Abstract

Recent work by Frigg *et. al.* (2014a, 2014b) and Mayo-Wilson (forthcoming) have called attention to a particular sort of error associated with attempts to model certain complex systems: structural modeling error (SME). The assessment of the degree of SME in a model presupposes agreement between modelers about the best way to individuate natural systems, an agreement which can be more problematic than it appears. This problem, which we dub "the system individuation problem" arises in many of the same contexts as SME, and the two often compound one another. This paper explores the common roots of the two problems in concerns about the precision of predictions generated by scientific models, and discusses how both concerns bear on the study of complex natural systems, particularly the global climate.

0.0 Introduction

Over the last century, the influence of collective behavior on the global climate system has increased to the point that anthropic factors are now among the most significant drivers of climate change. Science has at times struggled to keep pace with humanity's rapidly evolving impact. Constructing models of the global climate that are empirically adequate and predictively useful has proven a very difficult task, and climate science has emerged as one of the most rapidly progressing (and challenging) sciences in history. Understanding the behavior of the global climate well enough to make predictions that are precise enough to be useful in public policy deliberations means (among other things) integrating knowledge about a constellation of different physical systems into individual models, validating those models in a context where direct experimentation on the system being modeled is virtually impossible, and building computers that are sophisticated enough to derive results from our best models on time-scales that are relevant to human decision making. These tasks, daunting as they are, are further complicated by the *complex* nature of the global climate system. While complexity science has also begun to come into its own as an independent field over the last few decades, this has if anything only served to underscore the unique challenges associated with studying complex physical systems.

Complex systems in general--and the global climate in particular--ought to be of particular interest to philosophers of science. There are a number of novel foundational problems posed by science's increasingly common confrontations with complexity, and the advancement of complexity science has given us the tools (and motivation) to reexamine previously well-worn concepts: emergence, identity, lawhood, and many other central philosophical issues have, in recent years, been examined anew through the lens of complexity theory. In a more practical and immediate way, however, climate science has presented philosophers of science with a new set of challenges--how are we to understand the methodologies of contemporary climate science, and to what extent are we justified in taking the predictions made by climate models seriously? A large number[1] of philosophers from a variety of backgrounds have attempted to meet these (and other) challenges, and progress is steady.

Most recently, Frigg *et. al.* (2013, 2014a, 2014b) and Mayo-Wilson (forthcoming) have called attention to a separate problem for climate modelers (among others): the problem of structural model error. Frigg presents a strong argument suggesting that climate modelers are in a significantly worse epistemic position than that which they'd been previously taken to occupy--that climate modeling attempts are vulnerable to a distinct flavor of uncertainty-related issues that significantly restrict the level of precision in their predictions (even in principle). If true, this is worrying for obvious reasons: given the looming socio-political relevance of anthropogenic climate change, well-informed policy decisions in the next century will necessarily be informed by our best contemporary understanding of the future of the global climate. In the interest of making the most informed policy choices possible, it is vital that we understand what relevant models can and cannot do.

Structural model error (SME) is indeed a serious worry for climate modeling, and it remains to be seen if there might be some way to meet the challenge it poses. In this paper, we explore and develop some of the background implicit in Frigg's presentation of SME, and show how it relates to another problem in the foundations of complex systems. This problem, which we dub "the system individuation problem," helps explain some of the novel difficulties faced by scientists attempting to model complex systems. In **Section 1** of this paper, we outline the difference between SME and standard (Lorenzian) chaos, a similar (but distinct) problem faced

[1] An exhaustive list of philosophical work grappling with the foundations of climate science is impossible to present. For some representative (and particularly strong) examples, see recent work by Winsberg (2001; 2003; 2009; 2012), Frigg (2007; 2013; 2014a; 2014b), Parker (2006; 2010), Lloyd (2010), Lawhead (2014), Knutti (2008), Lenhard & Winsberg (2010).

by the geosciences. After introducing the basics of standard chaos theory, we explore how chaotic behavior does (and does not) constrain the sorts of predictions that we can make about the future of the global climate. We then describe SME, and see how it relates to the challenges posed by chaotic behavior. In **Section 2**, we introduce the system individuation problem, and explore some of the general issues associated with which it is associated before seeing how it bears on the discussion of climate science. Finally, in **Section 3** we consider how SME and the system individuation problem are related, and discuss the implications of both for the future of climate science and complexity.

1.0 Chaos vs. Structural Modeling Error

Standard (Lorenzian) deterministic chaos is relatively well understood, both conceptually and mathematically. However, there are a number of different ways of presenting the intuition behind standard chaos. Because we'll need a consistent, standardized notation for our discussion of structural chaos, let's start by reviewing some of the concepts behind Lorenzian chaos. Lorenz (1963) discusses a system of equations first articulated by Saltzman (1962) to describe the convective transfer of some quantity (e.g. average kinetic energy) across regions of a fluid:

$$\frac{dx}{dt} = \sigma(y - x) \qquad (1)$$
$$\frac{dy}{dt} = x(\rho - z) - y \qquad (2)$$
$$\frac{dz}{dt} = xy - \beta z \qquad (3)$$

In this system of equations, x, y, and z represent the modeled system's position in a three-dimensional state space[2] represents the intensity of convective motion, while σ, ρ, and β are parameters representing how strongly (and in what way) changes in each of the state variables are connected to one another.

The important feature of Lorenz's system for our discussion is this: the system exhibits chaotic behavior *only for some parameterizations*. That is, it's possible to assign values to σ, ρ, and β such that the behavior of the system has more in common with the behavior of (say) a

---

[2] Precisely what this means, of course, depends on the system being modeled. In Lorenz's original discussion, x represents the intensity of convective energy transfer, y represents the relative temperature of flows moving in opposite directions, and z represents the the degree to which (and how) the vertical temperature profile of the fluid diverges from a smooth, linear flow.

clock's pendulum than it does with global weather patterns: on some parameterizations of Lorenz's equations, initial conditions that begin close to one another in the system's state space remain close together as the system evolves over time. Moreover, even systems that exhibit chaotic behavior in general may contain regions in their state space in which average distance between trajectories decreases. This suggests that it isn't always quite right to say that *systems* themselves are chaotic. It's possible for some systems to have chaotic *regions* in their state spaces such that small differences in overall state not when the system is *initialized*, but rather when (and if) it enters the chaotic region are magnified over time[3]. That is, it is possible for a system's behavior to go from non-chaotic (where trajectories that are close together at one time *stay* close together) to chaotic (where trajectories that are close together at one time diverge)[4]. Similarly, it is possible for systems to find their way *out* of chaotic behavior. Attempting to simply divide systems into chaotic and non-chaotic groups drastically over-simplifies things, and obscures the importance of finding *predictors* of chaos—signs that a system may be approaching a chaotic region of its state space before it actually gets there[5].

Exactly how hard is it to predict the behavior of a system given a chaotic model? It's difficult to answer that question in any general way, and saying anything precise is going to require that we at least dip our toes into the basics of the mathematics behind chaotic behavior[6]. We've seen that state space trajectories in chaotic region diverge from one another, but we've said nothing at all about *how quickly* that divergence happens. As you might expect, this is a feature that varies from system to system: not all chaotic behavior is created equal. The rate of

---

[3] Indeed, it is possible for two models that putatively represent the same real-world system to differ with respect to the magnitude (or presence) of chaotic behavior. This is a point that we will return to in our discussion of structural model instability.

[4] The Phillips curve in economics, which describes the relationship between inflation and unemployment, is a good real-world example of this. Trajectories through economic state space described by the Phillips curve can fall into chaotic regions under the right conditions, but there are also non-chaotic regions in the space.

[5] For example, the appearance of a period-doubling bifurcation in the system's state space.

[6] In the discussion that follows, we present a simplified description of chaotic behavior, omitting many nuances, caveats, and details in the interest of clarity and concision. There are a number of qualification that could be made at several points of this exposition, but which do not bear directly on our central argument. Our purpose here is merely to familiarize the reader with some of the relevant basic ideas so that she can grasp the ways in which chaos is (and is not) an obstacle for prediction in order to set up a contrast case for structural modeling error and the system individuation problem. For a more careful and comprehensive discussion, we refer the interested reader to one of the many excellent references on this topic, including Smith (2007), Strogatz (2001), and Alligood, Sauer, & Yorke (2000).

divergence between two trajectories is given by a particular number—the Lyapunov exponent—that varies from system to system (as well as within the system, in many cases).[7]

Let's consider a very simple case to get a feel for the concepts here. Suppose we have two trajectories $x_0 \rightarrow x_t$ and $y_0 \rightarrow y_t$ in a system with a constant Lyapunov exponent λ. If the two trajectories are initially separated by some distance $\boldsymbol{\delta} Z_0$ then their separation at some later time t is:

$$\delta Z_t = e^{\lambda t}(\delta Z_0) \qquad (4)$$

The time-scales at which chaotic effects come to dominate the dynamics of the system, then depend in part on two factors: the value of the Lyapunov exponent, and how much divergence we're willing to allow between two trajectories before we're willing to consider it *significant*. If λ is small, divergence at short timescales will be very small, and will thus likely play little role in our treatment of the system (unless we have independent reasons for requiring very great precision in our predictions). Likewise, there may be cases when we care only about whether the trajectory of the system after a certain time falls into one or another *region* of state space, and thus can treat some amount of divergence as irrelevant.

This point is not obvious but it is very important. Let's spend some time thinking about what we can learn by playing around a bit with the toy system from above.

To begin, let D be some neighborhood on $\mathfrak{R}^n$ such that:

$$< x_0 \, , \, y_0 > \; \in D \leftrightarrow \; \delta Z_0 \leq \varepsilon \qquad (5)$$

That is, let D be some neighborhood in an n-dimensional space such that for all pairs of points that are in D, the distance between those two points is less than or equal to some small value epsilon. Let D_t be the region containing the points of the system at time t if the system were

[7] Because of this variation—some pairs of trajectories may diverge more quickly than others—it is helpful to also define the *maximal* Lyapunov exponent (MLE) for the system. As the name suggests, this is just the *largest* Lyapunov exponent to be found in a particular system. Because the MLE represents, in a sense, the "worst-case" scenario for prediction, it is standard to play it safe and use the MLE whenever we need to make a general statement about the behavior of the system as a whole. This is just one of the many real-world complications that we have chosen to omit from this discussion in the interest of clarity.

initialized a state in *D*. If $\Re^n$ is the state space our system with constant Lyapunov exponent λ, then combining (4) and (5) lets us deduce[8]

$$\forall (t > 0)\ [< x_t\,,\, y_t > \ \in Dt \rightarrow \ \delta Z_t \leq \ \varepsilon(e^{\lambda t})] \qquad (6)$$

Informally, this means that for all times after the initialization time, the size of the smallest neighborhood that *must* include the successors to some collection of states that started off arbitrarily close together will increase as a function of rate at which trajectories in the system diverge and the amount of time that has passed. That's a mouthful, but the concepts are fairly intuitive. In chaotic systems, the average distance between two trajectories through the state space of the system increases exponentially as time goes by—two states that start off very close together will eventually evolve into states that are quite far apart. How quickly this divergence takes place is captured by the value of the Lyapunov exponent. Generalizing from particular pairs of trajectories, we can think about defining a *region* in the state space. We can think about the relationship between our region's volume at one time and the smallest region encompassing the end-state of all the trajectories that started in that region at some later time. This size increase will be straightforwardly related to the degree of divergence of individual trajectories in the region, so the size of the later region will depend on three things: the size of the initial region, the rate at which paths through the system diverge, and the amount of time elapsed.[9] If $\lambda > 0$, then no matter how small we make our region the trajectories followed by the states that are included in it will, given enough time, diverge significantly.

How much does this behavior actually limit the practice of predicting what chaotic systems will do in the future? Consider two limit cases of the inequality in (6). First:

$$lim_{\varepsilon \rightarrow 0}\ \varepsilon(e^{\lambda t}) \ = 0 \qquad (7)$$

This is just the limiting case of perfect measurement of the initial condition of the system—a case where there's absolutely *no* uncertainty in our first measurement, and so the size of our "neighborhood" of possible initial conditions is zero. As the distance between the two points in the initial pair approaches zero, then the distance between the corresponding pair at

[8] Thanks to an anonymous reviewer for pointing out an error in the original formulation of (6) and suggesting a cleaner way to make the intended point.
[9] If we have some way of determining the largest Lyapunov exponent that appears in D, then that can stand in for the global MLE in our equations here. If not, then we must use the MLE for the system as a whole, as that is the only way of *guaranteeing* that the region at the later time will include all the trajectories.

time t will also shrink (again, given the simplifying assumption that λ is constant). Equivalently, if the size of the neighborhood is zero—if the neighborhood includes one and only one point—then we can be sure of the system's position in its state space at any later time (assuming no stochasticity in our equations). This highlights the fact that standard chaos is *deterministic* chaos; the practical difficulties associated with predicting their long-term behavior emerges from small uncertainties about the initial conditions. However:

$$lim_{\lambda \to 0}\, \varepsilon(e^{\lambda t}) = \varepsilon \qquad (8)$$

As λ approaches zero, the second term on the right side of the inequality in (6) approaches unity. This represents another limiting case—one which is perhaps even more interesting than the first one. Note that (8) is still valid for non-chaotic systems: if $\lambda = 0$, the distance between two trajectories will remain constant as those points are evolved forward in time.[10] More interestingly, think about what things look like if $\lambda > 0$ but still very small. No matter how small λ is, if $t \gg \frac{1}{\lambda}$ the distance between even two trajectories that begin arbitrarily close together will become arbitrarily large; even a very small amount of divergence becomes significant on long enough time scales. Similarly, if $t \ll \frac{1}{\lambda}$ then we can generally treat the system as if it is non-chaotic (as in the case of the orbits of planets in our solar system). The lesson to be drawn is that it isn't the value of either t or λ that matters so much as the *ratio* between the two values combined with our tolerance for error in the precision of our predictions.

1.1 Prediction and Standard Chaos

It can be tempting to conclude from this that if we know λ, ε, and t, then we can put a meaningful and objective "horizon" on our prediction attempts. If we know the amount of uncertainty in the initial measurement of the system's state (ε), the rate at which two paths through the state space diverge (λ), and the amount of time that has elapsed between the initial measurement and the time at which we're trying to make our prediction (t), then shouldn't we be able to *design* things to operate within the uncertainty by defining relevant macrostates of our system as being uniformly smaller than $\varepsilon(e^{\lambda t})$? If this were true, it would be very exciting—it

[10] If the Lyapunov exponent is *negative*, then the distance between two paths *decreases* exponentially with time. Intuitively, this represents the initial conditions all being "sucked" toward a single end-state. This is the case for dissipative systems (for instance, the case with a damped pendulum): all initial conditions eventually converge on the rest state.

would let us deduce the best way to construct our models from the dynamics of the system under consideration, and would tell us how to carve up the state space of some system of interest optimally given the temporal scales involved.

Unfortunately, things are not this simple. In particular, this suggestion assumes that the state space can be neatly divided into continuously connected macrostates, and that it is not possible for a single macrostate's volume to be distributed across a number of isolated regions. It assumes, that is, that simple *distance* in state-space is always going to be the best measure of qualitative similarity between two states. This is manifestly not the case. As a simple analogue, suppose you're making a measurement (in familiar, everyday space) and decide to round the outcome of your measurement to the nearest half inch. It's clear that this decision might have consequences ranging from completely innocuous to horribly disastrous, depending on the circumstances in which you're making your measurement. If you're reporting your height to the DMV to get a new driver's license, the half-inch rounding is inconsequential; if you're a neurosurgeon preparing to make the first incision, however, that same half-inch is likely to be the difference between a successful operation and a well-deserved malpractice lawsuit (not to mention patient death). Something very similar is true in state space: while a very short distance between two states tells you that those states are similar in *some* respect, whether or not they're similar *enough*--similar in the way that matters--is another matter entirely. Judging the second sort of similarity involves considering a whole host of other factors, none of which can be straightforwardly discerned from information about state space distance, no matter how precise that information is.

Might this just go to show that such a naive notion of "distance" isn't the appropriate one to work with here? Perhaps scientists, when building models, merely need to define a different conception of "distance" that actually tracks the factors that result in "significant" distance, whatever that might mean in a given context. After all, there's no rule that says we have to remain wedded to the intuitive standard distance metric that we're used to working with. The impulse behind this kind of objection is correct and solidly grounded, and the recognition that we can (and in many cases *must*) construct novel metrics for measuring distance across state spaces when working with mathematical models of dynamical systems is an important one. However, suggesting that this redefinition is a trivial (or even relatively easy) matter is a mistake. Mathematical modeling the natural world is *hard*, and constructing a model that is both tractable and accurate enough for our purposes is difficult enough already, and so modelers often (quite reasonably) choose distance metrics that ease calculation, computation, and derivations. These choices reflect a spectrum of practically motivated choices in model building, and there is no guarantee that distance metric chosen to reflect those ease of use criteria will correspond neatly with our distinct (albeit just as practical) interests in predicting the qualitative behavior of the system. Indeed, it is precisely this potential mismatch that we are interested in here.

Let's distinguish, then, between two related but distinct concepts: the quantitative notion of *distance* and the qualitative notion of *similarity*. Information about the distance between two states tells us something about the formal structure of the model used to generate the space but, by itself, won't reveal the presence of any sensible way to group those states into regions that share predictively useful behaviors in common. Without an independent measure of how to group regions of a state space together such that the states inside those regions are similar to one another, we have no way of guaranteeing that just because some collection of states falls within the bounds of the region defined by (6)--a region defined with reference to distance--they are alike in any significant way. Two states might be very close together in terms of distance, and quite far apart in terms of similarity. Failing to notice this fact can obscure interesting, important dynamical facts about the system.

Generalizing from this case, we can conclude that knowing λ, ε, and t is enough to let us put a meaningful cap on the distance resolution of future predictions (i.e. that they can be only as fine-grained as the size of the neighborhood given by $\varepsilon(e^{\lambda t})$) *only if* we stay agnostic about the presence (and location) of similar macrostates when we make our predictions. That is, while the inequality in (6) does indeed hold, we have no way of knowing whether or not the size and distribution of similarly interesting, well-behaved regions of the state-space will correspond neatly with distance-based size of the neighborhoods defined by that inequality.

To put the point another way, restricting our attention to the behavior of some system *considered as a collection of states* can distract us from relevant factors in predicting the future of the system. In cases where the dynamical form of a system can shift as a function of time, we need to attend to patterns in the formation of well-behaved regions (like those of thermodynamic macrostates)—including critical points and bifurcations—with just as much acumen as we attend to patterns in the transition from one *state* to another. This is a very important point, and one we shall return to later.

1.2 Structural Modelling Error

Similar predictive concerns have led Frigg *et. al.* (2014a) to worry about structural modelling error (SME). Just as with Lorenzian chaos, SME has to do with the rate at which predictions about a system diverge from one another. However, SME is explicitly a property of *ensembles of models* rather than of systems themselves. Informally, SME is present when the predictions made by multiple models of the same system diverge from one another exponentially over time, rendering predictions by those models useless from the perspective of practical decision-making. The most striking difference between Lorenzian and SME is that the latter can be present even in the case of perfect knowledge of initial conditions, a circumstance in which (as we saw in [7]) eliminates the predictive problems associated with standard chaos.

Frigg *et. al.* (2014a) frames the discussion of SME with a revamp to the familiar story of Laplace's Demon. In Frigg's retelling, the Demon has decided to hire two apprentices to help him with his work: a Senior Apprentice and a Freshman Apprentice. Recall the traditional account of Laplace's Demon. The Demon is an entity gifted with the ability to perfectly measure and predict the future of any physical system to which it turns its attention, and to do so very rapidly. Frigg describes the Demon as having the following three powers:

a. Computational omniscience: the ability to apply a set of deterministic equations of motion to calculate the future state of a given physical system with perfect precision, and to do so arbitrarily quickly.
b. Dynamical omniscience: the ability to discern the true deterministic equations of motion for the relevant physical system
c. Observational omniscience: the ability to determine the initial conditions of a given physical system, and to do so with perfect accuracy.

In the language of **Section 1**, we can say that for all cases, the Demon is able to measure the state of a physical system such that $\varepsilon = 0$, and so $\varepsilon(e^{\lambda t}) = 0$, irrespective of the value of λ. Sensitive dependence on initial conditions--Lorenzian chaos--is of no concern to the Demon, as for even the largest values of λ, the demon is able to determine the precise initial conditions of the system, so his application of the equations of motion will always generate the correct prediction, with zero uncertainty. Moreover, the Demon has access to the correct equations of motion for the system with which he's engaged. For classical physical systems, we can think of this as the Demon's having access to the Newtonian equations of motion. Finally, we can stipulate that the Demon is very careful in his calculations, and never introduces an error through the incorrect application of those equations.

The Demon's apprentices are not quite as lucky. The Senior Apprentice, being more advanced than the Freshman, has nearly all the Demon's powers; he only lacks power (c), observational omniscience. The Freshman is less advanced, and lacks not just power (c), but power (b) as well, and makes errors in both his measurement of initial conditions and in his discernment of the proper equations of motion to apply when generating his predictions. Frigg asks us to consider how these two apprentices will stack up against their master in generating predictions about a system's time-evolution. It seems clear that neither will fare as well as the Demon himself, as both are subject to the concerns described in **Section 1**, and thus susceptible to the limitations on predictive utility resulting from Lorenzian chaos. The Freshman, however, seems to be in markedly worse shape than the Senior: not only are his predictions vulnerable to Lorenzian effects, he can never be sure if they were generated using incorrect equations of motion. As Frigg describes, this can lead to cases where the Freshman's predictions are not just *useless* for making decisions about the future, but actively *misleading*. Even if the Freshman's model is structured very much like the Demon's, in cases where the system behaves chaotically it can happen that:

> [The Freshman's] probabilities are off track: he regards events that do not happen as very likely, while he regards what actually happens as very unlikely. So his predictions here are worse than useless: they are fundamentally misleading. Hence, simply moving an initial distribution forward in time under the dynamics of a model--even a good one need not yield decision-relevant evidence. Even models that yield deep physical insight can produce disastrous probability forecasts[11].

[11] Frigg *et. al.* (2014a), p. 39

Just as with Lorenzian chaos, the heart of the problem here is the possibility of extreme, difficult to foresee divergence between trajectories that are generated by initial conditions that are separated by an arbitrarily short distance from one another. Here, however, the problem seems much more severe--the "initial conditions" are not points in a dynamical state space separated by a confusion in measurement, but rather different *models* of the system represented by that state space. Frigg et. al. (2014a) have adopted the term "the hawk-moth effect" (c.f. "the butterfly effect") from Thompson & Smith (2013) to refer to this meta-level instability. Informally, the hawk-moth effect states that we might be arbitrarily close to the correct model to predict the future behavior of a particular dynamical system, and yet still generate predictions that make no contact with the *actual* behavior of the system. Mayo-Wilson (forthcoming) quite astutely points out that this is a mathematically precise way of articulating the classic problem of induction: a model might be close enough to the "correct" model to perfectly reproduce past states of the system (up to and including the present time), and yet still fail spectacularly to generate accurate predictions about the future of the system. This hawk-moth effect is in many ways more troubling than the butterfly effect, as it seems to stymie the traditional approach to dealing with Lorenzian chaos in dynamical systems--that is, the process of moving from precise "point predictions" about the future of the system, and embracing statistical ensembles of the system's most likely future behavior.

Given the apparent fact that climate models exhibit SME, Frigg et. al. (2014a) is deeply pessimistic about the possibility of model-driven policy recommendations.[12] Mayo-Wilson notes that giving coherent *explanations* of structurally chaotic systems is equally difficult. However, there are a number of questions that need to be addressed with respect to SME--Mayo-Wilson rightly calls identifies it as the birth of a new research program.

There are, however, a set of suppressed assumptions in Frigg's formulation of SME: a set of assumptions about the Demon's knowledge, and of the knowledge of his apprentices. Specifically, Frigg has assumed that both the Demon and his two apprentices agree on what

[12] This family of worries is first expressed in Frigg et. al. (2013), and fully elucidated in the context of structural model instability in Frigg et. al. (2014a)

counts as their target system in the first place, and that this assumption is both deeply correct and can be maintained over the timeframe in which they're interested. This assumption is innocuous enough in some cases, but may be far less so in other cases--especially those in which the three ur-scientists are attempting to deal with a highly complex system. This question of how to *individuate* a system is one that real-world scientists too must grapple with, and which has serious implications for the practice of modeling complex natural systems. Given Frigg's explicit interest in drawing lessons about the implications of SME for making decisions about how to respond to phenomena like anthropogenic global climate change, these considerations are particularly relevant.

2. The System Individuation Problem

Complex systems are partially characterized by the presence of many behavioral constraints operating at many different spatio-temporal scales. Lawhead (2014) argues that the degree to which the dynamics of a given system at one scale constrain the allowable dynamics at other scales serves as a good measure for *how* complex a given system is, and more traditional conceptions like the "effective complexity" outlined by Gell-Mann & Lloyd (1996) and the earlier "algorithmic information content" described in Kolmogorov (1963) also tie complexity in some way to the amount of information contained in a system. In many familiar systems, the dynamics at very different scales operate quasi-independently. If we (say) are interested in tracking the propagation of waves across the Pacific ocean after a major seismic event, the formalism of atomic physics is not the appropriate lens through which to view the system. The degrees of freedom that are relevant in atomic physics disappear at the relatively low-energy scale of ocean waves, and generating an accurate prediction of how macroscopic waves move and interact means ignoring some degrees of freedom in the system which would be central if we examined water at an atomic level. This is a familiar problem; Putnam (1975) notes that a very sophisticated knowledge of quantum electrodynamics (QED) isn't much help in predicting whether or not this square peg will fit in that round hole, despite the fact that the behavior of both the peg and the hole are consequences of the laws of QED. When we seek to model some system so that we can predict its behavior, we often choose to ignore other models that operate at scales that are wildly disparate from our scale of interest.

In some cases, as with wooden pegs and ocean waves, these cut-off scales are relatively well-defined, and allow scientists to investigate (say) the behavior of quarks without concerning themselves with the demand that their theories remain mathematically well-behaved when applied to, for instance, the motion of galaxies[13]. This seems very natural; it would strike us as incredibly odd if, for example, the construction of the Mars Rover depended sensitively on details about the dynamics of fermions, despite the fact that things like the Mars Rover are composed to no small degree of fermionic constituents.

Complex systems, however, often fail to present with such neatly demarcated scales. In many (if not all) complex systems, two striking features obtain: (1) states of the system's constituent parts are constrained by the state of the system as a whole, and (2) the constraint(s) on the system's constituent parts are not present if the parts are isolated from the rest of the system.[14] Part of the challenge in modeling the behavior of complex systems, then, lies in modeling how patterns operating at very different scales affect and constrain one another.

The observation that complex systems are often best viewed in terms of mutually-interacting constraints operating at highly disparate scales and levels of analysis is not novel. The implications for (in particular) complex adaptive *biological* systems has been well-explored.[15] Our concern here is not with rehashing this discussion, but rather applying these lessons to the problem of SME, and showing how that problem stems from more general issues in modeling complex adaptive systems.

2.1 Will The Real System Please Stand Up?

Frigg *et. al.* and Mayo-Wilson both seem to take it as a matter of course that there is a (single) correct model of a given system, and that scientific attempts to predict the future behavior of any system consist in attempts to hit on that model. The hawk-moth problem arises in cases where a model candidate that differs from the true model even in an arbitrarily slight way may lead to abysmally bad predictions. In that sense, the problem associated with SME is

[13] Hartmann (2001); Castellani (2002)

[14] If this strikes you of smacking of "downward causation" (of the type Kim [1992] and [2003] criticizes), you are quite correct. The structure of many complex systems does great violence to some of our cherished metaphysical beliefs about the natural world (and so much the worse for those beliefs!).

[15] See, e.g., Massio, Bich, and Moreno (2013); Mossio & Moreno (2010); Mossio, Saborido, and Moreno (2009), Collier (2008; 2011), Barandiaran, X., & Moreno, A. (2006), and many others.

far more serious than that associated with standard chaotic behavior: the error associated with Lorenzian chaos can be slowly whittled away with successively better measurements and models, and we can be confident that such a gradual procedure will converge on arbitrarily precise predictions if we are careful and diligent. In cases where SME is a possibility, we have no such assurances; neither improved measurements nor refined models are guaranteed to bolster the accuracy of predictions unless we have hit on the *correct* model of the system in question.

This is indeed a cause for concern (to put it mildly!), but this way of framing the problem strikes us as problematic. Attempting to discern the correct model of a system only makes sense as an endeavor once we've already agreed on how to individuate "the system," and the identity of a system--the answer to questions of the form "what counts as part of the global climate?"--is not something given to scientists by nature, fully formed and ready for our modeling attempts, nor is it something that can be discovered through careful observation and experimentation. Rather, as Cumming & Collier (2005) note:

> The role played by our subjective interest in the system is in many ways crucial to our system definition. If we ask different questions about the system that we are studying, we can expect different answers, and, for the same question, the answer might depend on our motivations for asking it.[16]

This concern is echoed in McAllister (2003) as part of his discussion of the problems associated with applying Gell-Mann's effective complexity[17] to the global climate system. McAllister points out that the data set associated with atmospheric temperature exhibits many different patterns at many different scales:

> These include a pattern with a period of a day, associated with the earth's rotation about its axis; patterns with periods of a few days, associated with the life span of individual weather systems; a pattern with a period of a year, associated with the earth's orbit around the sun; a pattern with a period of 11 years, attributed to the sunspot cycle; a pattern with a period of approximately 21,000 years, attributed to the precession of the earth's orbit; various patterns with periods of between 40,000 and 100,000 years, attributed to fluctuations in the inclination of the earth's axis of rotation and the eccentricity of the earth's orbit; and various patterns with periods of between 10^7 and 10^9 years, associated with variations in the earth's rate of rotation, the major geography of the earth, the composition of the atmosphere, and the characteristics of the sun

McAllister argues that such a plurality of signals makes the task of settling on a single value for the effective complexity of a climate data set impossible. There are clear parallels with the Cummings & Collier's point: the multiplicity of interesting patterns in complex systems can

[16] Cummings & Collier (2005), p. 29
[17] Gell-Mann & Lloyd (1996)

raise problems for traditional methods of modeling and analysis. This goes beyond "mere" perspectivalism in science. That is, the worry here is not just a rehashing of issues concerning the theory-ladenness of observation or the like; the problem is deeper than that. The task of *individuating* a system--picking it out as a thing to be studied, and separating it from a distinct (though interactive) ambient environment--and the task of *modeling* that system are wrapped up together in ways that magnify the difficulties of each.

Consider the global climate. How are we to specify what counts as "the global climate system," and what counts as exogenous forcings on the climate system? The standard definition of the climate (e.g. the IPCC glossary's reference to statistical weather patterns) is a useful individuation, but carving the world up in this way--making this very clearly purpose-driven and perspectival decision about where to draw the lines--has enormous practical implications for model building. In particular, the time-evolution some system (particularly if that system is a complex system) might result in behavior that requires us to redraw system boundaries--to *reindividuate* the system--if we we want to continue to make similar predictions about the future.

Hooker (2011) defines self-organized complex systems as those systems in which "dynamical form is no longer invariant across dynamical states but is rather a (mathematical) function of them."[18] That is, for many (if not all) complex systems, the processes that are supposed to be captured in our best models--details about how the system changes over time--result in the practice of modeling consisting in hitting a rapidly moving target. The model that's appropriate for forecasting the future behavior of (say) the global climate system today might be rendered inappropriate tomorrow, as new features come to dominate the system's dynamics. As a simple illustration, consider the difference between modeling the behavior of the paleoclimate and modeling the behavior of the contemporary global climate. In particular, consider the relatively recent importance of the relationship between the behavior of the global economy and the behavior of the global climate. It's clear that any model that hopes to make even reasonably precise predictions about the state of the global climate over the next 50 years will need to account for the interactions between human industry and the climate. It's equally

---

[18] Hooker (2011) p. 212

clear that modeling the paleoclimate *doesn't* involve accounting for similar interactions. In at least some cases, this problem can be solved by careful parameterization and tweaking of existing models. The range of carbon emission scenarios in the warming models discussed by the IPCC are an attempt to meet a very basic version of this challenge--to provide an answer to the question "how will our model outputs change if human civilization at large takes our model outputs seriously?" Still, this is only a very crude version of the deeper problem: at *some* point the coupling between the global economy and the climate may (if it has not already) become strong and intricate enough that, if we're to continue to make good predictions about the future of either system, we'll be forced to consider them *not* as coupled systems, but rather components of a single system. This reindividuation would require us to do far more than reparameterize existing models.

Similarly, consider the question of how to individuate *future* climate states such that the data we have now can be relevantly considered to generate the sorts of predictions we care about. Many economists are concerned about the impact that either climate change itself or attempts to forestall/mitigate climate change will have on the world gross domestic product (GDP). Many policy debates turn significantly on whether late-stage adaptation strategies or early-stage mitigation strategies will result in a larger decline in world GPD. If mitigation impacts world GDP more severely than adaptation, one line of reasoning goes, then it makes more sense *from a humanitarian perspective* to eschew mitigation policies like cutting fossil fuel consumption in favor of encouraging GDP growth now, then using the returns on that growth to adapt to a changed climate later. Whatever we think about this line of argumentation as a basis of global climate policies,[19] it is clear that to systematically evaluate it in the context of climate policy debates we must reconsider some assumptions about the "best" way to individuate future states of the planet. This point is reflected in the fact that there are two distinct ways in which we might object to the argument given above: we might accept this as a valid individuation of future states but reject the claim that mitigation strategies will lead to states with lower global GDPs than adaptation strategies, or we might reject this as an appropriate individuation entirely and

[19] Evaluating this class of arguments is beyond the scope of this discussion. See Stern (2006) and Tol & Yohe (2007) for more.

maintain that global GDP is simply not a valid factor to consider.

How do we know when this sort of reindividuation is appropriate, and when to simply reparameterize existing models to account for different exogenous forcings? We might think of this as an instance of what Cumming & Collier (2005) call "metamodeling," a practice that's motivated by exactly the challenges we've been discussing:

> At the heart of cohesive models of complex systems are a few issues that are extremely difficult to cope with in empirical investigation. Most complex systems are dynamic entities that span multiple spatial and temporal scales; the distinction between endogenous and exogenous dynamics is not always clear; and, because of their many components, the outcomes of manipulations of the system may differ depending on relatively small differences in starting conditions. [...] [Metamodels] are a step back from the immediate process of prediction...their value comes from the way in which they somehow capture the essential ingredients of many interrelated models in symbolic form.[20]

The implementation of this sort of metamodel reasoning is one of the things that makes the study of complex systems like the global climate challenging in a way that sets them apart from other systems science might study. There is an extra layer of inference here, and one that introduces a significant amount of new perspectivalism into the already value-laden practice of science. A particular approach to individuating a system for scientific study is informed by the predictive goals of the scientist, the possible application of those predictions, and other decisions reflecting the scientist's priorities. A natural approach to individuating a complex system is rarely uncontroversial and obvious, and there may be many approaches to individuation that outperform other approaches with respect to one set of predictive goals, and yet lag far behind with respect to different goals. Which individuation is best suited to a given predictive task cannot be simply read off of nature.

2.2 Similarity and Individuation

Let's continue with the example from above. In taking the economic argument against climate mitigation seriously, we implicitly agree on an at least partial metric for *similarity* of future states of the climate system: two states in which the global GDP is similar may be counted as similar states from this perspective, even if they differ in, e.g., global average temperature, a difference which may give them a rather large *distance* from one another in the space defined by

[20] Cumming & Collier (2005). p. 5

most climate models. This suggests a link between the system individuation problem and the distinction between similarity and distance outlined in **Section 1.**

To see how these two concepts are related, we need to think a little more carefully about what exactly we're doing when we individuate a system. Let's start with something simpler than the global climate: a box of warm gas. There are two ways of looking at this system that are likely to be interesting, each of which is associated with a collection of well-mapped dynamical laws that might be leveraged to make useful predictions. We might choose to adopt the perspective from which the "box of gas" is really not itself an individual, but rather a composite system of many small interacting individuals--molecules--each of which moves around in the way classical mechanics predicts. This two-part choice of individuation and model implicitly defines a state space for the system (in this case, the familiar six-dimensional position/velocity space used in statistical mechanics). On the other hand, we might choose to treat the gas as a whole as an individual, ignoring the dynamics of individual molecules and instead attending to features like temperature and entropy. Happily, this choice too is associated with a collection of well-mapped dynamical laws and so also implicitly defines a state space: that of thermodynamics. There are, of course, very many more possible individuations: we could treat the stuff on the left half and the stuff on the right half as individuals, for instance. The vast majority of possible ways of carving the system up into individuals aren't likely to be interesting, in the sense that they're unlikely to exhibit any useful, robust patterns that help make the sort of predictions that we care about making--what will happen if I throw a match into the box, say. Still, the fact that these two individuations are useful ones is emphatically *not* something obvious that's written into nature; we discovered that carving things up in that way was helpful only after long centuries of experimentation.

This example is, in a certain sense, special. We chose it in part because the relationship between the state spaces and dynamics associated with each individuation are related to one another in ways that are robust and well-explored. Quantities that are associated with the first individuation (like momentum) can be mapped on to quantities associated with the second (like temperature) in ways that help us understand both: this is the business of statistical thermodynamics. This is the first time that concepts like “microstate,” “macrostate,” and “similarity” enter into the picture; these terms only make sense when we’re comparing two (or more) individuations and looking for patterns linking them. If we’re attending only to a single way of individuating a system, the distinction between distance and similarity that we laid out in **Section 1** collapses for the simple reason that once we start to group states into regions that share what we described as “qualitative” features in common--once we start carving up a collection of microstates into macrostates--we’ve already begun to attempt to reindividuate the system. When we notice that some set of states in one space are all “similar” to one another in the sense that by grouping them together we can discover stable dynamical patterns that can be leveraged to make useful predictions, we’ve also discovered that there is another individuation worth exploring: the one where collections of points in the original space can all be associated with a single point in the new one--collections of microstates that all correspond to the same macrostate. What was similarity in our old space, then, becomes distance in our new space.

The choice of what to count as an interesting macrostate--how to carve the state space of a system up into volumes that are distinct from one another in significant and useful ways--is thus a choice about how we might reindividuate a system. The question or whether one microstate is similar to another is not just a question of whether those microstates are close to one another in the sense of being proximal in a given state space, but also a question of whether or not they fall into volumes corresponding to similar macrostates. The answer to this question--of which microstates are relevantly similar to one another such that they can appropriately be grouped together into a single macrostate--is non-trivial to discern, and depends on what our predictive goals are, what sorts of measurement tools we have access to, and also on the whole host of practical concerns that always go into model building. System individuation and model building are deeply linked, and both are foundational to the practice of science. Attending only to problems (or methodologies) associated with the latter means ignoring a significant part of what science consists in, and potentially failing to notice opportunities to improve the predictions made from the perspective of one individuation by noticing its relationship to another.

2.3. System Individuation in Practice

Consider again the task of predicting the future of the global climate. What are the criteria by which we divide the possible futures of the global climate into macrostates such that those macrostates are relevant for the kinds of decisions we need to make? That is, how might we *individuate* the global climate system so that we can notice the patterns that might help us predict the outcome of various climate policies? The answer to this question depends in part upon what we consider valuable; if we want to maximize long-term economic growth for human society, for instance, our set of macrostates will likely look *very* different than it would if we wanted to simply ensure that the average global temperature remained below a particular value. Both of those in turn may differ significantly from a set of macrostates informed by a desire to maximize available agricultural land. These different ways of carving possible future states up into distinctive macrostates do not involve changes to the underlying equations of motion describing how the system moves through its state space, nor does the microstructure of the system provide an obvious and uncontroversial answer to the question of which individuation we should choose. There is no clearly best way to go about reindividuating the world.

By comparing this situation to the one we found ourselves in when considering the box of gas, we can start to see why modeling complex systems is so difficult. In the case of the gas, there are a relatively small number of ways to individuate the system such that the state space we end up with is dynamically interesting. In the case of the global climate, there are a *tremendous* number of potentially interesting individuations, each associated with its own collection of models. The two problems are not mere difference of degree; they are difference in kind, and must be approached with that in mind. This may involve rather large changes in the way we think about the practice of science.

3. Conclusion

Frigg *et. al.* (2014a) write that "by assumption, the demon can compute the unvarnished truth about everything.[21]" This works well to illustrate their concern, but it is also not entirely

[21] Frigg *et. al.* (2014), p. 32

correct. In stating that Demon can compute the unvarnished truth about *everything*, Frigg *et. al.* tacitly assume that the only relevant predictions about the system's behavior to be had are those which are apparent given the system individuation that the Demon (along with his apprentices) have adopted: the fact that the Demon can predict with perfect precision precisely which state the system will end up in is taken to exhaust the interesting predictions to be had about the system. But this isn't necessarily the case--always knowing precisely which state the system will end up in exhausts the interesting predictions to be had about the system *given a particular single individuation.* This is a lot, but it certainly may not be *everything*: there may be interesting facts that can be discovered only by reindividuating the system, and those facts may be just as relevant--if not more relevant--to our practical decision making. The possibility of giving the Demon an extra power--macrostate omniscience--in which he has perfect knowledge of the best individuation simply does not arise, for there *is* no best individuation. Or, rather, there are very many best individuations, each reflecting a different set of priorities, values, and pragmatic choices.

None of this should be taken as an argument against the importance of considering SME as we evaluate scientific models. Frigg and Mayo-Wilson are both quite correct to argue that the problem of SME is distinct from the problem of prediction under Lorenzian chaos, and that in many cases it can be both more consequential and more difficult to remedy. If anything, the system individuation problem should be seen as an additional layer of difficulty that compounds the problem of SME. Both SME and the system individuation problem can be thought of as problems in the foundations of complexity science. Both worries stem from the observation that there may be some important scientific obstacles that cannot be overcome merely by making more precise measurements and constructing increasingly refined versions of the same models. These problems--and others like them--are likely to become increasingly salient as science pushes toward studying and modeling ever more complex systems. Keeping our eye on the system individuation problem serves as a powerful reminder of the difficulty inherent in attempting to engineer the future of complex systems, and also sheds some light on *why* that task is as difficult as it is. This underscores the importance of continued inquiry into the details of complex adaptive systems, and how scientific practice must be tailored to deal with complexity.

Finally, this discussion can be taken as a call for closer collaboration between scientists studying the behavior of complex systems like the global climate, and those with a stake in the decisions that such scientific study must eventually inform; in the case of climate science, that is a very big category indeed. Our practical decisions can and should be informed by our best contemporary science, but it is important to recognize that our best contemporary science--not to mention philosophy of science--is also informed by those decisions and the values they reflect. The more explicit this can be made, the more likely we are to make good decisions and to do good science.

.

Works Cited

Alligood, K. T., Sauer, T. D., & Yorke, J. A. (2000). Chaos: An Introduction to Dynamical Systems (Textbooks in Mathematical Sciences) (Corrected edition). Springer.

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, *14*(2), 171–185.

Bar-Yam, Y. (2004). Multiscale Complexity/Entropy. *Advances in Complex Systems* 7(1).

Biddle, J., & Winsberg, E. (2010). Value Judgements and the Estimation of Uncertainty in Climate Modeling. In P. D. Magnus & J. Busch (Eds.), *New Waves in Philosophy of Science* (pp. 172–197). Palgrave Macmillan.

Castellani, E. (2002). Reductionism, emergence, and effective field theories. *Studies in History and Philosophy of Science. Part B. Studies in History and Philosophy of Modern Physics* 33(2), 251-267.

Cumming, G. S., & Collier, J. (2005). Change and identity in complex systems. *Ecology and Society*, 10(1), 29

Frigg, R., Bradley, S., Du, H., & Smith, L. A. (2014a). Laplace's Demon and the Adventures of His Apprentices. *Philosophy of Science*, *81*(1), 31–59.

Frigg, R., Bradley, S., Du, H., & Smith, L. A. (2014b). Model error and ensemble forecasting: a cautionary tale. In G. C. Guo & C. Liu (Eds.), *Scientific Explanation and Methodology of Science: Proceedings of Sems 2012, Taiyuan, China* (pp. 58–68). Singapore: World Scientific.

Frigg, R., Smith, L. A., & Stainforth, D. A. (2013). The Myopia of Imperfect Climate Models: The Case of UKCP09. *Philosophy of Science*, *80*(5), 886–897.

Hooker, C. A. (ed.) (2011a). *Philosophy of Complex Systems*. Elsevier Science.

Hooker, C. A. (2011b). Introduction to philosophy of complex systems. *Philosophy of Complex Systems*.

Gell-Mann, M., & Lloyd, S. (1996). Information measures, effective complexity, and total information. *Complexity*, *2*(1), 44–52.

Hartmann, S. (2001). Effective Field Theories, Reductionism and Scientific Explanation. *Studies in History and Philosophy of Science. Part B. Studies in History and Philosophy of Modern Physics*, *32*(2), 267–304.

Kim, Jaegwon (1992). 'Downward causation' in emergentism and nonreductive physicalism. *Emergence or reduction* 119-138.

Kim, Jaegwon (2003). Blocking Causal Drainage and Other Maintenance Chores with Mental Causation." *Philosophy and Phenomenological Research* 67.1, 151-176.

Knutti, R. (2008). Should we believe model predictions of future climate change? *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *366*(1885), 4647–4664.

Kolmogorov, A. (1963). Sankhyā: The Indian Journal of Statistics, 25(4), 369-376.

Lawhead, J. (2014). Lightning in a Bottle. Columbia University Academic Commons. http://dx.doi.org/10.7916/D8348HW7.

Lenhard, J., & Winsberg, E. (2011). Holism and Entrenchment in Climate Model Validation. In *Science in the Context of Application* (pp. 115–130). Springer Netherlands.

Levy, A., & Bechtel, W. (2013). Abstraction and the Organization of Mechanisms. *Philosophy of Science*, *80*(2), 241–261.

Lloyd, E. A. (2010). Confirmation and Robustness of Climate Models. *Philosophy of Science*, *77*(5), 971–984.

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, *20*(2), 130–141.

Mayo-Wilson, C. (forthcoming). Structural Chaos. *Philosophy of Science*

McAllister, J. W. (2003). Effective Complexity as a Measure of Information Content. *Philosophy of Science*, *70*(2), 302–307.

Mossio, M., Bich, L., & Moreno, A. (2013). Emergence, Closure and Inter-level Causation in Biological Systems. *Erkenntnis. An International Journal of Analytic Philosophy*, *78*(2), 153–178.

Mossio, M., & Moreno, A. (2010). Organisational closure in biological organisms. *History and Philosophy of the Life Sciences*, *32*(2-3), 269–288.

Mossio, M., Saborido, C., & Moreno, A. (2009). An Organizational Account of Biological Functions. *The British Journal for the Philosophy of Science*, *60*(4), 813–841.

Parker, W S. (2006). Understanding Pluralism in Climate Modeling. *Foundations of Science*, *11*(4), 349–368.

Parker, Wendy S. (2010). Whose Probabilities? Predicting Climate Change with Ensembles of

Models. *Philosophy of Science*, *77*(5), 985–997.
Putnam, H. (1975). Philosophy and Our Mental Life. In H. Putnam (Ed.), *Mind, Language and Reality* (Vol. 2, pp. 361–362). Cambridge University Press.
Saborido, C., Mossio, M., & Moreno, A. (2011). Biological Organization and Cross-Generation Functions. *The British Journal for the Philosophy of Science*, *62*(3), 583–606.
Sales-Pardo, M., Guimerà, R., Moreira, A. A., & Amaral, L. A. N. (2007). Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(39), 15224–15229.
Saltzman, B. (1962). Finite amplitude free convection as an initial value problem-I. *Journal of the Atmospheric Sciences*, 19, 329-341.
Smith, L. (2007). Chaos: A Very Short Introduction (1 edition). Oxford University Press.
Stern, N. (Ed.). (2007). *The Economics of Climate Change: the Stern Review*. Cambridge University press.
Strogatz, S. H. (2001). Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering (Studies in Nonlinearity) (1 edition). Westview Press.
Thompson, E. L., & Smith, L. A. (2013). Consequences of the Hawkmoth Effect: Explicit subjective judgements about uncertain model-system relationships improve policy relevance of climate model output. *AGU Fall Meeting Abstracts*, *-1*, 1017.
Tol & Yohe (2006). "A Review of the Stern Review".*World Economics* 7(4): 233-50.
Winsberg, Eric. (2003). Simulated Experiments: Methodology for a Virtual World. *Philosophy of Science*, *70*(1), 105–125.
Winsberg, Eric. (2006). Models of Success Versus the Success of Models: Reliability without Truth. *Synthese*, *152*(1), 1–19.
Winsberg, Eric. (2010). *Science in the Age of Computer Simulation*. University of Chicago Press.
Winsberg, Eric. (2012). Values and uncertainties in the predictions of global climate models. *Kennedy Institute of Ethics Journal*, *22*(2), 111–137.