Causation Beyond Manipulation

Brett Park

BNP30@pitt.edu

Department of History and Philosophy of Science, University of Pittsburgh

June 5, 2023

Abstract

In this essay, I examine the mathematical underpinnings of the butterfly effect to interrogate its putatively causal status. Chaotic systems are mixing, meaning that bundles of initial conditions eventually spread out over phase space. This has two consequences. First, counterfactual dependence becomes ubiquitous between temporally distant states; slight changes anywhere in the system can lead to large changes anywhere else. Second, all events become probabilistically independent of one another. When we map these properties onto the butterfly effect, we notice a situation of counterfactual dependence and probabilistic independence. In this case, our two normal criteria for causation — the counterfactual and the probabilistic — contradict each other in an unexpected way. Rather than ruling in favor of one of these criteria, I argue that we should view the butterfly effect's causal status as indeterminate.

Acknowledgements: I would like to thank John Norton, Valia Allori, David Wallace, James Woodward, Jordan Olson, Gal Ben Porath, J.P. Gamboa, Jennifer Whyte, Tom Wysocki, and attendees of the LMP graduate conference for helpful feedback and discussions on various drafts of this work. The simulations found in the figures were made with Mathematica.

1 Introduction

In the philosophical literature, there are two traditions for describing causal dependence: probabilistic dependence and counterfactual dependence.[1] However, these two criteria are known to disagree. In cases of "mere associations" we have probabilistic dependence and counterfactual independence (e.g. a rooster crowing before the sunrise). We typically defer to the counterfactual criteria when ruling these cases as non-causal. But what about disagreement in the opposite direction: cases of counterfactual dependence and probabilistic independence? In this paper, I argue that chaos theory supplies such a case: the butterfly effect. The mathematical basis of chaos — mixing — implies that, over the long run, chaotic systems will have ubiquitous counterfactual dependence that cannot be tracked probabilistically.

This leaves the butterfly effect's causal status ambiguous. According to the popular manipulationist theories of causation, the pragmatic function of causation is to allow us to manipulate our environment. For this we want both counterfactual dependence and probabilistic dependence. However, the butterfly effect, particularly when compared to obvious cases of non-causation, calls into question whether probabilistic dependence is a part of the conceptual content of 'cause' or just a helpful association. Rather than rule decisively one way or the other, I argue that the butterfly effect's causal status is best thought of as indeterminate.

Because this is a paper about chaos theory and the butterfly effect, it is important to describe these ideas on their own terms. Thus, I will be operating within

[1]For examples of the former, see Mellor (1995) and Suppes (1970). For the latter, see Lewis (1974).

the classical framework of dynamical systems theory. The advantage of this approach is that dynamical systems theory offers an objective way to interpret two highly contentious topics in philosophy: probability and counterfactuals. However, not every probabilistic and counterfactual question will be answerable within the framework. We will proceed with these limitations in mind.

The paper is structured as follows. In §2, I describe the mixing property of chaotic systems and how that leads to probabilistic independence. §3 describes the butterfly effect. §3.1 describes how the butterfly effect exemplifies both probabilistic independence and counterfactual dependence. Then, in §3.2 I use a simple chaotic system — the baker's map — to illustrate. In §3.3, I describe how the butterfly effect differs from how we typically think about probabilistic independence. In §4, I argue that the extension of 'causation' is indeterminate in chaotic domains. I describe the purpose of causation in §4.1, and how this purpose leaves the concept indeterminate in §4.2. I answer various objections in §5.

2 Mixing and Probability in Dynamical Systems Theory

Chaos theory is part of dynamical systems theory. Dynamical systems are future-deterministic, meaning that every state of the system has one unique future under time evolution. For our purposes, we will focus on a subset of dynamical systems — *measure-preserving dynamical systems* — defined as a quadruple (Γ, μ, Σ, T). Γ is the *phase space* of the system, the set of all possible states the system can take. The state of a dynamical system is represented as a point x in its phase space. μ is a measure on Γ where $\mu(\Gamma) = 1$. Σ is a σ-algebra on Γ, defining the measurable sets. Measurable sets $A \in \Sigma$ are also called "events" and their negations are set compliments

$\neg A = \{x \in \Gamma : x \notin A\}$. $T_t : \Gamma \to \Gamma$ is a surjective map given by the system's dynamics: $T_t(x)$ is phase point $x \in \Gamma$ evolved forward by $t \in \mathbb{R}$ or $\mathbb{Z}$, depending on the system. For all $A \in \Sigma$, $T_t(A) = \{T_t(x) : x \in A\}$. A dynamical system is measure preserving iff for all measurable subsets $A \in \Sigma$, $\mu(T_t^{-1}(A)) = \mu(A)$, where $T_t^{-1}(A)$ are all the points that get mapped onto A.

Mixing is a property of chaotic systems.[2] J.W. Gibbs first introduces the concept of mixing using the analogy of a drop of ink in a glass of water (1902, 144-145). Eventually, the drop spreads out to uniformly fill the the glass. Similarly, for mixing systems, a bundle of solutions belonging to a small region of initial conditions will spread out to fill the phase space under time evolution. Formally, we would say that a system is *mixing* iff for any two subsets $A, B \in \Sigma$,

$$\lim_{t\to\infty} \mu\left(T_t\left(A\right) \cap B\right) = \mu(A)\mu(B). \tag{1}$$

This says that the measure of A that ends up in B is the product of the measures of A and B. In other words, every region A of phase space eventually evolves towards the same spread out distribution over every other region B. If the system's dynamics are mixing, they will effectively "stir" together various regions of phase space. Figure 1 is an example of mixing in the two-dimensional logistic map, which is a map from the unit

[2]Charlotte Werndl (2009) has argued that mixing is necessary and sufficient for chaos, while Belot and Earman (1997) describe it as merely necessary.

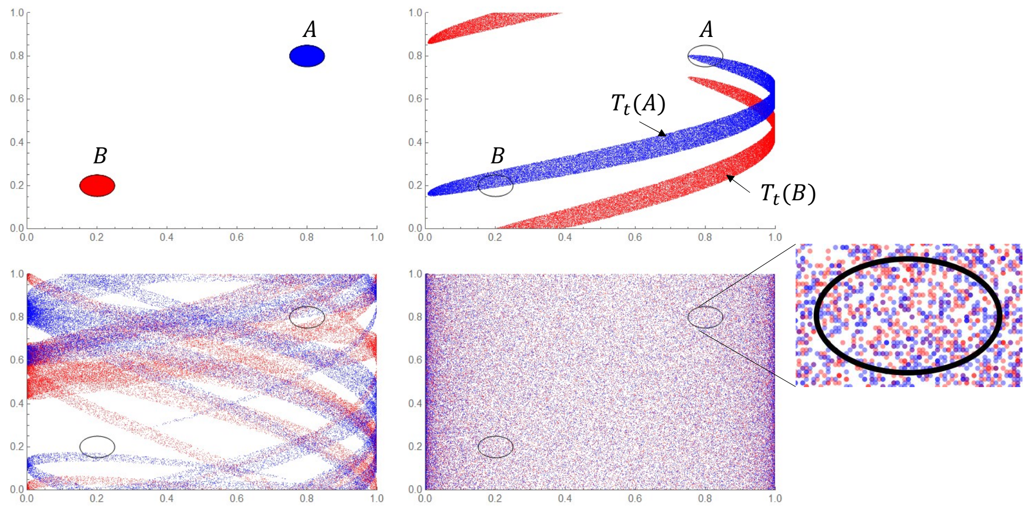

Figure 1: Simulation of mixing in the two-dimensional logistic map at $t = 0, 4, 7, 25$. Two regions of phase space, A and B, spread out under time-evolution, until they both approximate the time-invariant measure of the system.

square onto itself.[3] We can see how the spreading out of solutions displayed there implies the most recognizable property of chaos: sensitive dependence on initial conditions.

From Figure 1, it is easy to infer how mixing leads to probabilistic independence. However, let us walk through the formal result. In dynamical systems theory, it is typical to use the time-invariant, normalized measure as an objective probability

[3]2D logistic map:

$$(x_{n+1}, y_{n+1}) = \begin{cases} (4x_n(1-x_n), x_n + y_n) & \text{if } \; x_n + y_n < 1 \\ (4x_n(1-x_n), x_n + y_n - 1) & \text{if } \; x_n + y_n \geq 1 \end{cases}$$

distribution for the system. The prior probability of event $A \in \Sigma$ is given by the measure of that event,

$$\mu(A) = P(A) \text{ for all } A \in \Sigma. \tag{2}$$

This can be justified by the fact that mixing systems are ergodic, and so time averages equal space averages for the system. For any two events $A, B \in \Sigma$ occurring at time 0 and time t,

$$P(B_t \,\&\, A_0) = \mu(T_t(A) \cap B). \tag{3}$$

This just means that the probability of B occurring after A is measured by the "amount" of A that ends up in B. From (2) and (3), the definition of mixing (1) implies probabilistic independence:

$$\lim_{t\to\infty} P(B_t \,\&\, A_0) = P(A)P(B), \tag{4}$$

or using the standard ratio formula for conditional probability, $P(B|A) = P(A \,\&\, B)/P(A)$,

$$\lim_{t\to\infty} P(B_t|A_0) = P(B). \tag{5}$$

which says that every event becomes probabilistically independent of every other event as $t \to \infty$. By the definition of the limit, (5) also implies that for any $\epsilon > 0$ and for all $A, B \in \Sigma$, there exists a time t for which $|P(B_t|A_0) - P(B)| < \epsilon$. Charlotte Werndl has called this property "approximate probabilistic irrelevance" because conditionalizing on A will not *appreciably* raise or lower B's probability (2009, 214). Since ϵ grows arbitrarily small, I will not make much of the distinction between approximate probabilistic irrelevance and probabilistic independence in this context.

3 The Butterfly Effect

3.1 An Unstable Counterfactual

Since its introduction by Edward Lorenz in 1972, the butterfly effect has been both the most popular and misunderstood idea to come out of chaos theory.[4] Whenever the butterfly effect is responsibly depicted, it usually comes in the form of a counterfactual claim that closely resembles how philosopher's talk about "actual causation" — i.e. the causal relations between actually occurring events. For instance, Robert Bishop describes the butterfly effect as "the flapping of a butterfly's wings in Argentina could cause a tornado in Texas three weeks later. By contrast, in an identical copy of the world sans the Argentinian butterfly, no such storm would have arisen in Texas" (2008). Although it is frequently used as a metaphor for sensitive dependence on initial conditions, little attention has been paid to what exactly this counterfactual amounts to. In this section, we get to the bottom of this counterfactual claim.

Because the butterfly effect is a counterfactual claim coming out of dynamical systems theory, we will analyze it using dynamical systems theory. Dynamical systems theory provides a framework to carry out certain types of counterfactual reasoning. Because dynamical systems are future deterministic, any point x_0 in phase space will be mapped to another point $T_t(x_0) = x_t$ under time evolution. Thus, if we ask the question — "Would B happen at t if our initial conditions had been x_0?" — then there will be a determinate answer based on whether $x_t \in B$.

This point-to-point mapping is the typical dynamical view of the butterfly

[4]See Kellert (2009, 103-120) for discussion.

counterfactual. Jean Bricmont elaborates:

> The idea is that the flap of a butterfly's wings corresponds to a very small change in the initial conditions of the weather system that may lead to a big difference in the future: a tornado or no tornado in Texas (2022, 83).

Let us make this more precise. The weather system has a phase space Γ and dynamics $T_t : \Gamma \to \Gamma$, assumed to be mixing/chaotic. Say the weather system's "actual" initial state is $x_0 \in \Gamma$. Accordingly, its state at t is $T_t(x_0) = x_t$. Now change x_0 slightly to x'_0 as might correspond to the difference of a butterfly not flapping its wings, and find the future state at t by $T_t(x'_0) = x'_t$. Because the system is chaotic, we know that given enough time, x_t and x'_t will end up far apart, disagreeing on large macroscopic properties. Call 'Tornado' the region of phase space corresponding to the macro-event of the Texas tornado. Let us say we find that $x_t \in$ Tornado and $x'_t \notin$ Tornado. Thus, we have counterfactual dependence; changing x_0 to x'_0 changes the tornado's occurrence.

There are a few things to point out here. First, the future states x_t and x'_t will disagree on many more macroscopic events than just the tornado. Given enough time, the disagreement will be widespread over the entire global weather system. Thus, innumerable events in the future will be counterfactually dependent on the initial change. Second, the tornado's occurrence will be counterfactually dependent on many more changes to x_0. Lorenz points this out in his original lecture on the butterfly effect, stating that the tornado will be sensitive to "the flaps of the wings of millions of other butterflies, not to mention the activities of innumerable more powerful creatures, including our species" (1972, 1). What he means by this is that if you were to slightly change the initial conditions anywhere, you would also see large differences in future behavior. Thus, there is no special counterfactual connection between *that* Brazilian butterfly and *that* Texas tornado. Lastly, this counterfactual is highly unstable. If we were to change the initial state a little differently, a little more or less in some direction, we would find a trajectory that disagrees with the first two on many future macro-events.

Now let us consider the situation probabilistically. Call K the background conditions at the time of the flap, 'Flap' the event of the butterfly's flap at time 0, and 'Tornado' the event of the tornado's occurence at time t. Assuming large enough t so that mixing has dominated the dynamics, then the tornado will be probabilistically independent of the flap:

$$P(Tornado|Flap \,\&\, K) = P(Tornado|\neg Flap \,\&\, K) = P(Tornado) \tag{6}$$

Even holding the background conditions fixed, conditionalizing on 'Flap' does not raise or lower the objective probability of the tornadoes occurrence.[5] Although there are many changes to the butterfly's flap that might change the tornado's occurrence, those changes are not reflected as a correlation between 'Flap' and 'Tornado'. To see how this could be, let us turn to a simple example.

[5]Caution: if we fix the background conditions in K at singular values — akin to how David Lewis' similarity metric maximizes the perfect match of particular facts between worlds (1979, 472) — then the initial region will be measure zero $\mu(Flap \,\&\, K) = 0$. Therefore, the mixing condition will technically not imply probabilistic independence. However, we strongly suspect that probabilistic independence will still occur because the coupling of the relevant differential equations will smear out any initial differences over all of phase space.

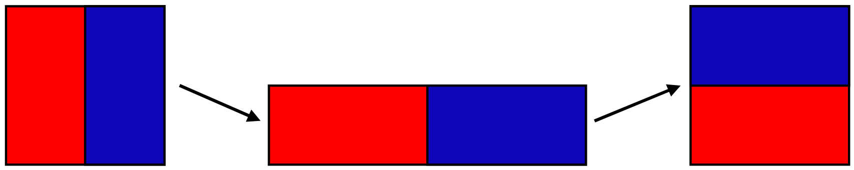

Figure 2: Single Iteration of the Baker's Map

3.2 Example: The Baker's Map

To visualize the situation, let us use a surrogate system: the time-discrete Baker's map. Because the chaos of time-continuous systems can typically be described by time-discrete Poincaré maps, the switch to time-discrete systems has little effect on the generality of the discussion (see Smith, 1998, 92-93). The Baker's map is a chaotic 2D map from the unit square onto itself whose invariant measure is the 2D Lebesgue measure. The map is defined as:

$$(x_{n+1}, y_{n+1}) = \begin{cases} (2x_n, y_n/2) & \text{if } 0 \leq x < \frac{1}{2} \\ (2x_n - 1, (y+1)/2) & \text{if } \frac{1}{2} \leq x \leq 1 \end{cases} \tag{7}$$

It is named the Baker's map because it mimics the process of kneading dough. It can be visualized as two separate operations (Figure 2). First the unit square is flattened out into a $2 \times 1/2$ rectangle. Then the rectangle is cut into left and right halves, and the right half is placed on top of the left half.

Say that we want to know whether our initial conditions end up on the top half of the square. The situation is shown in Figure 3. The blue dots represent initial conditions that end up in the top half after n iterates, and red dots are initial conditions that end up in the bottom half. As n increases, what we see is a procession

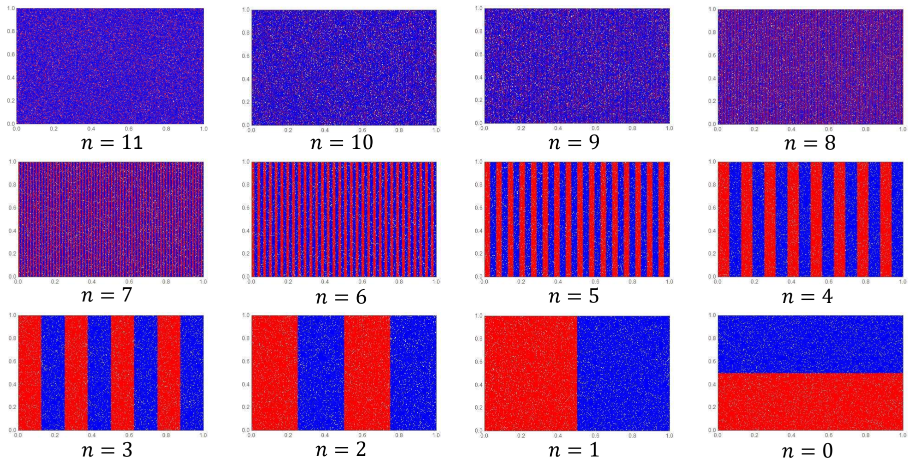

Figure 3: Blue dots are initial conditions that end up on the top half of the unit square after n iterations of the baker's map

of vertical blue stripes that get exponentially denser. By $n = 11$, the stripes have become too tightly packed to observe.

Compare the situation at $n = 2$ and $n = 11$ in Figure 4. In both cases, there are many changes to the initial conditions x_0 and y_0 that would result in the system evolving into the top half region. For $n = 2$, these changes can be imprecise. Judging from the amount of blue in each region, we can see that $P(\text{Top Half } |A) = 0$ and $P(\text{Top Half } |A') = 1$. By ensuring that the initial conditions start out in A' and not A, we guarantee that the system ends up in the top half. For $n = 11$, our final state is *hypersensitive* to practically all changes we could make in the initial conditions. Above a certain size, all regions B and B' will yield roughly the same $1/2$ conditional probability that the system ends up in the top half. From a practical perspective, every change we might make is as good as a coin flip for whether the system ends up in the top half. This is the primary way we should understand the conjunction of

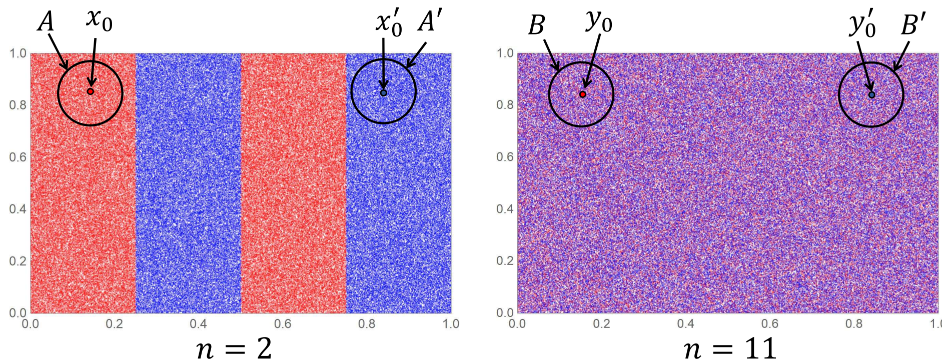

Figure 4: Magnified versions of $n = 2, 11$ above. Blue dots are again initial conditions that end up on the top half after n iterations. $n = 2$ shows counterfactual and probabilistic dependence, while $n = 11$ shows counterfactual dependence and probabilistic independence.

counterfactual dependence and probabilistic independence: *there are many changes to the initial conditions that would alter the occurrence of a future event, but this dependence becomes too fined grained to be reflected as correlations between events.*

So long as the system is mixing, the result will be the same in more complex examples, such as the global weather system in a high-dimensional phase space. Sensitive dependence entails that had the initial conditions been slightly different from their actual values — such as might correspond to a butterfly not flapping — then the resultant trajectories will show widespread disagreement on future events — such as a tornado's occurrence. However, the event of the butterfly's flap will not raise or lower the probability of the tornadoes occurrence. This is because the regions of 'Flap' and 'No Flap' will be filled with roughly the same proportions of initial conditions that end up in the 'Tornado' region. This circumstance is similar to a dice roll; although the relevant physics is modeled deterministically, the fairness of the roll is guaranteed by the fact that the outcome is far more sensitive to the initial conditions than the person rolling the dice.[6] In chaotic systems, this situation is ubiquitous over long timescales: any macro change we might make in the distant past would be effectively re-rolling the dice on the system's future. High probability events will probably still happen, and low probability events probably do not.

Accordingly, there is an additional way we could interpret counterfactual independence, beyond the fact the tornado's occurrence is sensitive to changes in the flap. Say in the actual world a butterfly flaps and a tornado occurs sometime much later. Tornadoes are very low probability events. Therefore, if we were to change only the butterfly's flap, the tornado almost certainly does not occur. However, conditionalizing on the butterfly's flap and background conditions once again does not alter the probability of the tornado. This interpretation is especially strange because the tornado's counterfactual dependence on the flap crucially relies on the tornado being a low probability event. A high probability event, e.g. sunshine in the Sahara desert, will look to counterfactually independent of the flap.

3.3 A Counterfactual Degeneracy

We have just described how the hypersensitivity of the butterfly-tornado counterfactual implies probabilistic independence between the two events. However, this is not the typical route to probabilistic independence. When we think about two events being probabilistically independent from one another, we typically imagine that they are counterfactually insensitive to one another as well.

[6]For an elaboration, see Strevens (2011) and Diaconis et al. (2007).

James Woodward gives the following example. Say Suzy throws a rock at a vase, causing it to shatter. According to Woodward, we expect that "If Suzy's rock strikes the vase in Boston at the moment at which someone sneezes in Chicago, then presumably if that person had not sneezed but the world had remained relevantly similar in other respects, the bottle still would have shattered" (2006, 5). The macro-event of the shattering is probabilistically and counterfactually independent of the sneeze; reasonably sized variations to the sneeze *cannot* change the shattering's occurrence. In this case, probabilistic independence is a good marker for counterfactual independence.

However, for chaotic systems, probabilistic independence is no longer a good marker for counterfactual independence. The events of the butterfly's flap and the tornado are probabilistically independent of one another, but sensitive dependence implies that varying the butterfly's flap *can* vary the tornado's occurrence. Insofar as tornadoes are objectively rare events, a slight variation in the butterfly's flap *probably would have* prevented a tornado that actually occurred from occurring. However, this counterfactual dependence is too fine-grained to be picked up as a correlation between butterflies and tornadoes.

This shows us that probabilistic independence exhibits a surprising counterfactual degeneracy. There are two distinct counterfactual routes to probabilistic independence: counterfactual *insensitivity* and *hypersensitivity*. If we treat probabilistic dependence as a necessary condition for counterfactual dependence, then we conflate these starkly different situations. From a causal perspective, we think the counterfactually insensitive situation (the sneeze-vase case) is non-causal. The vase breaking is predictably unaffected by changes to the sneeze. The butterfly effect represents a much different circumstance; small changes lead to large but unpredictable differences. But does this amount to causation? This will be the topic of the next section.

4 The Indeterminacy of 'Causation'

I have just shown how chaos produces widespread counterfactual dependence and probabilistic independence. How should we interpret these results in light of the larger debate on causation? Thus far, I have withheld judgement on whether the butterfly effect is a genuine causal relation. In this section, I will argue that chaotic systems describe a domain where the term 'causation' is indeterminate.

4.1 Causation and Manipulation

In philosophy of science, manipulability accounts of causation have gained considerable traction.[7] Manipulationists break with the philosophical tradition in how they think about causation. While many traditional philosophical theories treat causation as a metaphysical relation, something out there in the world, manipulationists think of causation as an invaluable pragmatic concept that has evolved to aid human beings navigate their environment.[8] Throughout the rest of this paper, I will assume this view of causation as a pragmatic concept is broadly correct. According to the manipulationist, both probabilistic and counterfactual dependence is required for causation to serve its purpose.

According to classic probabilistic theories of causation, causes always raise (or at

[7]For examples, see Woodward (2003, 2006, 2021b) and Halpern (2016)

[8]For a quintessential example of a metaphysical theory of the type described, see Lewis (1974).

least change) the probabilities of their effects.[9] However, a well known deficiency to these theories is the existence of “mere associations”: a class of relations which are probability raising but not causal. The rooster’s crow regularly precedes the sunrise but does not cause it. Probabilities are important for prediction, but human agents also need to understand what Nancy Cartwright calls “effective strategies”: ways they can manipulate their environment to bring about desired ends (1979, 419).

This is the province of counterfactual reasoning. Even though the rooster’s crow is correlated with the sunrise, I cannot prevent the sunrise by silencing the rooster. The sunrise is counterfactually independent of the crow. Thus, causal reasoning requires an understanding of counterfactual dependencies. This in turn provides us with a way of knowing which interventions we could successfully carry out to our advantage.

Manipulationists take both of these insights to heart. James Woodward describes his manipulability theory of causation as:

> **(M)** Where X and Y are variables, X causes Y iff there are some possible interventions that would change the value of X and if were such intervention to occur, a regular change in the value of Y would occur (2021a, 219).

This means that under certain background conditions, Y is counterfactually dependent on certain changes in X such that the change in Y is also probabilistically dependent on the change in X.[10] How we are to understand “possible interventions” is a point of some dispute, but regardless **(M)** is good at recovering our intuitions about standard

[9]See Mellor (1995, 67) and Suppes (1970).

[10]Manipulationist theories use directed acyclic graphs to represent the causal structure of different situations, with arrows between variables denoting counterfactual

cases of causation and non-causation.

We can see that **(M)** asserts that causes should, under the right circumstances, act like a lever to bring about their effects with some regularity. However, the event of the butterfly flap cannot do this. The butterfly effect is an intervention on X (changing the flap) which can change Y (the tornado), but that change is *irregular*. The only change that could bring about the tornado with regularity would be an impossibly precise positioning of the weather's initial conditions. Despite this, the butterfly effect is described in causal terms in both the physics and philosophy literature.[11] If **(M)** truly captured the conceptual content of 'causation', then the butterfly effect would clearly be non-causal, similar to the rooster-sunrise example. Yet it is not.

On the other hand, if one agrees with the manipulationist view of causation as a pragmatic concept, then we cannot assume that causation can be extended to all of the foreign contexts offered by physics. For the manipulationist, to count the butterfly effect as a counterexample would be to misunderstand the nature of their project. In doing so

---

dependence. There, the fact that probabilistic dependence always accompanies counterfactual dependence is implied by a fundamental assumption of the framework called the "Minimality Condition": if there is an arrow from X to Y then there must be a context in which X makes a probabilistic difference for Y (Hitchcock, 2021).

[11]For additional examples of this, see Smith (1990, 247), Frisch (2014, 212), and Hilborn (2004). Note that this is not a bald appeal to the authority of these philosophers and physicists. For this case, I am inclined to exclude the intuitions of laypeople who have only ever encountered the butterfly effect as a causal claim, and have no understanding of the underlying phase space dynamics from which it originates.

we would be taking the metaphysics of causation too seriously. Something has to give.

4.2 The Open Texture of 'Causation'

We can see a way out of this dilemma if we give up the referential determinacy of 'cause' in the case of hypersensitive counterfactuals. To understand what is going on, I will appeal to Friedrich Waismann's notion of "open texture concepts" (1945, 121). According to Waismann, a concept is open texture insofar as there exists potential contexts under which there is no correct answer as to whether it applies. This is because there are two or more definitions of the term that are coextensive within its ordinary domain of use, but disagree in the novel context. Consider a case given in Dennett (1987, 312). An isolated tribe of humans often encounter an explosive gas in their marsh they call "glug," which happens to be methane. Say we remove a member of this tribe from his marsh and introduce him to a novel explosive gas, acetylene, which he instinctively calls "glug." Has the tribesman erred? The properties the tribesman associated with 'glug' has hitherto always referenced methane, but now he is confronted with a different substance with similar properties. Perhaps after teaching the tribesman some chemistry, 'glug' might come to mean *either* 'methane' or 'explosive gas', but before this point, there is no fact to the matter.

The same thing is occurring when we are considering whether the butterfly flap "causes" the tornado. There are two definitions which fit our typical use of 'cause' equally well. One definition is closely tracked by **(M)**; Causation is counterfactual dependence plus probabilistic dependence under certain background conditions. Counterfactual dependence provides information about what listens to changes in what, and probabilistic dependence helps us predict those changes.

Another definition is that causation is just something like counterfactual dependence. On this reading, probabilistic dependence is not part of the meaning of ‘cause’ but part of its ordinary context. This is because the counterfactual dependencies which are readily identified and controlled are the ones which are predictable. Thus, probabilistic dependence has historically set the boundary for where the concept of causation is useful.

The butterfly effect provides a novel case where these two definitions come apart. It lies outside causation’s ordinary domain of use and exposes a hidden ambiguity. By ascribing open texture to ‘cause’ in this case, we can explain why we might encounter conflicting intuitions about the butterfly effect’s causal status. Just as modern chemistry confronts the tribesman with two possible definitions for ‘glug’ that do not decouple inside of his marsh, chaos theory provides us with two possible definitions for ‘cause’ which do not decouple in ordinary use.

Understanding causation to be open texture in this way is not necessarily at odds with the manipulationist. Recently, Woodward, Naftali Weinberger, and Porter Williams (2023) have argued that although causation is not a metaphysical concept of the type sought by traditional causal theorists, there are certain features of reality they dub “The Worldly Infastructure of Causation” that license and support causal reasoning. One of these features is that the macroscopic, coarse grained behavior of a system is largely independent of its microscopic realizers. In the butterfly effect, if we were to hold the background conditions fixed, then whether the tornado occurs will come down to the microscopic realization of the flap. Thus, there is a breakdown in some of the worldly infrastructure of causation. These authors argue that when such a breakdown occurs “such systems simply will not admit a straightforward causal interpretation, at least on anything like how we presently think about causation" (Weinberger et al., 2023, 34). All of this is consistent with 'causation' being indeterminate for the butterfly effect. Elsewhere, Woodward states that "causal notions are legitimate in any context in which we can explain why they are useful" (2007, 67). If we understand "useful" here in terms of our ability to control our environment, then the butterfly effect would be an illegitimate application of the concept. Even if we could fix the background conditions, we could not use the variable of the butterfly's flap to control the tornado. In applying the concept in this way, we have taken causation beyond its domain of useful application.

5 Objections

In framing the discussion, I have avoided some of the more philosophically and technically fraught issues of probability and dynamical systems theory. To head off potential objections, we shall return to them now.

5.1 Differing Measures

Early in the paper, I made the assumption in (2) that $\mu(A) = P(A)$ for all $A \in \Sigma$. In other words, our probability distribution over all events should mirror the system's invariant measure of those events. However, there are many measures we might want to adopt that differ from the invariant measure. If we relax the assumption in (2), does probabilistic independence still hold?

Mixing implies that probabilistic independence will occur for any generic measure. Call μ_0 any normalized measure that is absolutely continuous with respect to the invariant measure μ (i.e. $\mu(A) = 0 \implies \mu_0(A) = 0$ for all $A \in \Sigma$). Call the time-evolved measure μ_t, where $\mu_t(A) = \mu_0(T_t^{-1}(A))$ for all $A \in \Sigma$. If a dynamical system is mixing, then for all $A \in \Sigma$,

$$\lim_{t \to \infty} \mu_t(A) = \mu(A) \tag{8}$$

In other words, every measure that is absolutely continuous with the invariant measure will relax into the invariant measure over time.[12] From (1) and (8) we get:

$$\lim_{t \to \infty} \mu_t(T_t\,(A) \cap B) = \mu(A)\mu(B) \tag{9}$$

which can be used to once again derive probabilistic independence. Thus, no matter how localized our chosen measure is in phase space, so long as it is absolutely continuous with the invariant measure, then we will recover the same mixing behavior and probabilistic independence for large enough t.

What about non-generic measures that are not absolutely continuous with respect to the invariant measure? One way to create a measure that does not relax to the invariant measure is to assign positive measure to points in phase space. The simplest example is the Dirac measure:

$$\delta(A) = \begin{cases} 1\ ,\ x \in A \\ 0\ ,\ x \notin A \end{cases} \tag{10}$$

for every measurable set $A \in \Sigma$ and some fixed point $x \in \Gamma$. The Dirac measure is not absolutely continuous with the invariant measure (e.g. $\mu(x) = 0 \not\Longrightarrow \delta(x) = 0$), and it

[12]For additional details, see Cornfeld et al. (2012, 24-25)

will not relax into the invariant measure over time. This measure seems appropriate when conditionalizing on phase points $P(B_t|X_0 = x)$ because the invariant measure will yield an undefined value.[13] In this case, determinism will imply that for all events $B \in \Sigma$ and for all phase points $x \in \Gamma$, $P(B_t|X_0 = x) = 1$ if $T_t(x) \in B$, and $P(B_t|X_0 = x) = 0$ if $T_t(x) \notin B$. Thus, from a point-to-point perspective, counterfactual dependence will always be accompanied by probabilistic dependence.

This leaves us in the following situation. Say we want to describe the causal relationship between two events in a chaotic system. Insofar as we stipulate that probabilistic dependence needs to accompany counterfactual dependence, we must either restrict ourselves to events occurring over relatively short timescales before mixing occurs, or we restrict ourselves to talking about dependencies on points in phase space. In the latter case, we once again encounter a breakdown in the worldly infrastructure of causation where the exact microstate of the present is the only determiner of arbitrarily future macrostates. Accordingly, it may be that the only useful sense in which "causation" survives in chaotic systems over long timescales is that they are deterministic, not in the folk-sense that is built around relating macro-events.

5.2 Prior Examples of Ubiquitous Counterfactual Dependence/Probabilistic Independence

How new is the idea of widespread counterfactual dependence and probabilistic independence? David Lewis described causal histories as resembling a tree (1986, 215).

[13] $\mu(x) = P(X = x) = 0$, therefore, the ratio formula for conditional probability $P(B|X = x) = P(B \,\&\, X = x)/P(X = x)$ is undefined.

One effect is counterfactually dependent on many causes, who in turn are dependent on many more causes. As one goes farther back in time, there is a combinatorial explosion of causes for a single effect. In the limit, this means that knowing one of these causes occurred will not appreciably raise the probability of the effect because it is only one of many. Therefore, the idea that we could have ubiquitous counterfactual dependence and probabilistic independence is not so new.

Although Lewis was not thinking in terms of dynamical systems, the key difference between Lewis' discussion and mine is that Lewis is still presupposing a dependence between macro-cause and macro-effect. Presumably, if we could freeze the background conditions around one of the many causes in the distant past, then the macro-event of that cause could act like a lever for reliably bringing about some distant future effect. The butterfly effect means that all regions of phase space spread out under time evolution, leading to roughly the same distribution. This is why butterfly flaps cannot act like levers for tornadoes, not because of unknown background conditions. In other words, Lewis' tree-structure is a hugely complex network of causal relations between macrophysical events. The butterfly effect shows that, over the long run in chaotic systems, there is no causal structure linking macrophysical events. This is because counterfactual dependence has become hypersensitive to the precise initial conditions.

5.3 The Weather May Not Be Chaotic

In my analysis, I have focused on the butterfly effect because it is the canonical metaphor for chaos theory. However, we cannot be certain that the butterfly effect is actually representative of real-world weather dynamics. Lorenz's original model is a first-order approximation of the Navier-Stokes equations, consisting of three coupled ordinary differential equations which represent a single atmospheric convection roll (1963). As Peter Smith points out, the inference from the Lorenz model to butterflies and tornadoes is highly speculative (1998, 66-67). Despite this, numerical simulations of vastly more complex, contemporary meteorological models still exhibit chaos (Palmer et al., 2014).

As we have seen with the Baker's map, none of my conclusions particularly hinge on the weather being chaotic. We can view the weather as a stand-in for any chaotic system, and the butterfly and tornado as stand-ins for temporally separated events in the system. The analysis is still essentially the same. As long as chaos exists in certain systems, then we can import these conclusions there.

5.4 Quantum Mechanics is Not Chaotic

A deeper question is whether chaos is actually a fundamental component of the dynamics of the universe, or something emergent. This is brought to a head when we consider the fact that the basic evolution equation of quantum mechanics, the Schrödinger equation, is linear and thus not a candidate for chaos. For quantum mechanical systems, nearby initial conditions stay nearby indefinitely under Schrödinger evolution.[14] Nonlinearity enters orthodox quantum mechanics via the Born rule, which only applies in measurement contexts. Thus, how chaos emerges out of quantum mechanics is caught in the jaws of a deeper question: the quantum

[14]Formally, the Schrödinger equation is unitary, meaning vectors in Hilbert space preserve their inner product under time-evolution.

measurement problem.

Of the interpretations of quantum mechanics that offer a realist solution to the measurement problem, we have a range of descriptions of quantum chaos. On one end, hidden variables theories such as Bohmian mechanics leave chaos' essential character mostly intact (Dürr et al., 1992). On the other end, multiverse interpretations claim that classical chaos is an emergent phenomena that appears from linear dynamics plus environmental decoherence (Wallace, 2012, 64-102). On this view, it is the spreading out and subsequent branching of the wave function (a single microstate) across different macrostates that makes the future uncertain, not sensitive dependence on the initial conditions of microstates. The neutral conclusion to draw here is that we simply do not know whether chaos is a fundamental feature of the world's dynamics, or merely an emergent one. Thus, we do not know whether the butterfly effect, and the indeterminacy it creates in our concept of causation, is a feature of our world.

6 Conclusion

The butterfly effect is an example where one of our most fundamental concepts — causation — is indeterminate. Causation is intimately tied up with counterfactual dependence, which we expect to be accompanied by probabilistic dependence. However, the butterfly effect is a hypersensitive counterfactual; a counterfactual dependence that is too sensitive to be reflected probabilistically. In chaotic systems, this relation predominates over large timescales because both ubiquitous counterfactual dependence and probabilistic independence are a consequence of one of chaos' central traits: mixing. This unexpected separation is an example where the reference of 'causation' is indeterminate; there are two meanings of 'cause' — counterfactual dependence vs.

predictable counterfactual dependence — which are coextensive inside the concept's normal domain of use.

References

Belot, G. and Earman, J. (1997). Chaos out of order: Quantum mechanics, the correspondence principle and chaos. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics. 28*(2).

Bishop, R. C. (2008). What could be worse than the butterfly effect? *Canadian Journal of Philosophy. 38*(4), 519–547.

Bricmont, J. (2022). *Making Sense of Statistical Mechanics.* Springer Nature.

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

Cornfeld, I. P., Fomin, S. V. and Sinai, Y. G. (2012). *Ergodic Theory.* vol. 245. Springer Science & Business Media.

Dennett, D. C. (1987). *The Intentional Stance.* MIT press.

Diaconis, P., Holmes, S. and Montgomery, R. (2007). Dynamical bias in the coin toss. *SIAM review. 49*(2), 211–235.

Dürr, D., Goldstein, S. and Zanghi, N. (1992). Quantum chaos, classical randomness, and bohmian mechanics. *Journal of Statistical Physics. 68*(1), 259–270.

Frisch, M. (2014). *Causal Reasoning in Physics.* Cambridge University Press.

Gibbs, J. W. (1902). *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics.* C. Scribner's sons.

Halpern, J. Y. (2016). *Actual Causality.* MIT Press.

Hilborn, R. C. (2004). Sea gulls, butterflies, and grasshoppers: A brief history of the butterfly effect in nonlinear dynamics. *American Journal of Physics. 72*(4), 425–427.

Hitchcock, C. (2021). Probabilistic Causation In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. (eds). Metaphysics Research Lab, Stanford University. Spring 2021 edition.

Kellert, S. H. (2009). *Borrowed Knowledge: Chaos Theory and the Challenge of Learning Across Disciplines*. University of Chicago Press.

Lewis, D. (1974). Causation. *The Journal of Philosophy. 70*(17), 556–567.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*. pp. 455–476.

Lewis, D. (1986). Causal explanation In *Philosophical Papers Vol. II*, Lewis, D. (eds). Oxford University Press. pp. 214–240.

Lorenz, E. (1972). Predictability: Does the flap of a butterfly's wing in brazil set off a tornado in texas? American Association for the Advancement of Science, 139th meeting.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of atmospheric sciences. 20*(2), 130–141.

Mellor, D. H. (1995). *The Facts of Causation*. Routledge.

Palmer, T., Döring, A. and Seregin, G. (2014). The real butterfly effect. *Nonlinearity. 27*(9), R123.

Smith, P. (1990). The butterfly effect. Proceedings of the Aristotelian society. pp. 247–267.

Smith, P. (1998). *Explaining chaos*. Cambridge University Press.

Strevens, M. (2011). Probability out of determinism In *Probabilities in Physics*, Beisbart, C. and Hartmann, S. (eds). Oxford University Press. pp. 339–364.

Suppes, P. (1970). *A Probabilistic Theory of Causation*. North-Holland Pub. Co.

Waismann, F. (1945). Verifiability. *Proceedings of the Aristotelian Society, Supplementary Volume 19*. pp. 119–150.

Wallace, D. (2012). *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford University Press.

Weinberger, N., Williams, P. and Woodward, J. (2023). The worldly infrastructure of causation.

Werndl, C. (2009). What are the new implications of chaos for unpredictability? *The British Journal for the Philosophy of Science. 60*(1), 195–220.

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

Woodward, J. (2006). Sensitive and insensitive causation. *The Philosophical Review. 115*(1), 1–50.

Woodward, J. (2007). Causation with a human face In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Price, H. and Corry, R. (eds). Oxford University Press.

Woodward, J. (2021a). Downward causation defended . In *Top-Down Causation and Emergence*. Springer. pp. 217–251.

Woodward, J. (2021b). *Causation with a Human Face: Normative Theory and Descriptive Psychology*. Oxford University Press.