determinately equal probability and one which imposes no such constraints. But those are easily distinguished by general previsions: in the first case, $P(\mathbf{1}_{\{\omega_2\}} - \mathbf{1}_{\{\omega_3\}}) = \{0\}$, while in the second case $P(\mathbf{1}_{\{\omega_2\}} - \mathbf{1}_{\{\omega_3\}}) = [-1/2, 1/2]$.

tion lies within the interval. Now consider a rival representor that *additionally* contains a partial function undefined for that very proposition. I don't see how this expresses a different physical situation. As far as I can see, the same is true of all other cases: admitting partial functions into a representor simply reproduces a fact about physical laws that can already be expressed by representors of complete functions.

With urchance members ranging over all propositions, another reason one might favor partiality is to save countable additivity. I'll address this issue in subsection 2.2.6. The short answer is that primitive conditionality by itself already pressures us to relinquish countable additivity. So the principle can plausibly be saved only by abandoning both completeness *and* primitive conditionality, a trade-off we shouldn't accept given our independent reasons for treating chance functions as primitively conditional.

Moreover, abandoning completeness comes at independent cost. It would complicate the theoretical base: principles like Boundary Markov and Acyclic Chance Invariance (Chapter 1) would need constant caveats about well-definedness and almost every axiom would require exceptions to handle undefined expressions. Abandoning completeness would also significantly complicate the proofs in the appendices of Ch. 1, both because the Markov conditions become messier and because the arguments themselves rely on completeness at several points.

Since, so far as I can tell, partial-function representors yield nothing but bitter fruit, my account rejects them in favor of representors of complete functions. Whenever I speak of "probability functions" from now on, I mean complete ones.

2.2.5 "The (Cartesian) Square of the Boolean Algebra of Propositions"

An admission upfront: because I don't want to presuppose Booleanism—the view that only literally identical propositions can be logically equivalent—by "the Boolean algebra of propositions" I technically mean the Boolean algebra of *equivalence classes of logically* *equivalent propositions*.

There are two aspects to this one might question. First, why think that the arguments in each slot of a chance function form a Boolean algebra? The central idea that physical laws cannot tell apart logically equivalent propositions: whenever two propositions are equivalent, you can substitute one for the other inside any chance function *salva probabilitate*. In addition to the strong initial plausibility of this idea, chapter 3 details the myriad problems that follow if you abandon it. Chief among these problems is the lack of a well-worked-out alternative to classical probability theory on which to rebuild a theory of chance.

One might also question the assumption that the objects of chance are *propositions* rather than, say, *events*. I like the propositional account in part because I view probabilistic relations as generalizations of propositional logical relations. But another reason is the formalism's relative ideological flexibility. Whatever entities your preferred Boolean algebra consists of, it is likely just a special case of a propositional algebra. For example, the algebra of events is just (isomorphic to) the sub-algebra of propositions which are arbitrary Boolean combinations of propositions of the form $\ulcorner e$ occurs$\urcorner$, where e is an event. So, even if you disagree with my preference for propositions, the formalism developed here is general enough that it can likely be restricted to your preferred ontology without loss.

2.2.6 "Finitely Additive"

The final choice—finite additivity—is compelled by two of our previous choices. The first is the choice that probability functions are defined on the set of *all* propositions: only very special countably additive probability functions can have that feature. The second choice is that chances are primitively conditional probabilities: on pain of violating plausible symmetries, primitively conditional probability functions cannot generally satisfy countable additivity. Finally, there are arguably metaphysically possible scenarios where the physics alone forces finite additivity, irrespective of completeness and primitive conditionality.

Start with the first point. It is well-known that countably additive measures generally admit unmeasurable sets—see e.g. Vitali's (1905) construction of unmeasurable sets for translation-invariant measures on $[0, 1]$. Moreover, Banach & Kuratowski (1929) show that, if the continuum hypothesis is true, then the only countably additive measures on the power set of $[0, 1]$ are infinitely biased toward some singletons—*viz.,* they assign some singletons strictly positive numbers.[18] But it can be proved (e.g. via the Hahn-Banach theorem) that any (finitely or countably additive) probability measure, defined on some Boolean algebra over its sample space (of any cardinality), can be extended into a *finitely* additive probability measure on the sample space's power set. So, given plausible assumptions about the richness of propositions, whatever countably additive probability functions are granted on however rich an algebra of propositions, they are extendible to finitely additive, but not generally countably additive, functions on the algebra of *all* propositions.

The foregoing considerations hold even for merely derivatively conditional probability functions. For *primitively conditional* probability functions, symmetry considerations put additional pressure on countable additivity. An easy case is that of a fair spinner, illustrating the fact that one cannot have a countably additive translation-invariant probability conditional on any countable dense subset.[19] Being fair, the spinner should satisfy the following condition:

Rotational Symmetry: If A and A' are subsets of points with $A \cup A' \subseteq B$ such

[18] Ulam (1930) subsequently strengthens the result, showing that the consequent follows from the weaker assumption that neither the continuum itself nor any smaller cardinal is (weakly) inaccessible.

[19] This is different from Vitali's construction, which concerns the *unconditional* measure, of rather alien sets (constructed by picking points from an uncountable partition via the axiom of choice). While the domain restrictions solving the Vitali problem are arguably fairly modest, those which would be required to solve the problems from primitive conditionality are more severe, as the spinner and especially the sphere (two-dimensional spinner) show.

that a rigid rotation maps A into A', then

$$ch(A|B) = ch(A'|B),$$

provided either side is well-defined.[20]

Let $\mathbb{Q}$ be the set of rational-numbered degrees. Now, for a *primitively conditional* chance function, there should be a well-defined chance of the spinner's landing on $\mathbb{Q}$ conditional on its landing on $\mathbb{Q}$: specifically, a chance of 1. However, by Rotational Symmetry, for any rational-numbered degree there is an identical chance $\varepsilon \geq 0$ of landing on it, conditional on landing on $\mathbb{Q}$. Thus, by countable additivity, $ch(\mathbb{Q}|\mathbb{Q}) = \sum_{q \in \mathbb{Q}} ch(q|\mathbb{Q}) = \sum_{q \in \mathbb{Q}} \varepsilon$ is either 0 or diverges to $+\infty$—not 1. So, in our primitively conditional framework, countable additivity conflicts with Rotational Symmetry.

Another point worth mentioning is that, given primitive conditionality, countable additivity doesn't even give us all the aggregative principles we would hope for anyway. Specifically, Borel's paradox shows that *integral* versions of the law of total probability would still fail, on pain of violating extremely plausible symmetries. (The next four paragraphs go into more detail about this—they are slightly technical and may be skipped without losing the chapter's main thread.)

To illustrate, suppose you're given a *two-dimensional* fair spinner: that is, a sphere of unit area equipped with a uniform primitively conditional probability function P,[21] defined over some algebra $\mathcal{A}$ over the sphere's points. Designating some arbitrary great circle as the equator of zero latitude and some arbitrary orthogonal great circle G the *Greenwich great circle*, you're interested in the chance of the spinner's landing between $-\pi/4$ and $+\pi/4$ latitude—that is, closer to the equator than to either of the two poles—conditional on its landing on G.

[20] In an obvious short-hand, capital letters here denote sets of points as well as the respective proposition that the spinner *lands* on the given set of points. *ch* is the chance function conditional on the spinner setup.

[21] "Uniform" means that, for any subset A of the sphere, $P(A)$ equals A's *area.*

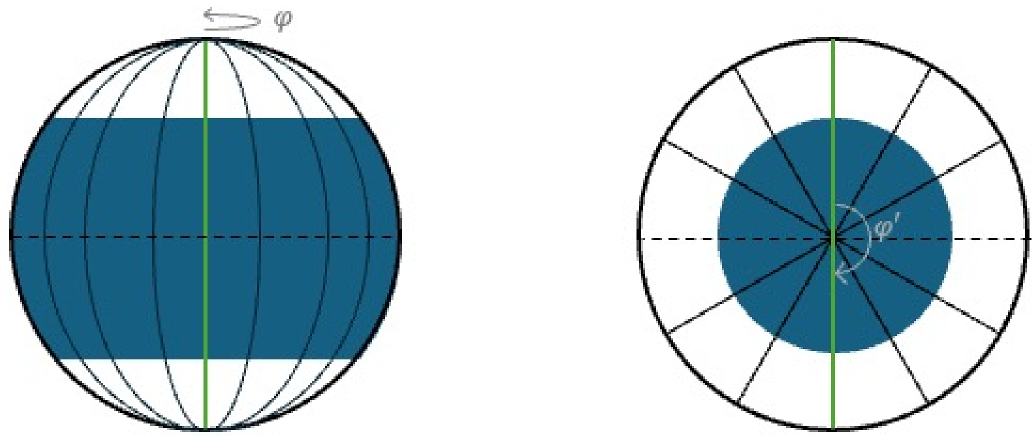


Figure 2.2: The green line indicates the Greenwich great circle G; the dashed horizontal line indicates the equator. The blue areas are A_1 (left) and A_2 (right).

There are obviously infinitely many sets which intersect G exactly along its$-\pi/4$ to $+\pi/4$ segment. Figure 2.2 indicates two such sets: A_1 is the region between $-\pi/4$ and $+\pi/4$ latitude and A_2 the disk whose radius goes from G's intersection with the equator to its intersection with the $+\pi/4$ latitude small circle. The figure also indicates two distinct partitions: one into the longitudinal great circles, and the other into all great circles intersecting the point of intersection of G and the equator. We may denote the elements of the two partitions by their corresponding azimuthal angles φ and φ' (see figure); correspondingly, we'll denote the partitions $\mathcal{L} := \{L_\varphi\}_{\varphi\in[0,\pi)}$ and $\mathcal{L}^* := \{L^*_\varphi\}_{\varphi\in[0,\pi)}$. Given the setup, the following two symmetry constraints are exceedingly plausible:

$$\text{for all } a, b \in [0,\pi)\colon P(A_1 \mid L_{\varphi=a}) = P(A_1 \mid L_{\varphi=b}), \text{ and} \tag{2.1}$$

$$\text{for all } a, b \in [0,\pi)\colon P(A_2 \mid L_{\varphi'=a}) = P(A_2 \mid L_{\varphi'=b}). \tag{2.2}$$

Now, every partition of the sphere into P-measurable subsets is associated with an instance of the law of total probability. For, given any $A \in \mathcal{A}$, and any partition $\mathfrak{B} = \{B_i\}_{i \in \mathcal{I}}$ of the sphere (for $\mathcal{I}$ some countable or uncountable index set), we can define the random variable

$$P(A||\mathfrak{B}) : \begin{cases} \Omega \to [0,1], \\ \omega \mapsto P(A|B_i) \quad \text{iff } \omega \in B_i. \end{cases}.$$

Provided $P(A||\mathfrak{B})$ is measurable,[22] then where $\mathcal{B}$ is the sub-algebra of $\mathcal{A}$ generated by $\mathfrak{B}$, the integral version of the law of total probability is simply this:

$$\text{for any } B \in \mathcal{B},\ P(A \cap B) = \int_B P(A||\mathfrak{B})(\omega) \cdot \mathrm{d}P, \tag{2.3}$$

where $\mathrm{d}P$ indicates that P is the integrating measure for the integral.[23] But eqs. 2.1, 2.2, and 2.3 jointly conflict. Briefly, the integral identity 2.3 entails conglomerability over both partitions, which together with the two symmetry constraints 2.1 and 2.2 forces $P(A_1) = P(A_2)$, which is false by elementary geometry.

[22]Measurability follows automatically if $P(A||\mathfrak{B})$ is the Radon-Nikodym derivative—but needn't hold if $P(\cdot|\cdot)$ is primitively conditional.

[23]If we drop countable additivity, the integral won't be the standard Lebesgue integral. But in a finitely-additive-measure space $(\Omega, \mathcal{P}(\Omega), P)$ with $P(\Omega) < \infty$ we still have a canonical notion of integration for bounded integrands like $P(A||\mathfrak{B})(\omega)$, based on *uniform* approximation by simple functions. *More details:* For any bounded f and $\varepsilon > 0$, define $s^f_\varepsilon(x) := \varepsilon\lfloor \frac{f(x)}{\varepsilon} \rfloor$—intuitively, dividing f's range into finitely many "levels" of size proportional to ε. Every s^f_ε is a simple function, i.e., it takes only finitely many values and is (obviously) measurable by P, which is defined on $\mathcal{P}(\Omega)$. Now, the sequence $(s^f_{1/n})_{n \in \mathbb{N}}$ *uniformly* converges to f: $\lim_{n\to\infty} \|s^f_{1/n} - f\|_\infty := \lim_{n\to\infty} \big(\sup_{x \in \Omega} |s^f_{1/n} - f|\big) \leq \lim_{n\to\infty} \frac{C}{n} = 0$, where C is some constant depending only on f's supremum and infimum. For a simple function s, we can define as usual

$$\int_A s\mathrm{d}P := \sum_{y \in \mathrm{range}(s)} y \cdot P(s^{-1}(y) \cap A).$$

Our canonical integral is then

$$\int_A f\mathrm{d}P := \lim_{n\to\infty} \int_A s^f_{1/n}\mathrm{d}P$$

whenever the right-hand side is well-defined, and undefined otherwise. The limit is independent of which uniformly convergent sequence we choose. Note that, since P is only finitely additive, the resulting integration operator $\int \cdot \mathrm{d}P$ lacks the usual niceties of Lebesgue integration, including dominated and monotone convergence.

More slowly: eq. 2.3 entails the following weak form of conglomerability (to see this, simply instantiate B with the entire sphere):

> If $P(A||\mathcal{B})(\omega)$ is constant c over the sphere, then $P(A) = c$.

Given conglomerability and the identity $P(A||\mathcal{L})(\omega)|_L = P(A|L)$ for all $L \in \mathcal{L}$, eq. 2.1 implies that

$$P(A_1|G) = P(A_1|L_{\varphi=0}) = P(A_1). \tag{2.4}$$

Meanwhile, given conglomerability and the identity $P(A||\mathcal{L}^*)(\omega)|_{L^*} = P(A|L^*)$ for all $L^* \in \mathcal{L}^*$, eq. 2.2 implies that

$$P(A_2|G) = P(A_2|L^*_{\varphi^*=0}) = P(A_2). \tag{2.5}$$

But $A_1 \cap G$ and $A_2 \cap G$ are the exact same sets. So we must have

$$P(A_1|G) = P(A_2|G).$$

Thus, from eqs. 2.4 and 2.5,

$$P(A_1) = P(A_2).$$

However, by elementary geometry,

$$P(A_1) = \frac{1}{\sqrt{2}} \approx 0.71,$$
$$P(A_2) = \frac{\sqrt{2}-1}{2\sqrt{2}} \approx 0.15.$$

So, eqs. 2.1, 2.2, and 2.3 jointly conflict.

As we saw earlier, some, like Easwaran (2008), take the Borel paradox to show that conditional probability is fundamentally partition-relative. Accepting two-place, primitively conditional probability, we disagree and instead live with the failure of the integral versions of total probability: *pace* eq. 2.3, we generally have[24]

$$P(A \cap B) \neq \int_B P(A||\mathfrak{B})(\omega) \cdot \mathrm{d}P.$$

So, this shows that countable additivity wouldn't even secure all the infinitary aggregation we would want. Primitive conditionality already spoils them by itself.

Onto the third point. In worlds with transfinite temporal structure, standard topological continuity at limit ordinals easily breaks countable additivity. We don't want to rule out such scenarios as *a priori* impossible, and so a fully general urchance formalism should be required to handle them.

Consider a physics implementing a symmetric random walk—for example, a world where particles have a constant chance, at the end of each second, to spontaneously jump to a neighboring energy state: $1/2$ chance to the next-highest and $1/2$ chance to the next-lowest. (Suppose the energy states are unbounded from above and below.)

With this chance dynamics, a natural candidate for the universe's temporal structure is the product $\alpha \times [0,1)$, with α some (possibly infinite) ordinal and the order topology induced by the lexicographic order. Suppose the universe's temporal structure is, specifically, $(\omega + 1) \times [0,1)$, with ω the first infinite ordinal; i.e., it has an additional second attached after "infinity". If a particle starts out, at time $(0,0)$, in energy state 0, then for

[24]Note that, if P is countably additive, the Radon-Nikodym theorem guarantees that, for any $A \in \mathcal{A}$ and any partition $\mathfrak{B}$ of the sphere, there is a random variable $X_{A,\mathfrak{B}} : \Omega \to [0,1]$ such that

$$\text{for all } B \in \mathcal{B} : P(A \cap B) = \int_B X_{A,\mathfrak{B}}(\omega) \cdot \mathrm{d}P,$$

where $\mathcal{B} \subseteq \mathcal{A}$ is the (sigma-)subalgebra generated by $\mathfrak{B}$. But a proponent of primitive conditionality simply wouldn't accept that for all partitions $\mathfrak{B}$ there is a $X_{A,\mathfrak{B}}$ which satisfies $X_{A,\mathfrak{B}}(\omega) = P(A|B)$ for all $\omega \in B$ whenever $B \in \mathfrak{B}$. The Radon-Nikodym theorem doesn't force one to interpret the integral relationship as an instance of the law of total probability. (Of course, if countable additivity fails, then the Radon-Nikodym theorem doesn't even get off the ground, and in general there may be *no* random variable $X : \Omega \to [0,1]$ such that, for all $B \in \mathcal{B}$, $P(A \cap B) = \int_B X(\omega) \cdot \mathrm{d}P$.)

large t, the chance that it is in state n at $(t,0)$ is[25]

$$p(S_{(t,0)} = n) \approx \exp(-\frac{n^2}{2t}),$$

if $t+n$ is even, and 0 otherwise. Thus, for any t, $p(S_{(t,0)} = n)$ tends to zero:

$$\lim_{t\to\infty} p(S_{(t,0)} = n) = 0.$$

So, provided the dynamics are continuous (in the lexicographic order topology) at time $(\omega, 0)$,

$$p(S_{(\omega,0)} = n) = 0 \tag{2.6}$$

for all n. That is, for any $n \in \mathbb{Z}$, the chance that the particle is in state n at $(\omega, 0)$ is zero.

But the particle is certainly *somewhere* at $(\omega, 0)$—suppose spontaneous particle annihilation is nomically disallowed. So,

$$p\big(\bigvee_{n\in\mathbb{Z}} S_{(\omega,0)} = n\big) = 1. \tag{2.7}$$

Eqs. 2.6 and 2.7 jointly violate countable additivity: $p\big(\bigvee_{n\in\mathbb{Z}} S_{(\omega,0)} = n\big) \neq \sum_{n\in\mathbb{Z}} p(S_{(\omega,0)} = n)$. In a universe with transfinite time, a probability distribution that spreads without bound over a countably infinite state space can remain continuous only by sacrificing countable additivity.

[25] More precisely, the chance is

$$p(S_{(t,0)} = n) = \begin{cases} \binom{t}{\frac{t+n}{2}} \cdot 2^{-t} & \text{if } t+n \text{ even and } n \leq t, \\ 0 & \text{else.} \end{cases}$$

For fixed n, the leading t-term of $\binom{t}{\frac{t+n}{2}}$, for $t+n$ even, grows like $\frac{2^t}{\sqrt{\pi t}} \exp(-\frac{n^2}{2t})$. Hence, for $t+n$ even, $p(S_t = n)$ tends to 0 like $\frac{1}{\sqrt{\pi t}} \exp(-\frac{n^2}{2t})$.

2.3 Boolean Axioms For Primitively Conditional Probability

So much for the defense of my urchance formalism. This final section covers the axiom base for primitively conditional probability, the formal underpinning of urchance member functions.

Popper (1968, Appendices *IV and *V) proposes an elegant axiomatization of complete, primitively conditional probability. The account makes no assumptions about the domains of probability functions, except that it is a set closed under a *meet* and a *complement* operation. From Popper's axioms it then follows that these operations satisfy the axioms of a Boolean algebra, relative to "=" being equality of probability conditional on all members of the set.

Now, *we* already assume that the domain of any urchance entity's member function is a Boolean algebra. This allows us to ditch one of Popper's axioms (Substitutivity—cf. Popper (1968, p. 349)). The remaining axioms are thus (I've changed the names of axioms 1 and 3):[26]

> **Axiom 0. Domain.** $p(\cdot|\cdot)$ is a real-valued function on $\mathcal{A} \times \mathcal{A}$, for some Boolean algebra $\mathcal{A}$.
>
> **Axiom 1. Unity.** $p(A|A) = p(B|B)$.
>
> **Axiom 2. Monotony.** $p(A \wedge B|C) \leq p(A|C)$.

[26] Exchanging axioms 1, 2, and 4, for the following axioms yields an equivalent system (likewise of mutually independent axioms), somewhat more reminiscent of the familiar Kolmogorov axioms:

> **Axiom 1*. Positivity.** $0 < p(A|A)$.
>
> **Axiom 2*. Non-Negativity.** $0 \leq p(A|B)$.
>
> **Axiom 4*. Additivity.** Whenever $A \wedge B = \emptyset$ and $p(D|C) \neq 1$ for some D, then $p(A \vee B|C) = p(A|C) + p(B|C)$.

I'll prove one direction of this equivalence in Appendix G. (The other direction isn't hard either; I'll omit it to save space.)

Axiom 3. Non-Triviality. There are $A, B, C, D \in \mathcal{A}$ such that $p(A|B) \neq p(C|D)$.

Axiom 4. Complementation. Whenever $p(C|B) \neq p(B|B)$ for some C, then $p(B|B) = p(A|B) + p(\neg A|B)$.

Axiom 5. Multiplication. $p(A \wedge B|C) = p(A|B \wedge C) \cdot p(B|C)$.

For consistency, consider the model $\Omega = \{w\}$ with $\mathcal{A}(\Omega) = \mathcal{P}(\Omega)$ and $p(\Omega|\Omega) = p(\Omega|\varnothing) = p(\varnothing|\varnothing) = 1$ and $p(\varnothing|\Omega) = 0$. I prove the axioms' mutual independence in Appendix F. In Appendix G, I prove that they verify the familiar properties of Popper functions:

- **Non-Negativity.** $0 \leq p(A|B)$.
- **Normalization.** $p(A|B) \leq p(B|B) = 1$.
- **Additivity.** Whenever $A \wedge B = \varnothing$ and $p(D|C) \neq 1$ for some D, then $p(A \vee B|C) = p(A|C) + p(B|C)$.
- **Explosion.** $p(A|C \wedge \neg C) = 1$.
- **Substitutivity.** If $p(A|C) = p(B|C)$ for all C$\in \mathcal{A}(\Omega)$, then $p(D|A) = p(D|B)$.

When I say that a function is a *complete, primitively conditional probability function*, I mean that it satisfies Axioms 0–5.[27]

[27] Rényi (1955) is another popular axiomatization of primitively conditional probability. The main difference to the present formalism is that Rényi functions are not complete (on the square of the first argument's Boolean algebra): in particular, $p(A|\varnothing)$ is never defined. Essentially, the difference is forced by Rényi's assumption of a version of Complementation which drops the condition that $p(C|B) \neq p(B|B)$ for some C; on pain of inconsistency, this version requires a restriction of the domain of p's second argument. More precisely, the following system is equivalent to a finitely additive version of Rényi's axiom system (proofs ommitted):

Axiom 0R. Domain$_R$. $p(\cdot|\cdot)$ is a real-valued function on $\mathcal{A} \times \mathcal{B}$, for some Boolean algebra $\mathcal{A}$ and $\mathcal{B} \subseteq \mathcal{A}$.
Axiom 1R. Unity$_R$. For all $B, B' \in \mathcal{B}$, $p(B|B) = p(B'|B')$.
Axiom 2R. Monotony$_R$. For all $A, B \in \mathcal{A}$ and $C \in \mathcal{B}$, $p(A \wedge B|C) \leq p(A|C)$.
Axiom 3R. Non-Triviality$_R$. There are $A, C \in \mathcal{A}$ and $B, D \in \mathcal{B}$ such that $p(A|B) \neq p(C|D)$.
Axiom 4R. Complementation$_R$. For all $A \in \mathcal{A}$ and $B \in \mathcal{B}$, $p(B|B) = p(A|B) + p(\neg A|B)$.

Axiom 5R. Multiplication$_R$. For all $A, B \in \mathcal{A}$ and $C \in \mathcal{B}$, $p(A \wedge B|C) = p(A|B \wedge C) \cdot p(B|C)$.

The formal difference between Rényi and Popper is rather shallow: given a probability functions satisfying Popper's axioms 0–5, one automatically obtains a function satisfying (a finitely additive version of) Rényi's axioms by restricting the second argument to those propositions B such that $p(A|B) \neq 1$ for some A—that is, by excluding all propositions interchangeable with logical contradictions *salva probabilitate*. Conversely, given a Rényi function with maximally large domain for the second argument—i.e., a domain that excludes only those propositions interchangeable with logical contradictions *salva probabilitate*—one obtains a Popper function by adding back the excluded second arguments, giving every proposition probability 1 conditional on them. Still, as a basis for a theory of urchance, Popper's formalism recommends itself for not requiring constant disclaimers about the second argument's consistency.

Chapter 3

Classical Probability Theory vs. Chance Invariance

3.1 Introduction

My chance theory is a *classical* probability theory: logical tautologies have chance 1 conditional on any non-contradiction and all logical contradictions have chance 0 conditional on any non-contradiction. Besides Berkovitz (2001) and Cusbert (2022)—whose arguments I discuss in footnotes 57 and 18 of Ch. 1—the only two previous discussions of chances on loops I'm aware of are Horacek (2005); Effingham (2020). They both hold that chance possibly obeys a *non-classical* probability theory: in the presence of causal loops, they say, logical contradictions can have positive chance conditional on some non-contradictions. This chapter argues against that strategy.

I obviously assume classical probability theory everywhere in Chapter 1; but in particular, it underlies the failure of Chance Invariance. My argument against Chance Invariance (or "Stability") shows that, in CIRCLE, Chance Invariance is logically inconsistent with classical probability theory and concludes that we should therefore reject it. Horacek and Effingham turn out to take the opposite route: they each propose a version of Chance

Invariance and argue that, because it necessarily holds, classical probability theory is possibly false.

Section 3.2 reproduces a time travel case from Effingham (2020) to focus our discussion on. In Section 3.3 I discuss Horacek's and Effingham's views, and in Section 3.4 I argue that we should reject them.

An aside before we start: I'll often speak of "logically impossible propositions" or "logically impossible worlds". This presupposes a propositional notion of logical necessity. For our purposes, let us simply note the following sufficient condition: a proposition is logically impossible *if* it is expressed by an English sentence of the form $\ulcorner p$ and it's not the case that $p\urcorner$.[1]

3.2 VEERING PARTICLES

Consider VEERING PARTICLES (or VP for short), a universe described as follows:

> "There are two wormhole mouths close together: Opening and Exit. Particle p will pass through the field at t_3. If p doesn't veer, it will enter Opening at t_5, sending it back in time to t_2 whereupon it'll emerge from Exit on a path perpendicular to that which it took before entering Opening. That perpendicular path is such that if the future version of p doesn't change course, it will strike its earlier self at t_4, causing its earlier self to move in such a way that it will never enter Opening (whether it veers or not)." (Effingham, 2020, 155-6)

The following figure of Effingham's (2020, 156) illustrates what's going on:

[1] For a much more extensive account of propositional logical modality, and an accompanying worldview, see Bacon (2020).

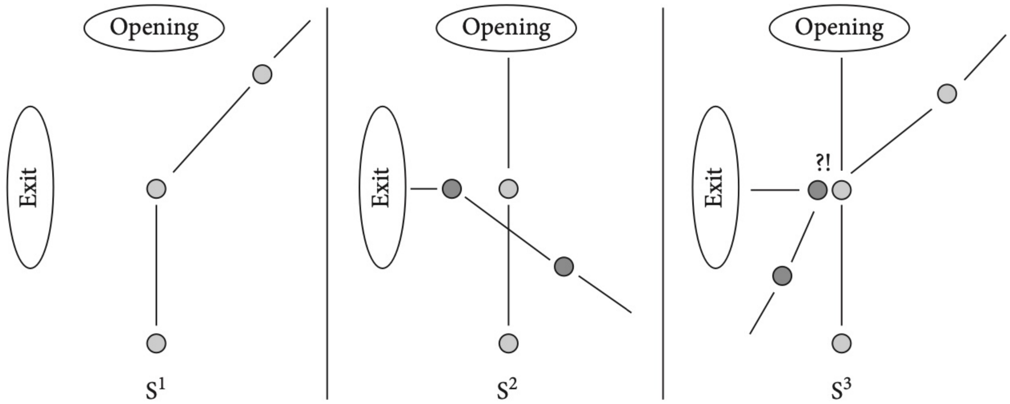


Figure 3.1: A sketch of Effingham's (2020, 156) case. (© 2023 Oxford University Press. Reproduced by permission. All rights reserved.)

Effingham describes the indicated scenarios S^1–S^3 as follows:

> "S^1: p initially veers to the right (at t_3). No time travel takes place.
>
> S^2: p's earlier version doesn't veer and so it travels through Opening. It then passes out of Exit at t_2, and its future self then veers (at t_3), missing its former self. This is metaphysically possible.
>
> S^3: p's earlier version doesn't veer and so travels through Opening. When it leaves Exit it, again, fails to veer. It strikes itself and prevents its earlier self from entering Opening. This is metaphysically impossible." (2020, p. 156)

For concreteness, suppose that S^2 is what actually occurs in VP.

The contradiction involved in S^3—that p's earlier self both travels through Opening and is prevented from doing so—becomes a strictly *logical* contradiction in conjunction with the principle that x's being prevented from ϕ-ing entails that x doesn't ϕ. In principle, one could formulate a theory of probability which assigns positive chance to S^3 but avoids positive counterlogical chances by assigning chance *less than* 1 to the aforementioned principle. ("Positive counterlogical chances" is what I'll henceforth call any situation in which a logical contradiction has positive chance.) But this maneuver isn't merely *ad hoc* but also faces many of the same issues as positive counterlogical chances, expounded in section 3.4, such as infecting time-travel-free worlds, lacking a well-established formal structure, and obscuring the concept of chance. In the following, I'll thus put this strategy aside and only consider chance functions which assign probability 1 to the aforementioned principle (conditional on any proposition); I'll also say that S^3 *entails* a contradiction.

Effingham's aim is to calculate the chance function Ch_{t_1} at time t_1, a time prior to any particle's emerging from the wormhole. In doing so, he implicitly assumes that Ch_{t_1} is conditioned on several auxiliary assumptions: namely VP's background geometry, the presence of the scattering field around time t_3, and the assumption that no particle other than p exits the wormhole. (The latter follows from Effingham's assumption that "S^1, S^2, and S^3 are the only states in the state space" (157) of Ch_{t_1}.) We can thus identify Ch_{t_1} in our framework as follows. Let urch_{VP} be the result of conditioning the urchance function on all three auxiliary assumptions. In our framework, Ch_{t_1} is then the function $\text{urch}_{VP}(\cdot|Q_1(H_{t_1}))$, where $Q_1(H_{t_1})$ is the temporal history of t_1 at VP.

Of particular interest is Effingham's claim about the value of $Ch_{t_1}(S^3)$. Since $Q(H_{t_1})$ is nomically possible and S^3 entails a contradiction, it follows immediately from my (classical) urchance axioms that

$$\text{urch}_{\text{VP}}(S^3|Q(H_{t_1})) = \text{urch}_{\text{VP}}(\varnothing|Q(H_{t_1})) = 0. \tag{3.1}$$

(Alternatively, we could of course derive this result more constructively via Boundary Markov: identify a thickly bounded region intersecting all closed causal curves—for example, a space-like hypersurface between t_2 and t_3. Then, using Boundary Markov, calculate the chance of S^3 conditional on all states of the region and its thick boundary compatible with t_1's temporal history—given Effingham's implicit assumptions, there are just two such states, corresponding to p_2 either exiting or not exiting Exit. Those calculations will yield chance 0 in both cases. By Multiplication, it then follows that the chance of S^3 is 0 conditional on t_1's temporal history.)

Horacek, meanwhile, is interested in the chances at times between wormhole entry and exit—his key example involves the chance of an attempted "autoinfanticide" succeeding shortly before the attempt. For an analogue of this, pick a time t^* between t_2 and t_3—i.e., after p_2 has exited the wormhole, but before it enters the scattering field. Working in the temporalist framework (cf. Chapter 1), Horacek assumes that even times intermediate to Opening and Exit have a privileged temporal history—for the sake of argument, let's grant that this is true. With the simplifying assumption that the wormhole ends are instantaneous (rather than persisting over long periods, as in Effingham's presentation[2]), we can then represent the situation in VP, with t^*'s privileged temporal history, via a 2D spacetime diagram as follows:

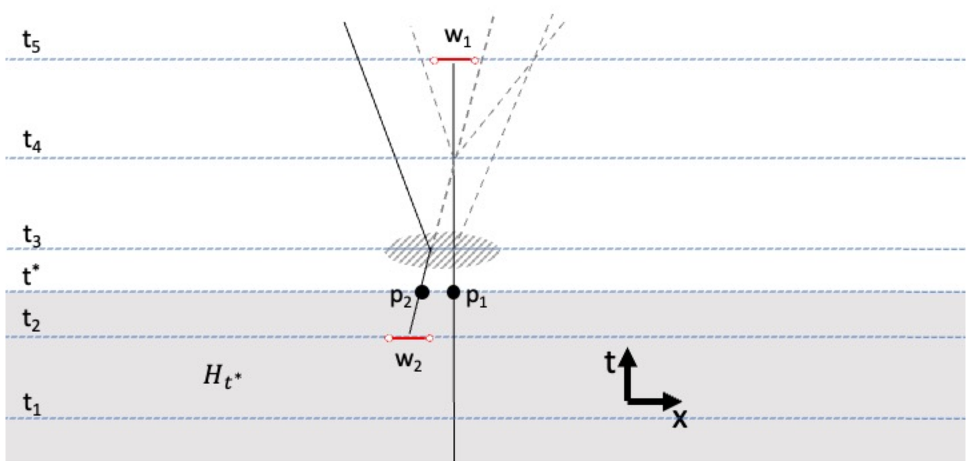


Figure 3.2: A spacetime diagram of Effingham's case with instantaneous wormholes.

The diagram also indicates alternative future trajectories resulting from veering, non-veering, and collisions.[3]

[2]This simplifying assumption could be dispensed with, at the didactic cost of a 3D diagram that would (given my artistic limitations) be significantly harder to read.

[3]The indicated hypothetical collision between p_1 and p_2 is evidently inelastic. This is in keeping with Effingham's own diagram: for in an elastic collision of two particles of identical masses, the (spatial) angle, in any given frame, between outgoing 3-velocities would have to equal the angle between incoming 3-velocities—which it does not in Effingham's diagram. (Momentum conservation gives $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}'_1 + \mathbf{v}'_2$ and energy conservation gives $|\mathbf{v}_1|^2 + |\mathbf{v}_2|^2 = |\mathbf{v}'_1|^2 + |\mathbf{v}'_2|^2$. Squaring the first equation and subtracting the second gives $2\mathbf{v}_1 \cdot \mathbf{v}_2 = 2\mathbf{v}'_1 \cdot \mathbf{v}'_2$.)

Let $Q^*(H_{t^*})$ be H_{t^*}'s intrinsic state in VP, and let V_1 [V_2] be the proposition that p_1 [p_2] veers on its next encounter with the scattering field. The conjunction of $Q^*(H_{t^*})$, $\neg V_1 \wedge \neg V_2$, and VP's background geometry entails a contradiction. Since the conjunction of $Q^*(H_{t^*})$ and VP's background geometry does not, and my theory is a classical probability theory, we thus have

$$\text{urch}_{\text{VP}}(\neg V_1 \wedge \neg V_2 | Q^*(H_{t^*})) = \text{urch}_{\text{VP}}(\varnothing | Q^*(H_{t^*})) = 0. \tag{3.2}$$

(Again, the result can also be more constructively derived via Boundary Markov.)

3.3 Two Previous Answers

Horacek (2005) and Effingham (2020) disagree with these results. *Pace* eq. 3.2, Horacek's account entails that, in VP, the chance of $\neg V_1 \wedge \neg V_2$ at t^* is greater than 0. And *pace* eq. 3.1, Effingham's account entails that the chance of S^3 at t_1 is greater than 0.

Start with Effingham. He derives his answer from the following principle:

> "IMPOSSABLE [sic] CHANCE: For any [proposition] φ ... $\text{Ch}_t(\varphi)$ is equal to the value $\text{Ch}_t(\varphi^*)$ would take were Λ the case". (2020, 152, notation adjusted)

Here,

1. φ is a proposition describing a "(physically possible) outcome[]" (ibid.) in a time-travel scenario at t,

2. Λ is "the proposition representing whatever it takes for [a] qualitatively similar, non-time travel, scenarios [to occur]"[4] (ibid.), and

[4]Effingham says "to be possible" instead of "to occur". This seems to be an error: the non-time-travel worlds he considers are physically possible at the time-travel world. Hence, Λ, thus understood, is already true at the time-travel world. But it's clear that Effingham wants to evaluate the value of $\text{Ch}_t(\varphi^*)$ at the *non*-time-travel world.

3. φ^* is "a proposition describing [an outcome of a scenario at t] saliently similar to that described by φ except that Λ is the case" (ibid.).

In other words, "in a time travel situation the objective chance of something coming about is exactly what it would be of that something coming about in a [qualitatively] similar non-time travel situation" (ibid.). This is a version of the chance invariance or "stability" thought we've encountered in Chapter 1.

Effingham doesn't say what exactly makes a non-time travel situation sufficiently "similar" to a time-travel situation, nor when two outcomes "correspond". But we can follow his specific application of the principle to calculate the chance function Ch_{t_1} in VP. Let ν be the "normal veer chance" (2020, p. 151), i.e. the "the objective chance of a particle veering in an intrinsically similar field absent the presence of wormholes" (2020, 157-8)—what we'd call the *acyclic* chance of veering. According to Effingham (where the V_i are as specified above):

> "$Ch_{t_1}(S^1) = Ch_{t_1}(V_1)$, $Ch_{t_1}(S^2) = Ch_{t_1}(\neg V_1 \wedge V_2)$, and $Ch_{t_1}(S^3) = Ch_{t_1}(\neg V_1 \wedge \neg V_2)$. $P(\psi \wedge \varphi) = P(\varphi|\psi)P(\psi)$ in cases where ψ depends on φ hence, since V_2 depends on V_1, $Ch_{t_1}(S^1) = \nu$, $Ch_{t_1}(S^2) = \nu(1-\nu)$, and $Ch_{t_1}(S^3) = (1-\nu)(1-\nu)$." (158, n.3; notation adjusted to add time index to Ch).

So, if $\nu < 1$, then $Ch_{t_1}(S^3) > 0$ according to Effingham. For example,

> "[I]f $\nu = 0.2$ ... then $Ch_{t_1}(S^1) = 0.2$, $Ch_{t_1}(S^2) = 0.16$, and $Ch_{t_1}(S^3) = 0.64$." (158)

But S^3 involves (as conjuncts) both that p is struck at t_4 and that p is not struck at t_4. So, where P is the proposition that p is struck at t_4,[5]

$$Ch_{t_1}(P \wedge \neg P) \geq 0.64. \tag{3.3}$$

[5]The rule that $p(A) \geq p(A \wedge B)$ follows from conjunction elimination alone—a rule which even the sort of paraconsistent logics Effingham favors validate (see below).

According to Effingham's account, VP is possible. Hence, by eq. 3.3, positive counterlogical chances are possible. Meanwhile, on our account we evidently have

$$\text{urch}_{VP}(P \wedge \neg P | Q_1(H_{t_1})) = 0.$$

So much for Effingham. Horacek (2005) has his own stability principle, which he calls "Lawful Chance principle":

> "LC: Suppose $\text{Chance}_{tw}(A) = x$ and w' matches w in both laws and initial conditions at at t. Then $\text{Chance}_{tw'}(A) = x$." (p. 428)

Note that, for LC to have a fighting chance, A can't be about any aspects of t's past not entailed by the laws and initial conditions at at t. Similarly for any aspects of the world's background geometry. For according to temporalism, at any world w, where Q_{tw} is H_t's intrinsic state at w, $\text{Chance}_{tw}(Q_{tw}(H_t)) = 1$. So, where $Q_{tw'}$ and $Q_{tw''}$ are two different states of H_t at worlds w' and w'', respectively, temporalism entails $\text{Chance}_{tw'}(Q_{tw'}(H_t)) = 1$ and $\text{Chance}_{tw''}(Q_{tw''}(H_t)) = 1$. Clearly, even if you think that contradictions *sometimes* have positive chance, you'll accept that, generally, $\text{Chance}_{tw'}(Q_{tw''}(H_t)) < 1$. But if w' and w'' agree with respect to initial conditions and laws at t, this contradicts LC, which requires $\text{Chance}_{tw'}(Q_{tw'}(H_t)) = \text{Chance}_{tw''}(Q_{tw'}(H_t))$. Similarly, Horacek implicitly assumes that the world's background geometry has chance 1 at any time, which entails that A in LC cannot concern any aspect of that background geometry that's not entailed by the laws and initial conditions at t. Let's therefore grant Horacek a restriction of LC to cases where A is purely about non-geometrical aspects of t's future.

To derive Horacek's account of the veer chances at t^*, consider the time-travel-free world, VP', in fig. 3.3, with the same scattering laws as VP. In VP', p_1's and p_2's scatterings are probabilistically independent. Given an acyclic veering chance of $\nu < 1$, we thus have

$$Ch_{VP't^*}(\neg V_1 \wedge \neg V_2) = Ch_{VP't^*}(\neg V_1) \cdot Ch_{VP't^*}(\neg V_2) = (1-\nu)^2 > 0. \qquad (3.4)$$

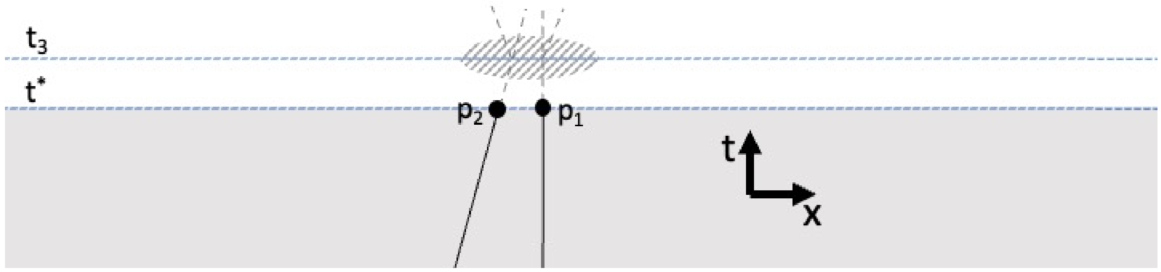


Figure 3.3: World VP′, with the same scattering laws as VP.

But VP and VP′ are intrinsically identical at t^*—in Horacek's terminology, they have the same "initial conditions" at that time. So, by LC and eq. 3.4,

$$Ch_{VPt^*}(\neg V_1 \wedge \neg V_2) = (1-\nu)^2 > 0, \tag{3.5}$$

in contrast to our eq. 3.2.

Horacek's account doesn't stop at eq. 3.5. His paper argues for the stronger conclusion that, in cases like VP, there is a positive chance of metaphysically impossible, since inconsistent, worlds. Specifically, where *autoinfanticide* is the permanent death of a baby at the hands of her older self, Horacek contends that his account "seems to be enough to guarantee the existence of a possible world where an autoinfanticide has a positive chance" (2005, 429), even though "autoinfanticide is metaphysically impossible", since worlds with autoinfanticide are not "internally consistent" (2005, 423).

We can arrive at this conclusion from eq. 3.5 as follows. Where GL specifies VP's background geometry and laws, we have, by Horacek's implicit assumption,

$$\begin{aligned} Ch_{VPt^*}(\neg V_1 \wedge \neg V_2) &= Ch_{VPt^*}(\neg V_1 \wedge \neg V_2 \wedge GL) \\ &= Ch_{VPt^*}(\neg V_1 \wedge \neg V_2 \wedge GL \wedge Q^*(H_{t^*})), \end{aligned} \tag{3.6}$$

where the second line follows from temporalism. Next, we grant an instance of the classical probability law that, if A (classically) logically entails B, then $p(A) \leq p(B)$: since $Q^*(H_{t^*}) \wedge GL$ (classically) logically entails $\neg P$, and $\neg V_2 \wedge Q^*(H_{t^*}) \wedge GL$ (classically) logically entails that P, $\neg V_2 \wedge GL \wedge \mathrm{Q}^*(H_{t^*})$ (classically) logically entails $P \wedge \neg P$;[6] so we infer

$$Ch_{VPt^*}(\neg V_1 \wedge \neg V_2 \wedge GL \wedge \mathrm{Q}^*(H_{t^*})) \leq Ch_{VPt^*}(P \wedge \neg P). \tag{3.7}$$

Eqs. 3.5, 3.6, and 3.7 together yield

$$Ch_{VPt^*}(P \wedge \neg P) > 0. \tag{3.8}$$

3.3.1 Effingham's Motivation

Instead of viewing them as a fatal blow to their theory, both authors embrace their conclusions.

Effingham suggests that the possibility of positive counterlogical chances is a natural consequence of another principle (cf. p. 152), "COUNTERMODAL PHYSICAL POSSIBILITY". According to this principle "([i]t's metaphysically possible that) for some φ, φ is metaphysically impossible but physically possible (where φ is about some concrete state of affairs ...)" (p.129). This principle, in turn, is supposed to be a natural consequence of a third commitment, namely "Countermodality". This is the combination of two ideas: that there are "[non-trivial] metaphysically impossible worlds" (p.118), and that counterfactuals with metaphysically impossible antecedents have non-trivial truth value (ibid.). Effingham thinks that Countermodality supports COUNTERMODAL PHYSICAL POSSIBILITY because, given Countermodality, it's natural to think that "[t]here are metaphysically impossible worlds *with the same laws of nature* as some metaphysically possible world" such that, in particular, "no laws of nature are breached at them" (132, my emphasis). If this is true, then the metaphysically impossible world is physically possible at the given metaphysically possible world, in virtue of the plausible principle that whenever a world has all the same laws as another world, then each is nomically possible at the other world.

[6]Note that $GL \wedge \mathrm{Q}^*(H_{t^*})$ already (classically) logically entails $\neg V_1$, hence the conjunct $\neg V_1$ isn't needed to classically derive $P \wedge \neg P$ from $\neg V_2 \wedge \mathrm{Q}^*(H_{t^*}) \wedge GL$.

To show that *logical contradictions* like $P \wedge \neg P$ are physically possible at VP, we must then establish that none of VP's laws are breached at the $P \wedge \neg P$-world. But on the ordinary understanding of "breach", a law L is breached at some world iff there's a truth A at that world such that $A \wedge L$ is a logical contradiction. So our $P \wedge \neg P$-world automatically breaches all laws, including VP's laws. Hence it is physically impossible at VP.

Effingham is aware of this, and responds by invoking non-classical logic. Specifically, he takes RM_3, a three-valued paraconsistent logic (Priest, 2008, 124-5), "to be the correct logic" Effingham (2020, p. 134). The three values are usually thought of as "true and only true", "false and only false", and "both true and false". The trivalence is supposed to help, according to Effingham, because now we can cash out "breach" in three different ways: φ is breached iff "the state of the world at t nomically determines that, at t^+, φ", and

1. ...φ is false and only false at t^+.
2. ...$\neg\varphi$ is true and only true at t^+.
3. ...it's not the case that φ is true and only true at t^+.

The laws at VP imply, for example, that $Q_1(H_{t_1}) \supset \neg(P \wedge \neg P)$. Is this law breached at the $P \wedge \neg P$-world? Dialetheists can say that the proposition $Q_1(H_{t_1}) \supset \neg(P \wedge \neg P)$ isn't merely false at the $\neg(P \wedge \neg P)$-world, but also still true. (She may endorse the general principle that necessarily, at a $P \wedge \neg P$-world, $P \wedge \neg P$ is both true and false.) In particular, then, neither is $Q_1(H_{t_1}) \supset \neg(P \wedge \neg P)$ false and only false, nor is its negation true and only true, at t^+. Hence, on the first and second interpretation of "breached", the law $Q_1(H_{t_1}) \supset \neg(P \wedge \neg P)$ remains unbreached at the $P \wedge \neg P$-world. Only on the third interpretation is it breached. About a similar case, Effingham notes that this is enough to ensure that "there's a perfectly fine disambiguation according to which [the laws at the time-travel world are unbreached at the contradiction world]" (136).

This is all wrong-headed. But before we dive into more critique, a note on Horacek's motivation.

3.3.2 A Note on Horacek

According to Horacek (2005) himself, he defends the same conclusion as Effingham: worlds that are not "internally consistent" (2005, 423) nonetheless (possibly) have positive chance. And we've seen how one can derive this conclusion from Horacek's principle LC. However, other things Horacek says contradict the conclusion.

Regarding the question of whether "autoinfanticide has a [positive] chance to succeed" (2005, 425) at some world w, Horacek disambiguates two readings. Where L_w is the conjunction of all laws at w and I_{tw} are the "initial conditions" at t (notation adjusted), the two readings are:

> "(1) Given (L_w & I_{tw}), what is $Ch_{tw}(\text{baby dies \& killer = baby})$?
>
> (2) Given (L_w & I_{tw} & killer = baby) what is $Ch_{tw}(\text{baby dies})$?" [ibid.]

Horacek takes (2) to be the correct reading of the question. About (1) he notes (in a related scenario) that the chance would be 0 since it is "the chance of an unsatisfiable conjunction" (2005, 425).

What does "Given..." mean here? If it's just ordinary Bayesian condititionalization, we should give the same answer to (2) as we give to (1): in particular, given (L_w & I_{tw} & killer = baby), $Ch_{tw}(\text{baby dies}) = 0$. This follows from the ratio formula provided only that, "given (L_w & I_{tw})", $Ch_{tw}(\text{killer = baby}) > 0$ which is evidently plausible since L_w and I_{tw} jointly leave it open that the killer doesn't succeed.[7]

That Horacek's "Given..." locution doesn't denote Bayesian conditionalization is further confirmed when he provides the following analogy:

[7] Throughout Horacek's discussion, "killer" really means "aspiring killer": "killer = baby" is *not* supposed to entail that the killing plan succeeds.

"Suppose A is a permanent bachelor, i.e., a man who never marries. What is the chance of A getting married? The question asked is either (3) or (4):

(3) Given (L_w & C_{tw}), what is $Ch_{tw}(A$ marries & A is a permanent bachelor$)$?

(4) Given (L_w & C_{tw} & A is a permanent bachelor$)$ what is $Ch_{tw}(A$ marries$)$?"
[ibid.]

Analogously to (1) and (2), he says that the answer to (3) is "zero" but that the answer to (4) may be "non-zero". Obviously, however, $Ch_{tw}(\text{A marries}|\text{A is a permanent bachelor})$ *is* zero. So Horacek's "Given ..." isn't Bayesian conditionalization.

A reading which makes sense of both of Horacek's examples is one where "Given ..." simply means "At a world where ...". It's clearly possible that, though A remains a permanent bachelor, at some point he had a non-zero chance of marrying. Similarly, one might think that, even if the aspiring killer is *in fact* the baby's future self, shortly before the murder there was a non-zero chance of him *not* being the baby's future self and of him succeeding at his murder attempt. On this account of the case, the correct answer to (2) is, indeed, "non-zero".

However, the account now fails to imply that *autoinfanticide* has positive chance. Autoinfanticide is, recall, a scenario where a future self kills her own baby self. On the present account, the worlds of non-zero chance *don't* contain any such scenario. Instead they only contain scenarios where someone *else* than the baby's future self kills the baby.[8] Horacek's desire to provide different answers to (1) and (2) is thus incompatible with the main claim of his paper. He contradicts himself here.

In the following, I'll understand under "Horacek's position" the position he *explicitly* advertises, i.e. that in some time travel worlds there are positive counterlogical chances.

[8]Cf. Smith (1997), who considers counterfactuals of the form "If I had succeeded in the murder, the victim would not have been me."

3.4 Problems Galore

A commitment to positive counterlogical chances generally, and Effingham and Horacek's implementations in particular, engender a flurry of problems. Here I'll list four.

3.4.1 Positive Counterlogical Chance Infects Ordinary Worlds

Their commitment to chance stability means that Effingham's and Horacek's positive counterlogical chances spread to time-travel-free worlds.

We've seen how Effingham's IMPOSSABLE CHANCE entails that $Ch_{t_1}(P \wedge \neg P) > 0$ (eq. 3.3), for P the proposition that p is struck at t_4. But now recall IMPOSSABLE CHANCE's letter: $\mathrm{Ch}_t(\varphi)$—the chance at t of a physically possible outcome φ in a time-travel scenario—equals "the value $\mathrm{Ch}_t(\varphi^*)$ would take were Λ the case", where Λ is "the proposition representing whatever it takes for [a] qualitatively similar, non-time travel, scenarios [to occur]", and φ^* is "a proposition describing [an outcome of a scenario] saliently similar to that described by φ except that Λ is the case" (*op. cit.*).

Let Λ be the proposition which specifies the same state of the world at t_1 as in VP but, embedded in a universe without time-travel-enabling wormholes. (Wormholes which connect points at the same time, or space-like separated points, might still be ok.) Now let VP^* be any world among the closest Λ-worlds to VP, and let φ be the proposition $P \wedge \neg P$. Since P merely says that p is struck at t_4, with no reference to the embedding geometry, $(P \wedge \neg P)^*$ is just the proposition $P \wedge \neg P$. But now, according to IMPOSSABLE CHANCE, $Ch_{VPt_1}(P \wedge \neg P)$ equals the chance $P \wedge \neg P$ would have at t_1 if Λ was the case. Since VP^* is among the closest Λ-worlds to VP, it follows—given standard world-theoretic truth conditions for counterfactuals—that

$$Ch_{VPt_1}(P \wedge \neg P) = Ch_{VP^*t_1}(P \wedge \neg P).$$

So, by Effingham's result, eq. 3.3,

$$Ch_{VP^*t_1}(P \wedge \neg P) > 0. \tag{3.9}$$

But, by stipulation, VP^* is a time-travel-free world, whose particle collision dynamics should thus be just like those in other time-travel-free worlds. In particular, they should involve no positive counterlogical chances. Eq. 3.9, then, is clearly false.

As for Horacek, we have already established that VP and VP' (fig. 3.3) have identical laws and initial conditions at t^*. Hence it follows immediately from eq. 3.8 and LC that

$$Ch_{VP't^*}(P \wedge \neg P) > 0.$$

So, both IMPOSSABLE CHANCE and LC entail that in time-travel-free worlds like VP' and VP*, logical contradictions have positive chance. But that's clearly false. So IMPOSSABLE CHANCE and LC are false.[9]

One might try to save IMPOSSABLE CHANCE and LC by altering the notion of "outcome" in IMPOSSABLE CHANCE (which LC could in principle adopt too) such that $P \wedge \neg P$ doesn't count as an outcome. This would be clearly *ad hoc*: we usually think that the conjunction of two physically compatible outcomes is an outcome—e.g. a coin's landing heads and its landing in my palm conjoin to the outcome of its landing heads in my palm. The fact that two outcomes are *logically* incompatible isn't a strong reason to break this rule if, as Effingham and Horacek have it, logical impossibilities can be physically possible.

But worse, the move settles Effingham (and Horacek) with an unacceptable dilemma: either classical probability theory is still violated in time-travel-free worlds or the *set-theoretic axioms* for probability—specifically, Additivity—are violated in *time-travel* worlds.

[9] The result for LC, though not the one for IMPOSSABLE CHANCE, relies on the assumption that VP is metaphysically possible. Horacek, though somewhat ambivalent (p. 433), admits he'd see the impossibility of time travel in indeterministic worlds as "quite dramatic". In any case, there are good reasons to think that time travel, and VP in particular, *are* metaphysically possible (cf. Ch. 5). Moreover, Ch. 1 offers an attractive account compatible with time travel's metaphysical possibility, all while avoiding Effingham's and Horacek's problems.

The latter would make it hard to recognize chance as a form of probability: even fans of non-classical probability theory will want to be able to construe probability functions as additive over *some* algebra, even if it's a "non-classical" set algebra, i.e. one whose union and complement operations aren't (sentential or propositional) conjunction and negation.

The dilemma arises as follows. Recall that, according to Effingham, S^1, S^2, and S^3 are the only members of Ch_{VPt_1}'s sample space. Moreover, S^1, S^2, and S^3 all verify $\neg P$, i.e. $\{S^1, S^2, S^3\} = \neg P$, and S^3 verifies P, i.e. $\{S^3\} \subseteq P$. By (set-theoretic) Additivity, it follows that

$$Ch_{VPt_1}(P) \geq Ch_{VPt_1}(S^3) = (1-\nu)^2,$$

and

$$Ch_{VPt_1}(\neg P) = Ch_{VPt_1}(S^1) + Ch_{VPt_1}(S^2) + Ch_{VPt_1}(S^3) =$$
$$= \nu + \nu(1-\nu) + (1-\nu)^2 = 1.$$

But both P and $\neg P$ are logically possible. So, on the amended version of IMPOSSABLE CHANCE, there'll be a time-travel-free world VP^* such that $Ch_{VP^*t_1}(P^*) = (1-\nu)^2$ and $Ch_{VP^*t_1}(\neg P^*) = 1$. But, since $\nu < 1$,

$$Ch_{VP^*t_1}(P^*) + Ch_{VP^*t_1}(\neg P^*) \geq (1-\nu)^2 + 1 > 1,$$

in violation of classical probability theory at VP*.[10] So, either Additivity is false at VP or classical probability theory is false at VP*. (The same argument works for LC and time t^*, provided S^1, S^2, and S^3 are still the only members of Ch_{VPt^*}'s sample space, with $Q^*(H_{t^*}) = \{S^2, S^3\} \subseteq \neg P$ and $\{S^3\} \subseteq P$.)

[10] One arrives at the additional conclusion that the *set-theoretic axioms* of probability are violated at VP* on the natural assumption that $Ch_{VP^*t_1}$ doesn't assign any positive probability to impossible worlds.

3.4.2 No Support for Countermodality

Countermodality, recall, is Effingham's claim that there are non-trivial metaphysically impossible worlds, and that some counterfactuals with metaphysically impossible antecedents have non-trivial truth value. His argument for the position goes like this. Where the "Paradoxer" is someone who thinks that time travel is metaphysically impossible, Effingham writes:

> "[T]he Paradoxer thinks there are cogent, sensible things to be said about how the world would be were the metaphysically impossible to obtain. They therefore rally against the traditional eschewing of taking such questions about impossibility seriously. Instead, the Paradoxers will side with more recent research which accepts metaphysically impossible worlds and the non-trivial truth values of counterpossibles". (p. 118)

The thought seems to be this: to convince her opponent, the Paradoxer can't just say anything—she will want to provide "cogent" arguments, based on things that are "sensible" even to her opponent. Naively, one might think that she achieves this as long as her argument is sound. But if Countermodality was false, all counterpossibles would be trivially true. Making a *sound* argument would then be extremely easy, from the perspective of the Paradoxer. For example, she'll think the following argument sound:

TRIVIAL:

1. If time travel were to occur, then $1 + 1$ would equal 3.
2. If 1., then time travel is metaphysically impossible.

$\therefore$ Time travel is metaphysically impossible.

This argument is valid,[11] the second premise is true since $1 + 1$ necessarily equals 2, and the first premise is true given the Paradoxer's views and the falsity of Countermodality.

But we never observe the Paradoxer making an argument like TRIVIAL. If she believed that all sound arguments are cogent and sensible, this would suggest that the Paradoxer thinks the argument unsound. But the argument is valid and the second premise near indisputable; so the Paradoxer must think that the first premise is false—and thus, that Countermodality is false.

The fallacy in this argument is evident: it's *not* the case that all sound arguments are cogent and sensible—nor should the Paradoxer believe this. To be cogent, an argument must also make an (reasonable) opponent *believe* in the argument's soundness. TRIVIAL clearly doesn't meet this bar: its first premise is dangerously close to question-begging, and thus extremely unlikely to be believed by anyone who thinks that time travel is metaphysically possible.

(Compare this to an argument like the following:

I. If time travel occurred, then it would be the case that I could kill my grandfather at t and that I couldn't kill my grandfather at t.

II. If I., then time travel is metaphysically impossible.

$\therefore$ Time travel is metaphysically impossible.

Here, premise I. has at least a chance of being believed by someone who isn't already convinced of time travel's metaphysical impossibility, perhaps after some reflection on what time travel means for human ability.)

So, the fact that the Paradoxer doesn't make an argument like TRIVIAL isn't evidence that she thinks that the argument is unsound and thus Countermodality is true, because that fact is already explained by the argument's being (close to) question-begging. So

[11] Validity follows by any normal modal logic for "metaphysically possible" conjoined with the rule $\vdash (A \mathrel{\Box\!\!\rightarrow} \neg B) \supset \neg(A \mathrel{\Box\!\!\rightarrow} B)$, which any standard logic of counterfactuals validates.

Effingham's diagnosis of the dialectical situation is fallacious: the dialectic is entirely compatible with everyone believing in the trivial truth of counterpossibles.

3.4.3 No Concrete Replacement

Neither Effingham nor Horacek provide a concrete replacement for classical probability theory.

What's the structure of the set algebra on which chance is defined? Presumably it's an algebra of propositions. But what are its join and complement operations?

Perhaps the propositions are sets of (possible and impossible) worlds, in which case the join and complement are set-theoretic intersection and complement. In this case, the nature of these operations, and their relationship to propositional conjunction and negation, depends on what impossible worlds are included—on which non-empty[12] non-classical logic regiments truth-at-impossible-worlds. But I am not aware of any cogent defenses of a particular non-classical logic to this effect. Indeed, there is a good argument for rejecting that any one non-empty non-classical logic could do the job: it would give rise to an unprincipled distinction between those impossibilities which occur at impossible worlds and those which don't (Nolan, 1997, pp. 546-8).

These questions about the structure of the underlying set algebra are critical, because surely any half-way promising theory of chance must posit *some* regularities about how chance plays with propositional conjunction and negation. [13]

Defenders of a non-trivial counterpossible semantics often tolerate the lack of specificity resulting from not defining a logic to regiment truth-at-impossible-worlds by leaning on our linguistic intuitions about counterpossibles. These might serve to indirectly—and vaguely and context-dependently—specify a set of impossible worlds (Nolan, 1997). This

[12] Don't say that *any* set of propositions is a world—if propositions themselves are sets of worlds, this is inconsistent on cardinality grounds!

[13] In principle, one could additionally tinker with the set-theoretic probability axioms themselves. But, as I noted earlier, this would only obscure the concept of chance even further.

may or may not be appropriate for the purpose of counterpossible semantics. But it is certainly insufficient for an account of fundamental chance, whose domain is neither inherently vague nor context-dependent.

The lack of a well-formed theory about the structure of chance's domain also destabilizes other areas of inquiry, like modal logic. Plausibly, whatever has positive chance is physically possible. But if so, positive counterlogical chances require a non-classical modal logic of physical possibility, one which allows for logical contradictions to be possibly true. Neither Effingham nor Horacek provide one.

3.4.4 Obscuring the Concept of Chance

Finally, it's natural to understand probability theory as a generalization of ordinary propositional reasoning. Indeed, this is how Popper (1968) develops it. Popper's approach is well motivated: the idea of probabilistic reasoning as an extension of ordinary propositional reasoning is arguably core to our conception of what probability *is*. But our ordinary propositional reasoning is *classical* propositional reasoning. So Effingham and Horacek must abandon this vital link: they can't hold probabilistic reasoning to be an extension of ordinary propositional reasoning.

Effingham and Horacek must also reject the canonical epistemic role of chance, as an expert which we necessarily expect to be at least as accurate as we are (cf. Levinstein (2023)). In fact, chance will look quite defective to us in time travel worlds. Probability measures satisfy the normalization axiom, i.e. the demand that the chance of the entire sample space is 1. Following Effingham (2020, Ch.12.3.1-2), let $\not\perp$ be the "proposition that no metaphysical impossibilities come about" (p.154). Suppose, with Effingham, that it's possible that urchance, for some metaphysically possible A, assigns a positive value to a logical impossibility conditional on A. Given normalization, this entails that the result of conditioning urchance on A is accuracy-dominated, over the set of metaphysically possible worlds, by the result of conditioning it on $A \wedge \not\angle$. So in worlds where time travel is possible, urchance tends to leave free accuracy on the table. By contrast, insofar as you're rational, you won't, under any circumstances, assign positive credence to the proposition that metaphysical impossibilities occur—after all, they are impossible! So the possibility of positive counterlogical chances conflicts with the usual understanding of chance as an expert expectedly at least as accurate as we are.[14]

Effingham's and Horacek's proposals leave us with an underspecified, ill-understood concept of chance, threatening to rear its ugly head even in nearby non-time-travel worlds. Simultaneously, their proposals are also undermotivated. It is best to leave them by the wayside.

[14] Effingham (2020, Ch.12.3.2) suggests replacing the standard Principal Principle, of which he provides the following special case:

$$Cr_0(A|E) = ch(A),$$

where E is the conjunction of all evidence admissible for ch (which, on the standard view, presumably includes the proposition that ch is a chance function). He suggests the following replacement:

$$Cr_0(A|E) = ch(A|\not\angle).$$

The present argument shows that the replacement doesn't do justice of the usual understanding of chance's epistemic role.

Part II

Causation

Chapter 4

Counterfactual Dependence Is Not Sufficient for Causation

Introduction

More than half a century ago, David Lewis published his seminal paper "Causation" (1973a). To this day, the project of analyzing actual causation in terms of counterfactual dependence remains alive and, by many accounts, well. The variety of approaches within this project are thought to be united by one common principle:

> "The currently most prominent approaches to defining actual causation are those within the counterfactual dependence tradition, which started with Lewis (1973a). *All of these approaches take as their starting point the assumption that counterfactual dependence is sufficient for causation*, but not necessary (Hitchcock (2001); Woodward (2003); Hall (2004a; 2007); Halpern & Pearl (2005); Halpern (2016); Weslake (2015) ...)." (Beckers & Vennekens, 2017, p. 2, my emphasis)

The italicized principle—the sufficiency of counterfactual dependence for causation—is appealing. If I hadn't hit the bullseye, I wouldn't have won. If my laptop didn't have enough juice, I'd soon be sitting in front of a black screen. If you weren't reading this sentence, you wouldn't know what it says. So, one concludes, my hitting the bullseye *causes* my winning, my laptop's ample charge *causes* its continued operation, your reading the sentence *causes* your knowing what it says. Generalizing: an event counterfactually depends on another only if it is caused by it.

Whether counterfactual dependence is sufficient for causation doesn't merely concern the metaphysician of causation. Causation play important theoretical roles all over philosophy, and hence a divorce of causation and counterfactual dependence—which I'll advocate for here—echoes far. At least since Kripke (1980), people have thought that causation plays an important role in reference fixing, where it's widely regarded key to resolving issues of semantic indeterminacy. But, one might ask: is it really *causation* which plays this role or—as an informational semanticist like Dretske (1981) would hold—*counterfactual dependence* which plays this role? Similarly, about theories of perception: is the relevant relation, connecting perceiver and the perceived, *causation* (as in Grice (1961)) or *counterfactual dependence* (as a natural development of Lewis's (1980) would suggest)? Finally, in causal decision theory, should we really care about our actions' *causal* impact, or instead about what would be the case if we performed them (cf. Hedden (2023))? Cases we'll discuss here will bear on these issues.

So far we've worked with an overly simplistic representation of our target position. Virtually nobody believes the sufficiency thesis as we have stated it. A more careful version adds room for various caveats:

> **Sufficiency:** Necessarily, if (c, e) is a suitable pair of actual events such that e wouldn't have occurred if c hadn't occurred, then c causes e.

The "necessarily" modal witnesses the fact that counterfactual reductivists consider the sufficient condition part of a *definition* of causation. The definition's caveats are accommodated by the restriction to "suitable" event pairs, which fends off easy counterexamples.

For example, my hitting the bullseye doesn't *cause* my hitting the board's center circle—they are the same event—yet if I hadn't hit the bullseye, I wouldn't have hit the board's center circle. One standard requirement on "suitability" is therefore that c and e be "distinct" events, meaning here that neither is a part of the other, and that neither's occurrence logically entails the other's occurrence (cf. Lewis (1973a; 1979)). We'll soon see other suitability requirements.

Another remark concerns the context sensitivity of subjunctive conditionals. Lewis (1979, p. 458) famously distinguishes between "standard" and "backtracking" contexts: suppose that earlier today you returned Susy's book, as you promised you'd do. Given that Susy is known to take promises seriously, the following conditional seems true:

(1) If I hadn't returned the book to her today, Susy would be disappointed in me.

Given (1), **Sufficiency** entails that your returning the book was a cause of your staying in Susy's good graces. So far so good. But now consider that you're extremely reliable and honest, known to never break a promise. With this in mind, you might have reasoned as follows:

(2) If I hadn't returned the book today, that would have been because Susy and I agreed on a later return date to begin with. So Susy wouldn't have been disappointed in me.

You could reasonably assert either (1) or (2), but not their conjunction. This indicates a context shift between (1) and (2). Following Lewis (1979), call the interpretation of the conditional triggered by (1) the "standard" interpretation and the interpretation triggered by (2) the "backtracking" interpretation. Given (2), **Sufficiency** entails that your returning the book today is a cause of your agreeing on today as the return date. But that's absurd: you have no such retrocausal powers. For this reason, it's important that the conditional in **Sufficiency** always be evaluated on its standard interpretation.

This paper argues that **Sufficiency**, and those counterfactualist approaches which rely on it, face an existential threat: the possibility of *synchronic laws*—laws which relate simultaneous distinct events[1] already explained by their individual dynamical histories. In Section 4.1, I explore previous challenges to **Sufficiency** and why, despite them, versions of **Sufficiency** have endured. In Section 4.2, I then provide two examples of synchronic laws: the first is Gauss's law of Maxwellian electrodynamics; the second involves constraints imposed by closed time-like curves on their pasts. I argue that each undermines even weak versions of **Sufficiency**. Section 4.3 argues that this immediately rebuts several reductivist theories, including Lewis (1973a); Hall (2007). After proving that, on the standard counterfactualist interpretation (due to Hitchcock (2001)), prominent reductions of causation to structural equations entail a version of **Sufficiency** (Appendix H), I conclude that these accounts fail too. No extant counterfactualist reduction of causation stands up to scrutiny.

4.1 Previous Arguments Against Sufficiency

Sufficiency has faced previous challenges. For example, some have argued that *omissions* aren't causes (e.g. Beebee, 2004): Julius Caesar's failure to water my plant isn't a cause of the plant's death. Yet, to **Sufficiency**'s apparent discredit, there is counterfactual dependence: if Caesar had watered my plant, it wouldn't have died. In reply, the **Sufficiency** lover may bite the bullet and insist that Caesar's botanical omission *is* a cause of my plant's death, and she might try to blunt the bullet's impact in various ways: she might attempt to partially reduce omissions to "positive" events (commissions) (Bernstein, 2014), or to explain the appearance of non-causation as mere infelicity (Schaffer, 2005). But, failing that, she can retreat and strengthen her notion of "suitability": she may stipulate that an

[1] Or relativistically speaking: not causally related events.

event pair is suitable only if its first element is a *positive* event.[2]

A demand for causal proportionality poses another challenge to **Sufficiency**. I greet my neighbor loudly, startling her. My *greeting loudly* causes her being startled, but my greeting *simpliciter* doesn't—my neighbor isn't *that* jumpy. Yet, if I hadn't greeted her *simpliciter*, my neighbor wouldn't have been startled. So counterfactual dependence isn't sufficient for causation. One might resist this line by reinterpreting the demand for proportionality: perhaps disproportionality incurs mere infelicity, not falsity—"A causes B" carries a pragmatic implicature that A is a *maximally specific* cause of B. Or one could try to separate causation from explanation and shift the burden of proportionality over to the explanatory side (cf. Weslake (2017)). But, failing that, the **Sufficiency** advocate can retreat and strengthen her notion of "suitability": she may stipulate that suitable event pairs consist of *proportional* events.

A third challenge emerges from David Lewis's own semantics for counterfactuals in deterministic worlds. On his miracles-based semantics, intended to model the subjunctive conditional on its "standard" resolution, the near past would still be different if the present was. According to Lewis, counterfactual antecedents are preferentially brought about by *small* miracles (cf. (Lewis, 1979)), but small miracles need time to snowball into big change. So, where antecedents dictate big differences to actuality, they imply a significant delay—a "ramping" period (Bennett, 2003)—from the occurrence of the miracle to the antecedent event. During this ramping period, the counterfactual world differs from actuality. This yet again raises the specter of retrocausation.

To illustrate by example: you throw a ball at me; I notice just in time and catch it. If I hadn't caught the ball, surely that wouldn't be because at the moment of impact a miracle instantly twisted my arm away from the ball. Instead, some macroscopic change would have occurred—perhaps I would have noticed the ball a little later, or you threw it a littler

[2]This restriction could also be subsumed under a ban on "overly disjunctive" events, as e.g. (Lewis, 1986b, D) discusses; see below.

harder, or a gust of wind deflected the ball outside of my reach. According to Lewis, any of these changes would be brought about by a small miracle—e.g. changes in neuronal firing patterns, or microscopic meteorological changes—which subsequently needs time to effect the big change. But surely my catching the ball causes neither my actual neuronal firing pattern nor the earlier atmospherical state.

Lewis (1979) offers a response. **Sufficiency** says that c causes e if e fails to occur in all closest worlds where c fails to occur. But (the thought goes) c might fail in a variety of ways in these worlds. Lewis hopes that, if e is in c's past, then some of those ways are bound to preserve e. That is, Lewis hopes that $\neg O(e)$ leaves open the specific content of the counterfactual ramping period:

> "[W]e should sacrifice the independence of the immediate past to provide an orderly transition from actual past to counterfactual present and future. That is not to say, however, that the immediate past depends on the present in any very definite way. There may be a variety of ways the transition might go, hence there may be no true counterfactuals that say in any detail how the immediate past would be if [some given event hadn't occurred]."[3] (Lewis, 1979, p.463)

This response works for our example. As we saw, there are all sorts of reasons I might not have caught the ball—decreased alertness, a harder throw, an altered wind pattern. The hope is that, for any actual positive event e^* preceding c, there is a closest possible way of filling out the ramping period in which e^* still occurs. Then $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e^*)$ is false for any such e^*.[4]

[3]The original quote ends with "...if the past were different". This isn't what Lewis needs here and makes the quote less plausible than my substitution. In some contexts, "if the past were different" will select very specific immediate pasts. For example, it might select the non-occurrence of an omission: "I didn't do my homework yesterday. But in a different past—i.e., *if the past were different*—I would have done it." (The original quote thus doesn't permit the first response to Vihvelin's (1995) objection below.)

[4]The hope requires a counterfactual semantics which permits multiple maximally close antecedent worlds. On a semantics like Stalnaker's (1968), in which a given counterfactual antecedent selects a *unique* closest antecedent world, there's a unique closest ramping period. This guarantees that some positive event e^* preceding c satisfies $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e^*)$. In response, the **Sufficiency** advocate should insist that, even

Vihvelin (1995) identifies two kinds of threats to this response. The first arises when the antecedent event is an *omission*: an omission's non-occurrence is just the occurrence of a positive event and thus typically entails a definite ramping period. But the weakening of **Sufficiency** to positive events is immune to this threat. The second worry stems from highly detailed past events. Let c be *my catching the ball at* t, for some time t, and let e be the totality of events during an open time interval with future boundary t. Given that c's non-occurrence would require a miracle to occur shortly before t, it requires the non-occurrence of e, and so we have $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$. But my catching the ball doesn't *cause* e.

One possible response follows Lewis (1986a) in banning overly "fragile" events. Those are events with extremely detailed essences—intuitively, events which could have very easily failed to occur. Lewis's justification for the ban is that our standard way of denoting event propositions, "standard nominalizations", isn't nearly detailed enough to pick out these events. Inspired by Lewis's move, the **Sufficiency** advocate may thus strengthen "suitability" yet again, additionally requiring that e be *robust*, i.e. not *overly fragile.*

Lewis's response is unlikely to convince everyone: it doesn't prove that *all* robust positive events can fail in a variety of equally-close ways. Moreover, Lewis doesn't independently characterize the distinction between the robust and the fragile. Some counterfactual reductivists have therefore proposed an alternative: retreating to a variant of **Sufficiency** in which the counterfactual conditional doesn't require ramping periods. I'll explain this alternative strategy in the following footnote.[5] Those in the structural

though the conditional is true, it's not *determinately* true. She may then retreat to a weakened version of **Sufficiency** which requires *determinate* truth of $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e^*)$ for causation. All relevant counterfactuals we'll encounter plausibly meet this bar, so our arguments will be unaffected by this.

[5] Glynn's (2013) account is an example of this. It's a combination of two ideas. First, it stipulates a technical meaning of the counterfactual conditional—let's denote it $\mathrel{\blacksquare\!\!\rightarrow}$ —according to which, when A is a proposition purely about particular matters of fact at the instant (or brief interval) t, "$A \mathrel{\blacksquare\!\!\rightarrow} B$" is evaluated using only miracles *at* t. That is, $A \mathrel{\blacksquare\!\!\rightarrow} B$ is true iff B is true at all worlds closest among those where (i) A is true, (ii) no miracles occur outside of t, (iii) everything prior to t is as it actually is, and (iv) everything after t evolves according to the actual laws of nature. Second, say that c *causes* e if there is some truth T solely about t such that $\neg O(c)$ and T are metaphysically compossible and $\neg O(c) \wedge T \mathrel{\blacksquare\!\!\rightarrow} \neg O(e)$. Intuitively, the role of T

equations tradition might have yet other means at their disposal to solving the ramping problem, e.g. via judicious choices of variables. We'll see that none of these responses will work against my objections. For the purpose of the essay, let us grant that some response to the ramping problem exists.

A *fourth* challenge results from *rejecting* counterfactual miracles, in favor of a view on which counterfactual worlds have the same laws as the actual world. If the laws are deterministic, this requires the counterfactual worlds to be different at all times, including *past* times (cf. Bennett, 1984; Loewer, 2007; Albert, 2015; Dorr, 2016). Now, in a world like ours, with continuous laws, the past will generally only have to differ *microscopically*, until very close to the antecedent's time (cf. Dorr (2016)). Still, ubiquitous retrocausation is implausible even if the effects are microscopic. In my view, the cleanest way for the **Sufficiency** advocate to circumvent the problem is not to get entangled in it to begin with: instead, they should stick with Lewis's miracles-based semantics (or variants of it which avoid ramping periods). Indeed, this is what the counterfactual dependence tradition in fact does (cf. Lewis (1979); Hitchcock (2001); Hall (2007); Glynn (2013)).[6]

is to suppress any unwanted effects which the $\neg O(c)$-realizing miracle would otherwise bring about—*viz.* consequences which affect e via causal routes bypassing c. Generally, T requires a highly complex miracle. For example, my bus is stuck in traffic, and my being on the bus right now (c) causes my being late to the meeting (e). Moreover, if I wasn't on the bus right now, I'd be on my bike ($\neg O(c)$), speeding through grid-locked traffic, and arriving on time. On the technical meaning of the counterfactual conditional introduced above, if I was on my bike right now, this would be because a miracle had quasi-instantly teleported me from the bus onto the bike. But such a miracle would have all sorts of undesired byproducts, affecting my arrival time independently of my newly acquired ability to cycle there. For concreteness, suppose the miracle would leave me extremely startled—so much so that I'd crash, thus not arriving on time. Still, we want to say that my being on the bus (c) causes my being late (e). Glynn's account achieves this as follows. Some T entail that I'm calm and not disoriented, thus negating the unintended consequence. Moreover, Glynn makes it plausible that we can find a T to do this for all unintended consequences (including consequences that arise from the miracles needed to bring about the various suppressors). If so, then if $\neg O(c) \wedge T$, then I'd arrive on time despite the sudden teleport. The account thereby secures the desired causal relation (that my being on the bus causes my being late) without the need for counterfactual ramping periods. The idea would then be to change **Sufficiency** as follows:

> **Sophisticated Sufficiency:** Necessarily, if (c, e) is a suitable pair of actual events such that, for some truth T, $\neg O(c) \wedge T \;\blacksquare\!\!\rightarrow \neg O(e)$, then c causes e.

Importantly, however, **Sophisticated Sufficiency** still falls to my counterexample, as I'll explain in fn. 32.

[6] Alternatively, the **Sufficiency** advocate could strengthen her notion of "suitability", restricting it to pairs whose second element (the putative effect) is a *macroscopic* event. The downside is that this is vulnerable to

A final constraint on suitability—obvious enough that it is often left implicit—is that (c, e) is suitable only if some counterfactuals with antecedent $\neg O(c)$ are false. A sufficient condition (indeed, on the standard view, a *necessary and sufficient* condition) for this is that c's occurrence is metaphysically *contingent*. Without this constraint, **Sufficiency** makes causation too cheap—events whose non-occurrences trivialize all counterfactuals needn't be omnicauses: for example, if *any* antecedents trivialize counterfactuals, logical contradictions are surely among them, yet some future region's being identical to itself doesn't cause (say) the Big Bang.

Sufficiency emerges weakened but still makes substantive predictions: the canonical examples of causation, which also motivate counterfactual reductions, tend to involve positive, proportional, not overly fragile, and contingent events—stone throws, hurling boulders, hurricanes, poisonings, and the like. So, the proponent of **Sufficiency** might still think of herself as occupying a true and substantive position. Unfortunately, as I'll argue, that appearance is illusory: even in its weakened form, **Sufficiency** is false. This is because, in the presence of *synchronic laws*, counterfactual dependence ceases to track causal dependence.

4.2 Synchronic Laws

4.2.1 Warm-Up: Mirror World

Before I get to the two main examples, I'll illustrate the concept of a synchronic law with a particularly simple toy example. Imagine a variant of John Hawthorne's (2007) *Mirror*

cases with non-continuous laws, where any counterfactual differences would, as a matter of law, have to be macroscopic. For an easy example, imagine a universe with Newtonian gravity but discretized masses. In response, the **Sufficiency** advocate could argue that the case for a fixed-law counterfactual semantics is weaker in such worlds, because one of its key motivations—that only *microscopic* adjustments to the past are required to accommodate a counterfactual antecedent—no longer applies. Be that as it may—since miracle-based semantics are standard within the counterfactualist tradition, I'll just stick with those.

World: a world split into two halves which, *by law*, are mirror images of each other.[7]

> **Mirror Law:** At all times, the matter configuration in one half is a mirror image of the matter configuration in the other half.

Additionally, each side obeys the same deterministic dynamics—say, Newton's equations.[8] I'll call a law like Mirror Law *synchronic*: it nomically ties simultaneous events even though each is already fixed by its own local past.

If you accept that the setup is possible, there's a simple argument against **Sufficiency**. Suppose I clap my hands at time t (and so does my mirror image). First premise: since Mirror Law is a law, the following counterfactual is true:

> (**M**) If I hadn't clapped my hands at t, my mirror image wouldn't have clapped at t either.

Given (**M**) (plus the plausible assumption that my and my twin's hand-clappings form a suitable event pair), **Sufficiency** implies that my clapping at t *causes* my mirror image's clapping at t. Second premise: my clapping at t *doesn't* cause my mirror image's clapping at t. Rather, my mirror image's clapping at t is already fully caused by events in its local past (via Newton's laws), and there is no causal overdetermination. The two premises entail that **Sufficiency** is false.

Of course, you might deny that Mirror World is possible. You might think that Mirror Law couldn't be an extra nomic layer on top of Newton's equations: rather, it is the result of symmetric initial conditions which are merely accidental. Or you might think that Mirror Law *could* be a law, but only through through non-local causation: it has to enforce its nomic status through action-at-a-distance. Luckily, we need not decide on the specific

[7] The difference to Hawthorne's example consists in the nomic necessity of the connection between the sides. See also Luzon (2024).

[8] That is, Newton's Second Law together with some local force law.

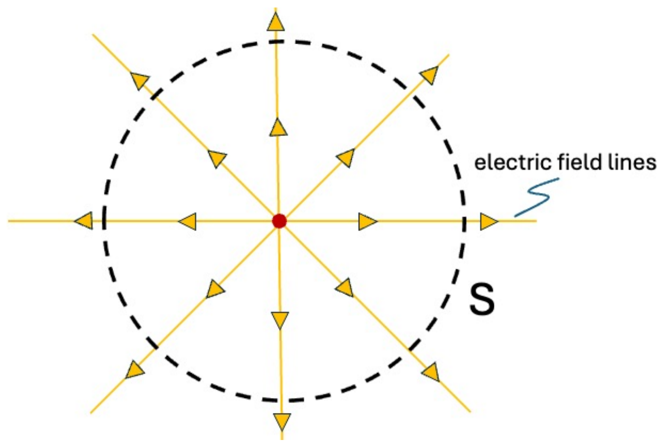


Figure 4.1: A 2D sketch of (three-dimensional) GAUSS

case of Mirror World. For in the following, I show that some worlds, with a nomic structure relevantly similar to Mirror World's, are clearly metaphysically possible.

4.2.2 Gauss's Law

Consider a world—GAUSS—with a single stationary proton at spatial location x.[9] Besides the particle, there is a static electric field, radiating outward from the proton—see fig. 4.1 for a 2D sketch.[10] Everything is governed by Maxwell's laws of electrodynamics. Otherwise the world is empty.

Maxwell's laws entail the following synchronic law:[11]

> **Gauss's Law:** At all times, the *electric flux* through the boundary of any (spatial) volume is proportional to the total electric charge enclosed within the volume.

Intuitively, the electric flux through a boundary is the difference between how much electric field, at the boundary, points *out* of the enclosed volume, versus how much points

[9] In the following, any references to spatial location, time, sphericalness, isotropy, and other Lorentz non-invariant properties, are relative to a fixed reference frame co-moving with the proton.

[10] To avoid discontinuity in the resulting electric field, assume that the proton has some very small but non-zero spatial extent and that the electric field density falls off continuously near x.

[11] Gauss's Law simply *is* one of Maxwell's laws. In its mathematically equivalent divergence form, it says that the electric field's spatial divergence at any location is proportional to the charge density at that location.

into the enclosed volume.[12] The electric flux through a boundary supervenes entirely on the electric field configuration *at* the boundary: no change in electric flux through a boundary without changing the electric field at the boundary.

Gauss's Law now says that, if the total electric charge inside some volume is positive (negative)—that is, there is more (less) positive electric charge inside the volume than negative electric charge—then the electric flux through the volume's boundary is positive (negative). Equally, if the total electric charge is zero—positive and negative charges are exactly equal—the electric flux is zero. Intuitively, Gauss's Law requires that the electric flux through the boundary "mirror", as it were, the value of the total electric charge within—similar to how Mirror Law requires that my counterpart's actions mirror mine.

In addition to Gauss's Law (and its analogue for the magnetic field) Maxwell's laws comprise two *diachronic* laws.[13] They make it so that, in a Maxwellian universe, everything is nomically determined by its local past—specifically, by (any spatial cross-section of) its *past light-cone*. Intuitively, an event's past light-cone is the union of all possible space-time trajectories via which a material particle could reach the event in question.[14]One implication of these laws (together with Gauss's Law) deserves highlighting:

> **Charge Conservation:** Electric charge is conserved—charge is neither spontaneously created nor annihilated.[15]

Here is now an argument against **Sufficiency**. Pick some arbitrary time t, and some

[12]Mathematically, the electric flux through an oriented spatial surface is the integral, over the surface, of the scalar product of electric field and the surface's normal vector.

[13]They are *Faraday's Law*—relating the magnetic field's time derivative to the electric field's spatial curl—and *Ampère's Law*—relating the electric field's time derivative to the magnetic field's spatial curl and the electric current. The details won't matter here.

[14]Throughout I'll thus understand an event's "light-cone" as a subset of the *manifold*, not its tangent bundle. Moreover I mean its *solid* light-cone—a 4D region—not merely its boundary. Thus what I call "past light-cone" physicists also commonly call "causal past". Since the latter terminology is potentially confusing in a discussion of the *philosopher's* concept of causation, I avoid it.

[15]Formally, Charge Conversation says that the net *electric current* through the boundary of any volume equals the (temporal) change in electric charge inside the volume. This is an immediate consequence of Gauss's Law and Ampère's Law (cf. fn. 13).

(a) (b)

Figure 4.2: Spatial snapshots of two charge-free solutions to Maxwell's laws (2D sketches). Figure (a): zero electric field throughout the universe. Figure (b): uniform electric field throughout the universe.

arbitrary (hollow) spatial sphere S, at all times centered on the proton. The following counterfactual is true:

> **(G_1)** If the proton wasn't present at x at t, there would be no charged particles at t.

But Charge Conservation would still hold at all times after t. So, from $(\mathbf{G_1})$,

> $(\mathbf{G_2})$ If the proton wasn't present at x at t, there would be no charged particles present *at all times after t*; in particular, S would enclose zero total electric charge at all times after t.

Likewise, Gauss's Law would still be true. So, from $(\mathbf{G_2})$:

> $(\mathbf{G_3})$ If the proton wasn't present at x at t, there would be *zero* electric flux through S at all times after t.

See fig. 4.2 for sketches of two solutions with zero electric flux through S. Since electric flux supervenes on electric field, in both cases the electric field is different from actuality. (For the purpose of the argument, we needn't settle here which of the two—if any—is closest to actuality.) Given $(\mathbf{G_3})$, **Sufficiency** implies that the proton's presence at x at t *causes* the positive electric flux through S at all times after t. But S was arbitrary here. In particular, if we choose some time t^+ after t, we can let S be so large that no light signal sent from the proton at x at t could reach S by t^+. It follows that the proton at t has faster-than-light causal influence on the electric field at S at t^+. But that's false. So **Sufficiency** is false.

To put this argument more formally, let p_G be the proton's presence at x at t and let e_G be the electric flux through S and t^+ being positive. In symbols, $(\mathbf{G_3})$ thus reads:

$$\neg O(p_G) \mathrel{\Box\!\!\rightarrow} \neg O(e_G).$$

The core of the previous argument can be validly expressed as follows (where the modalities are metaphysical):

1_G. If **Sufficiency** is true, (p_G, e_G) is a suitable pair of occurring events at GAUSS, $\neg O(p_G) \mathrel{\Box\!\!\rightarrow} \neg O(e_G)$ in GAUSS, and GAUSS is possible, then p_G causes e_G in GAUSS.

2_G. (p_G, e_G) is a suitable pair of events at GAUSS.

3_G. **Dependence**$_G$: $\neg O(p_G) \mathrel{\Box\!\!\rightarrow} \neg O(e_G)$ in GAUSS.

4_G. **Possibility**$_G$: GAUSS is possible.

5_G. **Non-Causation**$_G$: p_G does *not* cause e_G in GAUSS.

$\therefore$ **Sufficiency** is false.

Premise 1_G follows from the fact that a necessary truth is true in all possible worlds. What about Premise 2_G? We've seen several demands the **Sufficiency** advocate might place on "suitability". p_G and e_G are certainly *distinct*, occurring as they do in separate spacetime regions. They are also *proportional* to each other, both being simple configurations of physically fundamental properties. Likewise, neither event is overly "fragile": both the proton's being located at x at t and the electric flux through S at t^+'s being positive are specifiable by ordinary nominalizations (as we just did), and so aren't "fragile" in Lewis's intended sense. More generally, particle locations and electric fluxes seem like prime examples of physical facts we'd like to *causally explain*—any theory of causation which excluded them would be seriously incomplete.

I'll now turn to premises $3_G - 5_G$.

4.2.2.1 Defending *Dependence*$_{\mathbf{G}}$

We've already laid out the positive argument for **Dependence**$_G$: starting from $(\mathbf{G_1})$, we appeal to **Charge Conservation** to get to $(\mathbf{G_2})$, and then via **Gauss's Law** to $(\mathbf{G_3})$ (which is what **Dependence**$_{\mathbf{G}}$ says).

In conversations I've encountered two objections, one concerning the appeal to **Gauss's Law** in the last step, and one against the very first step, the assumption of $(\mathbf{G_1})$. Both objections claim that I'm misconstruing what goes on in the counterfactual, proton-less world. I'll take them in turn.

The first objection holds that, if the proton wasn't present at x at t, the electric field would be 0 at x at t, but *otherwise unchanged* (except perhaps for some local smoothing); thereafter a sphere of vanishing electric field, centered on x, would expand at light speed. This restores local counterfactual dependence, as the electric field outside of (t, x)'s future light-cone would remain unchanged at any time.

Because there would be no charges after t, this objection violates **Gauss's Law** at all times after t. It thus posits a temporally infinite counterfactual miracle, spanning t's entire future. However, the objector may reply, the miracle's infinity by itself can't be the problem. For **Dependence**$_G$ *also* requires an infinite counterfactual miracle, a spatial one: to *preserve* Gauss's law in the absence of charges, the electric field would have be different from actuality at every spatial radius from x at t. So an infinite miracle is required either way. What's still problematic, though, is that the counterfactualist tradition often explicitly rejects miracles reaching beyond the antecedent's time of occurrence. For example, Lewis (1979) wants the counterfactual world, post antecedent, to evolve

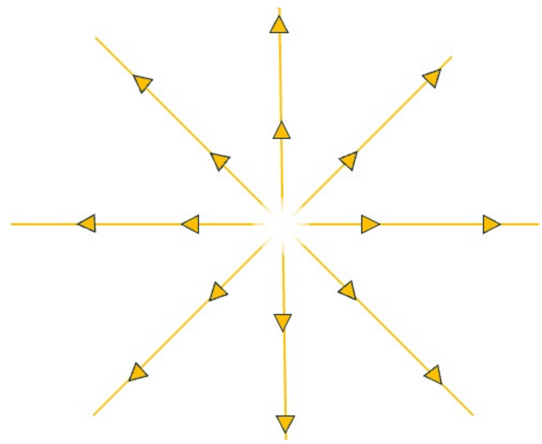


Figure 4.3: The future of the universe if the proton is deleted at (t, x) and Gauss's Law suspended forever after.

according to the actual laws of nature.[16] Glynn's (2013) account even explicitly confines miracles to the antecedent's time of occurrence. So, a temporally infinite future miracle seems incompatible with the letter of many counterfactualist approaches in a way that a spatially infinite miracles don't.

But there is a more devastating problem for the present objection: the envisioned scenario—zero electric field at (t, x), with a sphere of vanishing field expanding at light-speed thereafter—has no basis at all in Maxwell's laws. Once we suspend **Gauss's Law**—as the counterfactual scenario demands—the remaining three Maxwell equations entail that the electric field would simply be static forever after t. That is, with its configuration at t being as in figure 4.3, it would maintain that configuration forever after t. To see this, one has to dig into the equations, which I'll do in the following footnote.[17] The upshot is that

[16]"Often we do have the right sort of supposition [namely one involving a particular time], the standard resolution of [subjunctive] vagueness, and no extraordinary circumstances [like time travel, precognition, or tachyons]. Then Analysis 1"—which explicitly requires counterfactual worlds to evolve according to the actual laws after the antecedent time—"works as well as we could ask." (Lewis, 1979, p. 464)

[17]From the fact that **Charge Conservation** holds at all times after t and the fact that there are no charges present at t, it follows that there are no charges present at all times after t. (To make this step formally airtight, let t be an arbitrarily small open interval—we are free to make this choice because the presence of a proton at x *throughout* t is a *bona fide* event, just as much as if t was an instant. As a temporally differential equation, **Charge Conservation** holds *at* t's future boundary $\bar{t}^+$ only if it holds in some small open interval t^+ around $\bar{t}^+$; and with no charges present throughout t, this is the case only if there are no charges present throughout t^+. So, where t is an open interval, **Charge Conservation** in t's future ensures that an open subinterval of t's future contains no charges, and thus that t's entire future is charge-free.) *A fortiori*, there

the imagined scenario is physically unmotivated.

Ok, but what about the alternative scenario just described? Couldn't the world be as in fig. 4.3 forever after t? This would still avoid non-local counterfactual dependence: the electric field outside of (t, x)'s future light-cone would be unchanged. Now, granted, it inherits the same problems about future miracles as the previous objection does. But its conclusion can at least be motivated from the remaining Maxwellian laws. Unfortunately, however, this scenario is independently problematic for the counterfactualist. For it rules out non-local counterfactual dependence only at the price of ruling out *virtually all* counterfactual dependence. Outside of perhaps a small neighborhood around x, the electric field is counterfactually invariant to the proton's presence. This makes it hard to see how a counterfactualist account of causation could deliver the desirable result that the proton's presence at x at t is a cause of the electric field *in* (t, x)*'s future light-cone*.[18]

Onto the second objection. It proposes to reject $(\mathbf{G_1})$ in favor of the following subjunctive:

> $(\mathbf{G_1^*})$ If the proton wasn't present at x at t, there'd be some positive charge elsewhere at t.

One variant of this objection has the proton located somewhere else at t. Another variant has it replaced by a tiny hollow sphere of positive charge, centered on x at t, and expanding at light speed thereafter. (The latter makes a sphere of vanishing electric field, expanding at light speed, consistent with **Gauss's Law**.) But either variant already gives up on **Sufficiency**. For $(\mathbf{G_1^*})$ entails

is no *electric current* at any time after t. Moreover, by hypothesis, the electric field is radially symmetric around x at all times, and hence its curl vanishes at all times. Since the magnetic field vanishes prior to t, Faraday's law ($\text{curl}(\mathbf{E}) = -\frac{\partial \mathbf{B}}{\partial t}$) entails that it always vanishes. Hence, via Ampere's law for vanishing current ($\text{curl}(\mathbf{B}) = \frac{\partial \mathbf{E}}{\partial t}$), the electric field is static after t.

[18] Now, of course, everyone already concedes that counterfactual dependence isn't necessary for causation. But necessity violations are confined to cases of preemption or overdetermination—cases where, besides a given actual cause, there is a backup (actual or non-actual) cause, ready to cause the effect in the absence of the former. The present case is not like that.

$(\mathbf{G_1^{**}})$ If the proton wasn't present at x at t, it would not be the case that x's complement is neutrally charged everywhere at t.

From $(\mathbf{G_1^{**}})$ and **Sufficiency** (and the suitability of the relevant event pair[19]), it thus follows that the proton's presence at x at t causes x's complement's charge neutrality at t. But that's false—given Maxwellian electrodynamics, (subluminal) electric charges only affect their future light-cones (see also 4.2.2.3). So, **Sufficiency** is false. The objector's attempt to replace $(\mathbf{G_1})$ falls flat.

4.2.2.2 Defending *Possibility*$_{\mathbf{G}}$

I take it for granted that Maxwell's laws of electrodynamics are metaphysically possible (and possibly laws), as is the existence of a single proton and a static electric field. But are these things also *com*possible? A nomic reductivist might complain: from the perspective of a Humean best-system account of lawhood—the most popular nomic reductivist approach—Maxwell's laws are too complicated to be laws at GAUSS. Due to its simplicity, the actual particle-field configuration in GAUSS can be specified by a very economical system of propositions, which purchases much more strength than Maxwell's laws for comparable (or less) complexity.

But the Humean's worries are easily assuaged: simply add to GAUSS complex electrodynamical systems astronomically far away from our particle, with kinematics compatible with Maxwell's laws. The Humean will agree that, in this new world, Maxwell's laws are laws. Yet, the argument against **Sufficiency** goes through just as smoothly in the new world as in GAUSS.

[19] The pair is (the proton's being present at x at t, x's complement's being charge-neutral at t). This is clearly suitable: the events are distinct, they are proportional to each other, the former is positive, and the latter not overly detailed.

4.2.2.3 Defending *Non-Causation*$_G$

Non-Causation reflects received scientific opinion on Maxwellian electrodynamics. Two representative quotes, both from standard textbooks, on Maxwellian electrodynamics:

> "The displacement between causally related events is always timelike." (Griffiths, 1981, p. 531)
>
> "[I]f any change takes place in one of the interacting bodies, it will influence the other bodies only after the lapse of a certain interval of time. *It is only after this time interval that processes caused by the initial change begin to take place in the second body*." (Landau & Lifschitz, 1994)

There is a powerful argument for the received opinion: similarly to Mirror World, the actual electric field configuration over S at t^+ is completely determined by (any spatial cross-section of) its past light-cone—a light-cone which doesn't contain the proton-at-t. Thus, to deny Premise 4 would be to stipulate ubiquitous and theoretically costly (since non-local) causal overdetermination, without explanatory payoff.

4.2.3 Time Loops

So much for Gauss's Law. Our second example involves synchronic laws of a different character, grounded in part in the world's global spacetime geometry. The geometrical feature in question are *time loops*—roughly, trajectories through spacetime which travel back in time to their starting point.[20]

Consider a world full of marbles, some grey and others white. Upon contact, the marbles fuse with each other, and the shade of the outgoing marble (i.e., the fusion product) is determined by the shade of the incoming marbles, as follows:

[20]In relativistic spacetimes, these are *closed causal curves* (where "*causal*" has its technical physics meaning: *having everywhere either time-like or null tangent vector*).

- If exactly an **odd** number of incoming marble are white, the outgoing marble is white.

- Otherwise, the outgoing marble is grey.

Outside of fusions, a marble's shade is always preserved. Here is a 2D sketch of a world abiding by these rules (with fusion events indicated by jagged bubbles):

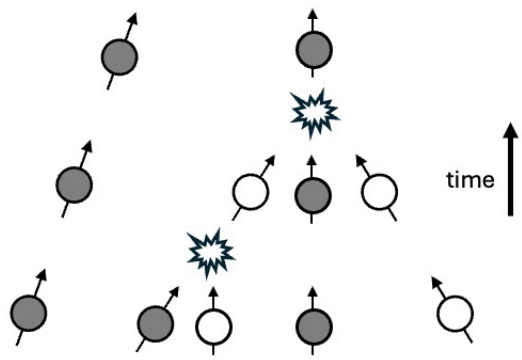


Figure 4.4: Collisions in a marble world.

We can construct a *wormhole* by starting with an ordinary spacetime and then "identifying" two regions of space at distinct times. More specifically, the resulting topology is such that any trajectory which enters the earlier region from the past exits at the later region, and any trajectory which enters the later region from the past exits at the earlier region. A 2D illustration of both sort of processes:[21]

[21] Depending on the initial spacetime structure, the choice of spatial regions may be highly non-unique. For example, if the underlying spacetime is Minkowskian, then any pair of (non-intersecting) duplicate bounded space-like surfaces with the same boundaries generates the same wormhole. In this case, the surfaces indicated in the sketch represent but one arbitrary choice of such a pair. For ease of illustration, we'll stick with that pair throughout. Note also that, to preserve manifoldness (*viz.*, local Euclidean topology and differential structure) the regions' (2D) boundaries are also removed (this is not explicitly indicated in the sketches).

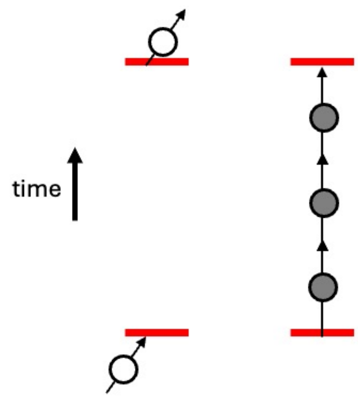


Figure 4.5: Trajectories through wormholes.

Now consider LOOP, a marble world with a wormhole, as follows:

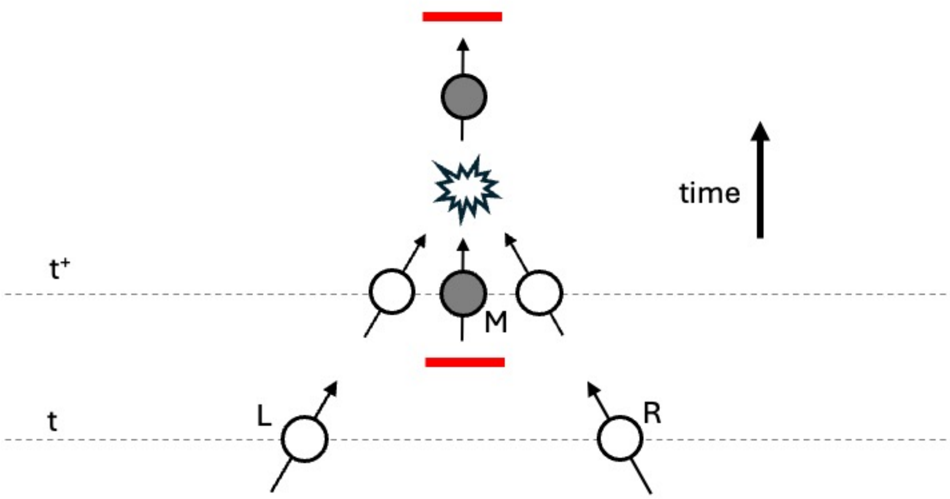


Figure 4.6: A sketch of LOOP.

That is, two white marbles approach the region between the wormhole ends. As they pass the earlier end, a grey marble exits. The three marbles eventually collide, fusing into a single marble, later entering the later wormhole end. Let's introduce three binary variables L, R, and M, to represent the shades of the left and right marble at t and the shade of middle marble at t^+, respectively (0 = grey, 1 = white). So, actually, $L = R = 1$ and $M = 0$.

The fusion law implies the following relationship (where $\overline{\phi}$ abbreviates $(1-\phi)$):

$$M = LRM + \overline{LR}M + \overline{L}R\overline{M} + L\overline{RM} \tag{4.1}$$

The equation has exactly four solutions:

L	R	M
0	0	0
0	0	1
1	1	0
1	1	1

Table 4.1: The four solutions of LOOP's fusion law

So, we have the following law:

> **Agreement:** The left marble is white iff the right marble is white, and grey iff the right marble is grey.

Like Gauss's Law and Mirror Law, Agreement is a synchronic law.

Again, we have a straightforward argument against **Sufficiency**, as follows. Since Agreement is a law, the following subjunctive holds:

> **(L)** If the left marble had been grey at t ($L = 0$), the right marble would have been grey at t ($R = 0$).

But the actual events ($L = 1$ and $R = 1$) clearly form a suitable event pair—they are distinct, proportional, $L = 1$ is positive, and $R = 1$ isn't overly detailed. So, by **Sufficiency**, the left marble's being white at t *causes* the right marble's being white at t. But that's false. So, **Sufficiency** is false.

As a valid argument:[22]

[22] Where the modalities are again metaphysical.

1_{L}. If **Sufficiency** is true, $(L = 1, R = 1)$ is a suitable pair of occurring events at LOOP, $\neg O(L = 1) \mathrel{\Box\!\!\rightarrow} \neg O(R = 1)$ in LOOP, and LOOP is possible, then $L = 1$ causes $R = 1$ in LOOP.

2_{L}. $(L = 1, R = 1)$ is a suitable pair of occurring events at LOOP.

3_{L}. **Dependence**$_{\mathrm{L}}$: $\neg O(L = 1) \mathrel{\Box\!\!\rightarrow} \neg O(R = 1)$ in LOOP.

4_{L}. **Possibility**$_{\mathrm{L}}$: LOOP is possible.

5_{L}. **Non-Causation**$_{\mathrm{L}}$: $L = 1$ does *not* cause $R = 1$ in LOOP.

$\therefore$ **Sufficiency** is false.

Premise 1_{L} again follows from the fact that a necessary truth is true in all possible worlds. Premise 2_{L} we've already covered. Let's take the remaining premises in order.

4.2.3.1 Defending *Dependence*$_{\mathrm{L}}$

Dependence$_{\mathrm{L}}$ follows from the conjunction of three counterfactual conditionals:

(a) If $L = 0$, the positions and velocities of all particles at t would be the same.

(b) If $L = 0$, then the dynamics in t's future would be the same.

(c) If $L = 0$, then the spacetime structure in t's future would be the same.

By the identity rule (i.e., $\vdash A \mathrel{\Box\!\!\rightarrow} A$) and Agglomeration,[23] (a), (b), and (c) jointly entail

(d) If $L = 0$, then it would be that $L = 0$, the positions and velocities of all particles at t would be unchanged, as would be the spacetime structure and dynamics in t's future.

[23] I.e., $(A \mathrel{\Box\!\!\rightarrow} B) \wedge (A \mathrel{\Box\!\!\rightarrow} C) \vdash (A \mathrel{\Box\!\!\rightarrow} (B \wedge C))$, part of any standard logic of counterfactuals, including Lewis (1973b) and Stalnaker (1968).

But (d)'s consequent logically entails that $R = 0$. Hence:

(d) If $L = 0$, then it would be that $R = 0$.

Why believe premises (a)–(c)? Start with (a). The **Sufficiency** lover can't plausibly hold that, if $L = 0$, then some particle's position or velocity would be different. For she would then be committed to the particles' current position or velocity being *caused* by $L = 1$—an implausible overdetermination we're seeking to avoid.

Premise (b) follows from canonical formulations of a miracle-based semantics (a semantics which, recall, **Sufficiency** advocates should and do embrace). On the canonical formulation due to Lewis (1979)—as well as alternatives like Glynn (2013)—miracles are confined to *no later* than the time of the antecedent.[24] This is for good reasons: miracles placed after the antecedent time can easily prevent relevant effects. In 1983, Stanislav Petrov prevented a nuclear war by correctly judging an incoming missile warning to be a false alarm. But we can easily generate a counterfactual world in which Petrov's judging otherwise nonetheless didn't lead to nuclear war, e.g. by suppressing the signal from nuclear button to missiles. If this world is among the closest button-pressing worlds, counterfactual dependence accounts of causation will struggle to vindicate that Petrov's correct judgment prevented nuclear war. But it's hard to see principled reasons for permitting post-antecedent miracles in LOOP which wouldn't also carry over to this case.

The argument for Premise (c) is much the same, for topology changes have similar history-altering power as miracles do—the signal in the nuclear cable could be swallowed up by an aptly placed temporary singularity in the cable. Again, it's hard to see a principled reason for avoiding this if post-antecedent topology changes were allowed in LOOP.

[24]Now, Elga (2000) has shown (in my view conclusively) that Lewis's (1979) particular "hierarchy of importance" fails to produce the desired asymmetry of miracles. So I'm focusing here on what Lewis professes to rather than actually delivers—that is, I grant that *some* semantics in Lewis's spirit (e.g. one on which the Past Hypothesis is a law) produces the desired asymmetry of miracles. There is no analogous problem with Glynn's (2013) semantics, which confines miracles to exactly the time of the antecedent.

4.2.3.2 Defending *Possibility*$_L$

As far as worlds with time loops are concerned, LOOP is nothing special: *if* time loops are metaphysically possible, so is LOOP. But I think there are strong reasons to think that time loops *are* metaphysically possible, and little reason to think they aren't.

The strongest reason for the possibility of time loops is the existence of well-understood spacetime models that contain closed time-like curves (time loops in the language of Lorentzian manifolds). One way to articulate this argument is via positive conceivability—roughly, conceivability that doesn't merely involve the absence of contradiction but also the presence of a "positive picture" of the scenario (Chalmers, 2002). The positive conceivability-possibility link holds that if a situation can be positively conceived, it is metaphysically possible. (Emphasis on *positive* conceivability sidesteps counterexamples that trouble simpler links; for instance, while the falsity of unprovable mathematical truths—Goldbach's conjecture, perhaps—may be naively conceivable, their truth is nonetheless necessary.) Fully interpreted spacetime models[25] with closed time-like curves are paradigm cases of positive conceivability, presenting as they do precise and detailed positive pictures.[26] This is a strong positive reason for the possibility of time loops.

In addition to considering conceivability in the abstract, note that the Einstein equations themselves have solutions with closed time-like curves. In Kurt Gödel's (1949) famous example (remarkably, homeomorphic to $\mathbb{R}^4$), suitably accelerated material bodies can travel along closed time-like curves. Other well-known examples of spacetimes with closed time-like curves include ones with rotating black holes (Kerr solutions, cf. Carter (1968)) and Van Stockum's rotating dust cylinders (van Stockum, 1938).[27] Physicists

[25] Such as Gödel (1949), Carter (1968), or van Stockum (1938).

[26] "Fully interpreted" is doing work here. For consider the debate on haecceitism in spacetime: do diffeomorphically equivalent Lorentzian manifolds represent genuinely distinct possibilities? (See Norton *et al.* (2023) for an overview of the debate) This question arises because it's not fully settled what swaps of mathematical points are supposed to represent. But the representational aspects of LOOP which matter to us—the meaning of spacetime trajectories looping back to their origin—*are* fully interpreted. That's all we need.

[27] See also Kajari *et al.* (2004).

have studied the physics of time loops even outside the general relativistic context (e.g., Echeverria *et al.*, 1991b; Deutsch, 1991; Novikov, 1992). Usually, we are suspect of *a priori* attempts to dismiss of serious scientific literature as concerned with the metaphysically impossible.

Moreover, initial worries *against* time loops have now been, to my mind, convincingly refuted. One of the more influential worries has traditionally stemmed from paradoxes involving *ability*. Let *autoinfanticide* be the act of a future self's (permanently) killing her own infant self. You can't possibly commit autoinfanticide. But if time loops are possible, then (it seems) you *can* possibly commit autoinfanticide: simply travel back in time, gun in hand. It would thus seem that, if time loops are possible, then a contradiction is possibly true: that you both can and can't commit autoinfanticide. So time loops aren't possible.

I find the standard reply to this, due to Lewis (1976), convincing. Seemingly contradictory yet individually acceptable utterances often indicate a *context shift*. So it is here: what you "can" do is highly context-sensitive. To quote Lewis's example: compared to a (non-human) ape, I *can* speak Finnish: I have sufficiently developed articulators. But compared to a Finnish speaker, I *can't* speak Finnish: I don't know any Finnish vocabulary or grammar. Similarly, I *can* kill the infant, insofar as "I have what it takes": I have a loaded gun, I'm a good shot, etc. But I *can't* kill the infant, considering that the infant is my younger self. There's no contradiction here: the second statement is evaluated at a different context.

Another objection concerns the "bootstrapping" aspect of time loops, with some arguing that their inexplicability renders them impossible (Al-Khalili, 1999). However, as Lewis (1976, p. 148) notes, inexplicability does not imply impossibility. The universe's initial state (if it has one), outcomes of stochastic processes, and God are all arguably inexplicable, yet metaphysically possible. For other versions of the no-bootstrapping worry, and rebuttals against them, see also Effingham (2020, Ch. 5.2.2).

4.2.3.3 Defending *Non-Causation*$_L$

Non-Causation$_L$ is motivated by a similar thought as its GAUSSian analogue: the right particle's shade at t is already fully causally explained by its shade at preceding times. To posit additional causal influence from the *left* particle at t would be to posit non-local causal overdetermination.

The situation in LOOP adds an additional twist. If you deny **Non-Causation$_L$**, you must think that present causal facts hinge on what happens in the far future—in particular, on whether the future contains time loops or not. This violates a plausible constraint on causation: at least where time has a linear order, what causes what up to a time is intrinsic to the world's history up to that time. More precisely:

> **Weak Intrinsicness:** Let w and w' be worlds with the same laws and with identical (linear) histories up to time t.[28] Then, for any events c and e in w occurring up to time t, if it's true at w that c causes e, then it's true in w' that c causes e.

Others before me have defended similar principles. Hall (2004a) proposes the following stronger principle:

> **Intrinsicness (Hall):** Let world w contain S, a "structure of events that consists of e, together with all of its causes [in S] back to some arbitrary earlier time t" (ibid., p. 239). Let c be some cause of e in w. Then, if w' has the same laws as w and contains S,[29] then c causes e in w'.

[28] Here "identical" means *numerically identical*—i.e., w and w' overlap up to t. One can also formulate a version of this principle in terms of qualitative duplication, which will be friendly to those who think that worlds don't overlap. That principle will just be slightly more cumbersome to state, but the relevant conceptual content the same.

[29] Hall also considers further strengthenings, where w' merely contains a structure "similar" to S. These are, of course, subject to the same counterexamples as the current principle.

Hall's Intrinsicness principle is stronger because it merely requires that w and w' share a small subset of e's history, namely e's causes up to some prior time t. As Hall (2007) elsewhere points out, his own principle produces awkward results in certain canonical scenarios. But my weaker principle avoids these results, as I explain in the following footnote.[30] Hall's cases thus undermine **Intrinsicness** but not **Weak Intrinsicness**.

But accepting **Weak Intrinsicness**, you should accept **Non-Causation$_{\mathbf{L}}$**. For consider the following loop-free world, sharing LOOP's history up to shortly after t:

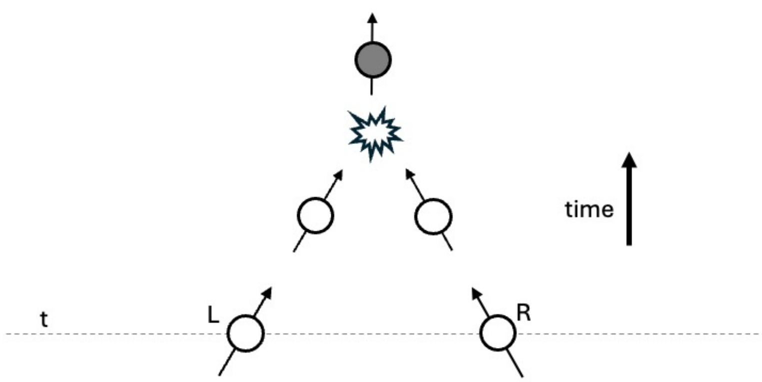


Figure 4.7: A loop-free collision world.

In this world, $L = 1$ obviously doesn't cause $R = 1$. But then it follows by **Weak Intrinsicness** that $L = 1$ doesn't cause $R = 1$ in LOOP either.

[30] Consider *Switching*. A cable carries current to a lamp. I can flip a switch that would divert the current along a second cable to the same lamp. Clearly my choice to leave the switch alone is not a cause of the lamp's remaining lit. Now compare a twin world where the second cable is grounded and only the first cable still reaches the lamp. There, leaving the switch alone *does* cause the lamp to light. But because the lamp, the first cable, and the switch are physically identical across the two worlds—and constitute the causes of the lamp's lighting in the twin world—Hall's **Intrinsicness** principle incorrectly rules that my inaction is a cause in both. The weaker principle avoids that mistake because it evaluates the whole temporal history, including which cables are connected. Similar comments apply to cases of *Threat Cancelation* (cf. Hall (2007)).

4.3 Troubles for Counterfactualist Accounts of Causation

4.3.1 Against Lewis (1973a) and Hall (2007)

So, there can be non-causal counterfactual dependence, even between distinct (and otherwise suitable) events. What does that mean for counterfactualist reductions of causation?

Firstly, any account which entails **Sufficiency** should be rejected. At least two such accounts come to mind. The most famous is Lewis (1973a). It says that c causes e iff there is a chain of actual events $d_1, ..., d_n$ with $d_1 = c$ and $d_n = e$ such that, for all $i = 1, ..., n-1$, (d_i, d_{i+1}) is a suitable event pair and $\neg O(d_i) \mathrel{\Box\!\!\rightarrow} \neg O(d_{i+1})$. In particular, then, if $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$ for a suitable pair (c, e) of actual events, c causes e, and so Lewis's account entails **Sufficiency**.

Now, our argument against Lewis's account joins the ranks of many previous objections raised against it—notably its failure to handle cases of late preemption and symmetric overdetermination. However, our argument also applies to successor theories which handle these cases. One of these successors is Hall (2007). According to it, c causes e iff (c, e) is a suitable pair of actual events and there is a "*reduction*" of the actual world in which c counterfactually depends on e. It needn't concern us what exactly a reduction is;[31] what matters here is that every world counts as a reduction of *itself* (Hall, 2007, p. 127). Thus we have again, that, where (c, e) is a suitable pair of actual events, c's counterfactually depending on e is sufficient for c's causing e—that is, we have **Sufficiency**. So Hall's (2007) account should be rejected too.[32]

[31] Just to give a flavor: roughly, it's a situation in which zero or more parts of the world that are actually in a "non-default" (or "deviant") state adopt their default state instead, while the rest is unchanged.

[32] Earlier I discussed Glynn's (2013) account (fn. 5). It's easy to see that it, too, succumbs to our two counterexamples. The counterfactual situation in GAUSS only has "late" miracles anyway—disappearing the proton and changing the field values exactly at t—with everything before and after t evolving according to the actual laws. So we straightforwardly have $\neg O(p_G) \mathrel{\blacksquare\!\!\rightarrow} \neg O(e_G)$.

As for LOOP, the existence of a global time order is a prerequisite of Glynn's account; so, for argument's sake, let's grant that we can identify times across the two strands of spacetime. Moreover, shrink A to a single point, so that it's part of a single time t; and let B occur strictly after t. In evaluating $A = 0 \mathrel{\blacksquare\!\!\rightarrow} \dots$ according to Glynn, we then only consider counterfactual worlds with miracles at t. But in the closest

4.3.2 Against Model-Theoretic Definitions of Causation

In part due to issues with preemption and overdetermination, the traditional Lewisian approach has nowadays largely been supplanted by more sophisticated analyses of causation in terms of *structural equation models* (SEMs). An SEM uses variables to represent sets of events and so-called "structural equations" to represent determination relations between these events. Structural equations are expressions of the form $\ulcorner X := f_X(Y_1, ..., Y_n)\urcorner$, where $X, Y_1, ..., Y_n$ are variables taking real values and f_X is a function from n-tuples of real numbers into real numbers. For any argument Y in which f_X is non-constant, the equation is interpreted to say that Y's value (partially) *determines* X's value—hence the asymmetric symbol ":=". Formally, an SEM is a pair $(\mathcal{V}, \mathcal{E})$ of a set of variables $\mathcal{V}$ and a set of structural equations $\mathcal{E}$ in those variables. Given a set $\mathcal{E}$, the set $\mathcal{V}$ bipartitions into a set $\mathcal{V}_{\text{ex}}$ of *exogenous* variables—those which aren't determined by any other variables in the model[33]—and into a set $\mathcal{V}_{\text{en}}$ of *endogenous* variables—those which *are* determined by other variables in the model.

A few more standard definitions are in order. A *solution* of an SEM $(\mathcal{V}, \mathcal{E})$ is an assignment of values to all variables in $\mathcal{V}$ that is consistent with the conjunction of all equations in $\mathcal{E}$. For any $X \in \mathcal{V}$ and value x, $\ulcorner(\mathcal{V}, \mathcal{E}) \vDash X = x\urcorner$ then says that, in all possible solutions of $(\mathcal{V}, \mathcal{E})$, $X = x$. Let $\ulcorner(\mathcal{V}, \mathcal{E})(X \leftarrow x)\urcorner$ denote the SEM with variable set $\mathcal{V}\backslash\{X\}$ and the structural equation set $\mathcal{E}(X \leftarrow x)$ resulting by deleting X's structural equation from $\mathcal{E}$ and replacing every remaining occurrence of X by x. When $\mathcal{V}_{\text{ex}} = \mathbf{v}$, we say that $C = c$ *depends on* $E = e$ *in* $\mathcal{M}$ iff $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e$ and $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c') \vDash E = e'$ for some $c' \neq c$ and $e' \neq e$. So, intuitively, $C = c$ depends on $E = e$ in $\mathcal{M}$ if manually setting E to some value other than e also changes

such world where $A = 0$ we must thus have $B = 0$: by stipulation, there is no miracle after B, and hence $A = 0 \wedge B = 1$ would lead to contradiction. So we have $A = 0 \;\blacksquare\!\!\rightarrow B = 0$, i.e. $\neg O(L = 1) \;\blacksquare\!\!\rightarrow \neg O(R = 1)$. So Glynn's account wrongly entails action at a distance in both GAUSS and LOOP.

[33] That is, exogenous variables appear on the left-hand side of a structural equation iff the equation's right-hand side is constant.

the value of C in the model. Finally, where $(\mathcal{V}, \mathcal{E})$ is an SEM with $V, W \in \mathcal{V}$, a *directed path from V to W in $(\mathcal{V}, \mathcal{E})$* is a sequence of variables $(X_1, ..., X_n)$ such that $X_1 = V$ and $X_n = W$ and, for all $i = 1, ..., n-1$, $f_{X_{i+1}}$ is non-constant in X_i—that is, $f_{X_{i+1}}$'s value depends non-trivially on X_i's value for some assignment of values to $\mathcal{V} \backslash \{X_i, X_{i+1}\}$.

Now, prominent SEM accounts of causation—e.g. Hitchcock (2001); Menzies (2004); Halpern & Pearl (2005); Halpern (2016)—entail the following sufficient condition for causation:[34]

> **Sufficiency in Adequate Models:** Necessarily, if (c, e) is a suitable pair of actual events and variables X and Z represent alterations of c and e, respectively: c is a cause of e if there is an *adequate* SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $X, Z \in \mathcal{V}$ such that (i) there's a directed path from X to Z in $\mathcal{M}$ and (ii) $Z = e$ depends on $X = c$ in $\mathcal{M}$.

"Adequate" is crucial here: to ground causal judgments, an SEM has to faithfully represent the world. What does that mean?

First of all, adequate SEMs can't assert outright falsehoods. Above we said that structural equations say that the right-hand side partially *determines* the left-hand side. "Determine" in what sense? Prominent accounts agree that it is *counterfactual* determination. Here is Hitchcock (2001, p. 280) (see also Hitchcock (2007, p. 500)):

[34] Two nuances: Halpern & Pearl (2005) officially only provide a definition of *endogenous* variables' being causes. But this seems like a defect of their account: adequate causal models should faithfully capture also exogenous variables' causal relationships to the rest.

Not a defect is Menzies's (2004) slight deviation from **Sufficiency in Adequate Models**: he evaluates dependence-in-a-model by contrasting c specifically with its *default* alternative. As a result, his account will entail only a weakening of **Acyclic Sufficiency** (see below)—one which additionally requires that c's non-occurrence is default. But this won't affect our argument: the *absence* of a proton is plausibly default in the requisite sense, as are both a marble's being white and a marble's being grey. So, for simplicity, I'll ignore this nuance in the following.

Finally, to see that Hitchcock's (2001, p. 287, 290) proposal entails **Sufficiency in Adequate Models**, note that a directed path $\langle X, Y_1, ..., Y_n, Z\rangle$ is "weakly active" in $\mathcal{M}$ if Z depends on X in $\mathcal{M}$ (Hitchcock, 2001, p. 290). (Merely some terminological differences remain: Hitchcock's "appropriate" is my "adequate", and his "causal route" is my "directed path".)

"[S]tructural equations encode counterfactuals. For example, $[Z := f_Z(X, Y, ..., W)]$ encodes a set of counterfactuals of the following form:

> If it were the case that $X = x$, $Y = y$, ... , $W = w$, then it would be the case that $Z = f_Z(x, y, ..., w)$."

Similarly, Menzies (2004, p. 822):[35]

> "[The equation $SH := ST$] asserts that if Suzy threw a rock, her rock [would] hit the bottle; and if she didn't throw a rock, her rock [wouldn't have] hit the bottle."

Similar quotes are found in Halpern & Pearl (2005, p.847) and Weslake (2015).

Where $\mathbf{X} = \{X_1, ..., X_k\}$, I write $f_Z(X_1, ..., X_k)$ also as $f_Z(\mathbf{X})$ and, where additionally $\mathbf{x} = \{x_1, ..., x_k\}$, I'll write $X_1 = x_1 \wedge ... \wedge X_k = x_k$ as $\mathbf{X} = \mathbf{x}$. (Henceforth let **bold-face** letters denote sets of variables or values.) I'll choose the convention where f_Z is always a function of *all* variables in $\mathcal{V}\backslash\{Z\}$, while generally being *non-constant* only in a select few of those.[36]

The demand that $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ be true is then just the following demand: whenever $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$, then for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$,

$$\mathcal{V}\backslash\{Z\} = \mathbf{v} \;\Box\!\!\rightarrow Z = f_Z(\mathbf{v}).$$

Moreover, an adequate SEM is supposed to be *exhaustive*: it should contain a structural equation for every variable in the model. (Recall that, as we've defined it, exogenous variables are those whose structural equations have constant right-hand side.) Here is Hitchcock:[37]

[35] Curiously, Menzies uses indicative conditionals here, even though he means them to be "counterfactuals"; I've thus substituted the subjunctive form.

[36] Hitchcock (2001, p. 281) chooses a different convention, where f_Z is a function only of those variables in which it is non-constant. Philosophically this makes no difference, but Hitchcock's convention would complicate the notation, especially in the proofs of the Appendix.

[37] Given his convention, his expression "$Z = f_Z(Y, ..., W)$ is not in $\mathcal{E}$" means what, on my convention, "f_Z is non-constant in X" means. For the same reason, I translate his "and then eliminating those variables..." as "f_Z will be constant in those variables...."

> "By the same token, [Z's equation] in $\mathcal{E}$ must always include as arguments any variables in $\mathcal{V}$ upon which Z counterfactually depends, given the values of the other variables. If, for some $x, x', y, z, ..., w$, $f_Z(x, y, ..., w) \neq f_Z(x', y, ..., w)$, then the value of Z does depend upon the value of X, and [f_Z is non-constant in X]. The correct equation for Z can be arrived at by expressing the value of Z as a function of *all* other variables in $\mathcal{V}$, and then [f_Z will be constant in] those variables whose values are redundant given every assignment of values to the other variables." (p. 281)

This demand for exhaustiveness is just the converse of the conditional at the start of this paragraph; putting them together, we get the following biconditional: $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ *iff*, for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$, $\mathcal{V}\backslash\{Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z(\mathbf{v})$.

As a final condition on adequacy, the counterfactualist should require that all pairs of variables in $\mathcal{V}$ be *suitable* and that all value assignments to $\mathcal{V}$ are metaphysically possible.[38] If both of these things are the case, say that $\mathcal{V}$ *is suitable*.

> **Counterfactual Adequacy:**[39] Necessarily, $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ is adequate only if: $\mathcal{V}$ is suitable and, for all $Z \in \mathcal{V}$, $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ *iff* for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$,
>
> $$\mathcal{V}\backslash\{Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z(\mathbf{v}).$$

In the Appendix I prove that, given **Counterfactual Adequacy**, **Sufficiency in Adequate Models** entails that counterfactual dependence is sufficient for causation in the absence of

[38] Technically, rather than the metaphysical possibility of all value assignments, what matters is the weaker demand that all value assignments to any $|\mathcal{V}| - 1$-element subset of $\mathcal{V}$ can serve as antecedents in non-vacuous counterfactuals. But the stronger demand does just fine here.

[39] It's worth noting that this notion of adequacy immediately satisfies the additional *minimality* condition Hitchcock (2001, p. 280) lays out:

> "Equations in $\mathcal{E}$ must always be written in minimal form: [if f_Z is the right-hand side of Z's structural equation and] for all values x, x' of X and $\mathbf{v}$ of $\mathcal{V}\backslash\{X, Z\}$, $f_Z(x, \mathbf{v}) = f_Z(x', \mathbf{v})$, then the value of Z does not depend upon the value of X at all" (p. 280, notation adjusted).

For if, for all $x, x', \mathbf{v}$, $f_Z(x, \mathbf{v}) = f_Z(x', \mathbf{v})$, then, by the left-to-right direction of **Counterfactual Adequacy**: for all x, x', z, if $X = x \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = z$, then $X = x' \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = z$.

cycles.[40] That is, **Counterfactual Adequacy** and **Sufficiency in Adequate Models** jointly entail the following principle:[41],[42]

> **Acyclic Sufficiency:** Necessarily, if (c, e) is a suitable pair of actual events, X and Z are variables representing alterations of c and e, respectively, and there is an adequate, acyclic SEM including X and Z: if e wouldn't have occurred if c hadn't occurred, c causes e.

Acyclic Sufficiency entails that, if there's *any* adequate acyclic model in GAUSS whose variables represent alterations of the proton's presence at x at t and of the electric flux's being positive at S at t^+, then the former causes the latter. But there better *be* such models: otherwise it's hard to see how to avoid the distrastrous conclusion that there are ubiquitous, genuine causal loops in GAUSS. Thus, all prominent SEM accounts of causation are committed to the false conclusion that the proton's presence at x at t causes the electric flux's being positive at S at t^+.

Note that, since **Counterfactual Adequacy** merely posits a necessary condition for adequacy, this argument cannot be avoided by further strengthening adequacy. In particular, no additional demands on the richness of an SEM—how many variables it ought to contains, or how fine-grained their values ought to be—will help.

[40]The proof assumes CEM—a natural assumption to make for SEM reductions of causation in the absence of cycles.

[41]Given **Counterfactual Adequacy**, an explicit demand for (c, e)'s suitability is technically redundant—I'll still include it in the statement of **Acyclic Sufficiency**, for easier comparison with the other principles.

[42]**Acyclic Sufficiency** is clearly weaker than full **Sufficiency**. Do **Counterfactual Adequacy** and **Sufficiency in Adequate Models** also entail full **Sufficiency**? A reasonable assumption is that any pair of variables representing alterations of suitable events can be embedded in an adequate SEM. If so, then the only obstacle to full **Sufficiency** is the requirement of acyclicity. As far as I know, it's an open question if and how one might dispense with it. The answer presumably depends on one's preferred SEM account of causation in the presence of cycles—I shall leave this to future work. For now let it be noted that **Acyclic Sufficiency** does its job for GAUSS. In the case of LOOP where, plausibly, cyclic models *are* adequate, we can independently confirm that **Counterfactual Adequacy** and **Sufficiency in Adequate Models** misfire.

L	R	M	N
0	0	0	0
0	0	1	1
1	1	0	0
1	1	1	1

Table 4.2: The same four solutions, with additional variable N

In contrast to GAUSS, LOOP plausibly *does* contain causal loops. Rich enough SEMs—ones whose variable sets contain at least two events within the loop region—will thus generally be cyclic. Since **Acyclic Sufficiency** only concerns the predictions of *acyclic* SEMs, one might hope that, by positing additional richness constraints on adequacy, one might be able to avoid trouble at least in LOOP.

But even this hope is in vain: while I don't have a general theorem extending to the cyclic case, it's easy to see that **Counterfactual Adequacy** and **Sufficiency in Adequate Models** still entail the troubling conclusion for LOOP. For consider the model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{L, R, M, N\}$, where N registers the middle particle's shade shortly after the collision. The only nomically permitted solutions are as follows:

Thus the counterfactual dependence of R on L obtains regardless of the values of M and N. In particular, if $L = 0 \mathrel{\Box\!\!\rightarrow} R = 0$, then also

$$L = 0 \wedge M = N = 0 \mathrel{\Box\!\!\rightarrow} R = 0. \tag{4.2}$$

But since actually $L = R = 1$ and $M = N = 0$, we have, by And-to-If,

$$L = 1 \wedge M = N = 0 \mathrel{\Box\!\!\rightarrow} R = 1. \tag{4.3}$$

So, by **Counterfactual Adequacy**, conditions 4.2 and 4.3 entail that f_R depends on L—specifically, $f_R(L, M, N) = L$. In a similar fashion, we'll obtain $f_N(L, R, M) = LRM + \overline{L}\overline{R}M + \overline{L}R\overline{M} + L\overline{R}\overline{M}$ and $f_M(L, R, N) = N$. It follows that $\mathcal{M}(L \leftarrow 0) \vDash R = 0$ and

$\mathcal{M}(L \leftarrow 1) \vDash R = 1$. So, by **Sufficiency in Adequate Models**, $L = 1$ is a cause of $R = 1$ in LOOP.

It's easy to convince oneself that adding additional variables won't substantively change the result. **Counterfactual Adequacy** ensures that whatever variables represent the shades of the left and the right particle just before the collision depend on each other in the resulting model. So, notwithstanding the lack of a general theorem for the cyclic case, SEM reductions of causation plausibly yield the wrong result in LOOP, no matter the richness conditions on adequate models.[43]

4.3.3 No Easy Fix

As we saw early in Section 4.1, other arguments against Sufficiency can be answered by retreating to weaker, still substantive, versions of the principle. Is a similar strategy available for defusing the threat of synchronic laws? I'll discuss two attempts at this strategy here, both of which fail.

First, one might hope to exclude counterfactual dependencies due to synchronic laws by only considering dependencies between *non-synchronic events.* Specifically, consider the

[43]Gallow's (2016) more sophisticated counterfactualist theory of adequacy also fails in cases of synchronic laws. Let ϕ be the selection function for your favorite semantics of counterfactual conditionals, mapping proposition-world pairs into sets of worlds. For any world w and any given set of variables $\mathbf{X}$, let the $\mathbf{X}$-*closure of w under ϕ* be the closure of $\{w\}$ under the set of functions $\{\phi(\mathbf{X}' = \mathbf{x}', \cdot) | \mathbf{X}' \subseteq \mathbf{X}$ and $\mathbf{x}$ is in the range of $\mathbf{X}\}$, i.e., the smallest set W such that: (i) $w \in W$ and (ii) if $w' \in W$ and $\mathbf{x}'$ is in the range of some subset $\mathbf{X}' \subseteq \mathbf{X}$, then $\phi(\mathbf{X}' = \mathbf{x}', w') \subseteq W$. Intuitively, the $\mathbf{X}$-*closure of w under ϕ* is exactly the set of worlds you can reach from w by repeatedly subjunctively supposing $\mathbf{X}' = \mathbf{x}'$—i.e., taking conditionals of the form $\mathbf{X}' = \mathbf{x}' \mathrel{\Box\!\!\rightarrow}$ ...—where $\mathbf{X}'$ is a subset of $\mathbf{X}$. Now, according to Gallow, given a selection function ϕ, an adequate SEM $\mathcal{M} = (\mathbf{U}, \mathbf{V}, \mathcal{E})$ contains a structural equation $(V := f_V(\mathcal{V} \backslash \{V\})) \in \mathcal{E}$ only if the (ordinary) equation $V = f_V(\mathcal{V} \backslash \{V\})$ is true throughout the actual world's $\mathcal{V} \backslash \{V\}$-closure under ϕ. The "only if" is strengthened to a biconditional if additionally all variables in $\mathbf{U}$ are mutually counterfactually independent throughout that closure (formally, if no SEM with the same variable set but "strictly more" determination relations (i.e., directed paths) satisfies the aforementioned property).

According to this semantics, $\mathcal{M}_G$, if it is to be an adequate SEM for GAUSS, must contain the structural equation $E := P$, where E represents positive electric flux at S at t^+ and P the proton's presence at x at t. However many subjunctive suppositions of the form $P = i \mathrel{\Box\!\!\rightarrow} \ldots$, for $i = 0, 1$, are nested, Gauss's law would still hold at all times after t. Similarly for $\mathcal{M}_L$: however many subjunctive suppositions of the form $L = i \mathrel{\Box\!\!\rightarrow} \ldots$, for $i = 0, 1$, are nested, the dynamics and topological structure downstream from L and R would be unchanged and so we'd still have that $L = R$. So Gallow's theory of adequacy does no better than Hitchcock's when faced with synchronic laws.

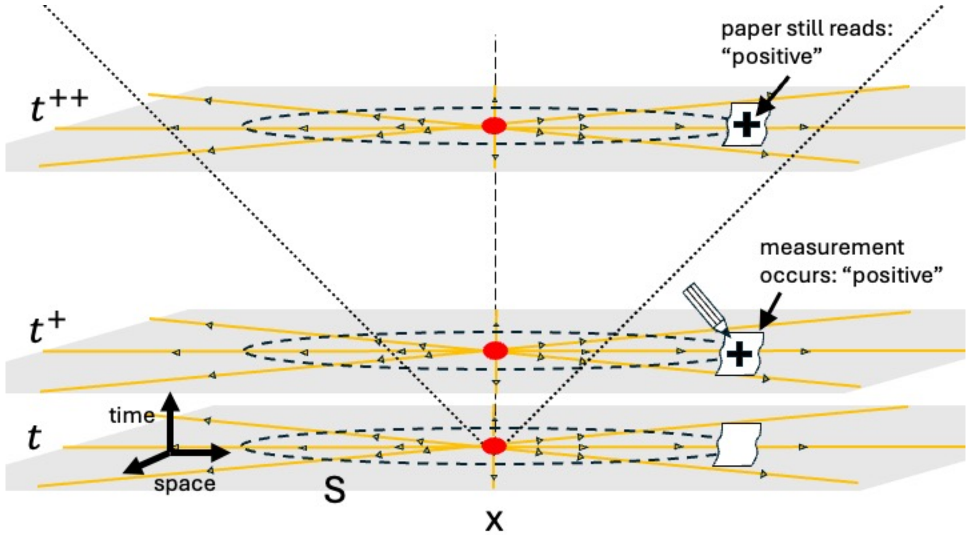


Figure 4.8: A sketch of the situation in **Measurement**.

following weakening of **Sufficiency**:

> **Sufficiency***: Necessarily, if (c, e) is a suitable pair of occurring events such that e wouldn't have occurred if c hadn't occurred, <u>and e occurs after c</u>, then c causes e.

Unfortunately, **Sufficiency*** still fails, because synchronic counterfactual dependencies can still induce dependencies between non-synchronic events. Consider the following modification of GAUSS:

> ***Case.* Measurement:** A measurement apparatus is placed somewhere along S at t^+, recording the local electric flux through S at t^+. It transcribes this result ("positive") on a piece of paper. Remaining at rest, the piece of paper will eventually be located inside of (t, x)'s future light-cone, say at time t^{++}.

See fig. 4.8 for a sketch. For concreteness, let's assume that, if the proton had not been located at x at t, then the electric field would have been *zero* everywhere on S at t^+.[44] So, if

[44]Now, recall from our discussion of Gauss's Law that there are zero-flux solutions with (everywhere) *non-vanishing* electric field. What if some such scenario is among the closest $\neg O(p_G)$-worlds? Then it'll not be guaranteed that any *particular* device in **Measurement** would read a different value in the proton's

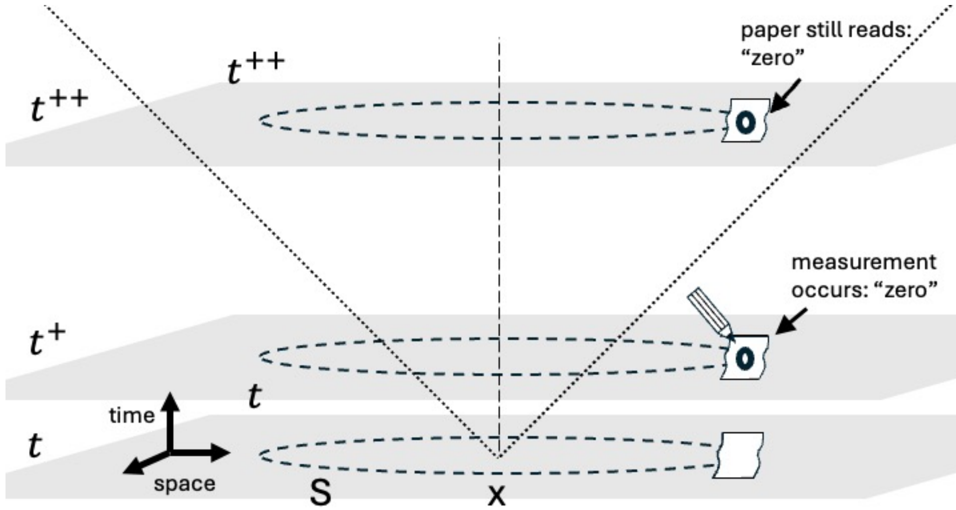


Figure 4.9: A sketch of the same situation had the proton not been present at x at t.

the proton hadn't been present at x at t, then the flux through S at t^+ would have been zero, and hence the paper at S at t^+ would have read "zero"; but then also the paper at S at t^{++} would read "zero"—see fig. 4.9. By **Sufficiency***, the proton's presence at x at t thus *causes* the paper's reading "positive" at t^{++}. But that's false—the paper is a record of a measurement that took place at t^+, *outside* of (t, x)'s future light-cone!

(An analogous counterexample can be constructed for LOOP: record the right particle's shade at t on a piece of paper. Assuming the underlying spacetime structure is classical, with an absolute notion of simultaneity, the piece of paper will immediately be located in the left particle's future.[45] **Sufficiency*** thus entails, wrongly, that the left particle's being white at t *causes* the right particle's being white at t.)

Now, another feature of synchronic laws is that they tend to generate *mutual* counterfactual dependence. This is clearest in LOOP, where not only $R = 1$ counterfactually

absence. What's then needed is a mild complication of the case: place a measurement apparatus at *every* point along S at t^+, with each result transcribed on a separate piece of paper. If the proton had been absent from x at t, at least one measurement apparatus would have to give a different reading, and so at least one piece of paper would be different at t^{++}.

[45]Otherwise, if the spacetime structure is relativistic, simply wait long enough for the piece of paper to be located in the future light-cone of the left particle at t.

depends on $L = 1$, but also (by an exactly symmetric argument) $L = 1$ counterfactually depends on $R = 1$. Likewise, we may grant (for argument's sake) that, in GAUSS, if the electric flux through S was 0, no proton would be present at x at t. So, as a second stab at the strategy, one might propose the following:

> **Sufficiency†**: Necessarily, if (c, e) is a suitable pair of occurring events such that e wouldn't have occurred if c hadn't occurred, *and* it's not the case that c wouldn't have occurred if e hadn't, then c causes e.

But the proposal succumbs to the same counterexamples. For Sufficiency† to avoid the wrong result in GAUSS, the following counterfactual would have to hold:

> **(?G†)** If the paper didn't read "positive" at t^{++}, then the proton wouldn't be present at x at t.

But t is much earlier than t^{++}, and—as we saw—the counterfactual dependence tradition broadly adopts Lewis's distinction between *standard* and *backtracking* contexts for subjunctive conditionals, cautioning us to evaluate the conditional in **Sufficiency** in the standard context. But, clearly, **(?G†)** is false in the standard context. To bring this out intuitively: suppose t^{++} occurs a week after t. Then, at t^{++}, you'd express **(?$G^\dagger$)** with the following subjunctive:

> **(??G†)** If the paper didn't read "positive" today, then the proton would be absent at x a week earlier.

A true reading of **(??G†)** requires explicit backward reasoning: if the paper didn't read "positive" today, that would have to be because the proton was already absent a week earlier. This parallels familiar backtracking cases: if I didn't return the book today, that would have to be because Susy and I agreed on a later date to begin with. On a miracles account—again something the counterfactual dependence tradition widely adopts (and should adopt)—the standard context instead only supports the following: if the paper had not read "positive" today, the far past would be unchanged; instead, a small miracle in the immediate past would have deleted the ink, or smeared it beyond recognition, or incinerated the paper, or something of that sort.

So, neither synchronic laws' synchronicity, nor their tendency to induce mutual counterfactual dependence, can be exploited for easy fixes of **Sufficiency**.

4.4 Conclusion

The preceding discussion suggests that there are two rather different sources of counterfactual dependence: first, counterfactual dependence due to *synchronic* laws (e.g. Mirror Law, Gauss); second, counterfactual dependence due to *diachronic* (or "dynamical") laws, such as Newton's Second Law ($\mathbf{F} = m \cdot \mathbf{a}$) or Faraday's and Ampére's Laws (cf. fn. 13).[46] Dynamical laws explain how systems evolve over time: given the state of a system at one moment, a dynamical law generates its future, or, a probability distribution over possible futures.[47] In contrast to the constraints imposed by synchronic laws, this temporal evolution is plausibly *causal*. If causation is to be reduced to counterfactuals, then, we have to separate this diachronic component from the synchronic one. The next chapter outlines some first steps.

We can now also see more clearly how a divorce of causation and counterfactual dependence could affect other areas of philosophy. In Mirror World, "Fido" refers to *my* dog, rather than my mirror image's dog, even though its use bears the same counterfactual relations to both. Causalist can explain this straightforwardly; counterfactualists, like Dretske (1981), might struggle.[48] Similarly, counterfactualists about perception (e.g. Lewis

[46]Other examples include the heat equation, the Navier-Stokes equations of fluid mechanics, and the Schrödinger equation.

[47]*Mathematically*, diachronic laws will tend to take the form of partial differential equations involving time derivatives.

[48]Luzon (2024) has recently made this point.

(1980)) will be pressed to explain how I *see* my own dog but not my mirror image's dog.[49] On the decision-theoretic side, meanwhile, our discussion arguably favors the counterfactualist. Suppose I, a Mirror World denizen, love chocolate cake, but that this love is trumped by my hate of the thought of my mirror image eating any. Should I eat a slice? Causal decision theory says yes, because my action causes only my own enjoyment. Counterfactual decision theory says no, because if I ate cake my mirror image would too. If the counterfactual verdict fits our preferences better—and my informal surveys, which have included several two-boxers, suggest it does—this lends fresh support to the counterfactualist approach.

[49] If you don't like Mirror World, parallel arguments could be set up for a world like GAUSS, where the question becomes how "electric flux" picks out the electric flux rather than the proton, or how I *see* the ink marks on the paper rather than the proton—given that my use of "electric flux" and my visual experience bear analogous counterfactual relationships to the flux as they do to the proton.

Chapter 5

Causation and Diachronic Dependence: Possible First Steps

The previous chapter showed that there are two different sources of counterfactual dependence, *synchronic* laws and *diachronic* ("dynamical") laws, of which the former generate non-causal counterfactual links between events. A satisfactory counterfactualist analysis of causation must exclude such links, in favor of the genuinely causal dependencies generated by the dynamical laws.[1] So, the goal is to carve out the dynamical component of counterfactual dependence and treat that—rather than counterfactual dependence *per se*—as the mark of causation.

What follows sketches possible first steps toward this goal. The chapter's scope is modest: it only looks at local and deterministic dynamical laws, and the causal relata are assumed to have determinate spacetime locations: specifically, they are all of the form $Q(R)$, where Q is an intrinsic property, and R is a spacetime region. We'll see that, even with these simplifications, isolating the diachronic counterfactual component is hard—the resulting account still fails in some fairly standard cases. Much more comprehensive future

[1] Besides Newton's Second Law, other examples of dynamical laws include laws about heat conduction, fluid mechanics, two of Maxwell's equations (Faraday's Law and Ampère's Law), Schrödinger's equation, and evolution equations of the Einstein field equations (obtaining in an initial value problem).

work will be needed to carry the project further.

A note on terminology: the physicist's name for a spacetime curve whose tangent vector is everywhere either time-like or null is *causal curve*. In this chapter I'll adopt this terminology too, cautioning the reader not to read into the term more than its narrow physics meaning.

5.1 Toward A New Sufficient Condition

To trace out an event's diachronic counterfactual impact, we want to propagate counterfactual changes forward using the diachronic laws, while appealing to the synchronic laws as little as possible.

When the diachronic laws are local and deterministic, the state of any instantaneous spatial region[2] fully determines the state of its *future domain of dependence* (FDOD):[3] the largest region such that all past-directed causal curves starting in it intersect the spatial region.[4] In a standard diagram of Minkowski spacetime, suppressing one spatial dimension, a spatial region's future domain of dependence looks like this:

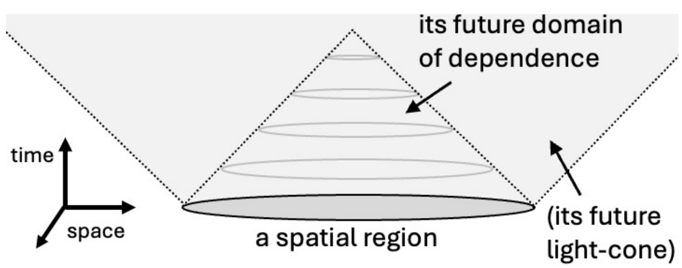


Figure 5.1: A spatial region's future domain of dependence (2+1-dimensional sketch, suppressing one spatial dimension).

[2]Technically, to ensure that their complete state define the derivatives (if any) required for solving the dynamical equations, we'll want spatial regions to have non-zero temporal extension. That is, we want them to be *thick*. But since we can choose this temporal extension to be as small as we like (so long as it's non-zero), it'll be safe to gloss over this detail in the following.

[3]A region R's FDOD is also commonly written $D^+(R)$; here I'll stick to the acronym for readability.

[4]To see how locality and determinism entail this: determinism ensures that the state of an inextendible space-like surface determine the universe's entire history; Parental Markov then immediately ensures that any *part* of said surface determines its future domains of dependence.

Let Q and Q^+ be possible intrinsic properties of R and R^+, respectively, and consider the events

$$c = Q(R),$$

$$e = Q^+(R^+).$$

Now assume that region R^+ is wholly inside R's FDOD. For this special case, we have a reasonably straightforward sufficient condition for diachronic dependence:

> *e diachronically depends* on c if, after minimally adjusting R to not be Q, and evolving the resulting state forward, via the diachronic laws, into R's FDOD, the FDOD's new state entails that R^+ isn't Q^+.

The relevant notion of "minimal adjustment" is just the standard counterfactual one: the state resulting after minimally adjusting R to not be Q is the maximal intrinsic state such that, had R not been Q', it would have been that state.[5] In particular, since the subjunctive here is just the standard one, in evaluating it we're respecting all synchronic laws *within* R.

This is *a* start. However, because the speed of light is so large, for everyday purposes future domains of dependence are *tiny*: at any given time my room has a future domain of dependence that, at its longest duration, measures about 10 nanoseconds. Realistic causal claims, involving everyday-sized spacetime regions, thus tend to range far beyond their FDODs. We thus need a way to extend diachronic influence futureward.

The obvious move to try is *chaining* regions. Starting with R's FDOD, pick a slice that intersects it; treat that slice as a new "initial" region, adjust its state based on R's FDOD's new state, evolve the resulting state forward, and repeat (cf. 5.2).

[5]Without any shorthands, we can write out the preceding condition as follows:

> *e diachronically depends* on c iff: where Q' is the maximal intrinsic state of R such that $\neg Q(R) \mathrel{\Box\!\!\rightarrow} Q'(R)$, and P the maximal intrinsic state of R's FDOD jointly entailed by $Q'(R)$ and the dynamical laws, the proposition that R's FDOD is P entails that R^+ isn't Q^+.

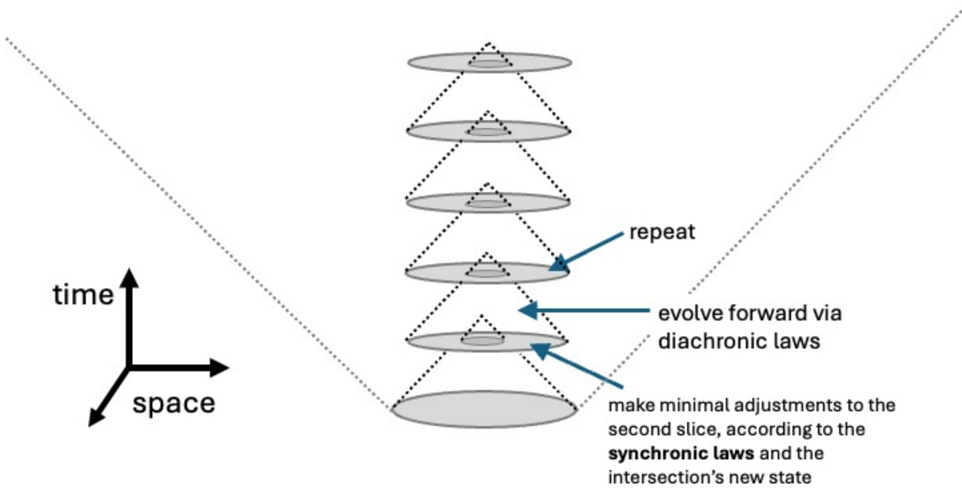


Figure 5.2: An FDOD chain.

To make this work, we can't just chain arbitrary spatial slices together. If we allowed future slices to grow without limit in spatial extent, they'd easily pick up spurious synchronic dependencies. Remembering that our aim is to capture the initial region's diachronic impact on the final region, the natural proposal is to limit members of a chain to the intersection of the initial element's future light-cone and the final element's past light-cone[6]—what physicist would call the *causal diamond* of the initial and final element. We thus define:

Def. **FDOD chain:** A sequence of spatial regions is an *FDOD chain* iff:

1. each region intersects its predecessor's FDOD, and
2. all regions lie in the initial region and final region's causal diamond.

Now, given an FDOD chain $\mathbf{R} = (R_1, ..., R_n)$ and an intrinsic property Q of R_1, we *trace Q's counterfactual impact through* $\mathbf{R}$ by following the previous procedure: we minimally adjust R_1 to not be Q and, using the diachronic laws, evolve the resulting state forward

[6]More precisely, in physicist speak: the intersection of the initial element's *causal future* and the final element's *causal past*.

into R_1's FDOD; this gives a new state for R_2's intersection with R_1's FDOD; we then minimally adjust R_2 to this new state,[7] evolve forward, and repeat. If R_n actually has some intrinsic property Q^+, but the state we end up with for R_n after tracing Q's counterfactual impact through **R** *isn't* Q^+, we say that $Q^+(R_n)$ *depends on* $Q(R_1)$ *via* **R**. We might then propose the following sufficient condition for diachronic dependence:

> *e diachronically depends* on *c* if there *exists* an FDOD chain from R to R^+ via which $Q^+(R^+)$ depends on $Q(R)$.

This yields good results for the flux-measurement case. Let R_1 and R_n be small spatial neighborhoods centered around x at times t and t^{++}, respectively.[8] Insofar as diachronic dependence patterns with *causal* dependence, we ought to obtain that both the presence of a proton at x in R_n and the electric field configuration throughout R_n diachronically depend on the presence of the proton at x in R_1. Our sufficient condition delivers exactly this; for consider the following FDOD chain:

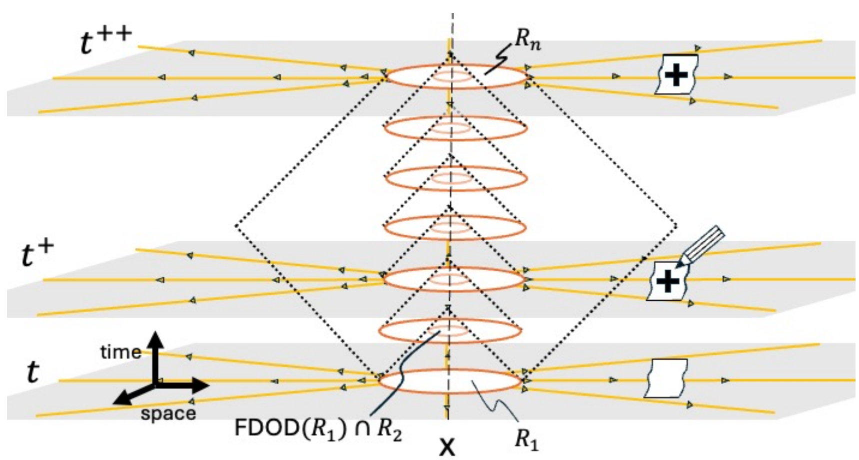


Figure 5.3: An FDOD chain transmitting the diachronic influence of the proton's existence at x in R_1 to R_n.

[7] To be explicit, where the new state of $R_2 \cap$ (R_1's FDOD)is Q_2, we're determining the maximal intrinsic state S of R_2 satisfying the following subjunctive: if $R_2 \cap$ (R_1's FDOD) had been Q_2, then R_2 would be S.

[8] Recall that all references to spatial location and other non-Lorentz-invariant notions are relative to a fixed coordinate system.

We can trace both the absence of both the proton and the electric field through every member of the chain. Let Q be R_1's actual maximal intrinsic state. To trace the proton's absence, we first determine the maximal intrinsic state Q_1 of R_1 satisfying the following subjunctive:

> If R_1 hadn't been Q, it would have been Q_1.

Since the subjunctive ought to respect the synchronic laws—in particular Gauss's Law—$Q_1(R_1)$ plausibly says that R_1 is empty of any particles and fields.[9] Evolving this forward via Maxwell's dynamical laws (Faraday and Ampère), we obtain that $\mathrm{FDOD}(R_1) \cap R_{i+1}$ is empty of particles and fields too. We then plausibly assert the following subjunctives for each $i = 1, ..., n-1$:[10]

> If $\mathrm{FDOD}(R_i) \cap R_{i+1}$ was empty of fields and particles, R_{i+1} would be empty of fields and particles.

Each empty R_{i+1} evolves forward, via the dynamical laws, into an empty FDOD(R_{i+1}). So, together with the subjunctive judgments, we obtain (as desired) that R_n's containing a proton and its containing a non-zero electric field diachronically depends on R_1's doing so.

So far, so good. What about the paper readout's diachronic dependence, at t^{++}, on R_1's state? Our sufficient condition correctly avoids classifying this as a case of diachronic dependence. Here is a representative FDOD chain connecting R_1 to a spatial neighborhood of the paper readout at t^{++}:

[9] More strongly than in Ch. 4, we're assuming here that the zero-*flux* condition (through any closed surface of R_1), suggested by Gauss's Law, is realized specifically by the zero field. This could be replaced by other synchronic-law-abiding conditions (like a uniform electric field), provided context calls for it. This would accordingly change some of the resulting causal judgments. For concreteness, I'll stick to the zero-field condition here.

[10] Again, we're assuming that zero *flux* is realized by a zero *field* here.

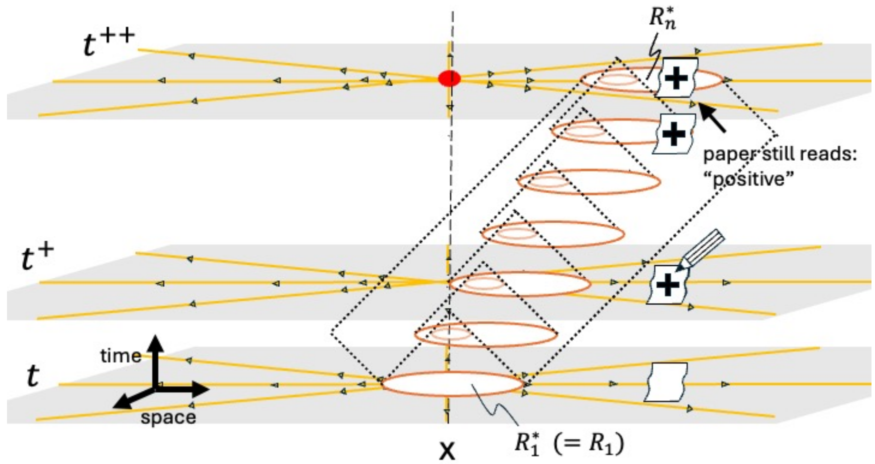


Figure 5.4: An FDOD chain transmitting the diachronic influence of the proton's existence at x in R_1 to region R_n^*: the paper readout is unaffected by flux changes at later times.

Consider the second-to-last region R_{n-1}^*; it's the FDOD chain's first member to contain the paper readout. Since the paper's content synchronically depends on the electric flux through S only at t^+, we have:

> If $\text{FDOD}(R_{n-2}^*) \cap R_{n-1}^*$ was empty, R_n^* would be empty except for a piece of paper, at S, reading "+".

The same analogous subjunctive goes through for R_n^*, and so R_n^*'s containing a paper reading "+" doesn't depend on R_1^*'s actual state via the FDOD chain $(R_1^*, ..., R_n^*)$.

But it's immediately clear that this is true for all other FDOD chains starting in R_1^*: by definition, none of its members reach outside of R_1^*'s future light-cone, and so none picks up the counterfactual dependence of the paper's readout on the proton's presence at t^*. We have successfully isolated the diachronic element of the counterfactual relationship between proton and paper readout.

Two clarifying remarks are in order. First, on the present account, the choice of the initial region affects whether e diachronically depends on c. If you carve out a large initial region at time t that is big enough for the later "paper region" (the bit of space-time containing the sheet of paper at t^+) to lie inside the region's FDOD, then the paper's reading "+" genuinely diachronically depends on the proton's being present in that initial region. By pushing this dependence along an FDOD chain we can carry the dependence all the way to any later region you care to track. But I think this is an acceptable result: the large slice's lacking the proton at x is a different event from the small slice's lacking the proton at x. The larger region already builds in the altered electromagnetic field that the missing charge would produce across a wider area. Given that extra built-in structure in the large slice, the mentioned diachronic dependence is exactly what one should expect.

Second, I'm not giving a new semantics for the ordinary subjunctive conditional. The worlds produced by propagating a counterfactual influence through an FDOD chain are rather alien worlds, with sudden cutoffs of electric fields and spontaneous proton creations even after the antecedent's time of occurrence (cf. fig. 5.2). They are nothing like what we would expect a "closest" possible $\neg Q(R)$-worlds to look like. Indeed, while FDOD propagation is designed to minimize the impact synchronic laws, full counterfactual dependence involves them fully: one of my main points in Ch. 4 was that synchronic laws *do* induce non-local counterfactual dependence—just one which doesn't amount to causal dependence. Now, you *could* treat the sufficiency condition above as defining a technical *diachronic counterfactual* operator, but recognize that it differs sharply from the everyday subjunctive.

5.2 Open Problems

5.2.1 Large FDOD Chains

Here is where the problems for the present theory start, to which I don't yet have good solutions. Consider again Mirror World (Ch. 4). Suppose that, at time t, you (and thus also your twin) set an alarm to go off the next morning, at t^+:

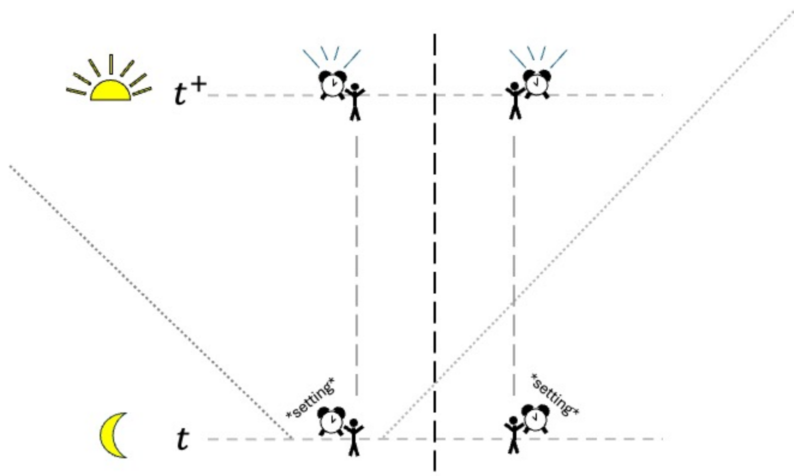


Figure 5.5: Setting an alarm in Mirror World.

Grant, for the sake of argument, that the characterization of the case in Ch. 4 is right: your twin's alarm's going off at t^+ depends merely counterfactually, not causally, on your setting yours at t. It's easy to see that there are FDOD chains which vindicate *your own* alarm's diachronic dependence on your setting it:

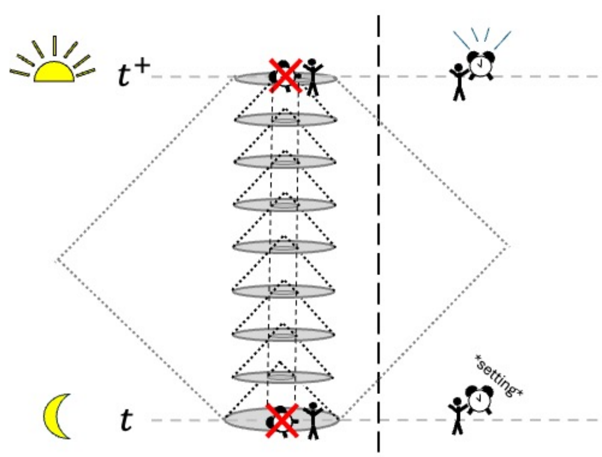


Figure 5.6: FDOD chain transmitting the diachronic influence of your inactive alarm on your morning alarm.

Meanwhile, an FDOD chain whose members are *sufficiently small* cannot carry the counterfactual impact of my setting the alarm at t to the mirror image's alarm the next morning:

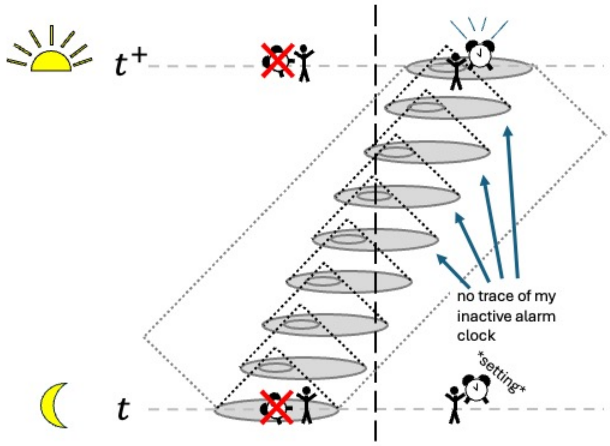


Figure 5.7: An FDOD chain of small regions leaves your twin's alarm intact.

Alas, "sufficiently small" is crucial here. Even within the causal diamond, some FDOD chains pick up on the synchronic connection between the alarms:

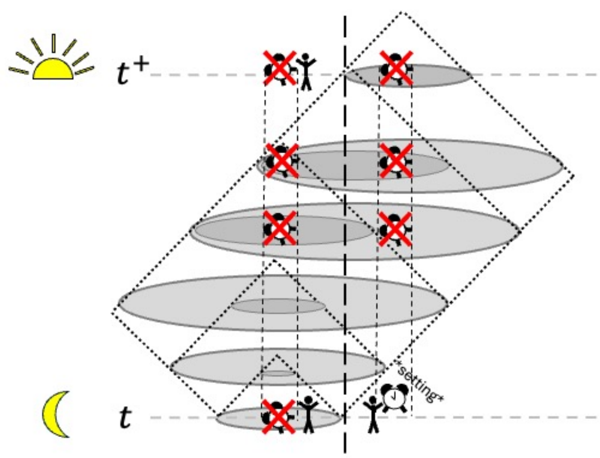


Figure 5.8: An FDOD chain of larger regions can still pick up unwanted synchronic dependencies.

So, our sufficient condition can't stand as is.

A more complete theory will need extra size constraints on FDOD chains. Maybe this can be done by limiting procedures, with only the "smallest" FDOD chains counting as genuine carriers of diachronic dependence. That strategy could dovetail with the structural-equation-model (SEM) approach discussed next.

5.2.2 A Semantics for SEM Adequacy?

Our efforts so far have focused only on providing a *sufficient* condition for causation—by itself that's a low bar, since sufficiency can always be bought by piling on enough constraints. Our hope, of course, is that our constraints will largely be retained, in some form, by a condition that's both necessary and sufficient—that our sufficient condition *points the way* toward such a condition.

What stands in the way of our condition being necessary for causation are the classic stumbling-blocks of rudimentary counterfactualist analyses of causation: preemption and overdetermination. A classic example:

> ***Example.*** **Stones.** At t_0, Susy and Billy simultaneously throw stones at a window. Billy throws from closer distance and his stone hits the window first, at t_1, breaking it. Susy's stone follows shortly after, at t_2, whizzing through the now empty window frame.

The window's being broken at t_2 diachronically depends on Billy's throw at t_0. Yet, there is no FDOD chain from Billy's throw at t_0 to the window (frame) at t_2 which, on its own, mediates that diachronic dependence. Here is a representative such chain:

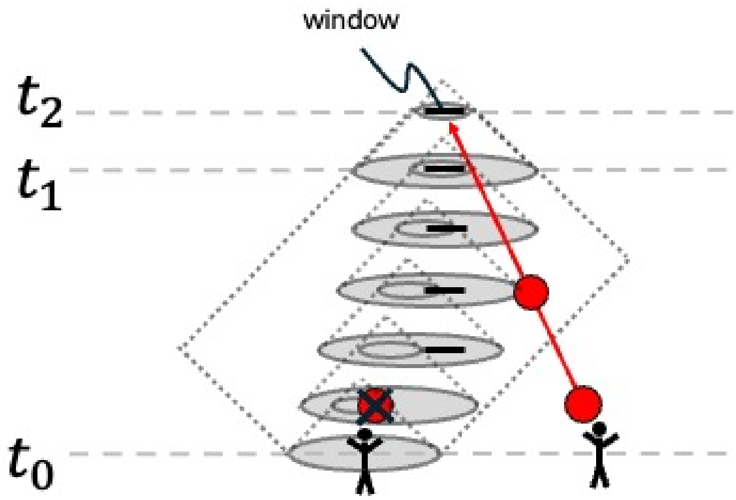


Figure 5.9: Susy's stone ensures that the window ends up broken anyway.

Susy's stone ensures that, at t_2, the window ends up broken, even if Billy's stone was absent.

Nowadays, the counterfactualist's standard remedy is to embed these scenarios in structural-equation models. As we have seen in Ch. 4, however, the standard counterfactualist semantics for model adequacy is unable to isolate diachronic dependence from synchronic dependence—indeed, it entails a version of the **Sufficiency** thesis. An FDOD-chain strategy potentially points to a different semantics: take each variable to represent a spacetime region, its values the region's intrinsic states, and draw an edge only when there is an FDOD chain from one region, bypassing every intervening region in the model.

Much work remains for the counterfactualist if she is to salvage her account's core idea from the threat of synchronic laws. This chapter has not established that the work *can* be done, but it has, I hope, cleared some ground on which future constructive attempts can start.

Appendices

Appendix A

Parental Markov Failure

A.1 Where Parental Markov Is False...

Consider the following one-dimensional spacetime, FORK (or F for short):

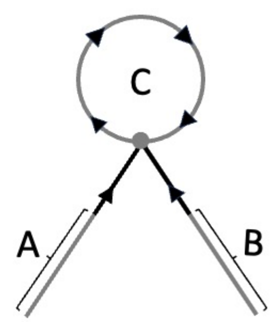


Figure A.1: A sketch of FORK.

Suppose the world is home to a scalar field, with deterministic dynamics. Specifically, suppose that the field everywhere takes the value 0 or 1, and that there are two kinds of spacetime points: *fork points*, where two or more lines converge (represented by the grey dot in the diagram above),[1] and *boring points*, all the others. In any interval consisting,

[1]Topologically, we can define a fork point as any point p such that there are two or more open lines containing p that do not share an open sub-line containing p.

except for possibly its initial point, of boring points, the field remains constant. At a fork point, meanwhile, the field's value is determined by the values on the incoming lines. If it is 1 on exactly an even number of incoming lines, it is also 1 at the fork point and on all outgoing lines; otherwise it is 0 at the fork point and on all outgoing lines.

The above figure indicates three disjoint segments, A, B, and C. If we introduce for each a homonymous binary variable whose value represent the field value taken throughout the segment, the given dynamics entails the following (where $\overline{\mathrm{X}} := (1 - X)$):

$$C = \overline{A}BC + A\overline{B}C + AB\overline{C} + \overline{ABC}. \tag{A.1}$$

If $A = B$, this becomes $C = \overline{\mathrm{C}}$—contradiction. Hence eq. A.1 entails that $A = \overline{B}$.

Now, the empty set is a pure thick parent of A, and A doesn't cause B. Hence, Parental Markov requires that the empty set screen off A from B:

$$\mathrm{urch}_F(A = 1|B = 0) = \mathrm{urch}_F(A = 1|B = 1).$$

But, since the dynamics requires $A = \overline{B}$,

$$\mathrm{urch}_F(A = 1|B = 0) = 1 \neq 0 = \mathrm{urch}_F(A = 1|B = 1).$$

So Parental Markov is false.[2]

A.2 ...Boundary Markov Is Still True

Yet Boundary Markov is still true. To see this, note that the fork point p partitions FORK into three disjoint segments: let $\mathcal{A}$ be the branch containing A, and $\mathcal{B}$ the branch containing B; the third segment is C itself, with $p \in \mathrm{C}$. Since $\mathcal{A}$ and $\mathcal{B}$ are connected segments consisting of boring points only, the field value at one point in the segment nomically entails the field

[2]If you are worried that this case relies on trickeries with empty sets, it's straightforward to change it into one involving *non-empty* thick parents—just introduce additional intervals prior to A, generating a non-trivial chance distribution over the field values in A.

values everywhere else in the segment. The same is true for C, where all points can be connected to each other via segments consisting of boring points only. Let now R be any region, and let D be a thick boundary of R. Note that, whenever R and $R^{\perp}$ have non-empty intersection with $\mathcal{A}$, D also has non-empty intersection with $\mathcal{A}$; the same goes for $\mathcal{B}$ and C and arbitrary unions of $\{\mathcal{A}, \mathcal{B}, C\}$.[3] (Where X and Y are regions, I'll abbreviate "The field values throughout X entail the field values throughout Y" by "X entails Y".)

- Case 1: Suppose $D \cap \mathcal{A} \neq \emptyset$ or $D \cap \mathcal{B} \neq \emptyset$. It follows that D either entails $\mathcal{A}$ or entails $\mathcal{B}$. Since the laws (eq. A.1) moreover entail $A = \overline{B}$, it follows that D entails $\mathcal{A} \cup \mathcal{B}$.
  - ▷ Case 1.1: Suppose $R \cap C \neq \emptyset$ and $R^{\perp} \cap C \neq \emptyset$. Then $D \cap C \neq \emptyset$ and so D entails C, and hence $\mathcal{A} \cup \mathcal{B} \cup C$. In particular, D entails R.
  - ▷ Case 1.2: Suppose $R^{\perp} \cap C = \emptyset$. Then $R^{\perp} \subseteq \mathcal{A} \cup \mathcal{B}$ and so D entails $R^{\perp}$.
  - ▷ Case 1.3: Suppose $R \cap C = \emptyset$. Then $R \subseteq \mathcal{A} \cup \mathcal{B}$ and so D entails R.
- Case 2: Suppose $D \cap \mathcal{A} = \emptyset$ and $D \cap \mathcal{B} = \emptyset$. Then $D \subseteq C$. It follows that either $\mathcal{A} \cup \mathcal{B} \subseteq R$ or $(\mathcal{A} \cup \mathcal{B}) \cap R = \emptyset$.[4]
  - ▷ Case 2.1: Suppose $\mathcal{A} \cup \mathcal{B} \subseteq R$. Then either $C \subseteq R$ or $D \cap C \neq \emptyset$. Since $R^{\perp} \subseteq C$, in either case D entails $R^{\perp}$.
  - ▷ Case 2.2: Suppose $(\mathcal{A} \cup \mathcal{B}) \cap R = \emptyset$. Then $R \subseteq C$ and either $R = \emptyset$ or $R \neq \emptyset$. If $R = \emptyset$, D triviallyentails R. If $R \neq \emptyset$, then either $C \subseteq R$ or $C \nsubseteq R$. If $C \subseteq R$, then $R = C$, and so $R^{\perp} = \mathcal{A} \cup \mathcal{B}$ and $D \cap (\mathcal{A} \cup \mathcal{B}) \neq \emptyset$. It follows that D entails $\mathcal{A} \cup \mathcal{B}$ and hence $R^{\perp}$. If $C \nsubseteq R$, then $R^{\perp} \cap C \neq \emptyset$ and so $D \cap C \neq \emptyset$. It follows that D entails R.

[3] In particular, note that if $R = \mathcal{A}$, then, due its thickness, D intersects both $\mathcal{B}$ and C. The same goes, *mutatis mutandis*, for $R = \mathcal{B}$ and $R = C$.

[4] For suppose otherwise, i.e. $(\mathcal{A} \cup \mathcal{B}) \cap R^{\perp} \neq \emptyset$ and $(\mathcal{A} \cup \mathcal{B}) \cap R \neq \emptyset$. Then $(\mathcal{A} \cup \mathcal{B}) \cap D \neq \emptyset$ and hence $D \nsubseteq C$.

So, in every case, D either entails R or entails $R^{\perp}$. In particular,

$$\mathrm{urch}_F(Q_1(R^{\perp})|Q_2(R) \cup Q_3(D)) = \mathrm{urch}_F(Q_1(R^{\perp})|Q_3(D)),$$

for any intrinsic properties Q_1 and Q_2, and maximal intrinsic property Q_3 such that $Q_2(R) \cup Q_3(D)$ are nomically possible. ■

Appendix B

Thick Neighborhoods = Neighborhoods of Closures

Let a *neighborhood of R* be any open superset of R. We've defined a *thick neighborhood N* as a neighborhood which satisfies the following additional condition: every continuous curve starting in $N^{\perp}$ and ending in R has a non-trivial subcurve in $N \backslash R$ before ever intersecting R. Here we prove that, in any space homeomorphic to $\mathbb{R}^n$, N is a thick neighborhood of R iff N is a neighborhood of R's closure, denoted $\overline{R}$. It follows that B is a thick boundary of R iff B is disjoint from R and $B \cup R$ contains a neighborhood of $\overline{R}$. We'll make ample use of this equivalence in Appendix C.

> *Theorem 1.* **Equivalence.** For any $R, N \subseteq \mathbb{R}^n$, N is a thick neighborhood of R iff N is a neighborhood of $\overline{R}$.

Proof: Right-to-left direction: Let N be a neighborhood of $\overline{R}$, and let c be a continuous curve which starts in $N^{\perp}$ and ends in R. Without loss of generality, assume $c : [0, 1] \to \mathbb{R}^n$. Since c is continuous, $c^{-1}(\overline{R}) \subseteq [0, 1]$ is closed and hence compact. Hence there is a first point $t^* \in [0, 1]$ such that $q := c(t^*) \in \overline{R}$ and for all $t < t^*$, $c(t) \notin \overline{R}$. Since N is open and $q \in N$, there is an open ball $B(q) \subseteq N$ around q. Since c is continuous, $c^{-1}(B(q))$ is open, and so there is a $t^- < t^*$ such that $]t^-, t^*[$ is an open interval in $c^{-1}(B(q))$. Since $B(q) \subseteq N$ and for all $t < t^*$, $c(t) \notin \overline{R}$, the subcurve $c|_{]t^-,t^*[}$ is a non-trivial subcurve of c in $N \backslash \overline{R}$ prior to ever intersecting $\overline{R}$, and so in particular a non-trivial subcurve in $N \backslash R$ prior to ever intersecting R.

Left-to-right direction: Suppose, for contradiction, that N is a thick neighborhood of R but not a neighborhood of $\overline{R}$. Then $\overline{R} \nsubseteq N$, and so there is a $q \in \overline{R} \backslash N$. We now construct a continuous curve starting in q such that every non-trivial initial segment of it intersects R. Since $q \in \overline{R}$, for every $r > 0$ the open ball $B_r(q)$ has non-empty intersection with R. For every $n \in \mathbb{N}_{>0}$, choose a point q_n in $B_{1/n}(q) \cap R$. Let $c : [0,1] \to \mathbb{R}^n$ be such that $c(0) = q$, for all $n \in \mathbb{N}_{>0}$ $c(1/n) = q_n$, and c maps $]1/(n+1), 1/n[$ continuously to the straight line from q_{n+1} to q_n, excluding endpoints. We now prove that c is continuous. The subcurve $c|_{]0,1]}$ maps $]0,1]$ into a concatenation of straight lines, and hence is continuous. It remains to prove that c is continuous at 0. Let $\{a_k\}_{k \in \mathbb{N}}$ be any sequence in $[0,1]$ converging to 0. Since balls in $\mathbb{R}^n$ are convex, for every $r \in [0,1[$, $c(r) \in B_m(q)$ where m is the largest integer such that $r < 1/m$. Since $\{a_k\}_{k \in \mathbb{N}}$ converges to 0, for every $\delta \in]0,1[$ there is a $k \in \mathbb{N}$ such that for all $l > k$, $a_l \in [0, \delta[$, and hence (by the foregoing) $c(a_l) \in B_{1/m}(q)$ where m is the largest integer such that $\delta < 1/m$. Let now $\varepsilon \in]0,1]$. Then there is a smallest integer $m \geq 2$ such that $1/(m-1) < \varepsilon$. Since $1/m \in]0,1[$, it follows from the foregoing that there is a $k \in \mathbb{N}$ such that for every $l > k$, $a_l \in [0, 1/m[$ and $c(a_l) \in B_{1/(m-1)}(q)$ (since $m-1$ is the largest integer such that $1/m < 1/(m-1)$). Since $1/(m-1) < \varepsilon$, it follows that $c(a_l) \in B_\varepsilon(q)$. So, for every $\varepsilon \in]0,1]$, there is a $\delta \in]0,1]$ (namely $\delta = 1/m$) such that for all l with $|a_l| < \delta$, $c(a_l) \in B_\varepsilon(q)$. This proves that c is continuous at 0. So c is continuous over $[0,1]$. Hence c is a continuous curve which starts in $N^\perp$, ends in R, and every non-trivial initial segment intersects R; in particular, it has no non-trivial subcurve in $R^\perp \supseteq N \backslash R$ before ever intersecting R, in contradiction with the assumption that N is a thick neighborhood of R. So N is a neighborhood of $\overline{R}$. ■

Appendix C

Parental Markov and Boundary Markov

Here we prove that, given plausible conglomerability and locality assumptions about urchance, Parental Markov already ensures, for sufficiently well-behaved regions, that they are screened off by their thick boundaries in Minkowski spacetime.

We'll first establish some auxiliary lemmas. As before, a *neighborhood* of A is any open superset of A. Where B is a thick boundary of R, let $B_R^+ := K^+(R) \cap B$, and $B_R^- := B \backslash B_R^+$ the rest of B. (To recall, $K^+(R)$ denotes R's proper causal future.) Note also the following elementary fact: if B is a thick boundary of R, then every continuous curve starting in $(R \cup B)^\perp$ and ending in R has a non-trivial subcurve in B before ever intersecting R.[1] Call a region *causally convex* iff it contains all causal curves starting and ending in it. ("Sufficiently well-behaved" will denote a mild strengthening of causal convexity.)

> ***Lemma 1*:** For any causally convex region R and any thick boundary B of R, B_R^- is a pure thick parent of R.

[1] The reverse implication fails, however. In $\mathbb{R}$, consider $B = \bigcup_{n=1}^{\infty}]-\frac{1}{n}, -\frac{1}{n+1}[\cup \{0\}$ and $R =]0, +\infty[$. Every continuous curve which starts in $(R \cup B)^\perp$ and ends in R has a non-trivial subcurve in B before ever intersecting R. But no *open subset* of $(R \cup B)$ has that property—that is, no open set $N \subseteq R \cup B$ is such that every curve starting in $N^\perp$ and ending in R has a non-trivial subcurve in N before ever intersecting R. Hence B isn't a thick boundary of R. The Equivalence theorem (Appendix B) relies on this extra strength in the definition of "thick boundary".

Proof of Lemma 1: Let c be a future-directed causal curve starting in $(R \cup B_R^-)^\perp$ and ending in R. Since c starts in $R^\perp$ and ends in R, c starts, specifically, in $K^-(R) \subseteq R^\perp$. Since R is causally convex, $K^+(R) \cap K^-(R) = \emptyset$. It follows that c starts in $(R \cup B)^\perp$ and ends in R. Since B is a thick boundary of R, c therefore has a non-trivial subcurve in B before ever intersecting R. Suppose, for contradiction, that there is a point q where c intersects B_R^+ and let $t \in [0,1]$ such that $c(t) = q$. Since q is in $K^+(R)$, there is a future-directed causal curve c' starting in R and ending in q. Concatenating c' and $c|_{[t,1]}$ thus yields a future-directed causal curve starting in R, intersecting B_R^+ and ending in R. Since $B_R^+ \subseteq R^\perp$, this contradicts R's causal convexity. So c doesn't intersect B_R^+. Hence c has a non-trivial subcurve in $B \backslash B_R^+ = B_R^-$ before ever intersecting R. So B_R^- is a thick parent of R. Finally, suppose for contradiction that R causes some point r in B_R^-. Then there is a future-directed causal curve c^* starting in R and ending in r. Since r is in $K^-(R)$, there is a future-directed causal curve c^{**} starting in r and ending in R. Concatenating c^* and c^{**} thus yields a future-directed causal curve starting in R, intersecting B_R^-, and ending in R. Since $B_R^- \subseteq R^\perp$, this contradicts R's causal convexity. Hence B_R^- is a pure thick parent of R. ■

> ***Lemma 2*:** If R is causally convex, $K^+(R)$ fully contains all future-directed causal curves starting in it; in particular, $K^+(R)$ is causally convex.

Proof of Lemma 2: Let R be causally convex and suppose for contradiction that there is a future-directed causal curve c starting in $K^+(R)$ and intersecting $(K^+(R))^\perp$. By the definition of $J^+(R)$, $J^+(R)$ contains all future-directed causal curves starting in it. Since $K^+(R) = J^+(R) \backslash R$, c thus intersects R in some point q; choose a $t \in [0,1]$ such that $c(t) = q$. Let p be c's starting point. Since $p \in J^+(R)$, there is a future-directed causal curve c^* starting in R and ending in p. Concatenating c^* and $c|_{[0,t]}$ yields a future-directed causal curve that starts in R, intersects $K^+(R) \subseteq R^\perp$, and ends in R, in contradiction with R's causal convexity. So $K^+(R)$ contains all future-directed causal curves starting in it. It follows that $K^+(R)$ is causally convex. ■

Let a *thick child* of R be any set C such that every future-directed causal curve starting in R and ending in $(R \cup C)^{\perp}$ has a non-trivial subcurve in C before ever intersecting $(R \cup C)^{\perp}$.

> ***Lemma 3*:** For any region R in Minkowski spacetime and any thick boundary B of R, if both R and $R \cup B$ are causally convex, then
>
> (i) B_R^+ is a thick child of R, and
>
> (ii) $R \cup B_R^-$ is a pure thick parent of B_R^+.

Proof of Lemma 3: (i): Let c be a future-directed causal curve starting in R and intersecting $(R \cup (B_R^+))^{\perp}$. Since R is causally convex, $K^+(R) \cap K^-(R) = \emptyset$, and so $K^+(R) \cap B_R^- = \emptyset$. Since also $R \cap B_R^- = \emptyset$, we have $J^+(R) \cap B_R^- = \emptyset$. But $J^+(R)$ contains c, so c doesn't intersect B_R^-. Since $B = B_R^+ \cup B_R^-$, c thus intersects $(R \cup B)^{\perp}$. Since B is a thick boundary of R, B is a thick boundary of $(R \cup B)^{\perp}$. Hence c has a non-trivial subcurve in B before ever intersecting $(R \cup B)^{\perp}$. Since c doesn't intersect B_R^-, it follows that c has a non-trivial subcurve in B_R^+ before ever intersecting $(R \cup B_R^+)^{\perp}$. So B_R^+ is a thick child of R.

(ii): In Minkowski spacetime, the causal future of a neighborhood of A's closure is a neighborhood of the closure of A's causal future.[2] Since B is a thick boundary of R, $R \cup B$ contains a neighborhood N of $\overline{R}$ (cf. Equivalence, Appendix B). Thus $J^+(N)$ is a neighborhood of $\overline{J^+(R)}$. Since $J^+(N) \subseteq J^+(R \cup B)$, $J^+(R \cup B)$ thus contains a neighborhood of $\overline{J^+(R)}$, and since $\overline{B_R^+} \subseteq \overline{J^+(R)}$, $J^+(R \cup B)$ contains a neighborhood of $\overline{B_R^+}$. Therefore, all continuous curves which start in $(J^+(R \cup B))^{\perp}$ and end in B_R^+ have a non-trivial subcurve in $J^+(R \cup B) \backslash B_R^+$ before ever intersecting B_R^+. Let c be a future-directed causal curve which starts in $(R \cup B)^{\perp}$ and ends in B_R^+. Since $(R \cup B)^{\perp} \subseteq (J^+(R \cup B))^{\perp}$, c has a non-trivial subcurve in $J^+(R \cup B) \backslash B_R^+$ before ever intersecting B_R^+. Suppose, for contradiction, that c

[2] To see this, note the following three facts about Minkowski spacetime (the first is true for any spacetime):
1. If $A \subseteq B$, then $J^+(A) \subseteq J^+(B)$.
2. Closure and causal future "commute", i.e. $J^+(\overline{A}) = \overline{J^+(A)}$ for any region A.
3. If A is open, $J^+(A)$ is open.

Let then N be a neighborhood of $\bar{A}$. Since $\bar{A} \subseteq N$ we have, by the first fact, $J^+(\bar{A}) \subseteq J^+(N)$. By the second fact, $\overline{J^+(A)} \subseteq J^+(N)$. Finally, by the third fact, $J^+(N)$ is open, and hence a neighborhood of $\overline{J^+(A)}$.

doesn't have a non-trivial subcurve in $R \cup B_R^-$ before ever intersecting B_R^+. Then c must intersect $\left(J^+(R \cup B) \backslash B_R^+\right) \backslash (R \cup B_R^-) = J^+(R \cup B) \backslash (R \cup B) = K^+(R \cup B)$. Let q be a point in $K^+(R \cup B)$ which c intersects, and choose a $t \in [0, 1]$ with $c(t) = q$. Since $q \in K^+(R \cup B)$, there is a future-directed causal curve c^* starting in $R \cup B$ and ending in q. Concatenating c^* and $c|_{[t,1]}$ thus yields a future-directed causal curve that starts in $R \cup B$, intersects $K^+(R \cup B) \subseteq (R \cup B)^\perp$, and ends in $B_R^+ \subseteq (R \cup B)$, in contradiction with $R \cup B$'s causal convexity. So c has a non-trivial subcurve in $R \cup B_R^-$ before ever intersecting B_R^+. Hence $R \cup B_R^-$ is a thick parent of B_R^+. Finally, suppose for contradiction that B_R^+ causes $R \cup B_R^-$. Since $B_R^+ \subseteq K^+(R)$ and $R \cup B_R^- \subseteq K^+(R)^\perp$, it follows that $K^+(R)$ causes $K^+(R)^\perp$. Since $J^+(R) = K^+(R) \cup R$ contains all future-directed causal curves which start in it, it follows that $K^+(R)$ causes R. Since R causes every point in $K^+(R)$, there is thus a future-directed causal curve starting in R, intersecting $K^+(R) \subseteq R^\perp$, and ending in R, in contradiction with R's causal convexity. Hence $R \cup B_R^-$ is a pure thick parent of B_R^+. ■

For any spacetime region X, let $\mathcal{R}(X)$ denote the set of all possible maximally specific intrinsic properties of X. For any $x \in \mathcal{R}(X)$, I'll also write $X = x$ instead of $x(X)$. For any $\mathbf{x} \subseteq \mathcal{R}(X)$, let $X \in \mathbf{x}$ denote the proposition that $X = x$ for some $x \in \mathbf{x}$. Let an *urchance candidate* be any function $\mathfrak{u}$ mapping pairs of propositions to functions on total, primitively conditional probability functions, such that, where u is such a function, $\mathfrak{u}(X, Y)(u) = u(X, Y)$. (In particular, then, urchance functions are urchance candidates.)

> ***Def.*** **Regional Conglomerability:** An urchance candidate $\mathfrak{u}$ is *regionally conglomerable* iff, for any spacetime region X, any $\mathbf{x} \subseteq \mathcal{R}(X)$, any propositions A and B, and any $a, b \in [0, 1]$ with $a \leq b$: if $a \leq \mathfrak{u}(A|X = x \wedge B) \leq b$ for all $x \in \mathbf{x}$, then
>
> $$a \leq \mathfrak{u}(A|X \in \mathbf{x} \wedge B) \leq b.$$

Where $\mathfrak{u}$ is an urchance candidate, let $\mathcal{R}(Z)_\mathfrak{u}$ denote the set of properties which are possible maximally specific intrinsic properties of Z *according to* $\mathfrak{u}$.[3] Let $(\mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$ denote the set of property *pairs* (y, z) such that: according to $\mathfrak{u}$, y and z are possible maximally specific intrinsic properties of Y and Z, respectively, *and* it's possible, according to $\mathfrak{u}$, that $Y = y \wedge Z = z$.

In addition to conglomerability, we need an additional locality assumption: information about a region is "nothing over and above" information about its parts. That is, necessarily, for any regions X, Y, maximally specifying X and Y, in a nomically compatible way, also *maximally* specifies their union $X \cup Y$ in a nomically possible way; moreover *every* nomically possible maximal property of the union can be so specified. Call this property *Separability*. Technically:

> ***Def.* Separability:** An urchance candidate $\mathfrak{u}$ is *separable* iff, for any spacetime regions X, Y, there is a one-to-one correspondence
>
> $$\phi : (\mathcal{R}(X) \times \mathcal{R}(Y))_{\mathfrak{u}} \to \mathcal{R}(X \cup Y)_{\mathfrak{u}},$$
>
> such that for any $(x, y) \in (\mathcal{R}(X) \times \mathcal{R}(Y))_{\mathfrak{u}}$, necessarily according to $\mathfrak{u}$, $(X = x \wedge Y = y) \leftrightarrow (X \cup Y = \phi(x, y))$.

Whenever such a ϕ exists I'll slightly abuse notation and write $(\mathcal{R}(X) \times \mathcal{R}(Y))_{\mathfrak{u}} = \mathcal{R}(X \cup Y)_{\mathfrak{u}}$, as well as $X \cup Y = (x, y)$ instead of $X \cup Y = \phi(x, y)$.

For any urchance candidate $\mathfrak{u}$ and for any regions X, Y, Z, let $(X \perp\!\!\!\perp Y|Z)_{\mathfrak{u}}$ denote that Z *screens off* X *from* Y *according to* $\mathfrak{u}$; that is:

$$(X \perp\!\!\!\perp Y|Z)_{\mathfrak{u}} \text{ iff: for all } \mathbf{x} \subseteq \mathcal{R}(X), \mathbf{y} \times \{z\} \subseteq (\mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}},$$
$$\mathfrak{u}(X \in \mathbf{x} | Y \in \mathbf{y} \wedge Z = z) = \mathfrak{u}(X \in \mathbf{x} | Z = z).$$

[3] In our system, a proposition A is *possible according to* $\mathfrak{u}$ iff $\mathfrak{u}(\neg A|A) < 1$, and A is *necessary according to* $\mathfrak{u}$ iff $\mathfrak{u}(A|\neg A) = 1$. These are dual provided all functions in $\mathfrak{u}$'s range agree on what's impossible; that is, for all u, u' in $\mathfrak{u}$'s range and all A, $u(\neg A|A) = 1$ iff $u'(\neg A|A) = 1$. (To see this: $\neg A$ is impossible iff $\mathfrak{u}(A|\neg A) \not< 1$, i.e. iff for some u in $\mathfrak{u}$'s range, $u(A|\neg A) = 1$. But given that all u agree on what's impossible, this is the case iff, for *all* u in $\mathfrak{u}$'s range, $u(A|\neg A) = 1$, i.e., iff $\mathfrak{u}(A|\neg A) = 1$.) Recall that, by assumption, all functions in an *urchance function*'s range agree on what's impossible—cf. fn. 14. Hence "possible according to urch" and "necessary according to urch" are dual if urch is an urchance function.

We have the following lemma (the names of the conditions follow Pearl's (1985) nomenclature for graphoids):

Lemma 4. **Spacetime Graphoid Theorems:** For any urchance candidate $\mathfrak{u}$, if $\mathfrak{u}$ is regionally conglomerable and separable, then for any regions X, Y, Z, and W:

- *Contraction:* If $(X \perp\!\!\!\perp Y|Z)_{\mathfrak{u}}$ and $(X \perp\!\!\!\perp W|Z \cup Y)_{\mathfrak{u}}$, then $(X \perp\!\!\!\perp W \cup Y|Z)_{\mathfrak{u}}$.
- *Weak Union:* If $(X \perp\!\!\!\perp Y \cup W|Z)_{\mathfrak{u}}$, then $(X \perp\!\!\!\perp Y|Z \cup W)_{\mathfrak{u}}$.[4]

Proof of Lemma 4:

Contraction: Suppose $(X \perp\!\!\!\perp Y|Z)_{\mathfrak{u}}$ and $(X \perp\!\!\!\perp W|Z \cup Y)_{\mathfrak{u}}$. That is: for any $\mathbf{y} \times \{z\} \subseteq (\mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$,

$$\mathfrak{u}(X \in \mathbf{x}|Y \in \mathbf{y} \wedge Z = z) = \mathfrak{u}(X \in \mathbf{x}|Z = z), \tag{C.1}$$

and for any $\mathbf{w} \times \{(y, z)\} \subseteq (\mathcal{R}(W) \times \mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$,

$$\mathfrak{u}(X \in \mathbf{x}|W \in \mathbf{w} \wedge Y = y \wedge Z = z) = \mathfrak{u}(X \in \mathbf{x}|Y = y \wedge Z = z). \tag{C.2}$$

By eqs. C.1 and C.2, for any $(w, y, z) \in (\mathcal{R}(W) \times \mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$,

$$\begin{aligned} \mathfrak{u}(X \in \mathbf{x}|W = w \wedge Y = y \wedge Z = z) &= \mathfrak{u}(X \in \mathbf{x}|Y = y \wedge Z = z) \\ &= \mathfrak{u}(X \in \mathbf{x}|Z = z). \end{aligned} \tag{C.3}$$

But by Separability, $(\mathcal{R}(W) \times \mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}} = (\mathcal{R}(W \cup Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$. Thus, by eq. C.3 and Regional Conglomerability, for any $\mathbf{v} \subseteq \mathcal{R}(W \cup Y)_{\mathfrak{u}}$ such that $\mathbf{v} \times \{z\} \subseteq (\mathcal{R}(W \cup Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$,

$$\mathfrak{u}(X \in \mathbf{x}|W \cup Y \in \mathbf{v} \wedge Z = z) = \mathfrak{u}(X \in \mathbf{x}|Z = z).$$

[4] The graphoid theorem *Decomposition*—if $(X \perp\!\!\!\perp W \cup Y|Z)_{\mathfrak{u}}$, then $(X \perp\!\!\!\perp W|Z)_{\mathfrak{u}} \wedge (X \perp\!\!\!\perp Y|Z)_{\mathfrak{u}}$—is also valid, and the proof is immediate. By contrast, only a restricted version of *Symmetry* is valid: *provided* that [for all $\mathbf{x} \subseteq \mathcal{R}(X)_{\mathfrak{u}}$ and $z \in \mathcal{R}(Z)_{\mathfrak{u}}$, $\mathfrak{u}(X \in \mathbf{x}|Z = z) > 0$], $(X \perp\!\!\!\perp Y|Z)_{\mathfrak{u}}$ implies $(Y \perp\!\!\!\perp X|Z)_{\mathfrak{u}}$. Likewise, only a restricted version of *Intersection* is valid: *provided* that [whenever $(w, z) \in (\mathcal{R}(W) \times \mathcal{R}(Z))_{\mathfrak{u}}$ and $(y, z) \in (\mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$, $(w, y, z) \in (\mathcal{R}(W) \times \mathcal{R}(Y) \times \mathcal{R}(Z))_{\mathfrak{u}}$], $(X \perp\!\!\!\perp Y|Z \cup W)_{\mathfrak{u}}$ and $(X \perp\!\!\!\perp W|Z \cup Y)_{\mathfrak{u}}$ imply $(X \perp\!\!\!\perp W \cup Y|Z)_{\mathfrak{u}}$. As none of these three properties will play a role in the following, I'll omit the proofs.

But this is just $(X \perp\!\!\!\perp W \cup Y | Z)_{\mathfrak{u}}$. ■

Weak Union: Suppose $(X \perp\!\!\!\perp Y \cup W | Z)_{\mathfrak{u}}$, i.e. for all $\mathbf{x} \subseteq \mathcal{R}(X)_{\mathfrak{u}}$ and $\mathbf{v} \times \{z\} \subseteq (\mathcal{R}(Y \cup W) \times \mathcal{R}(Z))_{\mathfrak{u}}$,

$$\mathfrak{u}(X \in \mathbf{x} | Y \cup W \in \mathbf{v} \wedge Z = z) = \mathfrak{u}(X \in \mathbf{x} | Z = z). \tag{C.4}$$

For any $(w, z) \in (\mathcal{R}(W) \times \mathcal{R}(Z))_{\mathfrak{u}}$, let $\top_{w,z} \subseteq \mathcal{R}(Y)$ be such that, according to $\mathfrak{u}$, $W = w \wedge Z = z$ entails $Y \in \top_{w,z}$. Then

$$\begin{aligned}
\mathfrak{u}(X \in \mathbf{x} | W = w \wedge Z = z) &= \mathfrak{u}(X \in \mathbf{x} | Y \in \top_{w,z} \wedge W = w \wedge Z = z) \\
&= \mathfrak{u}(X \in \mathbf{x} | Y \cup W \in \top_{w,z} \times \{w\} \wedge Z = z) \\
&\overset{\text{eq.C.4}}{=} \mathfrak{u}(X \in \mathbf{x} | Z = z).
\end{aligned} \tag{C.5}$$

But now, for any $\mathbf{y} \times \{w\} \times \{z\} \subseteq (\mathcal{R}(Y) \times \mathcal{R}(W) \times \mathcal{R}(Z))_{\mathfrak{u}}$,

$$\begin{aligned}
\mathfrak{u}(X \in \mathbf{x} | Y \in \mathbf{y} \wedge W = w \wedge Z = z) &= \mathfrak{u}(X \in \mathbf{x} | Y \cup W \in \mathbf{y} \times \{w\} \wedge Z = z) \\
&\overset{\textbf{eq.}\text{C.4}}{=} \mathfrak{u}(X \in \mathbf{x} | Z = z) \\
&\overset{\textbf{eq.}\text{C.5}}{=} \mathfrak{u}(X \in \mathbf{x} | W = w \wedge Z = z).
\end{aligned}$$

But, by Separability, $(\mathcal{R}(W) \times \mathcal{R}(Z))_{\mathfrak{u}} = \mathcal{R}(W \cup Z)_{\mathfrak{u}}$. So this is just $(X \perp\!\!\!\perp Y | Z \cup W)_{\mathfrak{u}}$. ■

The conglomerability and locality constraints on urchance are just this:

> ***Thesis.*** Necessarily, the urchance function is regionally conglomerable and separable.

This entails, by Lemma 4, that urchance validates Contraction and Weak Union.

As in the main text, when I speak of "Z **screens off** X **from** Y" (*simpliciter*)—denoted $X \perp\!\!\!\perp Y | Z$—I mean *screens off according to the urchance function, conditioned on a complete description of the world's geometry*. Finally, say that a region R is **tolerantly causally convex** iff it is causally convex and every thick boundary of R contains a thick boundary B' of R such that $R \cup B'$ is causally convex. (This is the aforementioned strengthening of causal convexity.) Many regions we typically consider are tolerantly causally convex—including any history segment (e.g., in Minkowski spacetime, the region between two Cauchy surfaces) and any light-cone segment (i.e., the intersection of a history segment and a past or future light-cone). We can now state the main result.

> ***Theorem.*** **Parental Markov & Boundary Markov.** In Minkowski spacetime, Parental Markov entails that, for all regions R, if R is tolerantly causally convex and B is a thick boundary of R, then B screens off R from $R^{\perp}$.

Proof of Theorem: Let R be causally convex and B^* a thick boundary of R. Since R is tolerantly causally convex, there is a $B \subseteq B^*$ such that B is a thick boundary of R and $R \cup B$ is causally convex. Once we show that $R \perp\!\!\!\perp R^{\perp} | B$, the desired result follows immediately by Weak Union: for, since $B^* \subseteq R^{\perp}$ and $B \subseteq B^*$, $R \perp\!\!\!\perp R^{\perp} | B$ entails $R \perp\!\!\!\perp R^{\perp} | B^*$.

Now, to prove $R \perp\!\!\!\perp R^{\perp} | B$, note the following three facts:

1. B_R^- is a pure thick parent of R.

2. There is a region S disjoint from $R \cup B$ such that $R \cup B_R^- \cup S$ is a pure thick parent of $K^+(R)$. ("S" stands for "spouse".)

3. B_R^+ is a thick child of R, and $R \cup B_R^-$ is a pure thick parent of B_R^+.

The first fact follows from Lemma 1 and R's causal convexity. The second fact follows like this: by Lemma 2, $K^+(R)$ is causally convex, and so, by Lemma 1 again, $K^+(R)$ has a pure thick parent. Since R is causally convex, $K^+(R)$ contains all causal curves which start in it, and so $K^+(R)$ doesn't cause $R \cup B_R^-$. Hence there is a pure thick parent P of $K^+(R)$ which contains $R \cup B_R^-$. Now simply choose $S := P \backslash (R \cup B_R^-)$. The third fact is just Lemma 3. The following sketch offers some orientation:

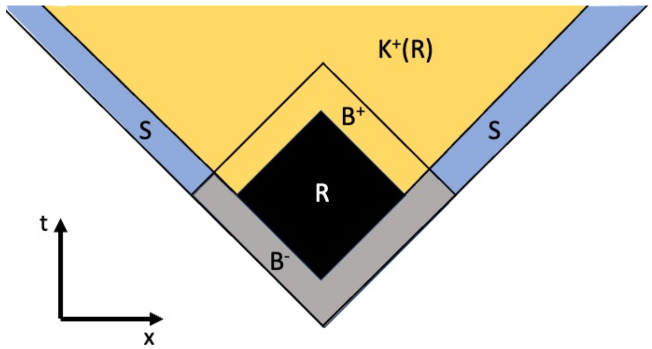


Figure C.1: 2D sketch of relevant regions near R.

Below we show the following three facts:

$$R \perp\!\!\!\perp R^{\perp} \mid B \cup S \cup (K^{+}(R) \backslash B), \tag{C.6}$$

$$R \perp\!\!\!\perp (K^{+}(R) \backslash B) \mid B \cup S, \tag{C.7}$$

$$R \perp\!\!\!\perp S \mid B. \tag{C.8}$$

Once these are proven, $R \perp\!\!\!\perp R^{\perp} | B$ follows: facts C.7 and C.8 entail, by Contraction,

$$R \perp\!\!\!\perp (K^{+}(R) \backslash B) \cup S | B.$$

Contracting again with fact C.6 gives

$$R \perp\!\!\!\perp R^{\perp} \cup (K^{+}(R) \backslash B) \cup S | B.$$

But $K^{+}(R) \backslash B \cup S \subseteq R^{\perp}$, hence

$$R \perp\!\!\!\perp R^{\perp} | B.$$

Proof of fact C.6: Recall that $R \cup B_R^- \cup S$ is a pure thick parent of $K^+(R)$. By Lemma 1, B_R^- is a pure thick parent of R. Since S is disjoint from R, it follows that $B_R^- \cup S$ is a thick parent of $R \cup K^+(R) = J^+(R)$.[5] Moreover, since S and B_R^- are disjoint from R and subsets of a thick parent of $K^+(R)$, and thus also disjoint from $K^+(R)$, $B_R^- \cup S$ is a *pure* thick parent of $J^+(R)$. Since $J^+(R)$ contains every future-directed causal curve starting in it, $J^+(R)$ doesn't cause its complement $(J^+(R))^\perp$. Hence, by Parental Markov, $J^+(R) \perp\!\!\!\perp (J^+(R))^\perp | B_R^- \cup S$. By Weak Union, $R \perp\!\!\!\perp (J^+(R))^\perp | K^+(R) \cup B_R^- \cup S$, and thus $R \perp\!\!\!\perp (J^+(R))^\perp \cup K^+(R) | K^+(R) \cup B_R^- \cup S$. By definition, $(J^+(R))^\perp \cup K^+(R) = R^\perp$ and $K^+(R) \cup B_R^- = B \cup (K^+(R) \backslash B)$. So $R \perp\!\!\!\perp R^\perp | B \cup (K^+(R) \backslash B) \cup S$.

Proof of fact C.7: Recall, again, that $R \cup B_R^- \cup S$ is a pure thick parent of $K^+(R)$. By Lemma 3(i), B_R^+ is a thick child of R. It follows that $B_R^+ \cup B_R^- \cup S = B \cup S$ is a thick parent of $K^+(R) \backslash B_R^+$.[6] We now show that $K^+(R) \backslash B_R^+$ doesn't cause $B \cup S$. Since R is causally

---

[5] This follows from the following general lemma (instantiate V with $K^+(R)$; P with $R \cup B_R^- \cup S$; T with B_R^-; and P' with R):

> **Lemma:** Let V be any region and P be a thick parent of V. Then, if T is a thick parent of some subset $P' \subseteq P$, then $T \cup (P \backslash P')$ is a thick parent of $V \cup P'$.

Proof of lemma: Let P be a thick parent of V, and T be a thick parent of a subset $P' \subseteq P$. Let c be a future-directed causal curve starting in $(T \cup P \cup V)^\perp$ and ending in $V \cup P'$. We show that c has a non-trivial subcurve in $T \cup (P \backslash P')$ before ever intersecting $V \cup P'$. Then c either ends in P' or ends in V.

Suppose c ends in P'. Since c starts in $(T \cup P')^\perp$ and T is a thick parent of P', there are $\tau_0, \tau_1 \in [0,1]$ with $\tau_0 < \tau_1$ such that $c(]\tau_0, \tau_1[) \subseteq T$ and $c([0, \tau_1[) \subseteq P'^\perp$. If $c([0,\tau_1[) \subseteq V^\perp$, then $c([0,\tau_1[) \subseteq (V \cup P')^\perp$, and so $c|_{[0,\tau_1[}$ is a non-trivial subcurve in T—and *a fortiori* in $T \cup (P \backslash P')$—which c has before ever intersecting $V \cup P'$. If instead $c([0,\tau_1[) \not\subseteq V^\perp$, $c^{-1}(V)$ has an infimum x in $[0, \tau_1[$. Since P is a thick parent of V, $c|_{[0,x[}$ contains a non-trivial subcurve in P before ever intersecting V. Since $c([0,x[) \subseteq P'^\perp$, it follows that $c|_{[0,x[}$ contains a non-trivial subcurve in $P \backslash P'$—and *a fortiori* in $T \cup (P \backslash P')$—before ever intersecting $V \cup P'$.

Suppose that c ends in V. Since c starts in $(V \cup P)^\perp$ and P is a thick parent of V, there are $s_0, s_1 \in [0,1]$ with $s_0 < s_1$ such that $c(]s_0, s_1[) \subseteq P$ and $c([0, s_1[) \subseteq V^\perp$. If $c([0,s_1[) \subseteq P'^\perp$, then $c|_{]s_0,s_1[}$ is a non-trivial subcurve in $P \backslash P'$—and *a fortiori* in $T \cup (P \backslash P')$—which c has before ever intersecting $V \cup P'$. If instead $c([0,s_1[) \not\subseteq P'^\perp$, $c^{-1}(P')$ has an infimum y in $[0, s_1[$. Since T is a thick parent of P', $c|_{[0,y[}$ contains a non-trivial subcurve in T before ever intersecting P'. Since $c([0,y[) \subseteq V^\perp$, it follows that $c|_{[0,x[}$ contains a non-trivial subcurve in T—and *a fortiori* in $T \cup (P \backslash P')$—before ever intersecting $V \cup P'$. ■

[6] This follows from the following general lemma (instantiate V with $K^+(R)$, P with $R \cup B_R^- \cup S$, C with B_R^+, and P' with R):

> **Lemma:** Let V be any region and P be a thick parent of V. Let C be a thick child of some $P' \subseteq P$. Then $(P \backslash P') \cup C$ is a thick parent of $V \backslash C$.

Proof of lemma: Let c be a future-directed causal curve starting in $((P \backslash P') \cup C \cup V)^\perp$ and ending in $V \backslash C$. Then c is, in particular, a future-directed causal curve starting in $V^\perp$ and ending in V. Since P is a thick parent of V, there are thus $a, b \in [0,1]$ with $b > a$ such that $c(]a,b[) \subseteq P$ and $c([0,b[) \subseteq V^\perp$. If $c(]a,b[) \subseteq P'^\perp$, then $c(]a,b[)$ is a non-trivial subcurve in $P \backslash P'$—*a fortiori*, in $P \backslash P' \cup C$—which c has before ever intersecting V, and hence before ever intersecting $V \backslash C$. If instead $c(]a,b[) \not\subseteq P'^\perp$, then the infimum x of $c^{-1}(P')$ is in $[0,b[$. Hence there is a $y \in]x,b[$ such that $c(y) \in P'$. But then $c|_{[y,1]}$ is a curve that starts in P' and ends in convex, $K^+(R)$ contains all future-directed causal curves starting in $K^+(R)$. In particular, $K^+(R)$, and hence $K^+(R)\backslash B_R^+$, don't cause $B_R^- \cup S$. Now suppose, for contradiction, that $K^+(R)\backslash B_R^+$ causes B_R^+. Then there is a future-directed causal curve c starting in $K^+(R)\backslash B_R^+ = K^+(R)\backslash B \subseteq (R \cup B)^\perp$ and ending in $B_R^+ \subseteq R \cup B$. Since c starts in $K^+(R)$, there is a future-directed causal curve c^* from R to c's starting point. Concatenating c^* and c yields a future-directed causal curve that starts in $R \cup B$, intersects $(R \cup B)^\perp$, and ends in $R \cup B$, in contradiction with $R \cup B$'s causal convexity. So $K^+(R)\backslash B_R^+$ doesn't cause B_R^+. Hence $K^+(R)\backslash B_R^+$ doesn't cause $B_R^+ \cup B_R^- \cup S = B \cup S$, and so $B \cup S$ is a pure thick parent of $K^+(R)\backslash B$. Finally, since $K^+(R)$ contains all future-directed causal curves starting in $K^+(R)$, $K^+(R)\backslash B$ doesn't cause R. So Parental Markov implies $R \perp\!\!\!\perp (K^+(R)\backslash B)|B \cup S$.

Proof of fact C.8: Recall that $R \cup B_R^-$ is a thick parent of B_R^+. Moreover, by Lemma 1, B_R^- is a thick parent of R. It follows that B_R^- is a thick parent of $R \cup (B_R^+)$.[7] Since R is causally convex, $J^+(R)$ fully contains all future-directed causal curves starting in it. Hence $R \cup B_R^+ \subseteq J^+(R)$ doesn't cause $B_R^- \subseteq J^+(R)^\perp$. Hence B_R^- is a pure thick parent of $R \cup B_R^+$. Since S is not caused by $R \cup B_R^+$, Parental Markov entails that $R \cup B_R^+ \perp\!\!\!\perp S|B_R^-$. By Weak Union, $R \perp\!\!\!\perp S|B$. ■

$V\backslash C \subseteq (P' \cup C)^\perp$. Since C is a thick child of P', $c|_{[y,1]}$ thus contains a non-trivial subcurve in C before ever intersecting $V\backslash C$. But $c([0,y]) \subseteq V^\perp \subseteq (V\backslash C)^\perp$, and so c itself contains a non-trivial subcurve in C before ever intersecting $V\backslash C$. In either case,$(P\backslash P') \cup C$ is a thick parent of $V\backslash C$. ■

[7]This follows from the following additional lemma (instantiate P with B_R^-, Q with R, and V with B_R^+):

> **Lemma:** Let Q be causally convex. Let P be a thick parent of Q and $Q \cup P$ be a thick parent of V. Then P is a thick parent of $Q \cup V$.

Proof: Let c be a future-directed causal curve starting in $(Q \cup P \cup V)^\perp$ and ending in $Q \cup V$. c either ends in Q or in V. We show that, in each case, c has a non-trivial subcurve in P before ever intersecting $Q \cup V$. First, suppose c ends in Q. Since P is a thick parent of Q and c starts in $Q^\perp$, c has a non-trivial subcurve in P before ever intersecting Q. Since Q is causally convex, $V \subseteq K^+(Q)$ doesn't cause Q, and so c intersects Q before ever intersecting V. Hence c has a non-trivial subcurve in P before ever intersecting $Q \cup V$; done. Second, suppose that c ends in V. Since $Q \cup P$ is a thick parent of V and c starts in $(Q \cup P \cup V)^\perp$ and ends in V, there are $a, b \in [0,1]$ such that $c(]a,b[) \subseteq Q \cup P$ and $c([0,b[) \subseteq V^\perp$. Suppose $c([0,b[) \subseteq Q^\perp$. Then $c|_{]a,b[}$ is a non-trivial subcurve in P which c has before ever intersecting $Q \cup V$; done. Suppose instead $c([0,b[) \not\subseteq Q^\perp$. Then there is a $q \in [0,b[\cap Q$. Since $c(0) \in Q^\perp$, $c|_{[0,q[}$ is a future-directed causal curve starting in $Q^\perp$ and ending in Q. Since P is a thick parent of Q, $c|_{[0,q[}$ thus has a non-trivial subcurve in P before ever intersecting Q. Since $c([0,q[) \subseteq c([0,b[) \subseteq V^\perp$, c thus has a non-trivial subcurve in P before ever intersecting $Q \cup V$. ■

Appendix D

Marginal Chances over CIRCLE

Abbreviate $\mathrm{urch}_C(\tau_i = j)$ as $[ij]$ and $\mathrm{urch}_C(\tau_i = j|\tau_{i'} = j')$ as $[ij|i'j']$. For every $i = 1, ..., n$ and $j = 1, ..., k$, eqs. 1.23 and 1.24 are then more compactly written as follows (where 0 in an index is identified with n):

$$[ij] = \sum_{l=1}^{k} [ij|(i-1)l] \cdot [(i-1)l],$$
$$1 = \sum_{j=1}^{k} [ij].$$

This yields the following $n \cdot (k-1)$ equations, one for every $i = 1, ..., n$ and $j = 1, ..., k-1$:

$$\begin{aligned}
[ij] &= \left(\sum_{l=1}^{k-1} [ij|(i-1)l] \cdot [(i-1)l]\right) + [ij|(i-1)k] \cdot \left(1 - \sum_{l=1}^{k-1} [(i-1)l]\right) \\
&= \sum_{l=1}^{k-1} \Big([ij|(i-1)l] - [ij|(i-1)k]\Big) \cdot [(i-1)l] + [ij|(i-1)k],
\end{aligned}$$

which can be rearranged to

$$[ij] + \sum_{l=1}^{k-1} \Big([ij|(i-1)k] - [ij|(i-1)l]\Big) \cdot [(i-1)l] = [ij|(i-1)k].$$

Writing this linear system as a matrix equation yields the following:

$$\mathbf{M} \cdot \hat{\mathbf{p}} = \hat{\mathbf{v}}, \tag{D.1}$$

where

$$
\begin{aligned}
\hat{\mathbf{p}} &= ([11], ..., [1(k-1)], [21], ..., [2(k-1], ..., [n1], ..., [n(k-1)])^{\mathrm{T}} \\
&= ([ij])^{\mathrm{T}}_{i=1,...,n;j=1,...,(k-1)}
\end{aligned}
$$

is a length $n(k-1)$ column vector of marginal probabilities ($\cdot^T$ denotes the transpose),

$$
\begin{aligned}
\hat{\mathbf{v}} =&([11|nk], ..., [1(k-1)|nk], [21|1k], ..., [2(k-1)|1k], ..., [n1|(n-1)k], ..., \\
&[n(k-1)|(n-1)k])^{\mathrm{T}} \\
=&([ij|(i-1)k])^{\mathrm{T}}_{i=1,...,n;j=1,...,(k-1)}
\end{aligned}
$$

is a length $n(k-1)$ column vector, and

$$
\mathbf{M} = \begin{pmatrix}
\mathbb{I}_{k-1} & 0 & 0 & 0 & \dots & 0 & \mathbf{P}^1_{k-1} \\
\mathbf{P}^2_{k-1} & \mathbb{I}_{k-1} & 0 & 0 & \dots & 0 & 0 \\
0 & \mathbf{P}^3_{k-1} & \mathbb{I}_{k-1} & 0 & \dots & 0 & 0 \\
\vdots & & & & & & \\
0 & 0 & 0 & ... & & \mathbf{P}^n_{k-1} & \mathbb{I}_{k-1}
\end{pmatrix},
$$

is a $n(k-1) \times n(k-1)$ matrix, where $\mathbb{I}_{k-1}$ is the $(k-1) \times (k-1)$ identity matrix and

$$
\mathbf{P}^i_{k-1} = \mathbf{Q}^i_{k-1} - \mathbf{R}^i_{k-1}
$$

is the $(k-1) \times (k-1)$ matrix such that

$$
\mathbf{Q}^i_{k-1} = \begin{pmatrix}
[i1|(i-1)k] & [i1|(i-1)k] & \dots & [i1|(i-1)k] \\
[i2|(i-1)k] & [i2|(i-1)k] & \dots & [i2|(i-1)k] \\
\vdots & & & \\
[i(k-1)|(i-1)k] & [i(k-1)|(i-1)k] & \dots & [i(k-1)|(i-1)k]
\end{pmatrix},
$$

and

$$
\mathbf{R}^i_{k-1} = \begin{pmatrix} [i1|(i-1)1] & [i1|(i-1)2] & \dots & [i1|(i-1)(k-1)] \\ [i2|(i-1)1] & [i2|(i-1)2] & \dots & [i2|(i-1)(k-1)] \\ \vdots & & & \\ [i(k-1)|(i-1)1] & [i(k-1)|(i-1)2] & \dots & [i(k-1)|(i-1)(k-1)] \end{pmatrix}.
$$

Both the matrix $\mathbf{M}$ and the enriched matrix $(\mathbf{M}|\hat{\mathbf{v}})$ generically have full rank $n(k-1)$, and so generically $\hat{\mathbf{p}}$ is unique.

To illustrate this further, consider CIRCLE. Because every color j only has itself and color $j+1$ as permissible successors, all entries of $\mathbf{Q}^i_{k-1}$ besides the first row are 0, and all entries of $\mathbf{R}^i_{k-1}$ besides the diagonal and the first lower diagonal are 0. In the simplest non-trivial case, $n = k = 2$, the loop is two days long, with two possible colors per day. In this case—call it SUPER SIMPLE CIRCLE, or SSC—we have

$$
\mathbf{M} = \begin{pmatrix} 1 & [11|22] - [11|21] \\ [21|12] - [21|11] & 1 \end{pmatrix}
$$

and

$$
\hat{\mathbf{v}} = ([11|22], [21|12])^{\mathrm{T}}.
$$

Note that $\mathbf{M}$ and $(\mathbf{M}|\hat{\mathbf{v}})$ both have rank 2 unless

$$
[21|12] - [21|11] = [11|22] - [11|21] = \pm 1,
$$

i.e. unless either[1]

$$
\begin{aligned}
\mathrm{urch}_{SSC}(\tau_2 = 1|\tau_1 = 2) = \mathrm{urch}_{SSC}(\tau_1 = 1|\tau_2 = 2) = 1, \\
\mathrm{urch}_{SSC}(\tau_2 = 1|\tau_1 = 1) = \mathrm{urch}_{SSC}(\tau_1 = 1|\tau_2 = 1) = 0,
\end{aligned}
$$

or

[1]We always assume that transition probabilities are precise.

$$\mathrm{urch}_{SSC}(\tau_2 = 1|\tau_1 = 2) = \mathrm{urch}_{SSC}(\tau_1 = 1|\tau_2 = 2) = 0,$$
$$\mathrm{urch}_{SSC}(\tau_2 = 1|\tau_1 = 1) = \mathrm{urch}_{SSC}(\tau_1 = 1|\tau_2 = 1) = 1.$$

In the first case, the particle is guaranteed to switch color every time. Any probabilistically coherent assignment of marginals respecting $\mathrm{urch}_{SSC}(\tau_1 = 1) = \mathrm{urch}_{SSC}(\tau_2 = 2)$ is a solution to the equations. In the second case, the particle is guaranteed to retain its color every time. Here, any probabilistically coherent assignment of marginals respecting $\mathrm{urch}_{SSC}(\tau_1 = 1) = \mathrm{urch}_{SSC}(\tau_2 = 1)$ is a solution to the resulting equations. These are the only two possible cases for SSC in which the dynamics fails to determine unique marginal chance distributions over the states of the loop.

Appendix E

Connected General Previsions Are Convex

To prove Theorem 2, we first recount that, for any bounded measurable random variable, the expectation functional is continuous on the space of probability functions (in any topology at least as fine-grained as the product topology).[1]

> **Lemma.** ***Expectation is continuous in the product topology.*** Let $\mathcal{A}$ be an algebra over sample space Ω. Then, for any bounded, $\mathcal{A}$-measurable random variable X over Ω, the map
>
> $$F_X : \begin{cases} \Delta(\mathcal{A}) \to \mathbb{R} \\ p \mapsto \mathbb{E}_p(X), \end{cases}$$
>
> is continuous, with respect to the product topology on $\Delta(\mathcal{A})$ (and the Euclidean topology on $\mathbb{R}$).

Proof: This is a standard result if $\mathbb{E}$ is defined in terms of the Lebesgue integral. But it's worth noting that it also holds for our "finitely additive" integral of bounded functions,

[1] Here and henceforth, by "product topology on the space of probability functions", I mean (more precisely) the *sub-topology* on the space induced by the product topology on the embedding Tychonoff cube.

defined in fn. 23. To see this, note that, for any simple function s, $\int s(x)p(\mathrm{d}x) = \int \sum_{k=1}^{m} s_k \cdot \mathbf{1}_{A_k} p(\mathrm{d}x) = \sum_{k=1}^{m} s_k \cdot p(A_k)$ is a linear combination of continuous functionals in p and hence itself continuous in p. Since X is bounded and measurable, there is a sequence $(s^n)_{n\in\mathbb{N}}$ of simple functions converging uniformly toward X (cf. fn. 23). Let thus $\sup_{x\in\Omega} |X(x) - s^n(x)| = \varepsilon_n$. Since $p(\Omega) = 1 < \text{const.}$ for all p, we have $|\int X(x)p(\mathrm{d}x) - \int s^n(x)p(\mathrm{d}x)| \leq \varepsilon_n$ for all p, and hence $\sup_p |\int X(x)p(\mathrm{d}x) - \int s^n(x)p(\mathrm{d}x)| \leq \varepsilon_n \overset{n\to\infty}{\longrightarrow} 0$. So, as functions of p, $(\int s^n(x)p(\mathrm{d}x))_{n\in\mathbb{N}}$ uniformly converge to $\int X(x)p(\mathrm{d}x)$. Since uniform limits of continuous functions are continuous, $\int X(x)p(\mathrm{d}x) = \lim_{n\to\infty} \int s^n(x)p(\mathrm{d}x)$ is continuous in p. ■

We can now prove:

> **Theorem 2.** ***Connectedness entails convexity.*** Let $\mathfrak{P}$ be a coherent general prevision on an algebra $\mathcal{A}$ over sample space Ω, with $A, B \in \mathcal{A}$. If $\mathbf{P}_{\mathfrak{P}}$ is connected (in the product topology on $\Delta(\mathcal{A})$), then it is convex.

Proof: Suppose $\mathbf{P}_{\mathfrak{P}}$ is connected in the product topology of $\Delta(\mathcal{A})$. By our previous lemma, we know that, for all bounded $\mathcal{A}$-measurable random variables X on Ω,

$$F_X : \begin{cases} \Delta(\mathcal{A}) \to \mathbb{R} \\ p \mapsto \mathbb{E}_p(X), \end{cases}$$

is continuous in the product topology. Thus, since $\mathbf{P}_{\mathfrak{P}}$ is connected in the product topology, the image of $\mathbf{P}$ under F_X is connected in $\mathbb{R}$. That is, $F_X(\mathbf{P})$ is an *interval* in $\mathbb{R}$. But, by definition,

$$F_X(\mathbf{P}) = \mathfrak{P}(X).$$

So, for any random variable X on Ω, $\mathfrak{P}(X)$ is an interval of $\mathbb{R}$.

Now pick any $p, q \in \mathbf{P}$ and $\lambda \in (0,1)$, and define

$$r := \lambda p + (1-\lambda)q.$$

Since expectation is a linear functional of probability functions,

$$\mathbb{E}_r(X) = \lambda \mathbb{E}_p(X) + (1-\lambda)\mathbb{E}_q(X),$$

for any X on Ω. Since $\mathbb{E}_p(X), \mathbb{E}_q(X) \in \mathfrak{P}(X)$ and $\mathfrak{P}(X)$ is an interval, it follows that $\mathbb{E}_r(X) \in \mathfrak{P}(X)$ for any X on Ω. That is, r is compatible with $\mathfrak{P}$, i.e. $r \in \mathbf{P}_{\mathfrak{P}}$. So $\mathbf{P}_{\mathfrak{P}}$ is convex.

■

Appendix F

Mutual Independence

For convenience, let's first reproduce the axioms:

> **Axiom 0. Domain.** $p(\cdot|\cdot)$ is a real-valued function on $\mathcal{A} \times \mathcal{A}$, for some Boolean algebra $\mathcal{A}$.
>
> **Axiom 1. Unity.** $p(A|A) = p(B|B)$.
>
> **Axiom 2. Monotony.** $p(A \wedge B|C) \leq p(A|C)$.
>
> **Axiom 3. Non-Triviality.** There are $A, B, C, D \in \mathcal{A}$ such that $p(A|B) \neq p(C|D)$.
>
> **Axiom 4. Complementation.** Whenever $p(C|B) \neq p(B|B)$ for some C, then $p(B|B) = p(A|B) + p(\neg A|B)$.
>
> **Axiom 5. Multiplication.** $p(A \wedge B|C) = p(A|B \wedge C) \cdot p(B|C)$.

To prove mutual independence, we construct, for each axiom, a model which falsifies that axiom and verifies all others. I'll leave proofs implicit except where they are non-trivial. In the following, the Boolean *meet* $\wedge$ and *complement* $\vee$ operations on sets are always understood to be intersection and set-theoretic complement, respectively.

Where $\mu : \mathcal{A} \to K$ is a function from some set $\mathcal{A}$ into field K, let $p_\mu(\cdot|\cdot) : \mathcal{A} \times \mathcal{A} \to K$ be such that

$$p_\mu(A|B) = \begin{cases} \frac{\mu(A \wedge B)}{\mu(B)} & \text{if } B \neq \emptyset, \\ 1 & \text{if } B = \emptyset. \end{cases}$$

- **Axiom 0:**
  - ▷ Axioms 1–5 are compatible with p taking irreal values in some larger ordered field, e.g. in the hyperreals. Where a is an infinitesimal number, let $\Omega = \{w_1, w_2\}$ and μ be a hyperreal-valued function over $\mathcal{P}(\Omega)$, with $\mu(\emptyset) = 0$, $\mu(\{w_1\}) = a$, $\mu(\{w_2\}) = 1 - a$, and $\mu(\Omega) = 1$. Then one may check that p_μ is a model of Axioms 1–5, but not of Axiom 0—in particular, it doesn't witness the real-valuedness of p.
  - ▷ Axioms 1–5 are also compatible with $\mathcal{A}(\Omega)$ not being a Boolean algebra over Ω. Let $\Omega = \{w_1, w_2\}$, $\mathcal{A}(\Omega) = \{\emptyset, \{w_1\}, \Omega\}$, and let μ' be a real-valued function over $\mathcal{A}(\Omega)$ with $\mu'(\emptyset) = 0$, $\mu'(\{w_1\}) = 1/2$, and $\mu'(\Omega) = 1$. Then one may check that $p_{\mu'}$ is a model of Axioms 1–5, but not of Axiom 0—in particular, $\mathcal{A}(\Omega)$ is not a Boolean algebra.
- **Axiom 1:** Let $\Omega = \{w\}$ and define $p : \mathcal{P}(\Omega) \to [0,1]$ such that $p(\Omega|\Omega) = p(\Omega|\emptyset) = 1$ and $p(\emptyset|\Omega) = p(\emptyset|\emptyset) = 0$. Then p is a model of Axioms 0 and 2–5, but not of Axiom 1—in particular, $p(\emptyset|\emptyset) = 0 \neq 1 = p(\Omega|\Omega)$.
- **Axiom 2:** Let $\Omega = \{w_1, w_2\}$ and μ be a function over $\mathcal{P}(\Omega)$, with $\mu(\emptyset) = 0$, $\mu(\{w_1\}) = 2$, $\mu(\{w_2\}) = -1$, and $\mu(\Omega) = 1$. Then p_μ is a model of Axioms 0, 1, and 3–5, but not of Axiom 2—in particular, $p_\mu(\{w_1\}|\Omega) = 2 > 1 = p_\mu(\Omega|\Omega)$.
- **Axiom 3:** Let $\Omega = \{w\}$ and define $p : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \{1\}$ to be constant. Then p is a model of Axioms 0–2, 4, 5, but not of Axiom 3.

- **Axiom 4:** Let $\Omega = \{w_1, w_2\}$ and μ be a function over $\mathcal{P}(\Omega)$, with $\mu(\emptyset) = 0$, $\mu(\{w_1\}) = \mu(\{w_2\}) = \mu(\Omega) = 1$. Then p_μ is a model of Axioms 0–3 and 5, but not of Axiom 4. To see that Complementation is falsified: $p_\mu(\emptyset|\Omega) = 0 \neq 1 = p(\Omega|\Omega)$, but $p_\mu(\{w_1\}|\Omega) + p_\mu(\{w_2\}|\Omega) = 2 \neq 1 = p_\mu(\Omega|\Omega)$. To see that Multiplication is satisfied: when $C = \emptyset$ or $A \wedge B \wedge C \neq \emptyset$, $p_\mu(A \wedge B|C) = 1 = p_\mu(A|B \wedge C) \cdot p_\mu(B|C)$, and when $C \neq \emptyset$ and $A \wedge B \wedge C = \emptyset$, $p_\mu(A \wedge B|C) = 0 = p_\mu(A|B \wedge C) \cdot p_\mu(B|C)$.

- **Axiom 5:** Let $\Omega = \{w_1, w_2\}$ and μ be a function over $\mathcal{P}(\Omega)$, with $\mu(\emptyset) = 0$, $\mu(\{w_1\}) = \mu(\{w_2\}) = 1/2$ and $\mu(\Omega) = 1$. Define $p : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to [0,1]$ such that

$$p(A|B) = \begin{cases} \mu(A) & \text{if } B \neq \emptyset, \\ 1 & \text{if } B = \emptyset. \end{cases}$$

  Then p is a model of Axioms 0–4, but not of Axiom 5. To see that Multiplication is not satisfied: $p(\{w_1\} \wedge \{w_2\}|\Omega) = p(\emptyset|\Omega) = 0 \neq 1/2 \cdot 1/2 = p(\{w_1\}|\{w_2\} \wedge \Omega) \cdot p(\{w_2\}|\Omega)$. To see that Complementation is still satisfied: if $p(C|B) \neq p(B|B) = 1$ for some C, then $B \neq \emptyset$ and so $p(A|B) = \mu(A)$ for all A. The only two sets of mutual complements are $\{\{w_1\}, \{w_2\}\}$ and $\{\emptyset, \Omega\}$, and we have $p(B|B) = 1 = 1/2 + 1/2 = p(\{w_1\}|B) + p(\{w_2\}|B)$ and $p(B|B) = 1 = 0 + 1 = p(\emptyset|B) + p(\Omega|B)$.

  ■

Appendix G

Properties of Popper Functions

Here we'll prove that our primitively conditional probability functions, defined by Axioms 0–5, satisfy the following canonical properties:

- **Non-Negativity.** $0 \leq p(A|B)$.
- **Normalization.** $p(A|B) \leq p(B|B) = 1$.
- **Additivity.** Whenever $A \wedge B = \emptyset$ and $p(D|C) \neq 1$ for some D, then $p(A \vee B|C) = p(A|C) + p(B|C)$.
- **Explosion.** $p(A|C \wedge \neg C) = 1$.
- **Substitutivity.** If $p(A|C) = p(B|C)$ for all $C \in \mathcal{A}(\Omega)$, then $p(D|A) = p(D|B)$.

The main purpose of this exercise is to create more familiarity with Popper functions. Except for the case of Substitutivity, the following proofs are found in similar form in Popper (1968, p. 357ff).

Proof of Non-Negativity: Suppose, for contradiction, that $p(A|B) < 0$. By Complementation, if $p(A|B) \neq p(B|B)$, then $p(B|B) = p(B|B) + p(\neg B|B)$, and thus $p(\neg B|B) = 0$; so, by Multiplication, $p(A \wedge \neg B|B) = p(A|\neg B \wedge B) \cdot p(\neg B|B) = 0$. Thus, by Monotony, if $p(A|B) \neq p(B|B)$, $0 = p(A \wedge \neg B|B) \leq p(A|B)$. But by assumption we have $p(A|B) < 0$;

thus $p(A|B) = p(B|B)$, and so $p(B|B) < 0$. But by Multiplication and Booleanism, for all $C \in \mathcal{A}$, $p(C|C) = p(C \wedge C|C) = p(C|C \wedge C) \cdot p(C|C) = p(C|C)^2$, hence either $p(C|C) = 0$ or $p(C|C) = 1$; in particular $p(C|C) \geq 0$ for all $C \in \mathcal{A}$. Contradiction. So, $p(A|B) \geq 0$. ■

Proof of Normalization: Above we've established that either $p(B|B) = 0$ for all $B \in \mathcal{A}$ or $p(B|B) = 1$ for all $B \in \mathcal{A}$. Suppose, for contradiction, that $p(B|B) = 0$ for all $B \in \mathcal{A}$. By Complementation, either $p(A|B) = p(B|B)$ for all A, or $p(B|B) = p(A|B) + p(\neg A|B)$ for all $A \in \mathcal{A}$. By Non-Negativity, the second disjunct implies that $p(A|B) = 0$ for all $A \in \mathcal{A}$. In either case, then, we have $p(A|B) = p(B|B)$ for all $A \in \mathcal{A}$. Since this holds for all $B \in \mathcal{A}$, we have a contradiction with Non-Triviality. Thus, $p(B|B) = 1$ for all $B \in \mathcal{A}$. Again, by Complementation, we have either $p(A|B) = p(B|B)$ for all A, or $p(B|B) = p(A|B) + p(\neg A|B)$ for all $A \in \mathcal{A}$. In either case, it follows that $p(B|B) \geq p(A|B)$ for all $A \in \mathcal{A}$. So, overall, $p(A|B) \leq p(B|B) = 1$. ■

Proof of Additivity: Let $A \wedge B = \perp$ and suppose $p(D|C) \neq 1$ for some D. By Normalization, $p(D|C) \neq p(C|C)$ for some D. Hence, by Complementation, $1 = p(C|C) = p(A \vee B|C) + p(\neg A \wedge \neg B|C)$; rearranging gives $p(A \vee B|C) = 1 - p(\neg A \wedge \neg B|C)$. By Multiplication and Complementation,

$$\begin{aligned}
p(\neg A \wedge \neg B|C) &= p(\neg B|\neg A \wedge C) \cdot p(\neg A|C) \\
&= (1 - p(B|\neg A \wedge C)) \cdot p(\neg A|C) \\
&= p(\neg A|C) - p(B \wedge \neg A|C) \\
&= p(\neg A|C) - p(B|C),
\end{aligned}$$

where in the last line we used the fact that, since $A \wedge B = \perp$, by Booleanism we have $B = B \wedge \neg A$. So,

$$\begin{aligned}
p(A \vee B|C) &= 1 - p(\neg A \wedge \neg B|C) \\
&= (1 - p(\neg A|C)) + p(B|C) \\
&= p(A|C) + p(B|C). \quad \blacksquare
\end{aligned}$$

Proof of Explosion: By Booleanism, $C \wedge \neg C = A \wedge C \wedge \neg C$, and so, by Normalization, $p(A \wedge C \wedge \neg C|C \wedge \neg C) = p(C \wedge \neg C|C \wedge \neg C) = 1$. By Multiplication and Normalization, $1 = p(A \wedge C \wedge \neg C|C \wedge \neg C) = p(A|C \wedge \neg C) \cdot p(C \wedge \neg C|C \wedge \neg C) = p(A|C \wedge \neg C) \cdot 1 = p(A|C \wedge \neg C)$. ■

Proof of Substitutivity: Suppose $p(A|C) = p(B|C)$ for all $C \in \mathcal{A}$. Then, by Normalization, $p(B|A) = p(A|A) = 1$ and $p(A|B) = p(B|B) = 1$. Suppose, for contradiction, that $p(D|B) = 1$ for all D but that $p(C|A) \neq 1$ for some C. By Additivity and Non-Negativity, $0 = 1 - p(B|A) = p(\neg B|A) \geq p(A \wedge \neg B|A) \geq 0$, and thus $p(A \wedge \neg B|A) = 0$. Hence, for all C,

$$\begin{aligned}
p(C|A) &\overset{\text{Norm.}}{=} p(C|A) \cdot p(A|A) \\
&\overset{\text{Mult.}}{=} p(C \wedge A|A) \\
&\overset{\text{Add.}}{=} p(C \wedge A \wedge \neg B|A) + p(C \wedge A \wedge B|A) \\
&\overset{\text{Mult.}}{=} p(C|A \wedge \neg B) \cdot p(A \wedge \neg B|A) + p(C \wedge B|A) \cdot p(A|A) \\
&= p(C \wedge B|A) \cdot p(A|A) \\
&\overset{\text{Norm.}}{=} p(C \wedge B|A) \\
&\overset{\text{Mult.}}{=} p(C|A \wedge B) \cdot p(B|A) \\
&= p(C|A \wedge B) \\
&\overset{\text{Mult.}}{=} p(C \wedge A|B)/p(A|B) \\
&\overset{\text{Ass.}}{=} 1
\end{aligned}$$

Contradiction. Analogously, we can derive a contradiction from the assumptions that $p(D|A) = 1$ for all D but that $p(C|B) \neq 1$ for some C.

So either (i) $p(D|B) = p(D|A) = 1$ for all D, or (ii) $p(C|B) \neq 1$ for some C and $p(C'|A) \neq 1$ for some C'. If (i), we are done. So assume (ii). By Additivity and Non-Negativity, we have, as before, $p(A \wedge \neg B|A) = 0$ and, by an exactly parallel argument, $p(\neg A \wedge B|B) = 0$; by Monotony, $p(D \wedge A \wedge \neg B|B) = 0$ for any $D \in \mathcal{A}$. Thus, for any

$D \in \mathcal{A}$,

$$\begin{aligned} p(D|B) &= p\big(D \wedge A|B\big) + p(D \wedge \neg A|B) \\ &= p(D|A \wedge B) \cdot p(A|B) + p(D \wedge \neg A|B) \cdot p(B|B) \\ &= p\big(D|A \wedge B\big) + p(D \wedge (\neg A) \wedge B|B) \\ &= p(D|A \wedge B). \end{aligned}$$

Analogously, $p(D|A) = p(D|A \wedge B)$, and thus

$$p(D|A) = p(D|B). \quad \blacksquare$$

Appendix H

Counterfactualist SEM Accounts & Sufficiency

Throughout this Appendix, I am concerned with acyclic SEMs. I also assume the validity of *Conditional Excluded Middle*—$\vdash (A \mathrel{\Box\!\!\to} B) \vee (A \mathrel{\Box\!\!\to} \neg B)$. This is the natural setting for acyclic SEMs, since all value assignments to exogenous variables have unique solutions.

Now recall the counterfactualist notion of SEM adequacy:

> **Counterfactual Adequacy:** Necessarily, $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ is adequate only if: $\mathcal{V}$ is suitable and, for all $Z \in \mathcal{V}$, $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ *iff* for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$,
>
> $$\mathcal{V}\backslash\{Z\} = \mathbf{v} \mathrel{\Box\!\!\to} Z = f_Z(\mathbf{v}).$$

Throughout the Appendix, "adequate" is assumed to satisfy **Counterfactual Adequacy**, which I'll abbreviate as **CA**.

H.1 Paths in Adequate Models

In a slight abuse of notation, where $\mathcal{E}$ is a set of structural equations, let $(X_1 \to X_2 \to ... \to X_n) \in \mathcal{E}$ denote the fact that, for each $i = 1, ..., n-1$, $f_{X_{i+1}}$ is non-constant in X_i

(for some assignment of values to $\mathcal{V}\backslash\{X_{i+1}, X_i\}$). Recall that, when $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ and $(X_1 \rightarrow X_2 \rightarrow ... \rightarrow X_n) \in \mathcal{E}$, I say that $\mathcal{M}$ *contains a directed path from* X_1 *to* X_n.

We now show that, if there's an *immediate* directed path from one variable to another in an adequate acyclic model, then adding an extra variable to the model either preserves that path or extends it by one variable, unless it creates cycles.

> ***Lemma A1.*** **Path Extension.** For any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$, any $C, E \in \mathcal{V}$, and any variable X: if $(C \rightarrow E) \in \mathcal{E}$ and $\mathcal{M}' = (\mathcal{V} \cup \{X\}, \mathcal{E}')$ is adequate and acyclic, then either $(C \rightarrow E) \in \mathcal{E}'$ or $(C \rightarrow X \rightarrow E) \in \mathcal{E}'$.

Proof of Lemma A1: Let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ be adequate with $(C \rightarrow E) \in \mathcal{E}$. By **CA**, there are values $c^*, c^{**}, e^*, e^{**}, \mathbf{v}^*$ with $c^* \neq c^{**}$ and $e^* \neq e^{**}$ such that

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \;\Box\!\!\rightarrow E = e^*, \text{ and} \tag{H.1}$$

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \;\Box\!\!\rightarrow E = e^{**}. \tag{H.2}$$

Let $\mathcal{M}' = (\mathcal{V} \cup \{X\}, \mathcal{E}')$ be adequate. By CEM, we have two cases—intuitively, they correspond to either X screening off all dependence of E on C or failing to do so:

Case 1 (No Screening). There are values $c^\dagger, c^\ddagger, e^\dagger, e^\ddagger, \mathbf{v}$, with $c^\dagger \neq c^\ddagger$ and $e^\dagger \neq e^\ddagger$, and a value x of X such that

$$C = c^\dagger \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v} \wedge X = x \;\Box\!\!\rightarrow E = e^\dagger, \text{ and}$$

$$C = c^\ddagger \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v} \wedge X = x \;\Box\!\!\rightarrow E = e^\ddagger.$$

By **CA**, it follows that $(C \rightarrow E) \in \mathcal{E}'$.

Case 2 (Screening). For every value x of X and $\mathbf{v}$ of $\mathcal{V}\backslash\{C, E\}$, there is a value $e_{x,\mathbf{v}}$ of E such that, for every value γ of C,

$$C = \gamma \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v} \wedge X = x \;\Box\!\!\rightarrow E = e_{x,\mathbf{v}}. \tag{H.3}$$

We first prove that $(X \to E) \in \mathcal{E}$. The rule *Conjunction Shift*—$(A \mathrel{\Box\!\!\rightarrow} (B \wedge C)) \vdash ((A \wedge B) \mathrel{\Box\!\!\rightarrow} C)$—is valid in any standard logic for counterfactuals (including Stalnaker (1968); Lewis (1973b)).[1]Given condition H.1, CEM, and Conjunction Shift, there is a value x' of X such that

$$C = c^* \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \wedge X = x' \mathrel{\Box\!\!\rightarrow} E = e^*. \tag{H.4}$$

By condition H.3,

$$C = c^{**} \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \wedge X = x' \mathrel{\Box\!\!\rightarrow} E = e^*. \tag{H.5}$$

Negation Transfer—$(A \mathrel{\Box\!\!\rightarrow} C), (A \wedge B \mathrel{\Box\!\!\rightarrow} \neg C) \vdash (A \mathrel{\Box\!\!\rightarrow} \neg B)$—is also valid in any standard logic for counterfactuals.[2] From H.2, H.5, and Negation Transfer,

$$C = c^{**} \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \mathrel{\Box\!\!\rightarrow} E = e^{**} \wedge X \neq x',$$

and so, by CEM and Conjunction Shift, there is a value $x'' \neq x'$ such that

$$C = c^{**} \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \wedge X = x'' \mathrel{\Box\!\!\rightarrow} E = e^{**}. \tag{H.6}$$

From H.6 and H.3,

$$C = c^* \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \wedge X = x'' \mathrel{\Box\!\!\rightarrow} E = e^{**}. \tag{H.7}$$

Finally, by **CA**, H.4, and H.7, $(X \to E) \in \mathcal{E}$.

Second, we prove that either $(C \to X) \in \mathcal{E}$ or $(E \to X) \in \mathcal{E}$. Let x be any value of X. Since $e \neq e'$, either $e_{x,\mathbf{v}^*} \neq e^*$ or $e_{x,\mathbf{v}^*} \neq e^{**}$. Suppose $e_{x,\mathbf{v}^*} \neq e^*$. Then, from conditions H.1 and H.3, Negation Transfer, and Agglomeration,

$$C = c^* \wedge \mathcal{V} \backslash \{C, E\} = \mathbf{v}^* \mathrel{\Box\!\!\rightarrow} (X \neq x \wedge E = e^*),$$

[1]Given Agglomeration (see below), Conjunction Shift is equivalent to *Cautious Monotonicity*, $(A \mathrel{\Box\!\!\rightarrow} B), (A \mathrel{\Box\!\!\rightarrow} C)) \vdash ((A \wedge B) \mathrel{\Box\!\!\rightarrow} C)$.

[2]Given CEM, Negation Transfer is the contrapositive of *Rational Monotonicity*, $((A \mathrel{\Box\!\!\rightarrow} B) \wedge \neg(A \mathrel{\Box\!\!\rightarrow} \neg C)) \supset ((A \wedge C) \mathrel{\Box\!\!\rightarrow} B)$.

Thus, by CEM, there is an $x^* \neq x$ such that

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \,\Box\!\!\rightarrow (X = x^* \wedge E = e^*). \tag{H.8}$$

Suppose, for contradiction, that

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \,\Box\!\!\rightarrow X = x^*.$$

Then, by H.2,

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \,\Box\!\!\rightarrow (X = x^* \wedge E = e^{**}). \tag{H.9}$$

By Conjunction Shift, H.8 and H.9 give us, respectively,

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge X = x^* \,\Box\!\!\rightarrow E = e^*), \text{ and}$$

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge X = x^* \,\Box\!\!\rightarrow E = e^{**},$$

in contradiction with condition H.3. So, there is a $x^{**} \neq x^*$ such that

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \,\Box\!\!\rightarrow X = x^{**}. \tag{H.10}$$

By Conjunction Shift, H.8 entails

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge E = e^* \,\Box\!\!\rightarrow X = x^*, \tag{H.11}$$

and H.10 and H.2 together entail

$$C = c^{**} \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge E = e^{**} \,\Box\!\!\rightarrow X = x^{**}. \tag{H.12}$$

H.11 and H.12 together entail that either $(C \to X) \in \mathcal{E}$ or $(E \to X) \in \mathcal{E}$.[3]

[3] For suppose that $(C \to X) \notin \mathcal{E}$; that is, for all $\mathbf{v}, e$, there is a $x_{\mathbf{v},e}$ such that, for all γ,

$$C = \gamma \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v} \wedge E = e \,\Box\!\!\rightarrow X = x_{\mathbf{v},e}.$$

Then H.12 entails

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge E = e^{**} \,\Box\!\!\rightarrow X = x^{**},$$

which together with H.11 entails $(E \to X) \in \mathcal{E}$.

Analogously, suppose that $(E \to X) \notin \mathcal{E}$; that is, for all $c, \mathbf{v}$, there is a $x[c, \mathbf{v}]$ such that, for all ε,

$$C = c \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v} \wedge E = \varepsilon \,\Box\!\!\rightarrow X = x[c, \mathbf{v}].$$

Then H.11 entails

$$C = c^* \wedge \mathcal{V}\backslash\{C, E\} = \mathbf{v}^* \wedge E = e^{**} \,\Box\!\!\rightarrow X = x^*,$$

which together with H.12 entails $(C \to X) \in \mathcal{E}$.

So, we have that $(X \rightarrow E) \in \mathcal{E}$ and either $(C \rightarrow X) \in \mathcal{E}$ or $(E \rightarrow X) \in \mathcal{E}$. In the first case, we have $(C \rightarrow X \rightarrow E) \in \mathcal{E}$, as desired. In the second case, we have $(E \rightarrow X) \in \mathcal{E}$ and $(X \rightarrow E) \in \mathcal{E}$, in contradiction with $\mathcal{M}'$'s acyclicity. So, $(C \rightarrow X \rightarrow E) \in \mathcal{E}$. ■

We say that there is a directed path from a *set* $\mathbf{X}$ to a variable Y iff there is a directed path from some member of $\mathbf{X}$ to Y. Lemma A1 entails the following theorem:

Theorem A1. Let $\mathbf{C} \subseteq \mathcal{V}$ and $E \in \mathcal{V}$ be such that $E \not\in \mathbf{C}$. If there are values $\mathbf{c}, \mathbf{c}', e$ such that $\mathbf{C} = \mathbf{c} \;\Box\!\!\rightarrow E = e$ and $\mathbf{C} = \mathbf{c}' \;\Box\!\!\rightarrow E \neq e$, then in any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathbf{C} \cup \{E\} \subseteq \mathcal{V}$, there is a directed path from $\mathbf{C}$ to E.

Proof of Theorem A1: Let $\mathbf{C}$ and E be such that $E \not\in \mathbf{C}$. We proceed by induction on the number of variables n in $\mathcal{V}$. Our base case is $n = |\mathbf{C}| + 1$. We want to show that, for any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathbf{C} \cup \{E\}$, there is a directed path from $\mathbf{C}$ to E. We prove this base case, in turn, by (sub-)induction on the number $k \leq n$ of variables in $|\mathbf{C}|$.

- <u>Sub-induction on k:</u> The base case $k = 1$ for our sub-induction is immediate: given CEM, $C = c' \;\Box\!\!\rightarrow E \neq e$ entails that there is an $e' \neq e$ with $C = c' \;\Box\!\!\rightarrow E = e'$, and given **CA**, it then follows from $C = c \;\Box\!\!\rightarrow E = e$ and $C = c' \;\Box\!\!\rightarrow E = e'$ that there's a path from $\mathbf{C} = \{C\}$ to E in $\mathcal{M}$. For the induction step, $k \mapsto k + 1 \leq n$, assume (as the induction hypothesis) that, for any $\mathbf{C}, E$ with $E \not\in \mathbf{C}$ and $|\mathbf{C}| = k$, and any adequate, acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathbf{C} \cup \{E\}$ such that there are value assignments $\mathbf{c}, \mathbf{c}', e$ with $\mathbf{C} = \mathbf{c} \;\Box\!\!\rightarrow E = e$ and $\mathbf{C} = \mathbf{c}' \;\Box\!\!\rightarrow E \neq e$, there is a directed path from $\mathbf{C}$ to E. Let now $\mathcal{M}' = (\mathcal{V}', \mathcal{E}')$ with $\mathcal{V}' = \mathbf{C}' \cup \{E'\}$ and $|\mathbf{C}'| = k + 1$ such that there are value assignments $\mathbf{c}, \mathbf{c}', e$ with $\mathbf{C}' = \mathbf{c}' \;\Box\!\!\rightarrow E' = e'$ and $\mathbf{C}' = \mathbf{c}' \;\Box\!\!\rightarrow E' \neq e$. Pick any $V \in \mathcal{V}' \backslash (\mathbf{C}' \cup \{E'\})$. It follows from the definition of suitability that, whenever a variable set $\mathbf{X}$ is suitable, any subset of $\mathbf{X}$ is suitable. Hence, by **Corollary A1**, if any model with variable set $\mathbf{X}$ is adequate, then for any $X \in \mathbf{X}$, there is an adequate model with variable set $\mathbf{X}\backslash\{X\}$. In particular, there is an adequate model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathcal{V}'\backslash\{V\}$. Hence, by the induction hypothesis, $(X_1 \to X_2 \to ... \to X_{n-1} \to X_n) \in \mathcal{E}$ for some $\{X_i\}_{i=1}^{n} \subseteq \mathcal{V}$ with $X_1 \in \mathbf{C}$ and $X_n = E$. **Lemma A1** entails that, for all $i = 1, ..., n$, either $(X_i \to X_{i+1}) \in \mathcal{E}'$ or $(X_i \to V \to X_{i+1}) \in \mathcal{E}'$. Hence, for every $i = 1, ..., n$ there is a path from X_i to X_{i+1} in $\mathcal{M}'$, and so there is a path from X_1 to E, and hence from $\mathbf{C}$ to E, in $\mathcal{M}'$.

This proves the base case.

For the (main) induction step, $n \mapsto n+1$, suppose (as the induction hypothesis) that in any adequate acyclic model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n \geq |\mathbf{C}| + 1$, if $\mathbf{C} \cup \{E\} \subseteq \mathcal{V}$ with $E \notin \mathbf{C}$, then $\mathcal{M}$ contains a directed path from $\mathbf{C}$ to E. Let $\mathcal{M}' = (\mathcal{V}', \mathcal{E}')$ with $|\mathcal{V}'| = n+1$ and $\mathbf{C} \cup \{E\} \subseteq \mathcal{V}'$ be an adequate acyclic model. Choose any $V \in \mathcal{V}'\backslash(\mathbf{C} \cup \{E\})$. Then, by the same reasoning as before, there is an adequate model $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \mathcal{V}'\backslash\{V\}$. By the induction hypothesis, $(X_1 \to X_2 \to ... \to X_{n-1} \to X_n) \in \mathcal{E}$ for some $\{X_i\}_{i=1}^{n} \subseteq \mathcal{V}$ with $X_1 \in \mathbf{C}$ and $X_n = E$. By **Lemma A1**, for all $i = 1, ..., n$, either $(X_i \to X_{i+1}) \in \mathcal{E}'$ or $(X_i \to V \to X_{i+1}) \in \mathcal{E}'$. Hence, for every $i = 1, ..., n$ there is a path from X_i to X_{i+1} in $\mathcal{M}'$, and so there is a path from X_1 to E, and hence from $\mathbf{C}$ to E, in $\mathcal{M}'$. ■

H.2 Counterfactual Dependence and Dependence in Acyclic Models

For any SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$, $\mathcal{V}_{\text{ex}}^{\mathcal{E}}$ and $\mathcal{V}_{\text{en}}^{\mathcal{E}}$ denote $\mathcal{M}$'s exogenous variables and $\mathcal{M}$'s endogenous variables, respectively. Where the model is obvious from context, I'll omit the superscripted $\mathcal{E}$.

For any SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $X \in \mathcal{V}_{\text{en}}$, let $\mathcal{M}_{\rangle X \langle} = (\mathcal{V}\backslash\{X\}, \mathcal{E}_{\rangle X \langle})$ be *the result of collapsing X in $\mathcal{M}$*: where $[X := f_X(\mathcal{V}\backslash\{X\})] \in \mathcal{E}$, $\mathcal{E}_{\rangle X \langle}$ is the result of deleting $[X := f_X(\mathcal{V}\backslash\{X\})]$ from $\mathcal{E}$ and replacing all other instances of X in $\mathcal{E}$ by $f_X(\mathcal{V}\backslash\{X\})$. A standard result about acyclic SEMs is that collapsing preserves entailment relations. It also turns out to preserve adequacy.

Lemma A2.1. **Endogenous Collapse:** Let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ be acyclic with $X \in \mathcal{V}_{\text{en}}$. Then $\mathcal{M}_{\rangle X \langle}$ is acyclic, adequate, and, for all $\mathbf{Y}, \mathbf{Z} \subseteq \mathcal{V} \backslash \{X\}$ and values $\mathbf{y}, \mathbf{z}$ of $\mathbf{Y}$ and $\mathbf{Z}$,

$$(\mathcal{M}(\mathbf{Y} \leftarrow \mathbf{y}) \vDash \mathbf{Z} = \mathbf{z}) \leftrightarrow (\mathcal{M}_{\rangle X \langle}(\mathbf{Y} \leftarrow \mathbf{y}) \vDash \mathbf{Z} = \mathbf{z}).$$

Proof of Lemma A2.1: Both $\mathcal{M}_{\rangle X \langle}$'s acyclicity and the biconditional are standard results. It remains to prove $\mathcal{M}_{\rangle X \langle}$'s adequacy.

Left-to-right direction: Suppose $\mathcal{M}$ is adequate. Let $Z \neq X$ with $[Z := f_Z^{\rangle X \langle}(\mathcal{V} \backslash \{X, Z\})] \in \mathcal{E}_{\rangle X \langle}$. Then $[Z := f_Z(\mathcal{V} \backslash \{Z\})] \in \mathcal{E}$ with f_Z either (i) constant in X or (ii) non-constant in X. Suppose (i). Then $f_Z(\mathcal{V} \backslash \{Z\}) = f_Z^{\rangle X \langle}(\mathcal{V} \backslash \{X, Z\})$. Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all values $\mathbf{v}$ of $\mathcal{V} \backslash \{Z\}$,

$$\mathcal{V} \backslash \{Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z(\mathbf{v}),$$

and hence, for all values x of X and $\mathbf{v}'$ of $\mathcal{V} \backslash \{X, Z\}$,

$$X = x \wedge \mathcal{V} \backslash \{X, Z\} = \mathbf{v}' \mathrel{\Box\!\!\rightarrow} Z = f_Z^{\rangle X \langle}(\mathbf{v}').$$

By the disjunction rule—$(A \mathrel{\Box\!\!\rightarrow} C) \wedge (B \mathrel{\Box\!\!\rightarrow} C) \vdash ((A \vee B) \mathrel{\Box\!\!\rightarrow} C)$—it follows that, for all values $\mathbf{v}'$ of $\mathcal{V} \backslash \{X, Z\}$,

$$\mathcal{V} \backslash \{X, Z\} = \mathbf{v}' \mathrel{\Box\!\!\rightarrow} Z = f_Z^{\rangle X \langle}(\mathbf{v}').$$

Suppose instead (ii). Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all values $\mathbf{v}$ of $\mathcal{V} \backslash \{X\}$,

$$\mathcal{V} \backslash \{X\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} X = f_X(\mathbf{v}). \tag{H.13}$$

Since $\mathcal{M}$ is acyclic, H.13 implies that, for all values $\mathbf{v}'$ of $\mathcal{V} \backslash \{X, Z\}$ and z of Z,

$$\mathcal{V} \backslash \{X, Z\} = \mathbf{v}' \mathrel{\Box\!\!\rightarrow} X = f_X(z, \mathbf{v}'). \tag{H.14}$$

By $\mathcal{M}$'s adequacy, we also have

$$X = f_X(z, \mathbf{v}') \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \,\Box\!\!\rightarrow Z = f_Z(f_X(z, \mathbf{v}'), \mathbf{v}'). \tag{H.15}$$

From H.14, H.15, and *Cautious Transitivity*—$((A \,\Box\!\!\rightarrow B) \wedge ((A \wedge B) \,\Box\!\!\rightarrow C)) \vdash (A \,\Box\!\!\rightarrow C)$—we obtain, for all values $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$, and z of Z,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \,\Box\!\!\rightarrow Z = f_Z(f_X(z, \mathbf{v}'), \mathbf{v}'). \tag{H.16}$$

Since $f_Z(f_X(z, \mathbf{v}'), \mathbf{v}')$ is constant in z, $f_Z(f_X(z, \mathbf{v}'), \mathbf{v}') = f_Z^{\rangle X \langle}(\mathbf{v}')$. Hence, from H.16, for all $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \,\Box\!\!\rightarrow Z = f_Z^{\rangle X \langle}(\mathbf{v}').$$

Right-to-left direction: Conversely, suppose that, for some function g and for all $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \,\Box\!\!\rightarrow Z = g(\mathbf{v}'). \tag{H.17}$$

Let $[X := f_X(\mathcal{V}\backslash\{X\})] \in \mathcal{E}$. Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all $z$of Z and $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$Z = z \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \,\Box\!\!\rightarrow X = f_X(z, \mathbf{v}'). \tag{H.18}$$

Either (i) $f_X(\mathcal{V})$ is constant in Z or (ii) $f_X(\mathcal{V})$ is non-constant in Z. Suppose (i). Then, by the Disjunction Rule and H.18,

$$\mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \,\Box\!\!\rightarrow X = f_X(z, \mathbf{v}'). \tag{H.19}$$

From H.17, H.19, Aggregation, and Conjunction Shift, for all z of Z and $\mathbf{v}'$ of $\mathcal{V}\backslash\{X, Z\}$,

$$X = f_X(z, \mathbf{v}') \wedge \mathcal{V}\backslash\{X, Z\} = \mathbf{v}' \,\Box\!\!\rightarrow Z = g(f_X(z, \mathbf{v}'), \mathbf{v}').$$

Since $\mathcal{M}$ is adequate, it follows by **CA** that $[Z := h(X, \mathcal{V}\backslash\{X, Z\})] \in \mathcal{E}$ with $h|_{X \in \text{range}(f_X)} = g$. Since $[X := f_X(Z, \mathcal{V}\backslash\{X\})] \in \mathcal{E}$ with f_X constant in Z, we thus have $[Z :=$

$h^{\rangle X\langle}(\mathcal{V}\backslash\{X,Z\})] \in \mathcal{E}_{\rangle X\langle}$. Since $h|_{X\in\text{range}(f_X)} = g$, we have $h^{\rangle X\langle}(\mathcal{V}\backslash\{X,Z\}) = g$, and thus $[Z := g(\mathcal{V}\backslash\{X,Z\})] \in \mathcal{E}_{\rangle X\langle}$.

Suppose instead (ii). Since $\mathcal{M}$ is acyclic, f_Z is constant in X. Thus, since $\mathcal{M}$ is adequate, we have by **CA** and the Disjunction Rule that, for all $\mathbf{v}'$ of $\mathcal{V}\backslash\{X,Z\}$ and x of X,

$$\mathcal{V}\backslash\{X,Z\} = \mathbf{v}' \mathrel{\Box\!\!\rightarrow} Z = f_Z(x,\mathbf{v}'). \tag{H.20}$$

From H.17, H.20, and $\mathcal{V}$'s suitability (cf. fn. 38), $f_Z(x,\mathbf{v}') = g(\mathbf{v}')$ for all $\mathbf{v}'$ of $\mathcal{V}\backslash\{X,Z\}$. Thus, $[Z := g(\mathcal{V}\backslash\{X,Z\})] \in \mathcal{E}$; since g is independent of X, $[Z := g(\mathcal{V}\backslash\{X,Z\})] \in \mathcal{E}_{\rangle X\langle}$.
■

Lemma A2.2. **Exogenous Substitution.** Let $\mathcal{M} = (\mathcal{V},\mathcal{E})$ be adequate and acyclic. Then, for any $U \in \mathcal{V}_{\text{ex}}$: $\mathcal{M}(U \leftarrow u)$ is acyclic; if moreover $U = u$, then $\mathcal{M}(U \leftarrow u)$ is adequate.

Proof of Lemma A2.2: Since substitution can only *remove* edges, acyclicity is immediate. It remains to prove adequacy.

For any function f_Z, let f_Z^u denote the result of replacing all instances of U in f_Z by u. Let $U = u$. We want to show that $[Z := f_Z^u(\mathcal{V}\backslash\{U,Z\})] \in \mathcal{E}(U \leftarrow u)$ iff, for all $\mathbf{v}$ of $\mathcal{V}\backslash\{U\}$,

$$\mathcal{V}\backslash\{U,Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}).$$

Right-to-left direction: Suppose that, for all $\mathbf{v}$ of $\mathcal{V}\backslash\{U,Z\}$,

$$\mathcal{V}\backslash\{U,Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}). \tag{H.21}$$

Suppose, for contradiction, that for some $\mathbf{v}^\dagger$ and some $z^\dagger \neq f_Z^u(\mathbf{v}^\dagger)$,

$$U = u \wedge \mathcal{V}\backslash\{U,Z\} = \mathbf{v}^\dagger \mathrel{\Box\!\!\rightarrow} Z = z^\dagger. \tag{H.22}$$

By H.21, H.22, Negation Transfer, and Conjunction Shift,

$$\mathcal{V}\backslash\{U,Z\} = \mathbf{v}^\dagger \wedge Z = z^\dagger \mathrel{\Box\!\!\rightarrow} U \neq u. \tag{H.23}$$

Since $U = u$, we have, by And-to-If,

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{v}^* \wedge Z = z^* \mathrel{\Box\!\!\rightarrow} U = u. \tag{H.24}$$

where $\mathcal{V}\backslash\{U\} = \mathbf{v}^*$ and $Z = z$. By **Theorem 1**, Conditions H.23 and H.24 entail that there's a directed path from $\mathcal{V}\backslash\{U\}$ to U, in contradiction with the assumption that $U \in \mathcal{V}_{\text{ex}}$. Therefore, we have, for all values $\mathbf{x}$ of $\mathbf{X}$ and $\mathbf{w}$ of $\mathcal{V}\backslash(\mathbf{X} \cup \{U\})$,

$$U = u \wedge \mathcal{V}\backslash\{U, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}). \tag{H.25}$$

Since $\mathcal{M}$ is adequate, by **CA** condition H.25 entails that $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ with $f_Z(u, \mathbf{v}) = f_Z^u(\mathbf{v})$ for all $\mathbf{v}$ of $\mathcal{V}\backslash\{U, Z\}$. Hence $[Z := f_Z^u(\mathcal{V}\backslash\{U, Z\})] \in \mathcal{E}(U \leftarrow u)$.

Left-to-right direction: Suppose $[Z := f_Z^u(\mathcal{V}\backslash\{U, Z\})] \in \mathcal{E}(U \leftarrow u)$. Then $[Z := f_Z(\mathcal{V}\backslash\{Z\})] \in \mathcal{E}$ with $f_Z(u, \mathcal{V}\backslash\{U, Z\}) = f_Z^u(\mathcal{V}\backslash\{U, Z\})$. Since $\mathcal{M}$ is adequate, it follows by **CA** that, for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{Z\}$,

$$\mathcal{V}\backslash\{Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z(\mathbf{v}),$$

and hence, for all $\mathbf{v}' \in \mathcal{V}\backslash\{U, Z\}$

$$U = u \wedge \mathcal{V}\backslash\{U, Z\} = \mathbf{v}' \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}'). \tag{H.26}$$

Suppose, for contradiction, that for some $\mathbf{v}^\dagger$ and $\mathbf{y}^\dagger$,

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{v}^\dagger \mathrel{\Box\!\!\rightarrow} Z \neq f_Z^u(\mathbf{v}^\dagger). \tag{H.27}$$

By Negation Transfer, conditions H.26 and H.27 entail

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{v}^\dagger \mathrel{\Box\!\!\rightarrow} U \neq u. \tag{H.28}$$

Since $U = u$, by And-to-If,

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{x}^* \mathrel{\Box\!\!\rightarrow} U = u, \tag{H.29}$$

where $\mathcal{V}\backslash\{U, Z\} = \mathbf{x}^*$. By **Theorem A1**, conditions H.28 and H.29 now entail that there is a directed path from $\mathcal{V}\backslash\{U, Z\}$ to U in $\mathcal{M}$, in contradiction with $U \in \mathcal{V}_{\text{ex}}$. So, for all values $\mathbf{v}$ of $\mathcal{V}\backslash\{U, Z\}$,

$$\mathcal{V}\backslash\{U, Z\} = \mathbf{v} \mathrel{\Box\!\!\rightarrow} Z = f_Z^u(\mathbf{v}). \quad \blacksquare$$

Theorem A2. **Counterfactual Dependence and Dependence in a Model:** Let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $C, E \in \mathcal{V}$ be adequate and acyclic. Then, if $\mathcal{V}_{\text{ex}}\backslash\{C\} = \mathbf{v}$ and $C = c \mathrel{\Box\!\!\rightarrow} E = e$,

$$\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e.$$

Proof of Theorem A2: If C and E are the same variable, then $C = c \mathrel{\Box\!\!\rightarrow} E = e$ entails that $c = e$ and so $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e$ immediately. So assume, in the following, that C and E are different variables.

We proceed by induction on the number n of variables in the model. For the base case, let $n = 2$ and let $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{C, E\}$ be adequate. Suppose $E \in \mathcal{V}_{\text{ex}}$. Then since $C = c \mathrel{\Box\!\!\rightarrow} E = e$, we have (by **CA**) that, $C = c' \mathrel{\Box\!\!\rightarrow} E = e$ for all values c' of C. Thus, by *Modus Ponens*—$A, A \mathrel{\Box\!\!\rightarrow} B \vdash B$, valid in any standard counterfactual logic—$E = e$. Hence $\mathcal{V}_{\text{ex}}\backslash\{C\} = \{E\}$ and $\mathbf{v} = e$, and so $\mathcal{M}(\mathcal{V}_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e$. Now suppose $E \notin \mathcal{V}_{\text{ex}}$. Then $\mathcal{V}_{\text{ex}}\backslash\{C\} = \varnothing$ and $[E := f_E(C)] \in \mathcal{E}$, with f_E non-constant in C. Since $C = c \mathrel{\Box\!\!\rightarrow} E = e$, we have $f_E(c) = e$, and thus $\mathcal{M}(C \leftarrow c) \vDash E = e$.

For the induction step, assume that for all adequate acyclic models $\mathcal{M}' = (\mathcal{V}', \mathcal{E}')$ with $C, E \in \mathcal{V}'$ and $|\mathcal{V}'| = n \geq 2$: for any $\mathbf{v}', c'$, and e', if $\mathcal{V}'_{\text{ex}}\backslash\{C\} = \mathbf{v}'$ and $C = c' \mathrel{\Box\!\!\rightarrow} E = e'$,

$$\mathcal{M}'(\mathcal{V}'_{\text{ex}}\backslash\{C\} \leftarrow \mathbf{v}', C \leftarrow c') \vDash E = e'.$$

Let now $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $C, E \in \mathcal{V}$ and $|\mathcal{V}| = n + 1$ be adequate and acyclic. Moreover, suppose $\mathcal{V}_{\text{ex}}\backslash\{C\} = \mathbf{v}$, and $C = c \mathrel{\Box\!\!\rightarrow} E = e$. Since $n + 1 \geq 3$, there is an $X \in \mathcal{V}\backslash\{C, E\}$. Either $X \in \mathcal{V}_{\text{en}}$ or $X \in \mathcal{V}_{\text{ex}}$.

Suppose $X \in \mathcal{V}_{\text{en}}$. By **Lemma A2.1**, $\mathcal{M}_{\rangle X \langle}$ is acyclic and adequate, and

$$(\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e) \leftrightarrow (\mathcal{M}_{\rangle X \langle}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e).$$

But $\mathcal{M}_{\rangle X \langle}$ has variable set $\mathcal{V}_{\rangle X \langle}$ with $|\mathcal{V}_{\rangle X \langle}| = n$. Thus, by the inductive hypothesis,

$$\mathcal{M}_{\rangle X \langle}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e,$$

and therefore

$$\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e.$$

Suppose instead $X \in \mathcal{V}_{\text{ex}}$. Then

$$\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) = (\mathcal{M}(X \leftarrow \mathbf{v}))\,(\mathcal{V}_{\text{ex}} \backslash \{C, X\} \leftarrow \mathbf{v}, C \leftarrow c).$$

But $\mathcal{M}(X \leftarrow \mathbf{v})$ has variable set $\mathcal{V} \backslash \{X\}$ with $|\mathcal{V} \backslash \{X\}| = n$ and exogenous variable set $\mathcal{V}_{\text{ex}} \backslash \{X\}$. Moreover, by **Lemma A2.2**, $\mathcal{M}(X \leftarrow \mathbf{v})$ is acyclic and adequate. Thus it follows by the inductive hypothesis that

$$(\mathcal{M}(X \leftarrow \mathbf{v}))\,(\mathcal{V}_{\text{ex}} \backslash \{C, X\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e,$$

and therefore

$$\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e. \ \blacksquare$$

When $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c) \vDash E = e$ and $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{C\} \leftarrow \mathbf{v}, C \leftarrow c') \vDash E = e'$ for some $c' \neq c$ and $e' \neq e$, say that *E depends on C in $\mathcal{M}$ relative to $\mathcal{V}_{ex}$*$= \mathbf{v}$; if additionally $\mathcal{V}_{\text{ex}} = \mathbf{v}$, say that *$E$ depends on C in $\mathcal{M}$* (simpliciter).

H.3 SEM Accounts Entail Sufficiency (in the Absence of Cycles)

Recall:

Acyclic Sufficiency: Necessarily, if (c, e) is a suitable pair of actual events, X and Z are variables representing alterations of c and e, respectively, and there is an adequate, acyclic SEM including X and Z: then, if e wouldn't have occurred if c hadn't occurred, c causes e.

We can now prove the following theorem.

Theorem A3. **Counterfactual Adequacy** and **Sufficiency in Adequate Models** jointly entail **Acyclic Sufficiency**.

Proof of Theorem A3: Let (c, e) be a suitable pair of actual events such that $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$. Let X and Z be variables representing alterations of c and e, respectively, such that there is an adequate, acyclic SEM $\mathcal{M} = (\mathcal{V}, \mathcal{E})$ with $X, Z \in \mathcal{V}$. To show that **Counterfactual Adequacy** and **Sufficiency in Adequate Models** jointly entail **Acyclic Sufficiency**, we now only need to show that they imply that c causes e.

Since $\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$, by CEM there is a unique closest possible $\neg O(c)$-world in which e doesn't occur. Where values $X = x$ and $Z = z$ represent c's and e's actual occurrence, respectively, let value $x' \neq x$ represent the non-occurrence of c, and $z' \neq z$ the respective non-occurrence of e. We then have $X = x' \mathrel{\Box\!\!\rightarrow} Z = z'$. By And-to-If, we moreover have $X = x \mathrel{\Box\!\!\rightarrow} Z = z$.

Let $\mathbf{v}$ be such that $\mathcal{V}_{\text{ex}} = \mathbf{v}$. By **Theorem A2** (which assumes **Counterfactual Adequacy**), we then have that $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{X\} \leftarrow \mathbf{v}, X \leftarrow x) \vDash Z = z$ and $\mathcal{M}(\mathcal{V}_{\text{ex}} \backslash \{X\} \leftarrow \mathbf{v}, X \leftarrow x') \vDash Z = z'$. Thus Z depends on X in $\mathcal{M}$. Moreover, by **Theorem A1** (again assuming **Counterfactual Adequacy**), there is a directed path from X to Z in $\mathcal{M}$. So, by **Sufficiency in Adequate Models**, c causes e. ■

Bibliography

Al-Khalili, Jim. 1999. *Black Holes, Wormholes and Time Machines*. 1st edition edn. Bristol, UK ; Philadelphia, PA: Taylor & Francis.

Albert, David Z. 2000. *Time and Chance*. Harvard University Press.

Albert, David Z. 2015. *After Physics*. Cambridge, Massachusetts London, England: Harvard University Press.

Arntzenius, Frank, & Hall, Ned. 2003. On What We Know About Chance. *The British Journal for the Philosophy of Science*, **54**(2), 171–179.

Bacon, Andrew. 2020. Logical Combinatorialism. *Philosophical Review*, **129**(4), 537–589.

Banach, Stefan, & Kuratowski, Casimir. 1929. Sur une généralisation du problème de la mesure. *Fundamenta Mathematicae*, **14**(1), 127–131.

Beckers, Sander, & Vennekens, Joost. 2017. The Transitivity and Asymmetry of Actual Causation. *Ergo, an Open Access Journal of Philosophy*, **4**.

Beebee, Helen. 2004. Causing and Nothingness. *Pages 291–308 of:* Paul, L. A., Hall, E. J., & Collins, J. (eds), *Causation and Counterfactuals*. MIT Press.

Bennett, Jonathan. 1984. Counterfactuals and Temporal Direction. *The Philosophical Review*, **93**(1), 57–91.

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. Oxford, New York: Oxford University Press.

Berkovitz, Joseph. 2001. On Chance in Causal Loops. *Mind*, **110**(437), 1–23.

Bernstein, Sara. 2014. Omissions as Possibilities. *Philosophical Studies*, **167**(1), 1–23.

Borel, Émile. 1924. A Propos d'un Traité de Probabilités. *Revue Philosophique de la France et de l'Étranger*, **98**, 321–336.

Briggs, Rachael. 2010. The Metaphysics of Chance. *Philosophy Compass*, **5**(11), 938–952.

Carter, Brandon. 1968. Global Structure of the Kerr Family of Gravitational Fields. *Physical Review*, **174**(5), 1559–1571.

Chalmers, David J. 2002. Does Conceivability Entail Possibility. *Pages 145–200 of:* Gendler, Tamar Szabo, & Hawthorne, John (eds), *Conceivability and Possibility*. Oxford University Press.

Cusbert, John. 2018. Backwards Causation and the Chancy Past. *Mind*, **127**(505), 1–33.

Cusbert, John. 2022. How Stable Is Objective Chance? *The British Journal for the Philosophy of Science*, **73**(3).

Deutsch, David. 1991. Quantum Mechanics near Closed Timelike Lines. *Physical Review D*, **44**(10), 3197–3217.

Dorr, Cian. 2016. Against Counterfactual Miracles. *The Philosophical Review*, **125**(2), 241–286.

Dretske, Fred I. 1981. *Knowledge and the Flow of Information*. Stanford, CA: MIT Press.

Earman, John. 1995. Recent Work on Time Travel. *Pages 268–310 of:* Savitt, Steven F. (ed), *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time*. Cambridge: Cambridge University Press.

Easwaran, Kenny. 2008. *The Foundations of Conditional Probability*. Ph.D. thesis.

Echeverria, F., Klinkhammer, G., & Thorne, K. S. 1991a. Billiard Balls in Wormhole Spacetimes with Closed Timelike Curves: Classical Theory. *Physical Review, D (Particles Fields); (United States)*, **44:4**.

Echeverria, Fernando, Klinkhammer, Gunnar, & Thorne, Kip S. 1991b. Billiard Balls in Wormhole Spacetimes with Closed Timelike Curves: Classical Theory. *Physical Review D*, **44**(4), 1077–1099.

Effingham, Nikk. 2020. *Time Travel: Probability and Impossibility*. New York: Oxford University Press.

Elga, Adam. 2000. Statistical Mechanics and the Asymmetry of Counterfactual Dependence. *Philosophy of Science*, **68**(3), 313–324.

Gallow, J. Dmitri. 2016. A Theory of Structural Determination. *Philosophical Studies*, **173**(1), 159–186.

Ghirardi, GianCarlo, Rimini, Alberto, & Weber, Tullio. 1986. Unified Dynamics for Microscopic and Macroscopic Systems. *Physical Review D*, **34**(D), 470–491.

Glynn, Luke. 2013. Of Miracles and Interventions. *Erkenntnis*, **78**(1), 43–64.

Gödel, Kurt. 1949. An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation. *Reviews of Modern Physics*, **21**(3), 447–450.

Grice, H. P. 1961. The Causal Theory of Perception. *Aristotelian Society Supplementary Volume*, **35**(1), 121–168.

Griffiths, David J. 1981. *Introduction to Electrodynamics*.

Hájek, Alan. 2003. What Conditional Probability Could Not Be. *Synthese*, **137**(3), 273–323.

Hall, Ned. 2004a. Two Concepts of Causation. *Pages 225–276 of:* Collins, John, Hall, Ned, & Paul, Laurie (eds), *Causation and Counterfactuals*. MIT Press.

Hall, Ned. 2004b. Two Mistakes About Credence and Chance. *Australasian Journal of Philosophy*, **82**(1), 93–111.

Hall, Ned. 2007. Structural Equations and Causation. *Philosophical Studies*.

Halpern, Joseph. 2016. *Actual Causality*. MIT Press.

Halpern, Joseph, & Pearl, Judea. 2005. Causes and Explanations: A Structural-Model Approach. Part I: Causes. *The British Journal for the Philosophy of Science*, **56**(4), 843–887.

Handfield, Toby, & Wilson, Alastair. 2014. Chance and Context. *In:* Wilson, Alastair (ed), *Chance and Temporal Asymmetry*. Oxford University Press.

Hawthorne, James. 2016. Logic of Comparative Support: Qualitative Conditional Probability Relations Representable by Popper Functions. *In:* Hájek, Alan, & Hitchcock, Christopher (eds), *The Oxford Handbook of Probability and Philosophy*. Oxford University Press.

Hawthorne, John. 2007. Craziness and Metasemantics. *The Philosophical Review*, **116**(3), 427–440.

Hedden, Brian. 2023. Counterfactual Decision Theory. *Mind*, **132**(527), 730–761.

Hitchcock, Christopher. 2001. The Intransitivity of Causation Revealed in Equations and Graphs. *The Journal of Philosophy*, **98**(6), 273–299.

Hitchcock, Christopher. 2007. Prevention, Preemption, and the Principle of Sufficient Reason. *The Philosophical Review*, **116**(4), 495–532.

Hitchcock, Christopher, & Rédei, Miklós. 2021. Reichenbach's Common Cause Principle. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, summer 2021 edn. Metaphysics Research Lab, Stanford University.

Horacek, David. 2005. Time Travel in Indeterministic Worlds. *The Monist*, **88**(3), 423–436.

Jäger, Jens. forthcoming. Loops and the Geometry of Chance. *Noûs*.

Joyce, James M. 2010. A Defense of Imprecise Credences in Inference and Decision Making. *Philosophical Perspectives*, **24**(1), 281–323.

Kajari, E., Walser, R., Schleich, W. P., & Delgado, A. 2004. Sagnac Effect of Gödel's Universe. *General Relativity and Gravitation*, **36**(10), 2289–2316.

Kolmogorov, A. N. 1950. *Foundations of the Theory of Probability*. New York: Chelsea Pub. Co.

Koopman, B. O. 1940. The Axioms and Algebra of Intuitive Probability. *Annals of Mathematics*, **41**(2), 269–292.

Kripke, Saul. 1980. *Naming and Necessity*. Harvard University Press.

Kyburg, Henry Ely. 1983. *Epistemology and Inference*. Univ of Minnesota Press.

Landau, E.M., & Lifschitz, L.D. 1994. *The Classical Theory of Fields*.

Levinstein, Benjamin A. 2023. Accuracy, Deference, and Chance. *Philosophical Review*, **132**(1), 43–87.

Lewis, David. 1973a. Causation. *Journal of Philosophy*, **70**(17), 556–567.

Lewis, David. 1973b. *Counterfactuals*. Malden, Mass.: Blackwell.

Lewis, David. 1976. The Paradoxes of Time Travel. *American Philosophical Quarterly*, **13**(2), 145–152.

Lewis, David. 1979. Counterfactual Dependence and Time's Arrow. *Noûs*, **13**(4), 455–476.

Lewis, David. 1980. Veridical Hallucination and Prosthetic Vision. *Australasian Journal of Philosophy*, **58**(3), 239–249.

Lewis, David. 1986a. Events. *Pages 241–269 of:* Lewis, David (ed), *Philosophical Papers Vol. II*. Oxford University Press.

Lewis, David. 1986b. Postscripts to 'Causation'. *In:* Lewis, David (ed), *Philosophical Papers Vol. Ii*. Oxford University Press.

Lewis, David. 1987. A Subjectivist's Guide to Objective Chance. *In: Philosophical Papers, Volume II*. Oxford University Press.

Loewer, Barry. 2007. Counterfactuals and the Second Law. *In:* Price, Huw, & Corry, Richard (eds), *Causation, Physics and the Constitution of Reality: Russell?S Republic Revisited*. Oxford University Press.

Loewer, Barry. 2019. The Mentaculus Vision. *In: Statistical Mechanics and Scientific Explanation*. World Scientific Publishing Co.

Luzon, Bar. 2024. *Explaining Content*. Ph.D. thesis, New York University.

Maudlin, Tim. 2019. *Philosophy of Physics: Quantum Theory*. Princeton: Princeton University Press.

Meacham, Christopher J. G. 2005. Three Proposals Regarding a Theory of Chance. *Philosophical Perspectives*, **19**(1), 281–307.

Mellor, D. H. 1995. *The Facts of Causation*. Routledge.

Menzies, Peter. 2004. Causal Models, Token Causation, and Processes. *Philosophy of Science*, **71**(5), 820–832.

Nelson, Kevin. 2009. On Background: Using Two-Argument Chance. *Synthese*, **166**(1), 165–186.

Nolan, Daniel. 1997. Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, **38**(4), 535–572.

Norton, John D., Pooley, Oliver, & Read, James. 2023. The Hole Argument. *In:* Zalta, Edward N., & Nodelman, Uri (eds), *The Stanford Encyclopedia of Philosophy*, summer 2023 edn. Metaphysics Research Lab, Stanford University.

Novikov, I. D. 1992. Time Machine and Self-Consistent Evolution in Problems with Self-Interaction. *Physical Review D*, **45**(6), 1989–1994.

Pearl, Judea. 1985. *Graphoids: A Graph-Based Logic For Reasoning About Relevance Relations*.

Popper, Karl. 1968. *The Logic of Scientific Discovery*. Routledge.

Popper, Karl R. 1959. The Propensity Interpretation of Probability. *The British Journal for the Philosophy of Science*, **10**(37), 25–42.

Priest, Graham. 2008. *An Introduction to Non-Classical Logic, Second Edition: From If to Is*. 2 edition edn. Cambridge ; New York: Cambridge University Press.

Reichenbach, Hans. 1956. *The Direction of Time*. Mineola, N.Y.: Dover Publications.

Rényi, Alfréd. 1955. On a New Axiomatic Theory of Probability. *Acta Mathematica Academiae Scientiarum Hungarica*, **6**(3), 285–335.

Russell, Bertrand. 1903. *Principles of Mathematics*. New York: Routledge.

Schaffer, Jonathan. 2003. Principled Chances. *British Journal for the Philosophy of Science*, **54**(1), 27–41.

Schaffer, Jonathan. 2005. Contrastive Causation. *Philosophical Review*, **114**(3), 327–358.

Schaffer, Jonathan. 2007. Deterministic Chance? *British Journal for the Philosophy of Science*, **58**(2), 113–140.

Sider, Theodore. 2011. *Writing the Book of the World*. Oxford University Press.

Smith, Nicholas J. J. 1997. Bananas Enough for Time Travel? *The British Journal for the Philosophy of Science*, **48**(3), 363–389.

Stalnaker, Robert. 1968. A Theory of Conditionals. *Pages 98–112 of:* Rescher, Nicholas (ed), *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*. Blackwell.

Tumulka, Roderich. 2006. A Relativistic Version of the Ghirardi-Rimini-Weber Model. *Journal of Statistical Physics*, **125**, 821–840.

Ulam, Stanisław. 1930. Zur Masstheorie in der allgemeinen Mengenlehre. *Fundamenta Mathematicae*, **16**(1), 140–150.

van Fraassen, Bas C. 1984. Belief and the Will. *Journal of Philosophy*, **81**(5), 235–256.

van Stockum, W. J. 1938. IX.—The Gravitational Field of a Distribution of Particles Rotating about an Axis of Symmetry. *Proceedings of the Royal Society of Edinburgh*, **57**, 135–154.

Vihvelin, Kadri. 1995. Causes, Effects and Counterfactual Dependence. *Australasian Journal of Philosophy*, **73**(4), 560–573.

Vitali, Giuseppe. 1905. Sul Problema Della Misura Dei Gruppi Di Punti Di Una Retta.

Voigt, Jürgen. 2020. *A Course on Topological Vector Spaces*. Compact Textbooks in Mathematics. Cham: Springer International Publishing.

Walley, Peter. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall.

Weslake, Brad. 2015. *A Partial Theory of Actual Causation*.

Weslake, Brad. 2017. Difference-Making, Closure and Exclusion. *Pages 215–231 of:* Beebee, Helen, Hitchcock, Christopher, & Price, Huw (eds), *Making a Difference: Essays on the Philosophy of Causation*. Oxford University Press.

White, Roger. 2009. Evidential Symmetry and Mushy Credence. *Oxford Studies in Epistemology*, **3**, 161–186.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford Studies in Philosophy of Science. Oxford, New York: Oxford University Press.