An antidote for hawkmoths: a response to recent climate-skeptical arguments grounded in the topology of dynamical systems

Lukas Nabergall, Alejandro Navas, and Eric Winsberg

June 22, 2017

Abstract

In a series of recent papers, two of which appeared in this journal, a group of philosophers, physicists, and climate scientists have argued that something they call the 'hawkmoth effect' poses insurmountable difficulties for those who would use non-linear models, including climate simulation models, to make quantitative predictions or to produce 'decision-relevant probabilites.' Such a claim, if it were true, would undermine much of climate science, among other things. Here, we examine the two lines of argument the group has used to support their claims. The first comes from a set of results in dynamical systems theory associated with the concept of 'structural stability.' The second relies on a mathematical demonstration of their own, using the logistic equation, that they present using a hypothetical scenario involving two apprentices of Laplace's omniscient demon. We prove two theorems that are relevant to their claims, and conclude that both of these lines of argument fail. There is nothing out there that comes close to matching the characteristics this group attributes to the 'hawkmoth effect.'

Contents

1 Introduction

In a series of recent papers[1], two of which appeared in this journal[2], Roman Frigg, Leonard Smith, Erica Thompson, and others[3], warned of the skeptical implications of what they have termed "the hawkmoth effect." Nowhere is the hawkmoth effect precisely defined, but whatever it is, it is a very very bad thing. It is a "poison pill" that "pulls the rug from underneath many modeling endeavors". It undermines the quantitative predictive power of almost all non-linear models and makes them incapable of producing "decision-relevant predictions" and "decision-relevant probabilities".

The LSE group has been primarily concerned with climate science, and in particular with the use of climate models to produce probabilities of future climate outcomes from initial conditions. Their conclusions are highly skeptical, arguing that the only trustworthy source of quantitative knowledge concerning the climate system comes from non-dynamical equilibrium models. Importantly, Winsberg and Goodwin (2015) have argued for several ways in which a 'hawkmoth effect,' even were it to exist, would not pose an epistemological threat to climate projections. In particular, they have pointed out that climate projections are meant to be independent of any knowledge of initial conditions. They are instead intended to determine what changes will befall the statistical averages of system properties under various forcing scenarios. This is unlike weather forecasting, which does depend on the initial conditions.

Here, we make the complementary point: there is no phenomenon, mechanism, or effect in the neighborhood of what the LSE group call a 'hawkmoth effect' that poses a threat to initial-condition-dependent predictions such as the predictions made in weather forecasting. We will argue that there are two ways in which the LSE group try to motivate a fear of the hawkmoth effect: the first by appealing to a family of mathematical results having to do with "structural stability," and the second by demonstrating an illustrative example. We argue that both of these attempts to motivate involve misunderstandings and that they fail to reveal what the LSE group claims they reveal.

1. (Smith, 2002), (Frigg et al., 2013a), (Frigg et al., 2013b), (Bradley et al., 2014), (Frigg et al., 2014), (Thompson, 2013).

2. By far the most detailed and prominent of these papers, "Laplace's demon and the adventures of his apprentices" (Frigg et al, 2014), was published in *Philosophy of Science*.

3. Hereafter referred to as the "LSE group".

2 What do the LSE group claim and how do they argue for it?

Before we proceed, we should give the reader a bit of background on the nature of the LSE group's claims, with an eye to answering two questions. First, just how skeptical are the worries that the LSE group are motivating? And second, what arguments do they use to motivate those worries? With respect to the second question, how important to their arguments are the "hawkmoth effect" and their claims about structural stability?

A. How skeptical are their conclusions and worries?[4]

In some places, the LSE group members suggest that the quantitative predictive power of all non-linear models is threatened by their arguments. If this is their intended scope, then not only would the most basic results of contemporary climate science—that the climate is changing as a result of human activity and will continue to do so—be cast under suspicion, but so too would most scientific modeling endeavors. They claim to have established that the combination of non-linear mathematical models with structural model error is a "poison pill" that "pulls the rug from underneath many modeling endeavors" (Frigg et. al., 2013a, p. 479). Since most mathematical models of interest in science are non-linear, and few of them can be expected to be free from all structural model error, it is supposed to follow from their argument, interpreted broadly, that any "probabilities for future events to occur" or "probabilistic forecasts" (Frigg et. al., 213a, p. 479) derived from such models cannot be trusted. Still, all is not lost, because the authors are willing to concede that, "not all the models underlying these forecasts are useless" (Frigg et. al., 2014, p. 57). This is because it is possible for a model that has been shown to be "maladaptive" for "quantitative prediction" (which is presumably what their argument establishes) to be "an informative aid to understanding phenomena and processes" (Frigg et. al., 2014, p. 48). In other words, mathematical models can be qualitatively informative in spite of the fact that they are not quantitatively trustworthy.

It is true that that in many places they are more circumspect, merely urging the use of caution in interpreting the "high resolution predictions out to the end of the century" (Frigg et. al. 2013b, p. 886) regarding the climate generated by one particular study. But there are two important points to be made here. The first, of course, is that merely because they sometimes make the explicitly weaker claim, they do not therein cancel the stronger claims. But the more important point is that even if the the weaker claims did cancel the stronger ones, these would be empty assurances. Their arguments do not have the finesse to establish the weaker claims without establishing the stronger ones. The arguments are not nuanced enough to allow for us to differentiate the high resolution from the low resolution, or the end of the decade from the end of the century from the end of the millennium. We will see that this is so in due course.

To be fair, we should also remind our readers that in most of their papers, the group offer assurances that their views are not climate skeptical. It is clear they do not want to be *seen* as climate skeptics. Appended to the very paragraph within which they consider whether or not "science [is] embroiled in confusion," they include a footnote with the reassuring claim that this, "casts no doubt on the reality or risks of anthropogenic climate change, for which

4. Most of this subsection follows (Winsberg and Goodwin, 2015). All the quotations in the subsection appear there as well.

there is evidence from both basic physical science and observations" (Frigg et. al., 2014, p. 48). There are two extremely important points to make here. The first is that whatever they say in the footnote, the consequences of their strongest claims, and of the arguments they offer (were they to be sound) *are what they are*. The second is that, even when they are trying to be reassuring, they do so in a way that continues to undermine the methods and practices that are in fact crucial to climate science. The assurance they offer is that we can find our evidence for "the reality or risks of anthropogenic climate change" in "basic [i.e. non-inclusive of non-linear-modeling] physical science and observations."

Would that we could be so sanguine. According to the IPCC, establishing the reality of anthropogenic climate change requires both *detecting and attributing* climate change. Detecting a change in the climate, based on observations (of roughly the weather), requires determining that "the likelihood of occurrence by chance due to internal variability alone ... is small" (Bindoff et. al., 2014, 872). This, in turn, requires an estimate of internal variability, generally derived from a "physically based model" (Bindoff et. al., 2014, 873). Furthermore, going on to attribute the detected change to a specific cause (such as human activity) typically involves showing that the observations are, "consistent with results from a process-based model that includes the causal factor in question, and inconsistent with an alternate, otherwise identical, model that excludes this factor"[5] (Bindoff et. al., 2014, 872). Indeed the authors of the IPCC report are quite clear that, "attribution is impossible without a model" (Bindoff et. al., 2014, 874). And the reasons for this are ones that should be quite familiar to philosophers of science: establishing or evaluating causal claims requires deciding how a system would have been different had things been otherwise; furthermore, in complex systems where multiple causal factors are at play, there is no 'basic physical science' that is capable of answering these modal questions. As the IPCC authors put it:

> We cannot observe a world in which either anthropogenic or natural forcings are absent, so some kind of model is needed to set up and evaluate quantitative hypotheses: to provide estimates of how we would expect such a world to behave and to respond to anthropogenic and natural forcings. (Bindoff et. al., 2014, 873).

Even if one takes the view of the IPCC to be controversial (full disclosure: we do not, and we would contradict the IPCC on a point like this only with the greatest caution), and one thinks as Frigg et al say, that "there is evidence from both basic physical science and observations," for the reality of anthropogenic climate change, it does not follow that undercutting model-based evidence—the only evidence that exists for identifying the relative strength of contributors to current changes in climate—"casts no doubt on the reality or risks of anthropogenic climate change." Undercutting some of the evidence obviously casts some doubt. Their arguments, in other words, are quite provocative. They don't just undermine prediction about regional climate in a hundred years, they undermine the most basic conclusions of contemporary climate science, at least as those conclusions are now established.

B. How important are results having to do with structural stability to their arguments? Don't they have other, stronger arguments to rely on?

5. This is what is famously known as 'fingerprinting'

Needless to say, therefore, it behooves anyone interested in securing the epistemic foundations of many, if not all, of the core claims of the IPCC from the skepticism in the LSE group's papers to thoroughly rebut their arguments. Their mere assurances that the IPCC's main claims are safe would not be enough if in fact their arguments were sound. So what are their arguments? We claim that they have two, and exactly two, lines of argument. Both involve the alleged existence of a "hawkmoth effect". The first line involves exhibiting a purported example of the hawkmoth effect, and arguing, by analogy, that climate models are likely to behave in a similar way. But they admit this is a weak argument. The second involves claiming that the generality of the effect they describe is secured by a set of mathematical results from the topology of dynamical systems related to the notion of "*structural stability*."

One might think that the importance of the structural stability arguments to their overall project could be downplayed and treated as a distraction.[6] Our sole interest here is securing the epistemic foundations of climate science and other instances of non-linear modeling—not in claiming any mathematical novelties.[7] So we should give some textual evidence for our claim that they have two, and exactly two, lines of argument:

1. An argument by analogy from an example that involves the logistic equation and a permutation of it (the "demon example"), and

2. an "analysis" of the mathematical literature on structural stability.

If we look at the structure of, for example, "Laplace's Demon and the Adventures of his Apprentices,"[8] what we find is this:

1. an illustration of the effect, using the demon example,

    - an admission that the illustration is not, by itself, very persuasive,

2. and finally the claim that the *real* argument is via mathematical results related to structural stability.

Textual evidence for this, from an article in this journal, is unequivocal:

> An **obvious line of criticism** would be to argue that the problems we describe are specific to the logistic map and do not occur in other systems. So **the question is, how general are the effects** we have discussed in the last section? **To answer this question we review a number of mathematical results about the structural stability** of dynamical systems. Our conclusion

6. Indeed an anonymous referee expressed the worry thusly: "I take the stuff the LSE group say about "structural stability" to be a bit of a distraction, and [the authors of this paper] show pretty comprehensively that they shouldn't have argued that way anyway."

7. One could certainly object that nothing we say in this paper is mathematically interesting or novel. We do not disagree. But given that the LSE group is relying on what they claim are mathematical results, it is important to bring clarity regarding them to the philosophical community. We are solely interested in the foundations of climate science and non-linear modeling generally.

8. We focus on this paper both because it is the most detailed and because it appeared in this jounal. There is nothing relevantly different to find in any of the other papers.

will be sober. There are special cases in which the above effects do not occur, but in general there are no such assurances. Not only are there no general stability results; there are in fact mathematical considerations suggesting that the effects we describe are generic. (Frigg et al, 2014, p.45, emphasis added).

Lest anyone think we are putting too much emphasis on one passage, we can assure the reader that many public discussions with the authors have involved them emphasizing that the structural stability results are the foundation of their arguments. They insist that anything that might be mistaken for having been an argument that they gave, if it wasn't based on the structural stability results, was in fact only intended to be an illustrative example—not an actual argument.[9]

3 What is a hawkmoth effect supposed to be?

So let us begin in earnest. The main argumentative thrust of all the papers is that the absence of structural stability displays a number of inter-related features: it gives rise to a "hawkmoth effect" whose power is exhibited in the demon example; it is underwritten by results in the topology of dynamical systems; and it is a close cousin of the butterfly effect. Indeed, nothing captures their intuitive idea of a hawkmoth effect better than the Miller analogy:

Butterfly Effect : Initial Conditions :: Hawkmoth Effect : Model Structure

The more technical term for the "butterfly effect," of course, is "sensitive dependence on initial conditions" (SDIC). The definition of SDIC can be formulated in a variety of ways, but two in particular bring out especially salient features. A third definition picks out a related property that is in the family of properties associated with chaotic systems: topological mixing.

Definition 1. *For a state space X with metric d, say that the behavior of a dynamical system $(\mathbb{R}, X, \phi)$ with time-evolution function $\phi : \mathbb{R} \times X \to X$ is sensitive to initial conditions to degree Δ if for every state $x \in X$ and every arbitrarily small distance $\delta > 0$, there exists a state y within distance δ of x and a time t such that $d(\phi(t, x), \phi(t, y)) > \Delta$.*[10]

9. See, for example, the question and answer period that occurs at the end of this lecture, during which Frigg and Smith both press this point: https://www.youtube.com/watch?v=qE7wpZ6t6Ts.

10. We take this definition almost verbatim from (Werndl, 2009) and (Mayo-Wilson, 2015) which also appeared in this journal. One small difference is that we have moved from a definition that applies to maps to one that applies to flows. (Crudely, a map is a function that we iterate to find a system's trajectory and a flow tells us what happens after a real-numbered value of time. The difference is discussed more formally below.) Both definition 1 and 2 can be converted from a flow-based definition to a map-based definition with ease. It is worth pointing out that talking about SDIC to degree "Δ" is very weak since it says nothing at all about how fast you need to get there and since it demands only that some state y near x, rather than that almost all states y near x, have the property. We use it for two reasons. One is that we are continuing a conversation in the pages of this journal that begins with (Frigg et al, 2014) and continues with a response to them from Mayo-Wilson. The second is that a maximally weak notion of SDIC is maximally favorable to the LSE group since it sets the bar maximally low for them.

Informally this says that a system exhibits weak sensitivity to initial conditions if: no matter the true initial state x, there is an arbitrarily close state y such that, if y had been the initial state, the future would have been radically different (to the degree Δ). We could also strengthen the definition of sensitivity to initial conditions to require that almost all such states have this property.

Definition 2. *For a state space X with metric d, say that the behavior of a dynamical system $(\mathbb{R}, X, \phi)$ is exponentially sensitive to initial conditions if there exists $\lambda > 0$ such that for any state $x \in X$, any $\delta > 0$, and all times t, almost all elements $y \in X$ satisfying $0 < d(x, y) < \delta$ are such that $d(\phi(t, x), \phi(t, y)) > e^{\lambda t} d(x, y)$.*[11]

Intuitively, not only does exponential SDIC allow you get to arbitrarily far away (subject to the boundedness of your dynamics) from where you would have gone by changing your initial state just a very small amount, but this definition requires that you be able to get there very fast. More precisely, it says that there will be exponential growth in error—that every $1/\lambda$ units of time (called the "Lyapunov time") the distance between the trajectories picked out by the two close-by initial states will increase by a factor of e. We assume that exponential SDIC is "strong," that it also requires that almost all the points near x take you there, not just one. Strong, exponential SDIC is what people usually have in mind when they talk about the butterfly effect.

Definition 3. *A time-evolution function ϕ is called topologically mixing if for any pair of non-empty open sets U and V, there exists a time $T > 0$ such that for all $t > T$, $\phi(\{t\} \times U) \cap V \neq \emptyset$.*

Informally, topological mixing (a crucial ingredient of chaos) occurs if, no matter how arbitrarily close I start, I will eventually be driven anywhere in the state space that I like.

It stands to reason, then, that the hawkmoth effect, if it exists, should involve a cluster of properties of dynamical systems that parallels the above three, but where the notion of two states being close is replaced with the notion of two *equations of evolution* being close. We will call making this kind of replacement a "lepidopteric permutation" of a definition associated with the butterfly effect because the hawkmoth and the butterfly are both Lepidoptera.

Definition 3 is the easiest of the three on which to perform the lepidopteric permutation, because it is topological, and topology is the natural domain in which to study families of dynamical systems (whose members do not have a natural metric between them). In

11. Note that this definition talks of "almost all" states, without there being specific mention of a measure. This is fairly standard; the reader is free to interpret them as either conditional on a specified metric or, as we more naturally intend it, as presupposing the Lebesgue measure, a standard practice in discussions of the state space of classical systems. Of course this is a significant problem, as we will see, for the LSE group, since there is no similarly natural measure on the space of evolution functions. Finally, just as the reader can easily convert either definition back and forth between a map and a flow, she can also convert them back and forth from being what is sometimes called a "strong" version (where the claim is about almost all nearby states) and a "weak" version (where the claim is about at least one nearby state). We have chosen to follow Mayo-Wilson and Werndl in giving Definition 1 a "weak" form but we have given definition 2 a "strong" form. Only strong versions of such definitions, obviously, require a measure, but only strong versions are usually taken to have strong epistemological consequences, since they are *likely* to produce error.

fact, Conor Mayo-Wilson[12] has come up with a reasonable candidate for a hawkmothified version of definition 3. In parallel with the notion of topological mixing, he calls it structural mixing. We do something similar by taking definition 3 and replacing the neighborhood of initial states with a neighborhood of evolution functions:

Definition 3*. *Let Φ be a space of time-evolution functions with metric δ, $\phi \in \Phi$, U be a non-empty open subset of X, and $\epsilon > 0$. Furthermore, for any time t, set $B_\epsilon(\phi)(t, x) = \{\phi'(t, x) \mid \delta(\phi, \phi') < \epsilon\}$. We say that Φ is structurally mixing at ϕ if for any state $x \in X$, there is a time T such that for all $t > T$, $B_\epsilon(\phi)(t, x) \cap U \neq \emptyset$.*

This definition uses a metric to pick out the preferred topology on the space of evolution functions, but it makes more sense to relax it to an arbitrary topology.

Definition 3.** *Let Φ be a space of time-evolution functions, $\phi \in \Phi$, U be a non-empty open subset of X, and V be a non-empty open subset of Φ (in the appropriate topology) containing ϕ. Furthermore, for any time t, set $V(t, x) = \{\phi'(t, x) \mid \phi' \in V\}$. We say that Φ is structurally mixing at ϕ if for any state $x \in X$, there is a time T such that for all $t > T$, $V(t, x) \cap U \neq \emptyset$.*

Definition 1 can also be finessed into a structural equivalent. Now we absolutely need a metric of distance between two evolution-specifying functions. We are of course free to pick one, but it is worth keeping in mind that in many cases there will be no natural choice. We would get a notion of "sensitive dependence on model structure to degree Δ" and define it roughly as follows (We use the "strong" version of Definition 1 since that is the epistemologically interesting version of SDIC):

Definition 1*. *Let Φ be a space of time-evolution functions with metric δ, $\phi \in \Phi$, and $\epsilon > 0$. We say that Φ is sensitively dependent on model structure to degree Δ at ϕ if for any state $x \in X$, there is a time t such that for almost all $\phi' \in \Phi$ satisfying $\delta(\phi, \phi') < \epsilon$, we have $d(\phi'(t, x), \phi(t, x)) > \Delta$.*

The reader can easily construct a weak version of this for herself.

Definition 2 is a definition on which it is very hard to see how to perform a lepidopteric transformation, because definition 2 gives a requirement in terms of an exponential growth in a single quantity: the distance of separation between nearby states, as the system evolves in time. We are already brushing under the rug the fact that 1* and 3* are assuming the existence of a metric of distance between evolution functions. But in a hawkmoth version of definition 2, unlike in the butterfly version, there is no single quantity that can grow in time. A hawkmoth version would have to relate model distance with state space distance. It would have to coordinate a metric of model structure distance with a metric of state space distance. This will turn out, we will see, to actually be even more challenging than it seems.

That the two above definitions should at least be considered to be in the right neighborhood of what a hawkmoth effect should involve is well motivated by the Miller analogy

12. (Mayo-Wilson, 2015) Our definition 3* is similar to and inspired by his attempt to capture one aspect of what the LSE group might mean by the hawkmoth effect, but we also emphasize that topological mixing is only one aspect of chaos. It happens not to be the feature of chaos, moreover, that is usually associated with the butterfly effect. And finally, in so far as one is looking for the lepidopteric analog of topological mixing, which is a purely topological notion, definition 3** makes the most sense, since it is also topological.

argument. But that is not the only reason for thinking what we have above captures the spirit of what the LSE group are keen to convince us that the hawkmoth effect can do.

We can also look at some of the claims that they make about the epistemological consequences of a hawkmoth effect. They claim that 1) the hawkmoth effect ruins the closeness-to-goodness rule, which suggests that a close-by model is a reasonably good model for producing forecasts, 2) the hawkmoth effect causes a fast growth in entropy, and 3) the hawkmoth effect is an underappreciated cousin of the butterfly effect with similar epistemological consequences. And they show pictures like figure 1. *Finally, and most importantly, if the hawkmoth effect does not have this cluster of properties, it is hard to see why anyone should think that the demon case is a reasonable illustration of it, or that it poses the kinds of epistemological challenges that the LSE group claims it does.*

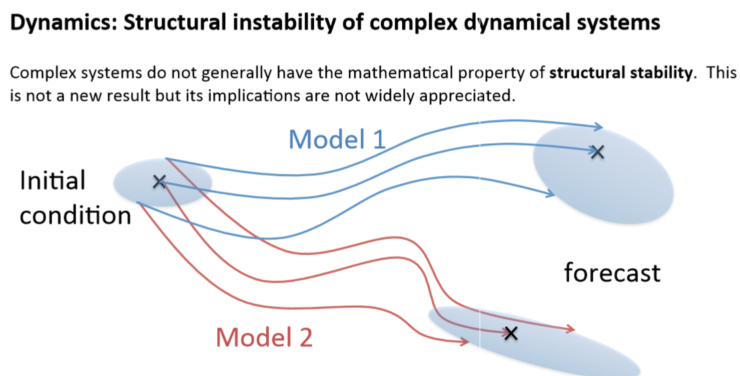

Figure 1: A visual representation of the hawkmoth effect, or lack of structural stability, from Thompson and Smith.[14]

It should be reasonably clear that figure 1 is a good depiction of what we have called sensitive dependence on model structure to degree Δ, as defined in 1*, and vice versa—if Δ represents roughly the distance between the two blue sets on the right. All of the above paint a picture in which we can imagine the following dialog taking place:

Peter is annoyed with the Weather Channel because their forecasts are never accurate beyond 10 days into the future. He goes to a public outreach meeting and demands that the Weather Channel produce longer term forecasts of greater accuracy.

Jessica the meteorologist explains that the butterfly effect makes this impossible. She explains that the Lyapunov time of a planetary atmosphere like our own is about 3 days. This means that any errors there are in our knowledge of the present conditions of the atmosphere will, after 3 days, have grown to about 3 times their present size. After 6 days, to 9 times, after 9 days to 27 times, and after 12 days, to 81 times the present degree of error. She explains that once initial condition uncertainty has grown that large, basic qualitative conclusions we try to draw from the data like the probability of rain have little value.

Everything the LSE group has written about the hawkmoth effect suggest that if a representative of their group were at the meeting, she would raise her hand and try to add

14. Source: http://www.lse.ac.uk/CATS/Talks and Presentations/Posters/Thompson-TheHawkmothEffect-LSEResearchFestival2014.pdf

the following claim: “Jessica, everything you say about the butterfly effect is true, but you have neglected an equally important part of the picture. Just as your knowledge of the present conditions is imperfect, so is your knowledge of the best model of the climate. And just as a small error in the initial conditions will blow up to a forecast-destroying level after 12 days, so will a small error in your knowledge of the correct model produce similar consequences—perhaps even on similar time scales.”

All of this is further reason to think that if a hawkmoth effect exists, it has to involve something like a cluster of properties that are the analogs of the 3 properties, above, that we associated with the butterfly effect and with chaos. We note, however, that we are already in trouble because a relevant analog of property 2 is lacking.

4 Two routes to a hawkmoth effect

So much for what a hawkmoth effect should be like. What reason is there for thinking that there is such a thing? More precisely, what reason do these authors have for thinking that very many non-linear systems, including, importantly, the atmosphere-ocean-earth system, are best modeled by dynamical systems that exhibit it? The answer to this question is clearly given in the passage from the “demon” paper that we quoted at the end of section 2. The LSE group try to motivate the worry that the hawkmoth effect affects many such systems in two different ways.

The first argument comes in the form of an appeal to a variety of previously known mathematical results associated with the phenomenon of structural stability. And the second is their (in)famous demon's apprentice example. As we have already remarked, they usually insist that the first argument for the hawkmoth effect is the only argument they mean to offer, while the second is only meant as an illustrative example of something whose existence is established by the first argument. Still, we should look at each one carefully. After all, even if the first argument fails to show that a large class of systems should be expected to be hawkmothish, the demon's apprentice example might show that hawkmoth behavior is still a danger to be reckoned with.

5 The structural stability route

The closest the LSE group come to giving a definition of the hawkmoth effect is in (Thompson, 2013). Indeed, in what might be considered the flagship hawkmoth paper (published in this journal) Frigg et al (2014) what passes for a definition is: “Thompson (2013) introduced this term in analogy to the butterfly effect. The term also emphasizes that SME yields a worse epistemic position than SDIC: hawkmoths are better camouflaged and less photogenic than butterflies.” (p.39)

Looking in (Thompson, 2013) we find the following:

> In this chapter I introduce a result from the theory of dynamical systems and demonstrate its relevance for climate science. I name this result, for ease of reference, the Hawkmoth Effect (by analogy with the Butterfly Effect). (p. 211)

and

> The term "Butterfly Effect" has greatly aided communication and understanding of the consequences of dynamical instability of complex systems. It arises from the title of a talk given by Edward Lorenz in 1972: "Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?".
>
> *I propose that the term "Hawkmoth Effect" should be used to refer to structural instability of complex systems.* The primary reason for proposing this term is to continue the lepidoptera theme with a lesser-known but common member of the order. The Hawkmoth is also appropriately camouflaged, and less photogenic. (Thompson, 2013, p. 213, emphasis added.)

Interestingly, the term "structural **in**stability" doesn't seem to appear very much in the mathematical literature. Nowhere is it described as producing an "effect." In fact, the discussion that we find in both (Thompson, 2013, "5.2.2 Identifying structurally stable systems", p. 214-215), and even more so (Frigg et al, 2014, "4. From Example to Generalization", p. 45-47) are clearly influenced by (Pugh and Peixoto, 2008)[15] and though the word "stability" appears 65 times in the above mentioned piece, the word "instability" does not appear even once. Nor does the word "unstable." If you search for it, you can find the occasional article with "structural instability" in the title, but the results one finds in them are always discussed in terms of the presence or absence of structural stability.

Why does this matter? It matters because structural stability does not have a complement with substantial features of its own, and the term "structural instability" suggests an overly close analogy to chaotic instability that no one in the mathematical literature ever had in mind. We can see why if we look closely at the notion of structural stability.

The first thing we should notice is that they are definitions of ways of guaranteeing to stay arbitrarily close. And failure to stay arbitrarily close is not the same thing as being guaranteed to go arbitrarily far. But in analogizing absence of structural stability to SDIC, the LSE group are engaging in exactly this conflation.

Take an early definition of structural stability in two dimensions due to Andronov and Pontrjagin as it is explained in Pugh and Peixoto. They consider dynamical systems of the form

$$\frac{dx}{dt} = P(x,y), \quad \frac{dy}{dt} = Q(x,y) \tag{1}$$

defined on the disc D^2 in the xy-plane, with the vector field (P, Q) entering transversally across the boundary ∂D^2.

Definition 4. *Let p and q be vector fields on D^2. We say that such a system is structurally stable (or "rough", as Andronov and Pontrjagin put it) if, given $\epsilon > 0$, there exists $\delta > 0$ such that whenever $p(x,y)$ and $q(x,y)$, together with their first derivatives, are less than δ*

15. Neither Thompson nor Frigg et al. list the Pugh and Peixoto as a work cited, but both of them follow Pugh and Peixoto's discussion very closely, as can easily be verified. We make this point only to assuage a possible concern on the part of the reader that we are not talking about exactly the same notion of structural stability as they are, or that they are appealing to results not covered by Pugh and Peixoto. We don't believe that any of the LSE group could object to us using Pugh and Peixoto as an authoritative reference, since the references in that piece are exactly the same ones we find in all the LSE group's papers (including two classic papers by Peixoto himself).

in absolute value, then the perturbed system

$$\frac{dx}{dt} = P(x,y) + p(x,y), \quad \frac{dy}{dt} = Q(x,y) + q(x,y) \tag{2}$$

is such that there exists an ϵ-homeomorphism $h : D^2 \to D^2$ (h moves each point in D^2 by less than ϵ) which transforms the trajectories of the original system to the trajectories of the perturbed system.

Intuitively, this definition says of a particular evolution function that no matter how ϵ-close to the trajectory of that evolution function I want to stay *for its entire history*, I can be guaranteed to find a δ such that all evolution functions within δ of my original one stay ϵ-close to the original trajectory—where δ is a measure of how small both the perturbing function and their first derivatives are. This is an incredibly stringent requirement.

When an evolution function fails to obtain such a feature, therefore, it is as perverse to call it "structurally unstable" as it is to talk about a system being insensitively dependent on initial conditions, or to describe a function that fails to have a particular limit as "unlimited." It is a perversion that conflates the following two sorts of claims.

1. You can't be guaranteed to stay arbitrarily close by choosing an evolution function that is within some small neighborhood.

2. Small changes in the evolution function are sure to take you arbitrarily far away.

Absence of structural stability in the sense of definition 4 gives you the first thing, but nothing anywhere near approximating the second thing. But to claim that absence of structural stability is an analog of SDIC is to suggest that absence of structural stability gives you the second thing. It is the second thing, moreover, that we concluded that a hawkmoth effect should ensure when we formulated definitions 1* and 3* above using the lepidopteric transformation of definitions 1 and 3. That, moreover, is just the first difference. Notice that 1*, in order to be the analog of strong SDIC, has to say "for almost any $\phi \in \Phi$". But absence of structural stability in the Andronov and Pontrjagin sense requires nothing of the sort. It only requires that, for each δ, one trajectory in the entire set δ-close trajectories be deformed by more than ϵ. (See figure 2.)

And notice, by the way, that it would be impossible to even formulate a useful definition of structural instability that required that something like "almost all" the nearby models diverge. That's because there is no natural measure over the models. The state space of a classical system has an obvious measure: the Lebesgue measure. So it is easy to say things like "almost all the nearby states have such and such property." But spaces of diffeomorphisms have no such measure.[16]

16. In reading some of their papers, and in conversation, one sometimes gets the impression that they think the result of Smale, which they summarize as follows "Smale, 1966, showed that structural stability is not generic in the class of diffeomorphisms on a manifold: the set of structurally stable systems is open but not dense," can stand in for a claim about the likelihood of a system being structurally stable. Or maybe of the likelihood of a nearby model being close if the first one is not structurally stable. This is the only thing we can find in the results they review about structural stability after they make the claim that "there are in fact mathematical considerations suggesting that the effects we describe are generic." (Frigg et al, 2014,

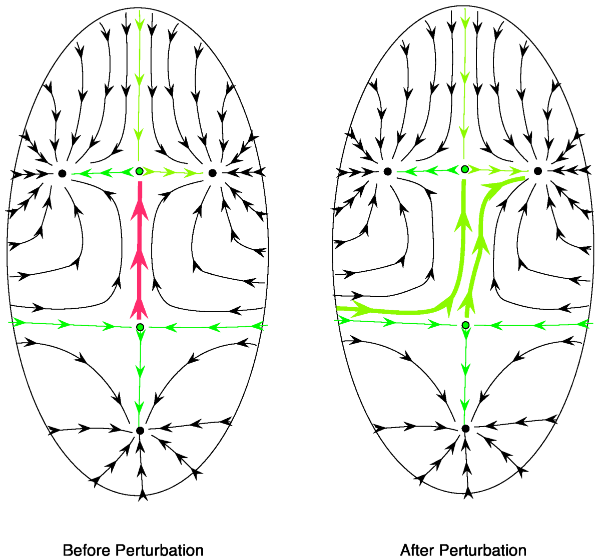

Figure 2: An example of a map that is not structurally stable. When a small perturbation turns the map on the left into the map on the right, we get two maps that cannot be smoothly deformed into each other. The key feature is this qualitative dissimilarity between the two maps, and not any metrical difference. (Source: Pugh and Peixoto, 2008)

This is what happens when you take the complement of a definition. Definition 4 is "strong." It requires that *not a single trajectory* diverge by more than ϵ. But when you take the complement of a strong definition you get a weak one (among other problems). Here, the complement of structural stability only requires that, for each closeness threshold δ, one trajectory go astray.

To put the point simply, absence of structural stability in the Andronov and Pontrjagin sense is much much weaker than either definitions 1* or 3*. It's weaker in two ways: rather than requiring that *most* trajectories (indeed as we have seen there is no coherent notion of "most" here) go *far* away, it only requires that *one* trajectory go *more than a very small epsilon* away. And hence it is much weaker than the LSE group or the hawkmoth analogy suggest.

We still haven't talked, moreover, about definition 2 of SDIC and the idea of exponential error growth, which is so fundamental to the epistemological impact of chaos. If we think back to the exchange between Peter, Jessica, and the representative of the LSE group, we realize that even more than 1* and 3* would be needed to underwrite the existence of a hawkmoth effect. We would need to formulate a definition that captured the idea that the small error in a model could grow very fast—indeed we would need something akin to

p. 57) But first, climate models are not likely to be diffeomorphisms, so that's not the relevant universality class. Second, and more importantly, density is not a measure-theoretic notion, it is a topological one. A set can be dense and have measure zero (think of the rationals in the real number line—the rationals are of course not generic in the real line). There are even nowhere-dense sets that have arbitrarily high measure in the reals. There is a general point here: all the relevant notions associated with structural stability are topological, and they provide no information about likelihoods, or genericness. This is again because there is no natural measure on the space of equations of evolution.

definition 2 that would allow us to calculate, as Jessica did with the butterfly effect, how soon a forecast would become useless given a certain amount of model structure uncertainty. But this looks unlikely for the case of failure of stability in the sense of Andonov and Pontrjagin. The reason is that in definition 4, the metric of model distance does not live in the same space as where the metric of state space distance lives. It would be strange and confusing to relate these two metrics in a single equation.

And things get even worse if we move from what Pugh and Peixoto call the "pre-history" of structural stability to its modern formulation. In the modern formulation, no metric is specified—neither between two different diffeomorphisms, nor between two trajectories.

The modern formulation of structural stability goes as follows:

Definition 5. *If D is the set of self-diffeomorphisms of a compact smooth manifold M, and D is equipped with the C^1 topology then $f \in D$ is structurally stable if and only if for each g in some neighborhood of f in D there is a homeomorphism $h : M \to M$ such that f is topologically conjugate to each nearby g. In other words, that*

$$\begin{array}{ccc} M & \xrightarrow{f} & M \\ h\big\downarrow & & \big\downarrow h \\ M & \xrightarrow{g} & M \end{array} \tag{3}$$

commutes, that is, $h(f(x)) = g(h(x))$ for all $x \in M$.

Definition 5 makes it clear that the most general formulation of it is not metrical at all. It is topological. It says nothing at all about diffeomorphisms that are "a small distance away". It talks about diffeomorphisms that are in topological neighborhoods of each other. And it doesn't talk at all about trajectories taking you some distance away. It talks about there being a homeomorphism (a topology preserving transformation that cares nothing about distances) between the two trajectories. Using the analogy of a rubber sheet that is often used to explain topological notions, it roughly says, intuitively, that if you replace f by any of the diffeomorphisms g in some neighborhood of f then the new entire statespace diagram of g will be one that could have been made just by stretching or unstretching (by deforming it in the way one can deform a rubber sheet without tearing it) the statespace diagram of f.

This is not, moreover, incidental—or a pointlessly abstract formalism. Some structural stability results have been achieved in which the relevant topological conditions are spelled out in terms of a topology that is provably not metrizable.[17] And so in fact, in some cases, structural stability tells you nothing at all, let alone anything strong enough, about how far away a slightly perturbed model will take you. It simply supplies no metrical information.

So let's review what it amounts to for a system defined by a diffeomorphic map f, applied to a manifold, to fail to be structurally stable. It means that if you look at the space of diffeomorphisms around f, you will be unable to find a neighborhood around f that is guaranteed not to contain a single other diffeomorphism g that is qualitatively different than f in essentially the following sense: that you cannot smoothly deform f into g. The Father Guido Sarducci version of what we have learned so far is this: *You don't get structural 'instability'*

17. Recall that this is possible because every metric picks out a topology, but the reverse is not true.

just by replacing "small changes in initial conditions" in SDIC with "small changes in model structure" because, both with regard to how much error you can get, and with respect to how many nearby trajectories will do it, SDIC says things are maximally bad, while structural instability merely says they will not be maximally good. And SDIC includes metrical claims, while its incoherent for structural stability to be given a metrical form.

Let us put this another way. Absence of structural stability is an incredibly weak condition on three dimensions: it need not take you far (the relevant notion is defined in the absence of any metric at all), it need not take you there at all fast, and there only needs to be one model in your entire neighborhood that does it. So absence of structural stability has no interesting consequences for the predictive power of a model, even in the presence of model structure uncertainty—so long as you are not interested in infinitely long predictions (the second dimension) that are topologically exact (the first dimension) and underwritten by mathematical certainty (the third dimension), rather than, say, overwhelming probability. Let that sink in: you could know for a fact that your model is structurally unstable, know for a fact that you had some small amount of stuctural model error, and still have it be the case that your model would not introduce more than an arbitrarily small amount of error for an arbitrarily long time. And it could still be overwhelmingly likely (no matter what measure you preferred on the model space) that it would introduce virtually no error at all.[18] Nor, by the way, do any of the results having to do with structural stability make any mention at all of non-linearity. Non-linearity is a red herring. There is no hawkmoth effect.

We should acknowledge that none of this is intrinsically newsworthy. But we believe it is worth clearing all of this up given some misleading claims that have been made in the philosophical literature in general and in this journal in particular: that the absence of structural stability is in any way analogous to a butterfly effect; that it (in anything like the normal cases) does something akin to what we see in figure 1; that it undermines the predictive capacity of nonlinear science; and that it undermines the capacity of the same to produce decision relevant probabilities. And that applies, inter alia, to the way in which that is done *both* in weather prediction and in climate projection. And as we will see, it is also misleading to suggest that the LSE group's famous demon's apprentice example in any way illustrates the typical effects of the absence of structural stability.

But what about the demon example itself? Doesn't it provide its own cautionary tale, irrespective of the epistemological import of structural stability considerations? Doesn't it do this given that figure 1 *does* seem to capture well what happens in the demon case? Doesn't it do this given that it seems to show that very nearby models can take a probability distribution over some relatively small set of initial conditions and very quickly drive that set into very different regions of state space? We turn to this question in the next section.

18. This is all of course because structural stability was a notion developed by people interested in achieving the mathematical certainty of proof while using perturbations—they were not interested in finding predictive accuracy. After all, they were studying the solar system. No one was worried that, until they found stability results for the solar system, its dynamical study would be embroiled in confusion or maladapted to quantitative prediction.

6 The demon's apprentice route

In (Frigg et al., 2014), the LSE group postulate the existence of a demon that is omniscient regarding the exact initial conditions of a given system, the true dynamical model of the system, and the computational output of such a model at any future time, for any initial conditions, to arbitrary precision. Such a demon also has two apprentices: a senior apprentice, who has omniscience of computation and dynamics yet lacks that of initial conditions, and a freshman apprentice, who has computational omniscience but not that of model structure nor initial conditions.

The problems of the senior apprentice can be overcome by Probabilistic Initial Condition Ensemble Forecasting (PICEF). In PICEF, instead of using a single point in the state space as the initial conditions for the dynamical system, we substitute in a probability distribution over the entire state space. In this way, the practical limitations of initial condition uncertainty can be mitigated: A point prediction for the state of the system in the future is replaced by a distribution over possible future states which may still inform practical considerations.

According to Frigg et al., however, there is no such solution for the freshman's ignorance. Aside from the initial condition uncertainties which reduce the precision of a system's trajectories, the freshman apprentice is beset by unreliability in the very dynamics by which initial states are evolved. This unreliability, they claim, is not easily resolved nor easily dismissed. And its consequences, they hold, are severe.

To illustrate this severity, et al. consider the logistic map, defined below, and a "similar" equation that represents the true model of some physical system. They show that, given enough time, the two equations evolve the same distribution of initial conditions to very different regions of state space. We've already seen that the absence of structural stability is not, in general, as severe as the results of the demon's apprentice example suggest. Absence of structural stability does not generally lead to wide divergence, nor does its absence imply anything about the majority of nearby models (see section 4). It only takes one stray model in the neighborhood to violate the definition. But we can still ask whether the demon example is at least *a* possible illustration of the absence of structural stability. And we can still ask if it provides a worthwhile cautionary tale of its own. The answers to both of these questions, alas, is "no."

Why is the demon example not a possible illustration of the absence of structural stability? To see why, we need to review some conceptual distinctions relevant to dynamical systems. We can start with the logistic equation, which is part of the demon example:

$$x_{t+1} = 4x_t(1 - x_t). \tag{4}$$

Notice that this is a dynamical system specified with an equation of evolution that lives in discrete time. This needs to be contrasted with dynamical systems specified with time-dependent differential equations, like the Lorenz model

$$\frac{dx}{dt} = \sigma(y - x), \quad \frac{dx}{dt} = x(\rho - z) - y, \quad \frac{dz}{dt} = xy - \beta z. \tag{5}$$

In the formalisms used to discuss structural stability, dynamical systems like the logistic equation are specified by maps: functions from a manifold onto itself; and those like the

Lorenz model by flows: a map from a manifold (state space) crossed with the real number line (a time variable) onto the manifold: $\phi : M \times \mathbb{R} \to M$.

Definition 5 gave us the definition of structural stability for a map, and the definition of structural stability for a flow is:

Definition 6. *If X is the set of smooth vector fields on a manifold M equipped with the C^1 topology, then the flow generated by $x \in X$ is structurally stable if and only if for each y in the neighborhood of x in X there is a homeomorphism $h : M \to M$ that sends the orbits of x to the orbits of y while preserving the orientation of the orbits.*

But notice that definition 5 does not apply to any old map. The definition only applies if the map is a diffeomorphism. To be a diffeomorphism, a map has to have some added conditions:

1. The map has to be differentiable.

2. The map has to be a bijection (its inverse must also be a function).

3. The inverse of the map has to be differentiable.

The obvious problem is that the logistic map is not a bijection! Every number other than 1 has two preimages. For example, both .3 and .7 map to .84. So .84 has no unique preimage and there is no function that is the inverse of equation 4. But this means that the logistic map isn't even the right category of object to be structurally stable or not.[19] Of course we are free to make up our own definition of structural stability that applies to all maps. But if we do, if we expand our model class beyond the space of diffeomorphisms to the space of all maps, then the very notion of structural stability becomes empty. You simply won't find many structurally stable maps on this definition. Consider the simplest map there is:

$$x_i = C \quad \text{for all } i \text{ and some constant } C. \tag{6}$$

This map is simple but, of course, not a bijection. Yet it also would not come out as "structurally stable" on our new definition. According to the definition we require all g in some neighborhood to satisfy $h(f(x)) = g(h(x))$. But if f is constant, $h(f(x))$ is constant, so $g(h(x))$ has to be constant for the definition to hold. But that condition is easy to violate with many of the g's in any neighborhood of f. But this means that the logistic map is no more "structurally unstable" than the function given by equation 6 is. Which

19. There are other nearby puzzles about what the LSE group could possibly have thought they were on about: the best model climate science could write down—that is the real true, partial-differential-equation-specified, undiscretized model—would have the form $\phi : M \times \mathbb{R} \to M$. It might or might not meet the additional criteria for being a flow, but it is certainly not of the form $\phi : M \to M$, which is the general form of a map. Once we start to think about a discretized model, however, the model does take the form $\phi : M \to M$. Even if we had the perfect climate model and it *were* a flow, a discretization of it (in time) would necessarily have the form of a map. And no map is in the right universality class—for purposes of structural stability—of a flow. In the sense relevant to structural stability, its simply a category error to ask if a discretization of a dynamical system of the form $\phi : M \times \mathbb{R} \to M$ is "nearby to" the undiscretized system. Climate models run on computers are all imperfect, but they don't live in the same universe of functions as the "true" model of the climate does.

means whatever the demon example illustrates, it actually has nothing at all to do with non-linearity. Technically, of course, equation 6 isn't linear. But its also not exactly what people have in mind when they think of non-linearity! And of course, once you open up the model class to non-diffeomorphisms, $f(x) = 3x$ (an obviously linear map) will also be structurally unstable.

So when Frigg et al (2014) write, "The relation between structural stability and the Demon scenario is obvious: if the original system is the true dynamics, then the true dynamics has to be structurally stable for the Freshman's close-by model to yield close-by results,"(p.47) they are saying something very misleading. In fact, the relation between structural stability and the demon is at best murky–because the logistic equation is not even a candidate for structural stability. And as we have seen, it is simply false that a model has to be structurally stable for a nearby model to produce nearby results. That is straightforwardly a misreading of the definition. And it is straightforwardly misleading to suggest that non-linearity is the culprit. If you open up the definition to include arbitrary maps, all kinds of incredibly simple maps become "structurally unstable." If you think absence of structural stability is the hawkmoth effect, and if you think the logistic equation (despite not being a diffeomorphism) displays the hawkmoth effect, then neccessarily you will have to say that $x_i = C$ displays the hawkmoth effect too. This is not a happy outcome.

Okay, but still, even if the logistic map is not a candidate for structural stability, surely the demon example still shows that two very nearby models can lead to radically different PICEF predictions, right? We saw in section 5 that structural stability and its absence did not underwrite what is depicted in figure 1. But surely the demon example does, right? This, after all, we can see with our own eyes in the Frigg et al (2014) paper. Not so fast. Let's look carefully at the two models in the example: the freshman apprentice's model and the demon's model.

$$x_{t+1} = 4x_t(1 - x_t) \tag{7}$$

$$\tilde{x}_{t+1} = (1 - \epsilon)4\tilde{x}_t(1 - \tilde{x}_t) + \frac{16\epsilon}{5}\left[\tilde{x}_t(1 - 2\tilde{x}_t^2 + \tilde{x}_t^3)\right] \tag{8}$$

Equation 8 is the demon's and senior apprentice's model, the "true" model in this scenario, and equation 7 is the freshman apprentice's model, the "approximate" model.

On first glance, these equations do not look very similar. But the LSE group argue that they are in fact similar. They argue this by arguing that the appropriate metric of similarity should be based on an output comparison of the two models over one timestep: Call the maximum difference that the two models can produce for any arbitrary input ϵ, the maximum one-step error. If ϵ is sufficiently small, then the two models can be said, according to the LSE group, to be very similar. Here, we argue that this is much too weak of a condition for considering two models similar. We also note, as we already did above, that it is not for nothing that the modern literature on structural stability is topological and not metrical. There just isn't anything sufficiently general and sufficiently natural to say about how to measure the distance between two models, two diffeomorphisms, or two flows. The model of equation 7 and the model of equation 8 are not appropriately similar for drawing the conclusions that the LSE group draw. We begin by proving a theorem (proofs of theorems can be found in the Appendix):

7 Small arbitrary model perturbations

Suppose we are given a difference equation of the form

$$x_{n+1} = f(x_n), \tag{9}$$

where $x_i \in \mathbb{R}$ and $f : A \to B$ is an arbitrary function from the bounded interval $A \subset \mathbb{R}$ into the bounded interval $B \subset \mathbb{R}$. Note that the logistic map takes this form with $f(x_n) = 4x_n(1 - x_n)$. Then we have the following result:

Theorem 1. *Given any function $g : A \to B$ and $\epsilon > 0$, there exists $\delta > 0$ and $\eta > 0$ such that the maximum one-step error of*

$$x'_{n+1} = \eta f(x'_n) + \delta g(x'_n), \tag{10}$$

from (13) is at most ϵ and $x'_{n+1} \in B$.

Observe that (8) takes the form of (10) with $f(x'_n) = 4x'_n(1 - x'_n)$, $g(x'_n) = (16/5)[x'_n(1 - 2{x'_n}^2 + {x'_n}^3)]$, $\eta = 1 - \epsilon$, and $\delta = \epsilon$. There are at least two ways in which this result undermines the claim that the demon's apprentice example demonstrates the existence of a hawkmoth effect which is an epistemological analog of the butterfly effect.

The existence of small arbitrary model perturbations demonstrated in the above theorem for first-order difference equations, of which the logistic map is an example, shows that the perturbation presented in the demon example is only one possible perturbation amongst the infinite space of admissible perturbations that are close to the logistic map under the maximum one-step error model metric. In fact, as the argument demonstrates, we can perturb our model in *any way* we wish and still remain as close to the initial model as desired. It should therefore be no surprise that we can find models close to the logistic map that generate trajectories in the state space vastly diverging from the logistic map over certain time intervals; indeed, we should expect to find nearby models that exhibit essentially any behavior we want, including some which vastly deviate from the logistic map across any given time interval and others which remain arbitrarily close to the logistic map for all future times. In particular, there is no a priori reason we should expect that the modified logistic map is an example of a commonly occurring small model error. The butterfly effect is so important because numerically small difference between the true value of a system variable and its measured value are absolutely common and normal. But what reason is there to think that climate scientists make mistakes about the order of the polynomials their models should have? Or that they sometimes write down exponential functions when they should have written down sinusoidal ones? Corollarily, what reason is there to think that small model errors of the kind we would expect to find in climate science, atmospheric science, and other domains of non-linear modeling will normally produce deviations on such short timescales as they do in the demon example? Why would we let such a weirdly concocted example do any "burden of proof shifting" of the kind the LSE group demand of us? Consequently, at the very least, the demon example only retains its relevance as evidence for the epistemic force of the so-called hawkmoth effect if it is accompanied by a strong argument showing how it is precisely this sort of perturbation which is often encountered when constructing weather forecasting models. But there may be good reasons to think this is not the case (see the next point as well as Section 10).

More generally, Theorem 1 indicates that the maximum one-step error metric is quite simply too easily satisfied and does not really get at what makes two models similar or close.[20] It would be difficult indeed to argue that the difference equations

$$x_{n+1} = 2x_n \quad \text{and} \quad x_{n+1} = 2x_n + .004e^{x_n} - .03\sin x_n$$

are highly similar and ought to be considered "close" in model space simply because after one time step they do not yet produce significantly diverging trajectories in state space—the newly added sinusoidal and exponential terms behave so differently from the linear term present in the original equation that we would certainly not want to call these two models "close". Furthermore, we would in particular definitely not expect to be able to predict well the long-term behavior of a physical system using both of these equations since the perturbations introduced in the second equation model entirely different physical dynamics.

8 Small polynomials on [0,1]

The astute reader might notice the following thing: in Frigg et al's demon example, they get a maximum one-step error between the freshman and senior apprentices of .005, and they do this with a value of ϵ of 0.1. Using the proof of theorem 1, you can calculate what value of ϵ the theorem guarantees will give you a maximum one-step that they achieve, (.005), and it is the relatively small number of .0025. But they achieve their result with a relatively large value of ϵ of .1.[21] What explains this? Is the measure of model closeness reasonable if we disallow overly small values of ϵ? The very small value of epsilon that our proof produces is sufficient, but doesn't seem to be necessary. Is it in fact us that is making a misleading case here?

No. Let $f(x)$ and $g(x)$ be the polynomials on $[0,1]$ underlying 7 and 8, respectively. Then we have

$$|f(x) - g(x)| = \left| 4x(1-x) - \left[(1-\epsilon)4x(1-x) + \frac{16\epsilon}{5}\left[x(1-2x^2+x^3)\right]\right]\right|$$

20. In addition to the evidence of theorem 1, we also offer the following anecdotal evidence that maximum one-step error is not a particularly robust measure of model closeness. It happens that in one of the many papers published by the LSE group on this topic, (Frigg et al., 2013(a)), they use an ever-so-slightly different version of the perturbation than they do in their other papers. In place of the function in equation 8 they instead used

$$\tilde{p}_{t+1} = 4\tilde{p}_t(1-\tilde{p}_t)\left[(1-\epsilon) + \frac{4}{5}\epsilon(\tilde{p}_t^2 - \tilde{p}_t + 1)\right] \tag{11}$$

What is interesting is that, like in (Frigg et al, 2014), they report in this paper that for this different perturbation, at $\epsilon = .1$, the maximum one-step error (relative to the standard logistic equation) is .005. But this is wrong. The small change in the equation makes the maximum one step error skyrocket to .04, for the same value of ϵ. We can think of no better anecdotal demonstration of how artificial the maximum one-step error is as a metric of model distance than the fact that the authors took themselves to be presenting the same perturbation twice, and it happened to differ on that metric by a factor of 10.

21. Note that there is still a similarly large discrepancy even if one folds the multiplicative factor 16/5 into the ϵ term for the purposes of comparison with Theorem 1.

$$
\begin{aligned}
&= \left| 4\epsilon x(1-x) - \frac{16\epsilon}{5}\left[x(1-2x^2+x^3)\right] \right| \\
&= \left| \epsilon \left(4 - \frac{16}{5}\right)\left[x(1-x) - x(1-2x^2+x^3)\right] \right| \\
&= \frac{4\epsilon}{5}\left|x - x^2 - x + 2x^3 - x^4\right| \\
&= \frac{4\epsilon}{5}\left|x^2 - 2x^3 + x^4\right|,
\end{aligned}
$$

and therefore $\sup|f(x) - g(x)| < \sup|x^2 - 2x^3 + x^4| = .0625$. This observation, that there is a polynomial with "large" coefficients that is approximately 0 on $[0, 1]$, in fact points to the existence of a whole space of such polynomials:

Theorem 2. *For all $\epsilon > 0$ and $1 > \delta > 0$, there exists an infinite set of polynomials $g : [0, 1] \to [0, 1]$, written*

$$g(x) = \alpha_n x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_{k+1}x^{k+1} + \alpha_k x^k,$$

such that

$$\min\{|\alpha_0|, |\alpha_1|, \ldots, |\alpha_n|\} \geqslant 1 - \delta \quad \textit{and} \quad \sup|g(x)| < \epsilon. \tag{12}$$

This explains why, in the demon case, it is possible to use a relatively large value of ϵ. While theorem 7 shows that maximum one-step error is a poor metric of model closeness, and in fact, provides a method of "cheating" this metric, theorem 8 shows that in certain spaces of polynomials, we don't need to use the theorem 7 cheat, since there are other resources for doing so. Both theorems, however, illustrate the same underlying fact. What are in fact large perturbations, can be made to look very small under the right constraints. Theorem 8, when we are restricted to polynomial space, is actually the more powerful cheat, as can be seen by looking at the two plots in Figures 3 and 4. They clearly show that freshman demon's perturbation is not at all small, even though its maximum value on the interval [0,1] is very small.

If this is right, then it suggests that the demon example is highly atypical in its ability to exhibit what looks like fast divergence in the trajectories given small maximum one-step error and a relatively large perturbation constant. It is doubly atypical, in fact, in just the ways discussed above. (It uses both the theorem 7 cheat and the theorem 8 cheat.)

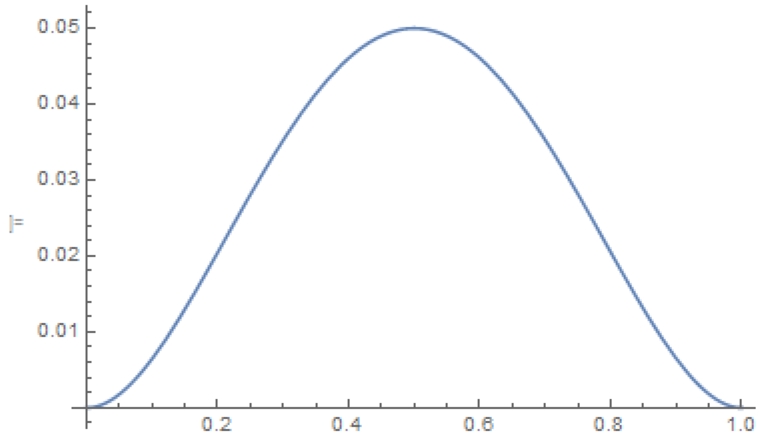

Figure 3: This is a plot of the size of the freshman demon's perturbation evaluated on the interval [0,1]

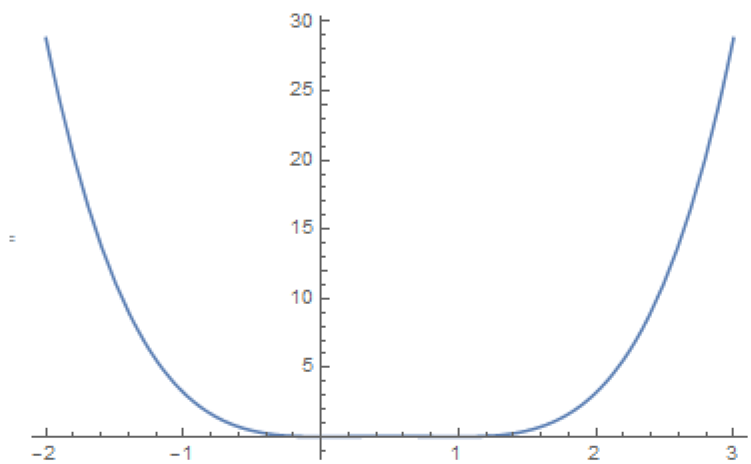

Figure 4: This is a plot of the size of the freshman demon's perturbation evaluated on the interval [-2,3] You can no longer even see the hump at 0.5.

9 A skeptical kernel?

Is there, nevertheless, a skeptical kernel of the hawkmoth papers that survives all of our arguments? Are we overstating how much of the LSE group's bundle of claims ought to be rejected? Some have expressed this worry. The alleged kernel goes something like this: we presumably don't have a model of the climate that captures "the true dynamics" of the climate system. Even if we are "nearby" the true dynamics, the ensemble distribution produced by our nearby model might differ from the "actual probabilities" that would have been produced by the true model. So decision making on the basis of model-probabilities might be dangerous. It's now up to us to find a way round this. Either find a way to get "close enough" or close in the right way so that the model distribution is guaranteed to track the true distribution, or find a way of generating decision relevant probabilities some other way. What the LSE group have shown is just that we can't just assume that the model probabilities will be decision relevant.[22]

Set aside, for the moment, that (Winsberg and Goodwin, 2015) already showed that this concern is completely misplaced vis. a vis. climate science, because climate science makes projections and not predictions, and projections are not conditional on "actual probabilities" of initial states that can be right or wrong. Here we are concerned with a more general point that is independent of the practice of climate science and of the distinction between prediction and projection.

But we are not sure along which of two possible dimensions this interlocutor thinks the kernel survives:

1. That the ensemble distribution might differ from the "actual probabilities" by a small amount, even if there is no hawkmoth effect to make it deviate by a large amount.

   or

2. That the ensemble distribution produced by the slightly wrong model might differ (significantly? by an important amount?) from the "correct" one even if point projections don't differ (significantly? by an important amount?) from their real values.

If the skeptical kernel is just (1) that a nearby model might produce a slightly different probability distribution, then this is of course correct. But this kernel of skepticism will not undermine the 'decision relevance' of any probability distributions produced by models that are "very nearby" to the real model, in any sense of "nearby" that is broached by any of the LSE groups arguments. And we needn't have a 'guarantee' that things will not go wrong. The skeptical kernel needs to be much stronger than this for it to have any policy implications at all, or for it to be anything stronger than ordinary scientific fallibilism.

If the skeptical kernel is (2), and the notion of "significance" or "important amount" is replaced by "decision relevant amount," then there are two obvious responses to anyone who continues to believe in this kernel.

A) What arguments are supposed to support that kernel? There are exactly two lines of argument that the LSE group give, and one depends on convincing us that the two apprentices' models are "nearby" and the other depends on their mis-characterizations of the structural stability results. There simply are no other arguments in their papers.

B) How could this be? A nearby model will only move an ensemble distribution far away if it moves point predictions far away. The one is parasitic on the other. And it will only do that if there is a hawkmoth effect. The whole point of the hawkmoth was that it could do to distributions what the butterfly effect could not, because the butterfly effect doesn't prevent you from getting from an exact initial condition to an exact final condition, and so it leaves probability distributions intact—even when there is uncertainty about the initial condition. But if the 'hawkmoth effect' (such as it *actually* is) is so much weaker than the

22. An anonymous referee posed something like this worry to us, and we have heard it elsewhere.

butterfly (as we have shown it is in three ways) for point predictions, there is no effect at all on the decision relevance of ensembles.

The only skeptical kernel that remains is the one that was obvious all along: a small error in your model is likely to produce a small error in your ensemble. There is no lepidopteric mechanism for turning the small error into a large one.

10 Conclusion

In the above, we have argued that there is no hawkmoth effect that generally plagues non-linear modeling. There is not one that accompanies the absence of structural stability, and there is not a ubiquitous feature of non-linear models that is highlighted by the demon example. (Indeed we have seen that the obsession with non-linearity by the LSE group is mysterious, since it does not figure in their arguments at all.) But it would be wrong to interpret us as pollyannaish about climate modeling and its epistemological pitfalls. We understand that climate modeling is hard, and it is fraught with many possible sources of error. Even our very best models of the climate are currently unsuitable for making some of the projections that important policy decisions might be sensitive to. We are not unaware of this, nor do we deny it. But the reasons for this have to do with the fact that some of the features of our climate system are poorly understood or poorly parameterized. Regional projections are particularly difficult because of the coarse graining of the globe's topography in global simulations, and other domain-specific features of global climate models. None of this has anything to do with arbitrarily small possible model errors.

Nevertheless, there is a significant difference between known inadequacies in a model that are the result of idealization: both "dynamical" (not accounting for e.g. turbulence, the biosphere, relativistic effects, etc.) and computational (discretization, parametrization, etc.), and the possibility of infinitesimally small structural errors. The former is a known problem, and climate scientists and the IPCC alike are deeply concerned with eliciting the best possible estimate of the degree of uncertainty that arises from these sources. On the other hand, so far we have seen no reason to believe that the the latter, as we have demonstrated via our arguments in the previous sections, produce any significant *decision-relevant* uncertainties.

We would further add that we have no *a priori* opposition to exploring the possible consequences of a phenomenon in the general vicinity of a hawkmoth effect. It might very well be the case that small model errors could have outsized impacts (relative to the size of the model error) on our predictions and projections, broadly construed. But a research program that was serious about exploring that question would need to be much more serious about two or three questions:

- What is the universality class (or model space) for a given physical system, e.g. Earth's climate? It seems safe to at least assume that the functions defining physical models are continuous, but could we go further? Perhaps all such functions are also differentiable or even smooth (infinitely differentiable); could they also necessarily be analytic? Maybe we can even rule out specific types of functions, e.g. logarithmic or exponential functions, and argue that a given physical system must be modeled by equation(s) involving only closed-form or algebraic expressions, or something far more restrictive, such as only polynomials of degree 2 or less. At the very least, it seems potentially too bold, and certainly unsubstantiated, to assert that the universality class is as large as the space of continuous functions. If the model space is more restricted, the impact of "small model error" could very well begin to disappear because model perturbations that don't substantially affect the predictions or projections that interest us could be far more common, perhaps even the norm.

- What is the right metric of distance? i.e. what characteristics make one model nearby to another? The kind of answer we give to this question might be rather different if the question is genuinely epistemological, rather than topological.

- If toy models are going to be used in a research program like the one that the LSE group want to conduct, what kinds of toy models are suitable? And what toy models produce idiosyncratic features like the ones we have pointed out, in sections 6 and 7, the logistic map suffers from?

But in any such research program, like any program in the philosophy of science, we believe the cardinal rule should be: do no harm. Wildly skeptical scenarios ("poison pills," and the like) about a scientific program with serious policy implications should be advanced only with the greatest possible care.

In "Probabilistic Forecasting: Why Model Imperfection Is a Poison Pill," the LSE group recommends that further research be devoted to finding an antidote for the 'poison pill' (Frigg, 2013a, p. 488). Climate models continue to be imperfect in a variety of ways that matter to policy making and decision support. But we consider this paper to be an antidote for hawkmoths.

11 Appendix A: Proof of Theorem 1

Theorem 1. *Suppose we are given a difference equation of the form*

$$x_{n+1} = f(x_n), \tag{13}$$

where $x_i \in \mathbb{R}$ and $f : A \to B$ is an arbitrary function from the bounded interval $A \subset \mathbb{R}$ into the bounded interval $B \subset \mathbb{R}$. Then given any function $g : A \to B$ and $\epsilon > 0$, there exists $\delta > 0$ and $\eta > 0$ such that the maximum one-step error of

$$x'_{n+1} = \eta f(x'_n) + \delta g(x'_n), \tag{14}$$

from (13) is at most ϵ and $x'_{n+1} \in B$.

Proof. Set δ and η such that

$$|\delta| \leqslant \left|\frac{\epsilon}{2\sup\{g(x_n)\}}\right| \quad \text{and} \quad |\eta - 1| \leqslant \left|\frac{\epsilon}{2\sup\{f(x_n)\}}\right|,$$

where $\sup\{f(x)\}$ denotes the supremum[23] of f over all $x \in \mathbb{R}$. Note that the suprema exist[24] because f and g are bounded. Since the one-step error is

$$|x'_{n+1} - x_{n+1}| = |\eta f(x_n) + \delta g(x_n) - f(x_n)|$$

23. Essentially, for sufficiently "well-behaved" functions, the maximum value.

24. That is, the suprema are finite.

$$= |(\eta - 1)f(x_n) + \delta g(x_n)|,$$

by the triangle inequality, the maximum one-step error is

$$\begin{aligned}
\sup\{|x'_{n+1} - x_{n+1}|\} &= \sup\{|(\eta - 1)f(x_n) + \delta g(x_n)|\} \\
&\leqslant \sup\{|\eta - 1||f(x_n)| + |\delta||g(x_n)|\} \\
&\leqslant \sup\left\{\left|\frac{\epsilon}{2\sup\{f(x_n)\}}\right| |f(x_n)|\right\} + \sup\left\{\left|\frac{\epsilon}{2\sup\{g(x_n)\}}\right| |g(x_n)|\right\} \\
&= \left|\frac{\epsilon \sup\{f(x_n)\}}{2\sup\{f(x_n)\}}\right| + \left|\frac{\epsilon \sup\{g(x_n)\}}{2\sup\{g(x_n)\}}\right| \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon,
\end{aligned}$$

as desired. □

12 Appendix B: Proof of Theorem 2

Theorem 2. *For all $\epsilon > 0$ and $1 > \delta > 0$, there exists an infinite set of polynomials $g : [0, 1] \to [0, 1]$, written*

$$g(x) = \alpha_n x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_{k+1}x^{k+1} + \alpha_k x^k,$$

such that

$$\min\{|\alpha_0|, |\alpha_1|, \ldots, |\alpha_n|\} \geqslant 1 - \delta \quad \textit{and} \quad \sup |g(x)| < \epsilon. \tag{15}$$

Proof. Let $0 \leqslant e^2(n - k)/\epsilon < k < n \leqslant 1/\delta$, where $e = 2.71828\ldots$ is Euler's constant, and define $\alpha_k, \ldots, \alpha_n$ by setting

$$\sum_{i=k}^{n} \alpha_i = 0, \quad |1 - |a_i|| < \delta, \quad \text{and} \quad |\alpha_{i+1} + \alpha_i| < \frac{\epsilon}{n - k} \tag{16}$$

for all $k \leqslant i \leqslant n$. Then since

$$\begin{aligned}
\sup |g(x)| &= \sup |\alpha_n x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_{k+1}x^{k+1} + \alpha_k x^k| \\
&\leqslant \sup |\alpha_n x^n + \alpha_{n-1}x^{n-1}| + \cdots + \sup |\alpha_{k+1}x^{k+1} + \alpha_k x^k|,
\end{aligned}$$

it suffices to determine the extrema of $g_i(x) := \alpha_{i+1}x^{i+1} + \alpha_i x^i$ for all $k \leqslant i \leqslant n - 1$. In that direction, note that the extrema of a function occurs either where that function's first derivative vanishes or at the end points. Thus, the possible maxima are $g_i(0) = 0$,

$$|g_i(1)| = |\alpha_{i+1} + \alpha_i| < \frac{\epsilon}{n - k},$$

by (16), and, since

$$0 = \frac{dg_i(x)}{dx} = x^i(\alpha_{i+1}x + \alpha_i)$$

$$
\begin{aligned}
&= ix^i(\alpha_{i+1}x + \alpha_i) + \alpha_{i+1}x^i \\
&= i\alpha_{i+1}x^{i+1} + (i\alpha_i + \alpha_{i+1})x^i \\
&= x^i(i\alpha_{i+1}x + i\alpha_i + \alpha_{i+1})
\end{aligned}
$$

has solutions at $x = 0$ and $x = -(i\alpha_i + \alpha_{i+1})/i\alpha_{i+1}$,

$$
\begin{aligned}
\left| g_i\left(-\frac{i\alpha_i + \alpha_{i+1}}{i\alpha_{i+1}}\right)\right| &= \left|\alpha_{i+1}\left(-\frac{i\alpha_i + \alpha_{i+1}}{i\alpha_{i+1}}\right)^{i+1} + \alpha_i\left(-\frac{i\alpha_i + \alpha_{i+1}}{i\alpha_{i+1}}\right)^i\right| \\
&= \left|(-1)^{i+1}\frac{\alpha_{i+1}(i\alpha_i + \alpha_{i+1})^{i+1}}{(i\alpha_{i+1})^{i+1}} + (-1)^i\frac{\alpha_i(i\alpha_i + \alpha_{i+1})^i}{(i\alpha_{i+1})^i}\right| \\
&= \left|\frac{\alpha_{i+1}(i\alpha_i + \alpha_{i+1})^{i+1} - i\alpha_i\alpha_{i+1}(i\alpha_i + \alpha_{i+1})^i}{(i\alpha_{i+1})^{i+1}}\right| \\
&= \left|\frac{\alpha_{i+1}^2(i\alpha_i + \alpha_{i+1})^i}{(i\alpha_{i+1})^{i+1}}\right|.
\end{aligned}
$$

Applying (16), the definition of k and n, and the well-known inequality $(1 + 1/i)^i \leqslant e$, we have

$$
\left| g_i\left(-\frac{i\alpha_i + \alpha_{i+1}}{i\alpha_{i+1}}\right)\right| < \frac{i^i(1+\delta)^i}{i^{i+1}\alpha_{i+1}^{i-1}} = \frac{(1+\delta)^i}{i(1-\delta)^{i-1}} \leqslant \frac{(1+1/i)^i}{i(1-1/i)^{i-1}} \leqslant \frac{e^2}{k} < \frac{\epsilon}{n-k}.
$$

Hence, after summing over all i, we conclude that $\sup |g(x)| < \epsilon$. Since the first inequality in (15) holds by definition, and there exist an uncountable infinity of coefficients $\alpha_k, \ldots, \alpha_n$ satisfying (15), this yields the desired result. □

References

[1] S. Bradley, R. Frigg, H. Du, and L. Smith, (2014) Model Error and Ensemble Forecasting: A Cautionary Tale, Scientific Explanation and Methodology of Science, *World Scientific* 58–66.

[2] R. Frigg, S. Bradley, R. Machete, and L. Smith, Probabilistic Forecasting: Why Model Imperfection Is a Poison Pill, New Challenges to Philosophy of Science, *Springer* (2013) 479–491.

[3] R. Frigg, L. Smith, and D. Stainforth, The Myopia of Imperfect Climate Models: The Case of UKCP09, *Philosophy of Science*, 80 (2013) 886-–897.

[4] R. Frigg, S. Bradley, H. Du, and L. Smith, Laplace's demon and the adventures of his apprentices, *Philosophy of Science*, 81 (2014) 31–59.

[5] C. Werndl, What Are the New Implications of Chaos for Unpredictability?, *The British Journal for the Philosophy of Science*, 60:1 (2009), 195–220.

[6] C. Mayo-Wilson, Structural Chaos, *Philosophy of Science*, 82:5 (2015) 1236–1247.

[7] C. Pugh and M. M. Peixoto, Structural stability, *Scholarpedia*, 3:9 (2008) 4008.

[8] L. Smith, What might we learn from climate forecasts?, *Proceedings of the National Academy of Sciences of the United States of America*, 99 (2002) 2487–2492.

[9] E. Thompson, "Modelling North Atlantic Storms in a Changing Climate." PhD diss., Imperial College, London (2013).

[10] E. Winsberg and W. Goodwin, The adventures of climate science in the sweet land of idle arguments, *Studies in History and Philosophy of Modern Physics*, 54 (2016) 9–17.