Chaos, causation, and describing dynamics

David Danks & Maralee Harrell

Carnegie Mellon University

Keywords: Chaos; Causal models; Dynamics; Level of description

1. Introduction

A standard platitude about the function of causal knowledge and theories is that they are valuable because they support prediction, explanation, and control. Knowledge of predator-prey relations enables us to predict future animal populations, as well as design policies or interventions that help influence those populations. If we understand the underlying biochemical mechanisms of some disease, then we can predict who is at risk for it, explain why it produces particular symptoms, and develop interventions to try to reduce its prevalence or the symptom severity. Of course, there are many situations in which one has, for practical reasons, only some of these desiderata; for example, control might be infeasible for technical or ethical reasons. But these remain, for many researchers, the ideal for why causal knowledge is a valuable end of scientific inquiry, including biological inquiry. There are, however, certain types of systems—in particular, chaotic systems—in which it appears that these ends are unattainable, and these systems appear to be widespread in the biological domain, broadly construed (e.g., Benincà et al., 2008; Cushing, Costantino, Dennis, Desharnais, & Henson, 2003; Guastello, Koopmans, & Pincus, 2009; Hastings & Powell, 1991; Skarda & Freeman, 1987; Tsuda, 2001). In this paper, we will show why it is natural to think that causal models of chaotic systems cannot satisfy any of the three functions. But we will also show why this natural thought is wrong: we can have usable causal knowledge about even chaotic systems. Moreover, the ways in which we can have such knowledge lead us naturally to rethink a standard understanding of how causal learning and modeling proceed. In particular, just as we often must find the appropriate variables for a causal system, we also must determine the proper level or granularity of description for the dynamics of that system.

Consider a paradigmatic case of learning and applying causal knowledge: a randomized clinical trial (RCT). Suppose that we have a biological theory **T** that implies that some drug *D* should be efficacious in treating symptoms *S*. We provide *D* to a randomly selected subset of our experimental participants, and either a placebo, some alternative treatment, or nothing at all to the other participants. We then measure *S,* as well as a number of other factors included in **T** (e.g., mediating processes), for an extended period of time. Using a range of statistical techniques, we can learn a variety of causal facts about the system (e.g., the efficacy of *D* to treat *S*; whether *D* functions as **T** predicts; etc.) by comparing the two groups. The causal facts represented by **T** (if **T** is correct) enable various predictions about, for example, likely future states of individuals if *D* were to be used or not (e.g., that giving *D* to a randomly chosen patient will make it less likely that the patient exhibits *S* in the future). And of course, RCTs are just one of the many ways in which we can come to have causal knowledge that can then be deployed for a variety of tasks.

A potential pitfall lurks, however, when we confront chaotic systems. We describe the potential issue in more detail in Section 2, but sketch it here. There is disagreement about how exactly to define 'chaos,' but widespread agreement that chaotic systems generally exhibit *sensitive dependence on initial conditions* (SDIC): roughly, trajectories with initial states that are arbitrarily similar rapidly grow to have arbitrarily different final states. But this means that obvious methods for prediction and/or control will not work, as we (seemingly) cannot simply measure the initial state *I*, plug it into our theory **T**, and generate predictions or explanations of a final state *F*. Given that we measure *I* with finite precision and accuracy, the true initial state *I** is likely to be slightly different. So, if the underlying causal structure exhibits SDIC, then predictions using *I* could be arbitrarily different from those using *I**. And to the extent that explanation requires using a theory **T** to show that some state of affairs was likely or inevitable, chaotic phenomena will seem to evade explanation. Similar considerations apply to attempts to control a chaotic causal system: SDIC implies that our manipulations must be essentially perfect in order to lead to the desired end-state, but in practice, our manipulations are never perfect and inevitably have some error or noise in them. Thus, it seems that causal knowledge about chaotic systems is useless (since we cannot actually predict, explain, or control with it). Moreover, it is arguably even unlearnable, as all standard theory discovery and confirmation methods involve (perhaps implicitly) comparison between theoretical prediction and empirical truth.

This line of thinking—causal knowledge is either useless or unattainable in chaotic systems—has a certain superficial plausibility.[1] If right, it would bode ill for the scientific desideratum of causal knowledge more generally, given that chaotic phenomena are widespread in all sciences, including physics (Abarbanel, 1996), chemistry (Zhabotinsky, 1991), biology and neuroscience (May, 1974; Skarda & Freeman, 1987; Tsuda, 2001), psychology (Guastello et al., 2009), ecology (Benincà et al., 2008; Cushing et al., 2003), and economics (Peters, 1994; Puu, 2003). In this paper, though, we argue that this line of thinking is wrong. We can and do have

[1] Indeed, this paper was prompted by a comment made at an earlier workshop to the effect that "System *S* is chaotic, so there is no point in thinking about causal knowledge about it."

usable causal knowledge about chaotic systems in a wide range of domains, though we have to be clear about exactly what we have causal knowledge *of*. Most notably, issues about the proper level or granularity of description become particularly salient for chaotic systems. Such issues are not novel to chaotic systems, but as we will see, those systems illustrate a novel way in which the granularity of description matters. In particular, we must find the proper granularity to describe not just the system's variables (or properties or features), but also the system's dynamics. In the next section, we provide a more precise discussion of chaos theory in order to clearly show the potential pitfall. In Section 3, we show how to continue to "think causally" about chaotic systems by changing the level of description. Finally, in Section 4, we explore the broader question of how to determine, for any system (chaotic or non-chaotic), the "right" granularity of description for the system's dynamics.

2. Why chaos might pose a problem

One of the first uses of the term 'chaos' (in Li & Yorke, 1975) was intended to describe the seemingly random behavior witnessed in some dynamical systems. Chaotic systems are particularly interesting, however, precisely because they are *not* random, but only appear that way (in some sense). A wide variety of definitions have been proposed for chaos, many of which try to capture the appropriate sense of "apparent randomness" in terms of information theory, or predictability, or some other notion (Bernardo, Frigg, & Kronz, 2006; Ford, Mantica, & Ristow, 1991; Frigg, 2006; Smith, 1998; Werndl, 2009). Thankfully, we do not need to adopt a precise definition of chaos, but can rather focus on the more specific feature that leads to the appearance of randomness: *sensitive dependence on initial conditions* (SDIC).

As a running example throughout this section, we will focus on the logistic equation, which can be used to model various population dynamics, including resource-constrained population change and a certain type of predator-prey interaction (May, 1974).[2] Let x_t denote the population in an area at time t, divided by the carrying capacity so x_t is normalized to lie in the [0,1] interval. If R encodes the growth rate, then the update equation is: $x_{t+1} = R(1 - x_t)x_t$. The (normalized) population size at the next time step is thus a complex function of the interaction between population size and how close it is to the carrying capacity of the area. For $R \leq 4$, x_t will always stay in the [0,1] interval; if $R > 4$, then x_t will eventually diverge to negative infinity. This is a very simple deterministic model, but nonetheless nicely illustrates SDIC.

In general, the state of a system (or perhaps more properly, a mathematical model of a system) can be described by the current values of the relevant variables in the system (e.g., position, velocity, density, etc.). In the case of the logistic equation, the state is fully captured by the population size at that time.[3] Since the system is dynamic, we can then talk about its *trajectory*: the sequence of system states x_1, x_2, x_3, … that follow from some starting point.[4] For the logistic equation, this is simply the sequence of population sizes over time. (Chaotic systems are standardly assumed to be deterministic,[5] so there is a single unique trajectory for each initial condition.) The system follows this trajectory through *state space*: the multidimensional space of all possible system states. In the case of the logistic equation, the state space is 1-dimensional and consists of the [0,1] interval; in general, state spaces can be multidimensional and substantially more complicated.

---

[2] There are various alternative formulations of the logistic equation (e.g., having the growth rate vary as a function of x_t or t itself), but we focus on the simplest version here.
[3] Since we are assuming that R is fixed for a particular system (and so cannot change), it is better understood as a parameter of the system, rather than a variable of it.
[4] For ease of exposition, we focus on discrete-time systems in which the dynamical equations for the system are stated in terms of discrete time steps (e.g., update equations). All of our general points transfer straightforwardly to continuous-time systems.
[5] There is a literature on stochastic chaos (e.g., Bjornstad & Grenfell, 2001; Kim & Reichl, 1996), but SDIC is the key for us, not determinism vs. stochasticity.

In systems without SDIC, close initial conditions typically follow close trajectories: for example, the path of a ball released to roll down a ramp does not change much if we start the ball 1 mm to the left. Similarly, if $R = 3.3$ in the logistic equation, then it does not exhibit SDIC; as shown in Figure 1, two close initial conditions—$x_1 = 0.37$ (grey line) and $x_1^* = 0.38$ (dotted line)—stay arbitrarily close to one another: they converge so closely that the lines in Figure 1 essentially overlap.

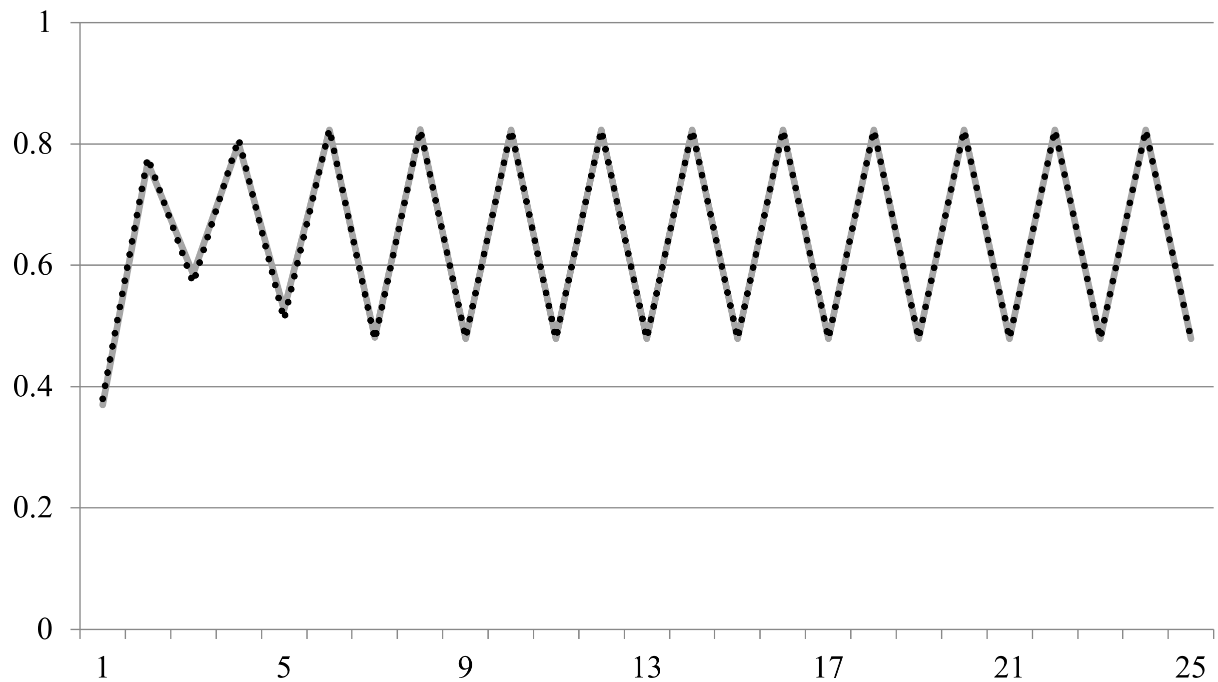

Figure 1: Logistic equation behavior for $R = 3.3$; grey line corresponds to $x_1 = 0.37$ and dotted line to $x_1^* = 0.38$

In contrast, a system exhibits SDIC just if two close-but-different initial conditions follow trajectories (through state space) that become arbitrarily different as time goes on. That is, even tiny differences can rapidly magnify to result in radically different trajectories. If $R = 4$ in the logistic equation, then it *does* exhibit SDIC. And as we show in Figure 2, the very same close initial conditions ($x_1 = 0.37$ and $x_1^* = 0.38$) lead to radically different trajectories.

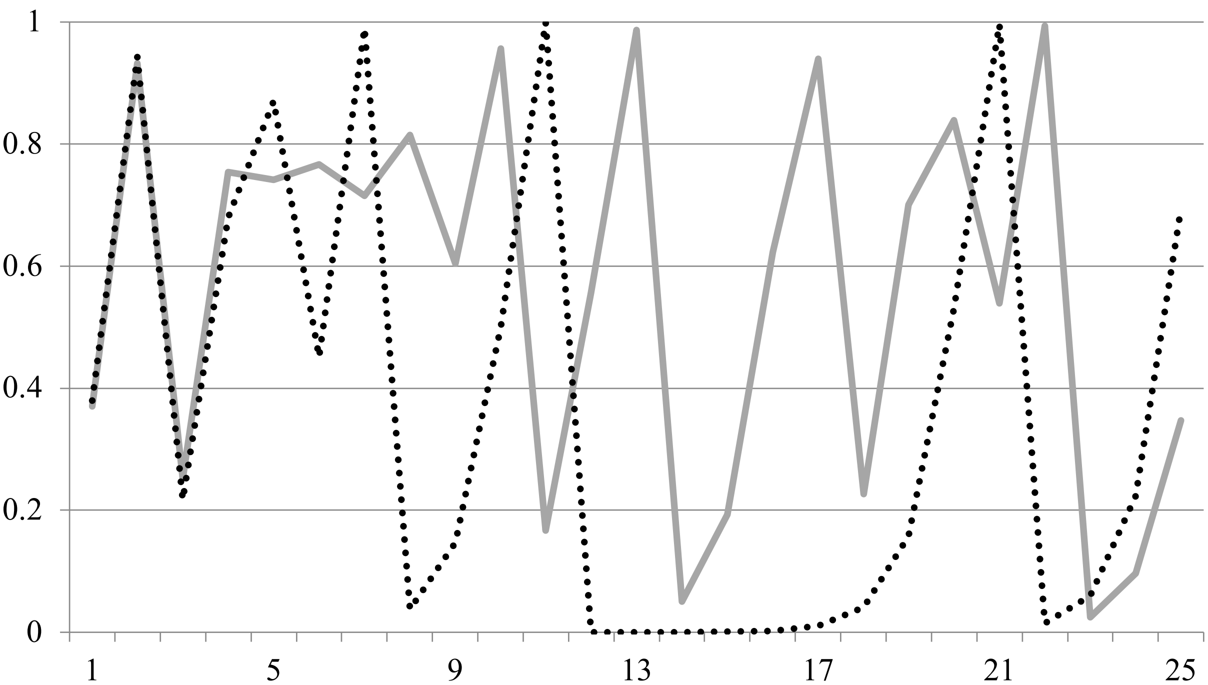

Figure 2: Logistic equation behavior for $R = 4$; grey line corresponds to $x_1 = 0.37$ and dotted line to $x_1^* = 0.38$

We can now pose the challenge for causal knowledge. As noted at the outset, one principal function of a causal theory **T** is to support prediction, explanation, and control. Control depends on our ability to use the causal theory **T** to generate complex hypothetical predictions about what would happen if we were to influence the system in some way. Similarly, the explanatory power of a theory **T** depends partly (though not wholly, as there are many aspects of explanation) on its ability to generate correct retrodictions—accounts of how the current state is implied by the truth of **T** and the earlier state(s) of the system. Thus, the latter two functions of causal theories depend in part on **T**'s predictive power, and so predictive power is plausibly a necessary condition for any particular theory **T** to satisfy these three functions of causal theories. As a result, predictive power provides a particularly powerful "lever" for the skeptic about causal knowledge for particular systems: if she can show that **T** has no (or insufficient) predictive power, then she can show in one fell swoop that our knowledge about the relevant system cannot satisfy *any* of the core functions of causal knowledge.

One way to think about prediction is that the theory **T** determines a conditional probability distribution over future states given the current state; that is, if m_t denotes that the system is measured to be in state m at time t, then **T**'s predictive power is captured by the (possibly quite complicated) conditional probability distributions $P_{\mathbf{T}}(m_{t+\Delta} \mid m^*_t)$ for all m^*, Δ. Depending on the structure of **T** (e.g., if it is non-Markov or there is significant noise in our measurements), these conditional distributions could vary if we condition on two measurements, or three, or more. If **T** is a deterministic theory with perfect measurement, then $P_{\mathbf{T}}$ will be a degenerate distribution giving probability one to a single state at each future time and zero to all others. If we conjoin **T** with an error distribution on our measurements, then these conditional distributions will be largely non-degenerate even if **T** is deterministic, since the conditional distributions are over measurements, not exact system states. For example, if we know that our measurements of population size are subject to some noise, then we can use the logistic equation plus that noise distribution to infer, for a given population measurement x_t, the probability of various future measurements.

The problem is that **T** arguably has little predictive power if it is a chaotic system. One way to think about SDIC is in terms of information: no matter how much information you have about the (sufficiently far back) past (i.e., no matter the precision or number of your measurements), you cannot improve your prediction of the current state (Werndl, 2009).[6] Thinking in terms of probability distributions, this says that we have the same probability distribution for (measurements of) the system's state, regardless of whether we condition on (measurements of)

[6] Alternately, "no matter how much information you have about the present, you cannot improve your predictions of the (sufficiently far ahead) future."

the system's states in the sufficiently far past. That is, $P(m_t \mid m^*_{t-\Delta^*}, m^{**}_{t-\Delta^{**}}, \ldots) \approx P(m_t)$, if the $t-\Delta^*$'s are sufficiently far back.[7] Moreover, this is true regardless of the noise distribution on our measurements; all that is required is that our measurements are not absolutely perfect, which is guaranteed by the finite precision of our instruments. Our inevitably noisy knowledge of the initial state of a system rapidly becomes useless, and does not support any strong inferences about the future state. But this just means that **T** does no predictive work at all: if the conditional distributions (for sufficiently past knowledge) collapse to the unconditional (i.e., observational) distribution, then the causal theory **T** is not providing any (predictive) information about the world. We would make identical predictions regardless of whether we had **T** or not. Thus, it seems that causal theories about chaotic systems cannot satisfy the necessary condition—predictive power—to fully satisfy any of their three core functions, and so such causal theories are arguably useless.

The potential problem gets even worse if we start to think about how to discover or confirm a causal theory of a chaotic system, as essentially all accounts of confirmation or discovery are based on similarity (of some sort) between predictions and observations. We seemingly can neither confirm nor discover causal theories about chaotic systems, as argued by Koperski (1998). To see the importance of predictive power in accounts of confirmation, consider Bayesian accounts of confirmation (Earman, 1992; Fitelson, 1999). At a high level, Bayesian accounts say that data D should lead us to update our subjective probabilities over possible theories using Bayes' Rule: $P(\mathbf{T} \mid D) = P(D \mid \mathbf{T})\, P(\mathbf{T}) \,/\, P(D)$. Confirmation is (roughly) an increase in probability; more generally, most Bayesians hold that one ought to believe the most probable theory at each point in time (*ceteris paribus*). The potential problem arises when one

[7] The qualification of 'sufficiently far back' is important: although the trajectories diverge exponentially, they do not diverge instantaneously.

tries to calculate the so-called "likelihood" term: $P(D \mid \mathbf{T})$. This term denotes the probability of getting data like D if **T** were actually true, and so presupposes (for essentially all causal theories) that we can generate predictions. That is, we need to be able to determine how likely it is that we would see the actual observed system trajectory if **T** were true, but we just saw an argument that shows (or purports to show) that such predictions do not depend on **T**. Thus, the Bayesian seemingly has no "raw materials" with which to update her subjective probability distribution over causal theories; all theories predict the same thing (i.e., the unconditional data distribution). Similarly, essentially all theories of causal discovery (e.g., Chickering, 2002; Pearl, 2000; Spirtes, Glymour, & Scheines, 1993) depend on finding causal structures that could have produced data like the observations.[8] All of these methods depend on having *some* way to generate predictions that can be checked against observations, but it seems that no causal theory of a chaotic system can have the required predictive power. As a result, causal discovery appears impossible for chaotic systems (though it may be possible for nonlinear systems more generally or when we measure fast enough, as in Chu & Glymour, 2008; Tillman, Gretton, & Spirtes, 2010). In summary, the potential problem is not just that causal theories of chaotic systems seem to be useless; it is that we seemingly cannot even (reliably) discover or confirm them in the first place.

[8] One might think that constraint-based causal learning algorithms (e.g., Spirtes, *et al*., 1993) could provide a way around this challenge since they depend on independencies and associations between variables, rather than explicit predictions (or likelihood functions). The problem is that independence and association are fundamentally about information: X and Y are associated just when information about one changes the distribution over the other. But there is no such information between (sufficiently far apart in time) datapoints from a chaotic system: x_t and $x_{t+\Delta}$ will be independent (for all sufficiently large Δ) precisely because neither carries information about the other.

3. How to (correctly) change the subject

The discussion in the preceding section took for granted that the exact system trajectory—that is, the description of the system state over time—is the only data or prediction that matters. But we are all familiar with the fact that predictive power can emerge when one changes level or granularity of description: for example, it is nearly impossible to predict the exact pattern of cracks in a wine glass when it is struck with a hammer, but predicting at a coarser grain of description (e.g., will it break or not?) is substantially easier. More generally, it is often easier to predict general features of a system's behavior (e.g., "*S* will mostly be in region *R* of state space") than to predict the exact trajectory that the system will follow. We suggest that this observation is the key to understanding the way in which causal theories about chaotic systems are discoverable, confirmable, and usable: rather than trying to predict the precise long-run system trajectory, one should focus on predicting either coarser-grain or shorter-run features of the system. If we do this, then we find that we can do exactly the comparisons of theoretical predictions and empirical observations that are required for all of these purposes. That is, if the data and theoretical predictions about the system *dynamics* are described at the correct granularity, then we can defeat the skeptical objection and thereby (potentially[9]) learn and confirm causal theories of chaotic systems that can support prediction, explanation, and control. The core idea is that, as Smith (1998) noted, chaotic systems have micro-unpredictability, but also some macro-predictability.

One way of coarse-graining the dynamics would be to describe the system as being in a particular region of the state space at a time, rather than at a particular point. That is, instead of focusing on the exact system state, we could try to predict only the rough location of the system in state space. In the case of the logistic equation, for example, we might characterize the system

[9] Predictive power is necessary for explanation and control, but of course not sufficient. Our focus is on defeating the skeptic's attack on a critical feature of causal theories, which does much, but not all, of the work in showing that causal theories of chaotic systems can satisfy all of their functions.

state (i.e., the normalized population) as being in either the [0.0, 0.5) or [0.5, 1.0] interval at time *t*, rather than in terms of its exact, real-valued state. There are many systems for which this is a natural way to coarse-grain: for example, one might not care about the exact location of a glass in a room, but only whether it is in the "on the table" or the "on the floor" spatial region.

Importantly, we do *not* pursue this strategy for chaotic systems.[10] Instead, we focus on the so-called *invariants of the motion*: properties of particular trajectories that hold for (almost[11]) all initial conditions, and so can be said to hold of the system in general. SDIC is one example of such a property: properly speaking, it is a property of a trajectory, but one can prove that either (almost) all trajectories for a system will exhibit SDIC or (almost) none will. As a result, we can speak of the system *as a whole* exhibiting SDIC. More precisely, the extent of SDIC is captured quantitatively in the *Lyapunov exponent* of a trajectory. A zero or negative value for the Lyapunov exponent indicates no SDIC; the more positive the Lyapunov exponent, the more sensitivity there is to the initial condition.[12] But (almost) all trajectories for a system will have the same Lyapunov exponent, and so we can justifiably talk about the Lyapunov exponent for the whole system.

Another invariant of the motion is the dimension of the attractor. An attractor for a trajectory is a collection of points in phase space to which the trajectory converges as time goes on. In the case of the logistic equation with $R = 3.3$, for example, the attractor is an oscillation between a pair of points (see Figure 1). The attractors for different trajectories of a system can differ, but (almost) all share certain topological properties. In particular, the attractors will all have the same dimension, and so we can sensibly speak of the *system's* attractor dimension. Non-chaotic systems have

[10] The problem is that the resulting "state space coarse-graining" would just be "for all future times, the system will be close to the attractor (in state space)." Attractors can be very large portions of state space, however, and so this coarse-graining is *too* coarse to be useful.

[11] Throughout this section, this qualification should be read in the measure-theoretic sense of "measure one" or "measure zero."

[12] Trajectories actually have multiple Lyapunov exponents, one for each degree of freedom. However, SDIC arises if *any* Lyapunov exponents are positive, so we can focus simply on the largest one.

attractor dimensions that are zero (corresponding to point attractors), one (ellipses), or some other whole number (tori). In contrast, systems exhibiting SDIC have *strange attractors* with fractional dimension. Strange attractors occupy some fractional part of the phase space, rather than lying on a simpler, lower-dimensional manifold.[13]

Invariants of the motion characterize some generalizable and predictable high-level properties of a system's dynamics. They can be calculated directly from a mathematical model. More surprisingly, and more importantly for theory confirmation, they can be estimated from real-world data. Suppose we have a large dataset of a population's size over an extended period of time, and we have reason to believe that the system might be chaotic; for example, we might suspect that the logistic equation governs the population size changes. The first step in analyzing this dataset is to reconstruct (something isomorphic to) the system trajectory in phase space from our data. In the case of population size, this is straightforward since we have only one dimension. Even if we have a multidimensional system (e.g., over variables X, Y, Z), we can reconstruct the trajectory—more properly, a mathematical object with the same properties as the system trajectory—using only the time series data from one variable of that system (Packard, Crutchfield, Farmer, & Shaw, 1980). Specifically, we reconstruct the trajectory as a sequence of multidimensional points, where the values of each are those of a single variable at a series of points in time.[14] In the X, Y, Z case, for example, although the "true" system trajectory points are the values of those different variables at each time t—(x_t, y_t, z_t)—we can also represent that trajectory using the points (x_t, $x_{t+\tau}$, $x_{t+2\tau}$) for a suitable time delay τ (Packard et al., 1980). Creating these new points is called "embedding the

[13] Depending on exact definitions, it seems that a system can have a strange attractor with no positive Lyapunov exponents (Grebogi, Ott, Pelikan, & Yorke, 1984), or a positive Lyapunov exponent with a non-strange attractor (Franaszek, 1987). We ignore that complication here.

[14] The dimension of the reconstructed point is the number of active degrees of freedom of the system. There are many techniques for discovering this number from our data, as well as learning the optimal time delay. While mathematically and scientifically interesting, the details of those techniques are tangential to our main point, so we omit them here. Abarbanel (1996) provides a useful overview of these techniques.

time series" using a "delay-coordinate map," or more simply, just "using a time delay embedding." This procedure takes advantage of the so-called *embedding theorem* (Mañé, 1981; Takens, 1981), which informally says: for (almost) all functions that assign real values to points in phase space, a delay-coordinate map of that function is structurally equivalent[15] to the original trajectory.

The invariants of motion such as Lyapunov exponents and attractor dimension[16] can be calculated from this reconstructed trajectory, and those are exactly the properties that we can attribute to the system as a whole. Thus, time delay embedding plus analysis of the reconstructed trajectory can give us substantial information about the dynamical properties of the system in a wide range of conditions. We have gone into some detail about how real-world data from a chaotic system can be analyzed because it is important that nothing is being swept under the rug here. All of these data analysis techniques are straightforward and mathematically defensible, and the net result is that we can empirically determine the invariants of motion for a given system *S*. Meanwhile, starting with a given causal theory **T**, there are numerous methods to determine these exact same invariant quantities for **T**, either analytically or through numerical simulation (see, e.g., Abarbanel, 1996). Thus, we *can* actually carry out suitable comparisons of predictions and observations for chaotic systems, as long as we stay at the correct granularity of description of the system behavior (e.g., the Lyapunov exponent). We have lost *some* predictive power relative to a non-chaotic system, as we cannot accurately predict exact system states in the distant future, but contrary to the skeptical concern, we have not lost *all* predictive power.

As an example of this process in action, consider the observational study of Benincà, *et al.* (2008). They measured the abundance of various microorganisms that are part of a common food

---

[15] That is, the reconstructed trajectory is a smooth one-to-one immersion to the original system trajectory. As such, all topological and differentiable properties of the original system trajectory are preserved in the reconstructed trajectory.
[16] There are other invariants of the motion (e.g., entropy), but we focus on these two, as they suffice to illustrate our points.

web in (an isolated patch of) the Baltic Sea over more than six[17] years. By isolating the water column from the broader Baltic Sea, they were able to ensure that there were relatively stable external conditions. Nonetheless, they observed significant variability in the abundances of the microorganisms; the absence of external sources of variability suggested that this variation was due to the various predator-prey relationships. Moreover, they found that the abundances were highly predictable in the short run of several days,[18] but were essentially completely unpredictable at the 15-30 day timescale. The data and background knowledge (e.g., Hastings & Powell, 1991; Huisman & Weissing, 1999) were both suggestive of chaotic phenomena, and so they investigated that possibility both empirically and theoretically.

On the empirical side, Benincà, *et al*. (2008) applied the time delay embedding technique separately to the time series data for each species in order to reconstruct the attractor for each species' abundance. These attractors were then used to calculate empirical estimates of the Lyapunov exponents for each of the abundances, all of which turned out to be significantly positive. The different Lyapunov exponents were also very similar in magnitude, suggesting that the food web is actually a single, connected chaotic system, rather than a collection of separable chaotic processes. Moreover, the empirically-derived Lyapunov exponents implied that the time series should be relatively unpredictable after 20-30 days, which fit with their earlier findings about predictability. On the theoretical side, Benincà, *et al*. (2008) developed a model of the food web based on *a priori* domain knowledge. They then derived the Lyapunov exponents for this theoretical model, and found that the theoretically-derived Lyapunov exponents were statistically indistinguishable from the empirically-derived ones (while still being significantly positive). From this fit (or rather, inability to reject a difference) between the empirical and theoretical Lyapunov

---

[17] They collected data over an eight-year span, but their data analysis only covered six years, four months.
[18] Specifically, they trained a neural net to predict species abundances several days in the future given the current species abundances. The resulting network predictions had $R^2 = 0.7 – 0.9$.

exponents, they concluded that we have reason to think that the proposed food web model captures something important about the causal structure of the interactions between the different species in this environment, even though those interactions produced chaotic trajectories. That is, they goth provided ample evidence that the observed system is chaotic, and also were able to generate model predictions and compare those to data, albeit at a coarser grain (i.e., Lyapunov exponents) than the raw time series data.

More generally, we are not limited to prediction and confirmation of causal chaotic models, but can also think about controlling or manipulating chaotic systems.[19] There is a large body of research on how to control chaotic systems (largely starting with Ott, Grebogi, & Yorke, 1990; Singer, Wang, & Bau, 1991; Vincent & Yu, 1991); much of this work has focused on shifting a chaotic system into a non-chaotic regime, usually for the purely practical reason that the precise system state (rather than coarse-grained properties) is easier to control in a non-chaotic system. A practical example of trying to move a system *towards* chaos arises in epilepsy treatment. There is evidence that epileptic seizures involve synchronized (i.e., regular and non-chaotic) cortical activity, including during the between-seizure periods (Gotman & Marciani, 1985). Thus, various chaos control techniques have been proposed as a way to keep the brain in its "normal" (i.e., chaotic) state so as to try to minimize the occurrence of epileptic seizures (Schiff et al., 1994; Slutzky, Cvitanovic, & Mogul, 2003). These techniques have been validated in the lab, though they have not been used in human patients for both practical and ethical reasons. Nonetheless, they show that it could potentially be both possible and valuable to push a system *into* a chaotic regime (or move it around within that regime). Moreover, this research focuses on controlling various coarse-grained system features, rather than controlling the exact system trajectory. Causal models of chaotic

[19] However, the appropriate model for answering control questions might be quite different than the appropriate one for prediction. We briefly revisit the possibility of multiple appropriate models at the end of the next section. Thanks to Frederick Eberhardt for emphasizing that point in this context.

systems can thus support control, albeit of a coarser-grain feature than one might have initially thought. Of course, we have not shown that *all* causal models of chaotic systems support explanation and control; rather, we have shown that the predictive challenges inherent in a chaotic system do not pose a principled barrier to such capacities.

4. Lessons for causation beyond chaos

This general lesson—the stability and reliability of our causal knowledge and models can depend on how we understand the target phenomena—is not unique to chaotic systems. However, chaotic systems show us that the lesson applies in ways that have not previously been recognized. In general, previous work on finding the proper level or granularity of description for a causal system (Danks, 2015; Glymour, 2007; Ludwig, 2016; Woodward, 2016) has focused on finding the right characterization of the *variables*: which ways of carving up the space of possible states of some aggregate (e.g., a brain, a weather system, etc.) are defensible as "causally appropriate"? So, for example, although the motion of no single atom causes me to feel warm, the motion of the collective (in the air) does. No individual neuron causes my behavior, but a collection of neurons can be a perfectly sensible cause (Glymour, 2007). No particular ocean molecule causes the weather elsewhere, but the collective El Niño phenomenon leads to increased snowfall in the U.S. Rocky Mountains (Ropelewski & Halpert, 1996). Groups of people can be causes of something (say, global warming) where it is at best unclear whether any particular individual is causally responsible (Sinnott-Armstrong, 2005). In all of these cases (and many others, as argued by Strevens, 2003), we need to find the "right" variables (or properties or features or…) for understanding the causal structure of the world, where "rightness" might depend on many factors (Danks, 2015; Ludwig, 2016; Woodward, 2016). And the standard view in the causal modeling literature has been that, once we get the variables right, it is a relatively straightforward matter to get the dynamics. That is, if we can find the right variables, then learning the dynamics is "just" a matter of estimating the parameters (e.g., linear coefficients in a structural equation modeling).

Chaotic systems show us that this view is much too simple: in many cases, we must also find the right granularity of description for the *dynamics*. Similar observations could perhaps be made about other dynamical systems, including non-stationary or highly non-linear time series, though we do not explore those cases here (see also Butterfield, 2012). To return to the earlier example, the challenge for Benincà, *et al*. (2008) was not to find the proper "causal variables"; the species abundances were clearly the correct way to think about the causal relata (and note that those variables are already coarse-grained characterizations of the aggregates composed of individuals of each species). The issue in their case was instead how to characterize the dynamics of the food web in a way that revealed a stable causal structure that could be used for prediction, explanation, and (perhaps someday) control. More generally, we suggest that most modeling of non-chaotic causal dynamics takes for granted that the goal is to write down the "update equations" (possibly non-linear) that express the way that the system state depends on the previous time step. And while this is potentially a goal even in chaotic systems if the measurements occur fast enough,[20] it is clear that an understanding of the fine-grained dynamics will often not be possible. Instead, just as we can sensibly ask about the "proper" granularity for causal variables, we can ask about the "proper" granularity of causal dynamics.

We propose a pragmatic (in the cost-benefit sense) criterion for selecting the granularity of description of the dynamics: namely, that the "right" granularity to describe the behavior of a

[20] Recall that Benincà, *et al*. had moderate short-run predictive success with their chaotic system. Of course, they achieved that by using a complex neural network predictor, so did not necessarily have simple update equations.

system is any that has sufficient stability, reliability, and reproducibility that it can be used for prediction, explanation, and control. There are, of course, many ways to change the granularity of the dynamics; we focus on our proposed evaluation criterion, and leave aside the question of how algorithmically to find (one of) the best granularity change(s). In general, there will be a trade-off between granularity and stability: *ceteris paribus*, coarser-grained dynamics will be more stable. We propose that the "proper" level of granularity is the finest-grain that still supports the requisite predictive power. In this regard, our proposal is closely related to Franklin-Hall's (this volume) account of scientific explanations in terms of cost-benefit trade-offs. Our proposal is also complementary to Glymour's (2007) suggestion that proposed variables for a causal model (in his case, features of aggregates) should be judged pragmatically by whether they yield stable and reliable causal models (see also Danks, 2015 and Woodward, 2016, though Ludwig, 2016 points towards other pragmatically relevant features). That is, these authors argue that variables are appropriate for causal models just when models built using such variables can be used reliably and repeatedly. We here contend that dynamical properties should be judged by the same criteria.

One typical feature of cost-benefit trade-offs is that there may be multiple ways to maximize. In this context, there may be different granularities of description that provide an acceptable balance between "cost" of description and "benefit" of stability. A decision between different granularities will then depend on other interests or goals of the scientists. For chaotic systems, the update equations provide a good trade-off for (very) short-run prediction, but do *not* provide a good trade-off for long- or even medium-run prediction, explanation, and control. In contrast, the Lyapunov exponents, dimensions of reconstructed attractors, and so forth are robust and reliable in just the right way. In fact, they are called "invariants of motion" precisely because they are *invariant* over (almost) all conditions: regardless of when we start tracking the system, or how far out we want to predict features of the system, these invariants will continue to hold. Thus, they provide the stability that is necessary for successful prediction, explanation, and control. At the same time, the precise nonlinear equations will ultimately not provide us with the necessary prediction and control stability (except indirectly, as we can generate the coarser-grained quantities from the update equations, either analytically or by numerical simulation).

Another way of thinking about our proposal is in terms of Woodward's (2006) discussion of the idea that we prefer working with insensitive causal relations (or at least, that our thinking about a causal relation depends in part on its sensitivity). Roughly speaking, a causal relation "*C* causes *E*" is *sensitive* if the "core" counterfactuals—"If *C* had occurred, then *E* would have occurred" and "If *C* had not occurred, then *E* would not have occurred"[21]—are sensitive, and a (true) counterfactual is sensitive just when that counterfactual is not true in nearby "close" possibilities. In general, all counterfactuals involve some relativity to the actual world, as evaluation of the counterfactual assumes the constancy (between the actual world and the counterfactual one) of a range of background conditions $B_1, \ldots, B_n$. For example, the truth of the counterfactual "If I had flipped the switch, then the lights would have turned on" depends in part on the (assumed) constancy of the background conditions "The house has power", "The bulb filament is unbroken", and so forth. A counterfactual is sensitive if it is true given $B_1, \ldots, B_n$ but false if these background conditions are changed (and the degree of sensitivity depends on the plausibility and feasibility of those changes). And so a causal relation is sensitive if it fails to hold under small (by some measure) changes in the background conditions. Thus, a professor's

---

[21] Of course, one of these is presumably a description of the actual world, not a counterfactual one.

recommendation letter causing some future individual *I* not to be born[22] (in Lewis, 1986) is a classic example of an exceptionally sensitive causal relation, while a hammer strike causing a wine glass to break is relatively *in*sensitive.

Our proposal about the proper modeling granularity for causal dynamics can thus be restated (roughly) as: the dynamics of a causal system should be expressed in a way that makes the resulting predictions and explanations as insensitive as possible. Thinking back about chaotic systems, the "update equation" characterization of the dynamics yields highly sensitive predictions: for example, if our measurement of the state of the system is close to, but not identical with, the true system state, then our predictions (using the update equations) of the future system states will be incorrect. However, insensitive dynamics emerge when we look at coarser-grained features of the dynamics, such as the invariants of the motion. Even under a wide range of changes in the relevant background conditions (including changes in the relevant measurement noise sources), the system will still exhibit the same invariants of motion, and so predictions of that aspect of the system's dynamics are quite insensitive. Woodward (2006) focused on binary causal claims—either *C* causes *E* or not—and so this proposal can be viewed as a generalization of his ideas to more complicated types of dynamics and causal relations.

We have presented this account of the appropriate granularity for describing the dynamics as complementary to the problem of finding the correct granularity for the variables in order to emphasize the contrast with prior work on the proper granularity of description for causal systems. But in fact, these two are deeply intertwined with one another, as there are important interactions between the variables used in a causal model and the appropriate dynamics for that

[22] At the least, there can be counterfactual dependence: getting a job can depend counterfactually on the quality of the recommendation letter; meeting one's (future) spouse can depend on one's job; marrying one's spouse can depend on meeting him or her; and having a particular child can depend on marrying one's spouse. So, if the letter had been different, then (possibly) one's child would never have been born.

model. Consider, again, a wine glass being shattered with a hammer. If the cause and effect variables in this case are simply *Hard Hammer Strike* and *Glass Shatters*, respectively (both either "yes" or "no"), then the causal relation—that is, the dynamic "equation"—can presumably be captured by a close-to-extremal conditional probability distribution. But if the variables are instead multidimensional ones such as *Exact Hammer Strike* (including the location, velocity, angle, etc.) and *Exact Locations of Cracks in Glass*, then the dynamics will presumably not be fruitfully captured in a standard "update" equation. In this latter case, the robust, reliable dynamics are presumably at a coarser grain, such as a mapping from (overlapping) regions of the *Exact Hammer Strike* state space to (overlapping) regions of the *Exact Locations* state space.

This dependence of "granularity of dynamics" on "granularity of variables" is mutual: coarser-grained dynamics could potentially be used to suggest or justify different variables such that our causal models are more reliable or robust. More generally, there are (conceptually) three different possibilities. First, we might have independent reasons to prefer a particular set of variables, and then the challenge is to find the appropriate granularity for the dynamics. The Benincà, *et al*. (2008) case study fits into this scheme, as they do not entertain other variable sets; they assume (for good reasons) that the species abundances are the proper variables and then try to find the appropriate dynamics granularity. Second, we might instead have independent reasons to prefer a particular granularity of dynamics, and then the task is to determine the appropriate variables. The time series analysis of Chu & Glymour (2008) has exactly this form. They aim to model the causal structure of climate teleconnections—long-distance influences of one large part of the ocean (e.g., El Niño/La Niña) on another ocean region—and they assume that month-to-month, nonlinear update equations are the proper granularity for the dynamics. Their challenge (of relevance to our paper) is then to determine whether the "proper" causal variables are aggregations of sea surface pressure and temperature measurements obtained from domain scientists, or those discovered by a machine learning clustering algorithm (see also Glymour, 2007).

A third possibility is arguably the most interesting: namely, that the proper granularities for the variables and dynamics might emerge only through a back-and-forth "negotiation" in which a set of variables is tentatively suggested and a dynamics granularity found, at which point the shortcomings of the dynamics prompts a search for different variables that (perhaps) have a more stable and reliable set of causal relations. This coevolution of variables and dynamics is arguably the most common process in scientific practice, as we rarely have sufficient *a priori* domain knowledge to fix in advance (and hold fixed throughout the scientific investigation) either the variable set or the granularity of the dynamics. At the same time, we often have background information that can help to provide a starting point for this coevolution.[23] For example, we might expect the causal relations to be linear, or we might privilege (on metaphysical, epistemic, or instrumental grounds) some variables as most likely to be appropriate. There is thus a difficult "joint estimation" problem here in terms of finding the best variable set and granularity for describing the dynamics, but it is rarely a problem that we must tackle *de novo* without any constraints on the possibility space. A characterization of constraints and methods for this coevolutionary process is, to our knowledge, a completely open question.

One intriguing question that is raised by our position is how many different, mutually supporting (and interesting or useful) pairs of descriptions of variables and dynamics exist for a particular system. At least in some particular cases, there can be multiple, defensible variable constructions that are arguably mutually incompatible with one another (Danks, 2015; Fancsali, 2013; Spirtes, 2009; Woodward, 2016). And it seems possible, at least in theory, that there could

[23] Thanks to Jim Woodward and Roberta Millstein for raising this point.

be a *huge* number of variables-dynamics pairs, each of which is an equally acceptable (or at least, incommensurable) way of capturing both the variables and dynamics of a system.[24] We know of no studies or arguments that shed light on this question, probably because the issue of the granularity of descriptions of the dynamics has been relatively understudied. If there are many such pairs for some systems, then scientists potentially face significant challenges in determining which pair best satisfies their goals in a particular investigation. More generally, we have focused on finding the granularity for describing time series dynamics, but these lessons should apply to any account of causal relations, even when those relations are not dynamic. We suggest that this challenging aspect of "theory choice" (broadly construed)—determining the proper granularity for descriptions of causal relations, not just the relata—is more common in scientific practice than has previously been recognized or discussed in the philosophy of science literature.

5. Conclusions

There has historically been significant debate about exactly what lessons to draw from chaotic systems, both from scientific investigations of them and from their very existence. As we have argued in this paper, there is a superficially plausible argument that we cannot use causal knowledge to predict, explain, and control chaotic systems. Given the widespread occurrence of such systems in nature and the importance of these functions for causal models, such an argument would imply that we cannot have useful causal knowledge about much of the world. This argument is based, however, on an implicit assumption—widespread in much of the causal learning and modeling literature—that the causal system's dynamics should be captured at the fine-grained level of a time series of exact system states. We have tried to show that chaotic

[24] Thanks to Carl Craver for emphasizing the possibility that there might be a large number of such pairs.

systems are interesting in part because their dynamics do not show sufficient predictability or stability at this fine-grained level. Rather, we should characterize them in terms of coarser-grained features, such as the invariants of motion. These are the aspects of the causal system that are relatively insensitive, and so provide the required stability for prediction, explanation, and control. Chaotic systems thus do provide an important lesson about causal learning and modeling: not that it is impossible, but rather that issues about the proper granularity of description arise not just about the causal variables, but also about the causal dynamics.

Acknowledgments

This paper benefited greatly from extensive discussion at the Causation in Biology II workshop at the Minnesota Center for Philosophy of Science, particularly the helpful comments and criticisms of Carl Craver, Laura Franklin-Hall, Ben Jantzen, Roberta Millstein, Naftali Weinberger, and Jim Woodward. Thanks also to Frederick Eberhardt and Clark Glymour for feedback on earlier drafts of this paper. The first author was partially supported by a James S. McDonnell Foundation Scholar Award.

References

Abarbanel, H. D. I. (1996). *Analysis of observed chaotic data*. New York: Springer-Verlag.

Benincà, E., Huisman, J., Heerkloss, R., Jöhnk, K. D., Branco, P., Van Nes, E. H., et al. (2008). Chaos in a long-term experiment with a plankton community. *Nature*, *451*(7180), 822–825.

Bernardo, Frigg, R., & Kronz, F. (2006). The ergodic hierarchy, randomness, and Hamiltonian chaos. *Studies in History and Philosophy of Modern Physics*, *37*, 661–691.

Bjornstad, O. N., & Grenfell, B. T. (2001). Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, *293*, 638–643.

Butterfield, J. (2012). Laws, causation and dynamics at different levels. *Interface Focus*, *2*(1), 101–114.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, *3*, 507–554.

Chu, T., & Glymour, C. (2008). Search for additive time series causal models. *Journal of Machine Learning Research*, *9*, 967–991.

Cushing, J. M., Costantino, R. F., Dennis, B., Desharnais, R. A., & Henson, S. M. (2003). *Chaos in ecology: Experimental nonlinear dynamics*. San Diego, CA: Academic Press.

Danks, D. (2015). Goal-dependence in (scientific) ontology. *Synthese, 192*, 3601-3616.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: The MIT Press.

Fancsali, S. E. (2013). *Constructing variables that support causal inference*. (Doctoral dissertation). Carnegie Mellon University, Department of Philosophy.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, *66*, S362–S378.

Ford, J., Mantica, G., & Ristow, G. H. (1991). The Arnol'd cat: Failure of the correspondence principle. *Physica D*, *50*, 493–520.

Franaszek, M. (1987). Chaotic, nonstrange attractors in the presence of external, random noise. *Physical Review a (General Physics)*, *35*(7), 3162–3165.

Frigg, R. (2006). Chaos and randomness: An equivalence proof of a generalized version of the Shannon entropy and the Kolmogorov–Sinai entropy for Hamiltonian dynamical systems. *Chaos, Solitons and Fractals*, *28*, 26–31.

Glymour, C. (2007). When is a brain like the planet? *Philosophy of Science*, *74*, 330–347.

Gotman, J., & Marciani, M. G. (1985). Electroencephalographic spiking activity, drug levels, and seizure occurence in epileptic patients. *Annals of Neurology*, *17*(6), 597–603.

Grebogi, C., Ott, E., Pelikan, S., & Yorke, J. A. (1984). Strange attractors that are not chaotic. *Physica D: Nonlinear Phenomena*, *13*(1-2), 261–268.

Guastello, S. J., Koopmans, M., & Pincus, D. (Eds.). (2009). *Chaos and complexity in psychology: The theory of nonlinear dynamical systems*. New York: Oxford University Press.

Hastings, A., & Powell, T. (1991). Chaos in a three-species food chain. *Ecology*, *72*(3), 896–903.

Huisman, J., & Weissing, F. J. (1999). Biodiversity of plankton by species oscillations and chaos. *Nature*, *402*, 407–410.

Kim, S., & Reichl, E. L. (1996). Stochastic chaos and resonance in a bistable stochastic system. *Physical Review E*, *53*(4), 3088–3095.

Koperski, J. (1998). Models, confirmation, and chaos. *Philosophy of Science*, *65*, 624–648.

Lewis, D. (1986). Postscript C to “Causation”: (Insensitive causation). In *Philosophical papers*

(Vol. 2, pp. 184–188). Oxford: Oxford University Press.
Li, T.-Y., & Yorke, J. A. (1975). Period three implies chaos. *American Mathematical Monthly*, *82*, 985–992.
Ludwig, D. (2016). Ontological choices and the value-free ideal. *Erkenntnis, 81*, 1253-1272.
Mañé, R. (1981). On the dimension of the compact invariant sets of certain non-linear maps. In D. A. Rand & L.-S. Young, *Dynamical systems and turbulence, Warwick 1980* (pp. 230–242). Berlin: Springer.
May, R. M. (1974). Biological populations with nonoverlapping generations: Stable points, stable cycles, and chaos. *Science*, *186*, 645–647.
Ott, E., Grebogi, C., & Yorke, J. A. (1990). Controlling chaos. *Physical Review Letters*, *64*, 1196–1199.
Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical Review Letters*, *45*(9), 712–716.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.
Peters, E. E. (1994). *Fractal market analysis: Applying chaos theory to investment and economics*. New York: Wiley & Sons.
Puu, T. (2003). *Attractors, bifurcations, and chaos: Nonlinear phenomena in economics* (2nd ed.). Berlin: Springer-Verlag.
Ropelewski, C. F., & Halpert, M. S. (1996). Quantifying Southern Oscillation-precipitation relationships. *Journal of Climate*, *9*, 1043–1059.
Schiff, S. J., Jerger, K., Duong, D. H., Chang, T., Spano, M. L., & Ditto, W. L. (1994). Controlling chaos in the brain. *Nature*, *370*, 615–620.
Singer, J., Wang, Y.-Z., & Bau, H. H. (1991). Controlling a chaotic system. *Physical Review Letters*, *66*(9), 1123–1125.
Sinnott-Armstrong, W. (2005). It's not my fault: Global warming and individual moral obligations. In W. Sinnott-Armstrong & R. B. Howarth, *Perspectives on Climate Change: Science, Economics, Politics, Ethics* (pp. 285–285). Amsterdam: Elsevier.
Skarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, *10*, 161–195.
Slutzky, M. W., Cvitanovic, P., & Mogul, D. J. (2003). Manipulating epileptiform bursting in the rat hippocampus using chaos control and adaptive techniques. *IEEE Transactions on Biomedical Engineering*, *50*(5), 559–570.
Smith, P. (1998). *Explaining chaos*. Cambridge: Cambridge University Press.
Spirtes, P. (2009). Variable definition and causal inference. In C. Glymour, W. Wang, & D. Westerstahl, *Proceedings of the 13th international congress of logic, methodology, and philosophy of science*. College Publications.
Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search* (1st ed.). Berlin: Springer.
Strevens, M. (2003). *Bigger than chaos: understanding complexity through probability*. Cambridge, MA: Harvard University Press.
Takens, F. (1981). Detecting strange attractors in turbulence. In D. A. Rand & L.-S. Young, *Dynamical systems and turbulence, Warwick 1980* (pp. 366–381). Berlin: Springer.
Tillman, R. E., Gretton, A., & Spirtes, P. (2010). Nonlinear directed acyclic structure learning with weakly additive noise models. In *Advances in neural information processing systems 22*.

Tsuda, I. (2001). Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Sciences*, *24*, 793–847.
Vincent, T. L., & Yu, J. (1991). Control of a chaotic system. *Dynamics and Control*, *1*(1), 35–52.
Werndl, C. (2009). What are the new implications of chaos for unpredictability? *British Journal for the Philosophy of Science*, *60*, 195–220.
Woodward, J. (2006). Sensitive and insensitive causation. *Philosophical Review*, *115*(1), 1–50.
Woodward, J. (2016). The problem of variable choice. *Synthese, 193*, 1047-1072.
Zhabotinsky, A. M. (1991). A history of chemical oscillations and waves. *Chaos*, *1*, 379–386.