Studies in History and Philosophy of Modern Physics 54 (2016) 9–17

Contents lists available at ScienceDirect

Studies in History and Philosophy of Modern Physics

journal homepage: www.elsevier.com/locate/shpsb

The adventures of climate science in the sweet land of idle arguments

Eric Winsberg*, William Mark Goodwin

University of South Florida, United States

ARTICLE INFO

Article history:
Received 20 May 2015
Received in revised form
26 January 2016
Accepted 8 February 2016

Keywords:
Climate Science
Models
Non-linear science
Chaos

ABSTRACT

In a recent series of papers Roman Frigg, Leonard Smith, and several coauthors have developed a general epistemological argument designed to cast doubt on the capacity of a broad range of mathematical models to generate "decision relevant predictions." The presumptive targets of their argument are at least some of the modeling projects undertaken in contemporary climate science. In this paper, we trace and contrast two very different readings of the scope of their argument. We do this by considering the very different implications for climate science that these interpretations would have. Then, we lay out the structure of their argument—an argument by analogy—with an eye to identifying points at which certain epistemically significant distinctions might limit the force of the analogy. Finally, some of these epistemically significant distinctions are introduced and defended as relevant to a great many of the predictive mathematical modeling projects employed in contemporary climate science.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Modern Physics*

In a recent series of papers (Frigg, Bradley, Machette, & Smith, 2013a; Frigg, Smith, & Stainforth, 2013b; Frigg, Bradley, Du, & Smith, 2014a, 2014b) Roman Frigg and Leonard Smith, as well as several coauthors, have developed a general epistemological argument designed to cast doubt on the capacity of a broad range of mathematical models to generate "decision relevant predictions" (Frigg et al., 2014a, p. 31). The presumptive targets of their argument are at least some of the modeling projects undertaken in contemporary climate science. The form of their argument is an argument by analogy: they demonstrate that a particular, imperfect mathematical model fails to produce decision relevant predictions of a certain sort, diagnose this failure, then argue that a broad but indeterminate range of additional imperfect modeling projects, with their associated predictions, would fail for the same, or similar, sorts of reasons. The philosophical interest and scientific significance of an argument of this sort depends crucially on its scope: the more modeling projects of scientific or philosophical interest that are subject to doubt on the basis of the argument, the more interesting and important the argument is.

Unfortunately, the scope of their argument is not clear. In some places, they suggest that the quantitative predictive power of all non-linear models is threatened by this argument. 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. On the other hand, in 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, and suggesting that scientists and philosophers devote attention to the predictive challenge that they have identified.

In what follows, we first trace and contrast the very different philosophical and scientific implications of the two interpretations —broad and provocative, or narrow and modest—of the scope of their argument. We do this by considering the very different implications for climate science that these interpretations would have. Then, we lay out the analogical structure of their argument, with an eye to identifying points at which certain epistemically significant distinctions might limit the force of the analogy. Finally, some of these epistemically significant distinctions are introduced and defended as relevant to a great many of the predictive mathematical modeling projects employed in contemporary climate science.

* Corresponding author.
E-mail address: eric.winsberg@gmail.com (E. Winsberg).

http://dx.doi.org/10.1016/j.shpsb.2016.02.001

1. The broad and provocative interpretation

In their more provocative moments, the authors 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., 2013a, 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., 2014a, 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., 2014a, p. 48). In other words, mathematical models can be qualitatively informative in spite of the fact that they are not quantitatively trustworthy.[1]

Now one might wonder what the implications of taking this conclusion seriously would be for contemporary climate science. How much of what climate scientists claim to know about the state of the current climate and its future possible trajectories would be undermined should one accept the strong reading of this argument? Though a precise answer would depend crucially on a more detailed parsing of the qualitative vs. the quantitative predictions of models, it is safe to say that much of what climate scientists claim to know would have to be regarded as untrustworthy. The authors seem aware of this risk because in the very paragraph where 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 there is evidence from both basic physical science and observations" (Frigg et. al., 2014, p. 48)[2]. However, 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., 2013, 872). This, in turn, requires an estimate of internal variability, generally derived from a "physically based model" (Bindoff et al., 2013, 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" (Bindoff et al., 2013, 872). Indeed the authors of the IPCC report are quite clear that, "attribution is impossible without a model" (Bindoff et al., 2013, 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., 2013, 873).

Even if one takes the view of the IPCC to be controversial, 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 chance, 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. The broad reading of the argument, in other words, is quite provocative. It doesn't just undermine prediction about regional climate in a hundred years; it undermines the most basic conclusions of contemporary climate science, at least as those conclusions are now established.

2. The narrow and modest interpretations

While it would certainly be earthshaking if this argument had managed to establish that most of contemporary mathematical science was "embroiled in confusion" or "maladaptive," the sheer implausibility of such a result suggests looking for a narrow, more modest reading of the scope of the argument. In their more modest moments, the authors do characterize the scope of their argument more narrowly, but they do this in two distinct ways. In some cases they infer from the precision and locality of a prediction that it must be subject to doubt on the basis of their argument, but in other cases they suggest that the applicability of their argument depends on how a probabilistic prediction was generated from the underlying mathematical model. Ideally, the class of modeling projects subject to doubt by this argument *identified by the details of the content of their predictions* and the class subject to doubt *in virtue of their method of generation* would line up, and it would be obvious why they line up. Whether this is so is something that will be considered after these two ways of characterizing the scope have been explicated.

The authors repeatedly use the same example when they want to establish the applicability of their argument to the sorts of modeling projects in which climate scientists are actually engaged. The example is the United Kingdom Climate Projections 2009 project, or UKCP09, which aspires to produce decision relevant "high-resolution forecasts of twenty-first-century climate" (Frigg et al., 2013b, p. 887). From the fact that UKCP09 forecasts, "the hottest day in August in a particular year" (Frigg et al., 2013a, p. 490) or "the temperature on the hottest day in central London in 2080," it is supposed to follow that their argument is "not just a hobbyhorse for academic philosophers" (Frigg et al., 2014a, p. 50). Evidently, it is the fact that the UKCP09 project assigns probabilities to climate variables in relatively small regions far off in the future that makes it obvious that this modeling project falls within the scope of their argument.[3] Indeed, regional climate projections are not generally considered to be as trustworthy as their global counterparts, and UKCP09 is not cited in the latest Working Group I report from the IPCC. However, if these authors are right, there is a new, in principle, argument establishing the impossibility of the kinds of projections that this group hopes to

[1] They also acknowledge that there might be some cases where some quantitative insight can be derived from such models, perhaps by switching to "non-probability odds." See (Frigg et al., 2014a, p. 57) and (Frigg et al., 2013a, p. 490).

[2] They are also clear, at least at some points, that climate models fall within the scope of their argument: "The problems arise if models are non-linear and imperfect …[w]ithout question, climate models have both of these properties" (Frigg et al., 2013a, p. 488).

[3] The authors characterize these sorts of predictions as predictions of "finely defined specific events" (Frigg et al., 2013b, p. 888), which seems to conflict with their characterization as climate variables (cf. Werndl, forthcoming).

produce. That would still be a result of considerable significance for, and with a considerable impact (one would hope) on, contemporary climate science.

The second, narrow way of characterizing the modeling projects subject to doubt by this argument involves identifying a method of generating probabilistic predictions of a particular sort from a mathematical model, and then claiming that predictions of that sort, generated in that way, from a non-linear, imperfect model are subject to doubt. Given that the argument is developed by analogy, the way to characterize the relevant method is as being relevantly similar to the case on which their argument from analogy rests. Thus, both in order to get clearer about the scope of the argument when it is characterized by method of generation, and to evaluate, eventually, the significance of the argument for climate science, we need to get clearer about the argument from analogy.

The base case of the argument from analogy is a dynamical system whose time evolution is given by the logistic map. We can think of this as a function that relates the population of fish in a pond to its population in the previous week. This system is subject to both the *butterfly effect*, where arbitrarily close initial conditions will follow very different trajectories, and what the authors have begun to call the *hawkmoth* effect, where small structural differences between two nearby time evolution functions can result in very different trajectories even when beginning with identical initial conditions. Consequently, the model can produce “fundamentally misleading” probability distributions over final states. The authors show that in this case a particular strategy[4] for managing the epistemic uncertainty coming from the butterfly effect will not work in the presence of the hawkmoth effect. Because the system is subject to the butterfly effect (and the initial conditions cannot be specified exactly), it is not, generally, possible to make informative point predictions about the state of the system at future particular moments in time. It is, however, possible to manage the epistemic uncertainty resulting from the butterfly effect by evolving a probability distribution over the initial states of the system into a probability distribution over the final states of the system at some future time. By following this strategy, though one is no longer able to predict the precise future state of the system, one can instead produce some probability distribution over possible future states. Unfortunately, as the authors demonstrate in this case, if the true time evolution is taken to be a slightly perturbed version of the logistic map, then the probability distributions over final states at some future time obtained by using the imperfect model can differ substantially from the “true” probability distribution over those final states. This shows that the suggested strategy for managing the uncertainty due to the butterfly effect will not work in this context. The suggested strategy gives rise to enormous second order uncertainty about how much uncertainty there is due to initial condition error plus the butterfly effect. It is this second order uncertainty that undermines the predictions generated by the model.

3. Generalizing from the base case and the application to climate science

Even if we grant everything the authors hope to have demonstrated about the implications of the hawkmoth effect for the base case (the particular sorts of pdfs generated by the default strategy from the dynamical system with the time evolution given by the logistic equation), it is appropriate to consider the conditions under which other modeling projects might—because of their similarity to the base case—be subject to concerns about the hawkmoth effect. For this purpose, it is useful to think of a modeling project as having four dimensions across which it could be compared to the base case. First, there is the time evolution function itself, which may or may not behave in the way the logistic equation does in response to small deviations from the ‘true’ time evolution. Only if the time evolution of a modeling project is subject to ‘structural instability’ in a manner relevantly similar to the logistic equation might the predictions of that model be undermined by the hawkmoth effect.[5] Second, there is the time scale of the desired prediction. Like the butterfly effect, the hawkmoth effect degrades certain sorts of predictive accuracy, and it may indeed do it at a very fast rate, but it still takes time, and so the degree of epistemic danger that it poses will always depend on how far out into the future one wants to predict. Third, there is the manner in which predictions are generated from the dynamic system, which may or may not be relevantly similar to the ‘default’ strategy for generating predictions that the authors employed in the base case. The default strategy was, according to the authors, a way to manage the butterfly effect, which was, in turn, shown to be undermined by the hawkmoth effect. Unless the strategy for managing the butterfly effect employed by a modeling project is similarly undermined by the hawkmoth effect, then that project would not be relevantly similar to the base case. Lastly, there is the kind of prediction[6] that one hopes to obtain from a modeling project. In the base case, the authors consider probability distributions over precise complete states of the system and how they evolve, precisely, as time advances, instant by instant. But there are many other kinds of predictions that one might make about a dynamic system. We are sometimes interested in statistics that a system will exhibit over an infinite lifespan. We might also be interested in statistics that are exhibited over periods of time that are less than infinite but that are averaged over intervals of time that are large compared to rate at which, for example, weather data are sampled. Thus, the kind of prediction that is generated in a modeling project must be relevantly similar to the distributions over final states considered in the base case in order for the authors concerns about the hawkmoth effect to apply to that project. Overall, then, a modeling project would be plausibly undermined by the hawkmoth effect based on this argument, if its time evolution, method, type of prediction, and duration are relevantly similar, in turn, to the logistic equation, the default strategy, and probability distributions over final states of the system at some future time, and if that future times is sufficiently far out into the future for the effect to become sufficiently severe as to be considered “undermining”.

The authors do proceed to argue that the hawkmoth effect, which they have demonstrated in the case of the logistic map, applies to many, or most, other nonlinear, imperfect mathematical models. In other words, they contend that most time evolutions are relevantly similar to the logistic equation. As a result, insofar as they are right, the strategy of managing the uncertainty due to the butterfly effect by evolving probability distributions over the initial conditions won't work in these other models either. The most detailed version of this argument introduces the notion of the structural stability of a model, which, they suggest, is the feature required of a dynamic system if it is not to be subject to the

[4] The strategy is also referred to as the “default position” (Frigg et al., 2013a, p. 488).

[5] As we will see below, the situation is slightly more complex than this, since structural stability is not a property of a model by itself, but rather of a family of nearby models.

[6] Indeed, as we will see later, in the language of the IPCC, and of the UKCP, some kinds of what we might loosely call “predictions” are not predictions at all, but are in fact what, in their vernacular, are called “projections.” We will come back to this distinction shortly.

hawkmoth effect. Additionally, they contend that there are some plausible reasons to believe that the sort of mathematical models scientists are typically interested in are not structurally stable, and so, if a scientist wants to pursue the suggested strategy for dealing with the butterfly effect, *the burden of proof should be on them* to establish that their model is structurally stable, and thus that the hawkmoth effect doesn't apply and undermine their strategy. This argument will be considered in more detail later, but for now it is important to point out that this argument does not in anyway address the other three dimensions—the method, the type, and the duration of a prediction—along which a modeling project would have to be similar to the base case in order for the concerns about the hawkmoth effect to apply to it.

The importance of these other three dimensions can be brought out by considering the implications of this argument, interpreted narrowly and modestly, for contemporary climate science. To consider the scope of the argument as characterized methodologically—as applying to just those modeling projects that are relevantly similar to the base case—we must consider how common it is for the modeling projects of climate scientists to attempt to deal with the butterfly effect by evolving probability distributions over initial conditions into probability distributions over the states of the model at future times (hereafter: using the "default strategy" to generate point predictions). These are the cases that, if the rest of the analogy holds, will have been unproblematically shown by the argument to be doomed to failure. The first thing to note on this point is that the probabilities produced by the UKCP09 project are not produced by the default strategy[7]; furthermore, the authors of the argument know this. They say, after describing the strategy for managing the butterfly effect characterized above, "UKCP09 probabilities are formed in a more complicated manner, combining outputs from multiple (imperfect) models using Bayesian methods" (Frigg et al., 2013a, p. 483). This establishes that there is no straightforward identity between the scope of their argument characterized in terms of the content of the predictions (localized probability statements about the future), and the scope characterized methodologically. While the UKCP09 projections are relatively precise and local, they were not produced using the default strategy and furthermore they are not point forecasts about the atmosphere at some future time (this point will be developed more in what follows). Still, one might hope to show that, ultimately, any probability statements about the climate of localized regions in the distant future must depend upon attempts to manage the butterfly effect by the default strategy, or are subject to similar worries due to the hawkmoth effect. The authors provide no such argument and merely suggest that it is unclear why the strategy used in UKCP09 would make the concerns about the hawkmoth effect "go away." However, as far as we can tell, they provide no reason, in the first place, for thinking that concerns about the hawkmoth effect are relevant to the sorts of probabilities that are generated in this study. Nowhere do they purport to show that either the method used in the UKCP09 or the kinds of projections that it produces are relevantly similar to the default strategy or probability distributions over states of the system at some future time.

As a way of bringing this out more clearly, it is worth considering some distinctions that are made in a chapter of the latest IPCC report called "Near-term Climate Change: Projections and Predictability". The point of the chapter is to assess recent attempts to make use of "observation-based initial conditions" (Kirtman et al., 2013, p. 960, our emphasis) in order to predict near term climate. The contrast is with what had been the main focus of assessments of future climate in previous IPCC reports: climate projections. Climate projections are "climate change experiments with models that do not depend on initial conditions but on the history and projection of climate forcings" (Kirtman et al., 2013, p. 958, our emphasis). Simulations run in a climate projection are "not intended to be forecasts of the observed evolution of the system but to be possible evolutions that are consistent with the external forcings" (Kirtman et al., 2013, p. 959). That is, they are projections, rather than predictions, in the language of both the IPCC and the UKCIP. The IPCC statements about climate change up to this point have been largely based on experiments of this sort. Insofar as the IPCC has made probabilistic claims about the future climate, then, it does not seem that those probabilistic claims have been generated using the "default strategy"—by evolving probabilities reflecting uncertainty about initial conditions through their imperfect models—those would be "predictions". Instead, it seems that the predominant strategy for dealing with uncertain initial conditions has been to ask questions of the models that do not depend, at least very much, on initial conditions, or even on probability distributions over initial conditions—that is, to produce climate "projections." This is not to establish or defend the legitimacy of the techniques[8] used to do this, but rather to point out that the probabilities are not produced by the strategy attacked by the authors of this argument—the "default strategy." Furthermore, even the climate predictions, for which initial conditions are relevant, "do not provide forecasts of the detailed day-to-day evolution of future weather. Instead, they provide probabilities of long-term changes to the statistics of future climatic variables" (Kirtman et al., 2013, p. 964). And so, the kinds of predictions produced (and of course the projections as well) in typical climate science modeling projects are not probability distributions over the state of the system at some future time.

Again, this suggests that nothing like the default strategy is involved in generating the probabilities produced in such climate modeling projects, and furthermore the sorts of outputs produced by both climate predictions and climate projections are not the same sorts of probabilistic predictions as those generated in the base case of the argument from analogy. Overall, then, it is not clear whether, or how many of, the modeling projects in contemporary climate science are either generated by a strategy that is relevantly similar to the default strategy, or produce relevantly similar types of output. Furthermore there are plausible reasons to suppose that the strategies and outputs employed by climate scientists are—on their face—importantly distinct from those employed in the base case of the argument from analogy, and so the burden of proof falls to the authors of the argument to show that their argument has any bearing at all on contemporary climate science.

[7] The UKCIP09 report does mention using initial condition based prediction (though not via the default strategy), and only as a possible future experimental technique, strangely enough, citing one of the co-authors of the Frigg et al papers (Leonard Smith) as one of the proponents of this method.

[8] It is not the case that it has been established or experimentally shown (it would cost too much computational power) that there is no dependence on the initial conditions. There is a lot of uncertainty about the exact degree of initial conditions uncertainty. But it should be noted that "initial condition uncertainty" about which there is legitimate concern, is not uncertainty about what the correct initial conditions are. There can be no such concern vis a vis a projection (as opposed to a prediction). It is uncertainty about what effect, if any, the (usually randomly chosen) initial conditions that are in fact used are having on the outcome of a forcing experiment. That is, it is uncertainty about whether what one is observing in a forcing experiment is a result of the forcing, or of the initial conditions that were tried. (To make this point clear: if we had the computer time to explore every possible initial condition, we could completely eliminate so called "initial condition uncertainty" in a forcing experiment, without any knowledge of the actual initial conditions.)

4. Three questions

There are, then, three questions we can ask about the strength of the argument by analogy that the authors offer. (1) What is the strength of the analogy for casting doubt on the capacity of a broad range of mathematical models to generate "decision relevant predictions" when those modeling efforts use the "default strategy." (2) What is its strength for casting doubt on the capacity of mathematical models for producing the kinds of local and precise projections that the UKCP09 engaged in. (3) And what is the strength of the analogy for establishing the more broad and provocative claims that Frigg et. al. make.

In order to probe the answers to these questions, we need to draw a few distinctions. The first is a distinction to which those authors draw a fair bit of attention. This is the distinction between uncertainty due to ignorance of the precise initial condition of a system, on the one hand, and uncertainty due to ignorance of the precise form of the dynamical equations that would make the best model for a particular system. The latter source of ignorance might be reducible to ignorance of the value of some parameter in those equations, or it might not. The first source of uncertainty is sometimes called "initial condition uncertainty"[9], and the latter is sometimes called either "parameter uncertainty" or "model structure uncertainty" (depending, of course, on whether the source of the uncertainty is best understood as ignorance about the precise value of a parameter or not). Let us simplify just a little bit and call the second source of uncertainty "model uncertainty."

Another important distinction is one to which Edward Lorenz drew attention in his (1975) paper, "Climate predictability." Lorenz distinguishes climate prediction from weather prediction. Weather prediction is the process of determining how the weather will change as time advances. Climate projection, canonically, is the process of determining what the statistics of an atmospheric system will be over its entirely lifespan if some external variable, (such as a forcing variable like carbon concentration) is changed. But climate projection can also involve predicting the statistics of an ensemble of states over a "long but finite time span." (Lorenz, 1975, 132). He calls the latter climate predictions of the first kind, while the predictions associated with infinitely long times are climate prediction of the second kind. He also notes that climate scientists are interested in both kinds of prediction, and that with respect to climate prediction of the first kind, they might be interested in periods as short as decades or as long as millennia or more.

In order to generalize a little bit beyond the cases of weather and climate, let us call it "synchronic prediction" when we are interested in predicting a value of one of a model's variables at some time in the model's future, and all the other kinds of predictions we might want to make about average values, maximum values, or topological features of the dynamics, "diachronic predictions". Thus, in a typical General Circulation Model, a prediction of a weather variable for a partial day is a synchronic prediction. Diachronic predictions, we should note, will then fall into at least three categories. They might be about something genuinely qualitative concerning the system's dynamics: the fractal dimension of an attractor; the overall location of an attractor; the existence of heteroclinic or homoclinic connections between attractors; or any other topological feature of the dynamics. These are, presumably, the kinds of qualitative predictions for which Frigg et al concede non-linear models might be suitable. But they might concern the average value some variable will take over the life of the system (in the case of the atmosphere, what Lorenz would call a climate prediction of the second kind). Or they might concern the precise maximum value that some variable of the system would exhibit over a relatively long period of time (be it decades or millennia). All three of these kinds of prediction tasks would count as diachronic, insofar as they not analogous to what Lorenz calls "weather prediction."

So much for distinctions. But before we proceed to answer any of our three questions, one issue would be extremely helpful to settle: What is the relationship between uncontrollable initial condition uncertainty (chaos) and uncontrollable model uncertainty? In the language of Frigg et al, what is the relationship between the *butterfly* effect and the *hawkmoth* effect? Does the hawkmoth effect always accompany the butterfly effect?

Here is what Frigg et al are keen to suggest that their simulations reveal, or at least exemplify: if we think we know the precise form that the best model of a system should take, and that model happens to be chaotic, then it is overwhelmingly likely, and the burden of proof should be on anyone who wants to use the model to show otherwise, that the model is structurally unstable and will be maladapted to quantitative prediction of any kind.

4.1. First question

In a nutshell, Frigg et al. really want to warn us that structural instability is a deeper threat to climate prediction than ordinary chaos, and that the two go hand in hand. So one extremely important question is: is the latter claim true? Does structural instability always accompany ordinary chaos? The situation is a bit more complex than the authors let on, and no univocal answer to this question is possible.

In the first place, 'structural instability' is not a property that a single time evolution function, considered in isolation, exhibits. We can only say that such and such a function is unstable relative to a family of nearby functions, and relative to a metric defined between each pair of such functions. The complexity of this kind of question becomes quite clear if we look at a recent paper by Conor Mayo-Wilson, who was trying to get a more precise handle on the questions that Frigg et al. had raised.

Mayo-Wilson starts by defining a chaotic model as a model whose time-evolution function is topologically mixing, where a such a function is topologically mixing just in case, for any of the non-empty open sets U and V in the state space, there exists a number $N>1$, such that

$$\varphi^n(U) \cap V \neq 0$$

for all $n \geq N$.

This is a useful definition because it tells us that, no matter which two regions you pick in the state space, if you wait long enough, some portion of the first region (no matter how small it is) will evolve into some portion of the second region (no matter how small it is). This is exactly the kind of behavior that gives rise to the butterfly effect.

Mayo-Wilson then tries to define a concept that he takes to be the relevant analog of chaos when it comes to model structure. That is, he tries to find the relevant concept he will call "structural chaos" that appropriately parallels ordinary chaos, in the just that way that model structure uncertainty parallels initial condition uncertainty.

He then claims that the relevant definition of structural chaos should involve the concept of structural mixing:

> "Structural mixing" should capture the idea that similar models can produce different trajectories through the state space given the [exact] *same* initial conditions.

[9] See footnote 8, however, for the important distinction between initial condition uncertainty in a prediction and initial condition uncertainty in a forcing experiment or climate projection.

He then offers the following definition[10]:

Given a set of time-evolution functions $\Phi \subseteq X^X$ and a particular model $\varphi \epsilon \Phi$, we can then define a set Φ as structurally chaotic at φ just in case for all $x \epsilon X$, all $\varepsilon > 0$, and all non-empty open sets $V \subseteq X$, there is some natural number N such that

$$f_{x,n}(B_{\varepsilon}(\varphi) \cap \Phi) \cap V \neq \varnothing,$$

$$n \geq N.$$

The intuitive idea of this definition is that it captures the idea that small differences between the estimated model and the true one can lead to radically different predictions after the same interval of time even if the initial conditions are known precisely. It captures the idea that, even for a precise initial condition, no matter where you want to go in the state space, there is a time-evolution function that is arbitrarily near to the one you started with that will take you there, from that precise initial condition.

Mayo-Wilson then offers some useful results and discussions. The first thing he shows is that, modulo some minor exceptions, almost all dynamical systems that exhibit chaos also exhibit structural chaos as he defines it. So far so good for Frigg et. al. But he also surveys some results that show that many systems that are, in his vocabulary, structurally chaotic, are also "structurally stable." Structural stability is, according to Mayo-Wilson, "intended to formalize the idea that small changes to the model do not result in large differences in the model's trajectory."

How can this be? How is it that a formal property of systems designed to capture the informal notion that similar models can produce very different trajectories through the state space given the *same* initial conditions, be coextensive, in a large number of cases, with a formal property designed to capture the informal notion that small changes to the model do not result in large differences in the model's trajectory? The possibility of overlap here hinges on a difference between the definitions, regarding what are included among the allowable "nearby" trajectories (what is sometimes called the "universality class") and the metric used to define the notion of two evolution functions being epsilon distance apart. The definition of the notion of structural stability that Mayo-Wilson is talking about depends on equations of evolution being continuous and differentiable. But his definition of structural chaos allows for any arbitrary function. In this respect, Mayo-Wilson's result regarding structural chaos is extremely weak —structural chaos, as he defines it, is not a terribly epistemically worrying feature for a model to exhibit.

And it is of course interesting that Frigg et al's simulations, the base case in their argument, are two functions, one of which is a polynomial of 5th order. Thus, in so far as we might be inclined to believe that the Earth's climate, being a physical system governed by physical laws, is best modeled by second order equations, it is not at all clear that the analogy between the pond of fish and the earth's climate holds, even with respect to the conclusion one might draw from it concerning its ability to cast doubt on the capacity of climate models to generate "decision relevant predictions" when those modeling efforts use the "default strategy."

This perhaps explains why many climate scientists will tell you that their models *are* stable under small perturbations.[11] What they mean is: when they make changes in their models to the "nearby" models in what *they* consider to be the relevant family of possible models, they don't see large changes in output. We should be clear that this does not prove that their models are indeed structurally stable (certainly not on any arbitrarily specified universality class and metric of closeness). They could be getting what appear to be stable results because they are not making the relevant perturbations.[12] Or they could be getting those results because they are not waiting sufficiently long for the instability to manifest itself.

This provides us with the resources to answer question 1, regarding the "default strategy": The strength of the analogy for casting doubt on the capacity of a broad range of mathematical models to generate "decision relevant predictions" when those modeling efforts use the "default strategy" is indeterminate insofar as there are far too many open questions about climate modeling. We simply do not have a clear grip on what the right universality class is for climate models, nor on what the relevant metric of similarity ought to be.

But even if it were to turn out that climate models *are* structurally unstable, in the relevant sort of way, it is entirely unclear what the relevant *time scales* would be for such instabilities to manifest themselves. Even chaotic models with initial condition error can make good precise predictions in the short run, and, in similar fashion, even structurally unstable models with small model error can make good precise predictions in the short run.[13] And when it comes to climate models, even if we suppose that there are good a priori reasons for thinking the relevant family of models are structurally unstable (or that the burden of proof should be to show otherwise), such an a priori argument would not provide us with anything like the resources for determining what the relevant notion of "short run" is. Thus, as far as we can tell, no a priori argument can tell us who ought to be considered to have the 'burden of proof.' In some empirical contexts, the default strategy might still be a viable one, even when the relevant models are chaotic or even structurally unstable—if the prediction task we are interested in does not get derailed very much by small model errors over the time frame that interests us.

When we move outside the domain of toy models and toy contexts where clean mathematical results are available, the best we can ever do is to experiment with our models and see what small perturbations do to our predictions. It is extremely unrealistic to insist that modelers have the burden of proving that their models fail to have some feature that we only know how to study carefully in certain very rarified contexts.

Fortunately, when climate scientists do study, empirically, the effects of small perturbations on the class of predictions that interest them, they don't find dramatically divergent results. But more importantly, climate science rarely or never uses the default strategy. That's because climate scientists are not canonically interested in forecasting probability density distributions over

---

[10] Here is the relevant notation: Let the distance between two time-evolution functions be the supremum/maximum distance between the two functions after one unit of time, where the maximum is taken over all possible starting states.

For any $\varepsilon > 0$, let $B_{\varepsilon}(\varphi)$ denote all models within distance ε of φ. Next, for any natural number n and any point x in the state space X, define a map

$$f_{x,n}: P\left(X^X\right) \to P(X)$$

as follows:

$$f_{x,n}(\Phi) = \{\varphi^n(x): \varphi \epsilon \Phi\}$$

where $P(x)$ is the power set of X.

[11] See for example (Doblas-Reyes et al., 2013).

[12] And of course, reasonable people will be able to disagree about what the relevant family of time evolution functions (across which one can make perturbations) is. And what the relevant one for meeting some particular mathematical definition is might vary from the relevant one for some particular epistemological question. The point here, in some sense, is that the dimensions of this kind of discussion are vast and not fully explored, and Frigg et al have done little to sort them out.

[13] The same is true, incidentally, in some cases, in the very long run. It actually turns out in many cases that the hawkmoth effect poses its greatest risk only in the "medium run."

precise states of the atmosphere for precise future times.[14] And that's what the default strategy would give you. That's the sort of predictive task that climate modeling would have to be engaged in —producing a probability distribution function for, say, 3:45pm on September 23rd 2098, which would give the probability of the earth being in any tiny region of massively-large-dimensional configuration space that defines the earths climate—for the analogy to the base to hold.

4.2. Second question

This brings us to question 2: we have noted that Frigg et. al. often argue for the applicability of their argument by analogy, not on basis of the fact that the "default strategy" has been used, but on the basis of the "precision and locality" of a prediction task being tackled. We have already noted that, despite the fact that the UKCP09 makes what are, for Frigg et al., canonically "precise and local predictions" (or rather projections), it does not employ the "default strategy." So we now turn to the question of strength of the analogy for casting doubt on the capacity of a broad range of mathematical models to generate precise and local predictions or projections, regardless of the strategy employed.

One thing to note is that it is very easy to equivocate between a prediction being precise and local, in the sense that these authors use the term, and a prediction being dependent on initial conditions. Even very precise, local, and relatively short-term (order of a decade) climate projections are not weather predictions. Let us define weather predictions as those that predict the state of the system conditional on some precise initial condition, moment by moment as it evolves in time. A prediction concerning, for example, the average value of a variable over the life of the system, to a precision of one part in one million, is highly precise, but needn't be dependent on initial conditions, in our sense. Even a precise prediction of the highest temperature that the city of London will reach in the summertime in the decade beginning in 2010, while highly "local" in some intuitive sense, could also be a climate projection (a forcing experiment), that is independent of knowledge of the present conditions. At worst, it would count as a climate prediction of the first kind with a relevant time period that presses up against the boundary of what Lorenz took to be reasonable.[15] Indeed, as we have noted above, on most definitions of what it is to be a "climate variable,"[16], and what it is to make a "projection" as opposed to a "prediction," making a projection of a climate variable is generally inconsistent with making a prediction conditional on knowledge of present conditions.

Many climate modeling strategies, as we have seen, involve methods that avoid dependence on a particular set of initial conditions. And thus it is only predictions (as opposed to projections) which might depend on the "default strategy" in order to manage the initial condition uncertainty. So, only an equivocation between predictions on the one hand and highly local and precise projections, on the other, would lead one to think that all precise and local 'predictions' depend on the default strategy and hence (setting aside our doubts about the answer to question 1) are subject to worries related to the hawkmoth effect. This is why even the UKCP09, despite the fact that it makes highly "precise and local" projections, needn't rely on the default strategy.

At one point in one paper, the authors flirt with dealing with the kind of distinction we are drawing here, but they insist that many of what they call the "predictions" (which are in fact projections) of the UKCP09 are "not so different from weekly predictions in the fish model" (489) and so the distinction they discuss "does not help circumvent the difficulties we describe." (490). But the relevant point of comparison between the two cases of prediction/projection is not whether they *seem* different or not so different, but the degree to which they depend on initial data and concern the determinate time after the time to which the initial conditions are indexed. The fish prediction does, and the UKCP09 projections do not. And in that respect, the projections of the UKCP09 are *not at all* like the weekly predictions in the fish model. One is a point forecast, and the other averages over massive numbers of degrees of freedom in the model. Put crudely, a weekly prediction in the fish model is a weather prediction, not a climate projection.

Is there another route from the hawkmoth effect to doubts about the kinds of highly local and precise projections made by the UKCP09, one that does not depend on the role of initial data? Again the question is highly complex, and depends on details of the definition of structural stability being employed and on the timeframe of the prediction. We have seen already that it is not usually helpful to talk about a particular model being structurally unstable. We always have to ask whether a family of models is stable under a class of perturbations. Similarly, once we move away from interest in pointwise synchronic predictions (the kind that play very little role in climate science) we have to ask of that class of perturbations whether a particular kind of *projection statistic* is stable under it. It will often be that case that some are and others are not. And it will almost always be the case that the more degrees of freedom a prediction statistic averages over, the less likely it is to be unstable under a class of perturbations.

Moreover, as Fillion and Corless (2014) have argued, almost every model that has any degree of empirical confirmation will have *some statistics* that are relevantly stable under any class of perturbations. It is useful to look back closely at Mayo-Wilson's definition of structural chaos. Here, one notices that his definition of a system being structurally chaotic tells us nothing at all about how good a model with small structural error will be at forecasting any diachronic features of the dynamics. His definition has only to do with where the model will take you after n steps. It is perfectly possible for a model's synchronic predictions to be unstable under a class of perturbations, while at the same time allowing for certain other kinds of predictive tasks to be stable under that same class. This is another respect in which Mayo-Wilson's results are extremely weak, since they only apply to pointwise predictions.

Thus, while we certainly do not want to insist that only synchronic predictions ever fall under the scope of the hawkmoth effect, we do want to caution that whether or not structural instability of some family of models (even when it does in fact exist) "pulls the rug out" from under some field of enquiry will depend a great deal on the complex interplay between the nature of the predictive task to which the model is being put and the time scale of the prediction.

The relevant question for climate science, then, (and for any science that depends on non-linear modeling) is: how do we decide whether a predictive enterprise is undermined by the possible existence of the hawkmoth effect? Here, we repeat a point we made earlier: that the relevant questions for climate science lie outside the domain of the study of toy models, and that the best we can do is to play around with our models and see which predictions are stable under the perturbations that we

[14] An anonymous referee has challenged this claim. We believe we have given textual support of this claim from the IPCC report. Still, we admit a negative claim of this sort is hard to prove. But it is interesting that the referee in question provided only two examples (Daron & Stainforth, 2013; Smith, 2002) and both of them were papers by people who are among the co-authors of the very series of papers by Frigg et all that we are discussing.

[15] "The question arises as to the appropriate length for the time span through which a given climate is supposed to last. There appears to be no unique answer—climates defined in terms of widely different time spans all constitute different aspects of the total problem. Typical conditions over millennia or longer are important in the study of ice ages; conditions of decades are important to agriculture. Personally I find it difficult to think of oscillations which complete a full cycle within a year or two as "climatic" fluctuations." (Lorenz, 1975, 132)

[16] See Werndl (in press) for a variety of definition of climate and a discussion of their strengths and weaknesses.

think ought to concern us. The likelihood of the question being settled by mathematical results is small.

So, even when Frigg et al hedge the stronger claims one finds in the rhetoric of their abstracts,[17] introductions and the like, by making repeated mention of the UKCP09, the hedges are off the mark. We do concede, with regard to the point made earlier about climate variables being chaotic, that it is more likely that regional and decadal climate variables are chaotic than that global and centenary variables are chaotic. But none of this is known or established by Frigg et al. Thus we do not think these are the best grounds for criticizing the UKCP09.

Why, then, might one agree (and why should anyone agree) with the authors that, e.g., the UKCP makes overly fine-grained projections?

If we want to get a good grip on the limits of the predictive accuracy of our best models, we are better off looking carefully at the actual relations that we believe they bear to the real world, than we are looking at their abstract mathematical properties. When we do look at the former set of features of our best models, they reveal many good reasons for skepticism about overly fine-grained projections. If our best models are currently unsuitable for making the fine-grained projections of the kind we find in, e.g., UKCP09, we believe, the reasons 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. It is not, as far as we can tell, because the models need to be infinitely precise in order to do this job well. We do not think there is any good reason to believe that last claim. And the factors that do in fact make climate modeling so difficult do not comprise a “poison pill.”

Climate models give rise to uncertainty, even with regards to global, diachronic features of the climate, because they contain structural problems that are much more than microscopic: they leave out elements of the earth's climate system and they include others that are poorly understood; they need to be discretized in order to be solved on a computer, and these discretizations can contribute to unknown errors; they contain parameterizations that may be far less than perfect, and they depend on assumptions about future scenarios (energy use, population, economic development, etc.) that are insecure. These are serious problems that our best models have, and it is a difficult problem to determine how much uncertainty these problems give rise to with regard to diachronic forecasts of different kinds and different degrees of fine-grainedness. This is the problem of determining the various models' adequacy for various purposes about which Knutti (in press), Parker (2009) and others have written some very useful things. But these are questions that can only be settled by carefully studying the real-world system of interest (the climate!) and by studying how well our models track the processes that appear to do the heavy lifting in generating the climate that we observe in the real world. They are not questions that can be settled by looking at toy models in the way that Frigg et. al. do. And the danger of supposing that they *can* be settled by those kinds or arguments is that those arguments, if taken seriously, are too abstract to be anything other than nuclear—they are insufficiently precise for locating the border between the class of predictions for which we should take our best models to be reliable and the class for which we should not. They either leave everything we actually do untouched, or they obliterate (“pull the rug out from under”) everything! Locating where the real border lies between genuinely sanctionable forecasts and “the sweet land of idle dreams” (Frigg et al., 2014a, p. 32) is a difficult and complex *empirical* task—not one that can be settled a priori or by analogy to toy models.

This concludes our answer to question 2, regarding “local and precise predictions”: The strength of the analogy for casting doubt on the capacity of a broad range of mathematical models to generate “local and precise predictions” that are “decision relevant” is poor insofar as the phrase “local and precise” as Frigg et al use it, describes projections which have no special status *vis a vis* the hawkmoth effect. There is no good reason, (or at a minimum, there is no good reason offered by Frigg et al) to think that the Hawkmoth effect undermines the projections of the UKCP09, but not those of, e.g., climate sensitivity studies.

4.3. Question 3

We turn, finally, to our third question: What is the strength of the argument by analogy for establishing the more “broad and provocative” claims that Frigg et al make? These claims can be best summarized using a phrase from the title of one of their papers: 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.” We believe we have already shown that these claims are implausible. Since we have already shown that the power of the hawkmoth effect depends on the model and its universality class, the predictive task, the time scale, and the method of prediction, it is perhaps useful to ask a broader question:

Are there qualitative and/or diachronic analogs of the hawkmoth effect that could potentially pose the kind of general danger to climate science that Frigg et al are worried about?

We think that there are. Indeed, Stewart (2011) provides a nice catalog of the kinds of uncertainty that can arise from dynamical systems with various sorts of properties over and above the “butterfly effect.” Dynamical systems can sometimes radically alter their attractors in response to tiny changes in the parameters that specify them. They can also live for a very long time on what appears to be a robust attractor, only to abruptly jump to another attractor after a long period of time, etc. We freely admit that there indeed exists the possibility that the earth’s climate system might exhibit one or more of *these* sorts of degrees of “high sensitivity”. And if they did, then even a model that was very close, on some relevant metric, to a perfect model of the climate might make very bad probabilistic diachronic forecasts. Such a system might even give rise to massive uncertainty regarding the most qualitative forecasts. This is a real worry. In this respect, Frigg et. al. are perfectly right to draw attention to these effects. When climate scientists worry about the possibility of abrupt climate change, they are sometimes worrying about this very sort of possibility. We would be the last people, moreover, to dismiss the seriousness of worries about abrupt climate change. But none of this amounts to an argument for burden of proof shifting. None of this amounts to an argument that the burden of proof, for example, falls on the IPCC to show that the best model of the climate does not exhibit any of these relevant sorts of properties. And if there is a good argument for this claim, it will have nothing to do with either the kinds of computer simulation that Frigg et al show us, nor with the more precise mathematical proofs that Mayo-Wilson offers. It would have to come from a very close examination of the particular properties possessed by the specific dynamical systems that are thought to best represent our climate. Studying models and discovering and categorizing the different kinds of uncertainty that they can give rise to (in the way that, e.g. Stewart does) is extremely useful in that it can raise awareness about things like

[17] Many people have tried to make the point to us that these papers are not making an argument by analogy—rather that it is a formal mathematical argument—but we do not understand this claim, and, in fact, in the original appearance of the arguments, Smith (2002), the argument is specifically labeled as an argument by analogy, both in the main text and in a section heading.

the possibility of abrupt climate change. But the hawkmoth effect in particular tells us nothing of earth-shattering epistemological importance in climate science, and no result we know of in the study of toy models does anything akin to shifting any burdens of proof in the epistemology of climate science.

References

Bindoff, N. L., Stott, P. A., Achuta, K. M., Rao, M. R., Allen, N., Gillett, D. . . Zhang, X. (2013). Detection and attribution of climate change: From global to regional. Climate Change 2013: *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Stocker, T. F., Qin, D., Plattner, G. -K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. M. (Eds.). Cambridge, United Kingdom, New York, NY, USA: Cambridge University Press.

Daron, J., & Stainforth, D. (2013). On predicting climate under climate change. *Environmental Research Letters*, *8*, 1–8.

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., Mochizuki, T., Rodrigues, L. R. L., & van Oldenborgh, G. J. (2013). Initialized near-term regional climate change prediction. *Nature Communications*, *4*, 1715. http://dx.doi.org/10.1038/ncomms2704.

Fillion, N., & Corless, R. M. (2014). On the epistemological analysis of modeling and computational error in the mathematical sciences. *Synthese*, *191*(7), 1451–1467.

Frigg, R., Bradley, S., Machette, R., & Smith, L. (2013a). Probabilistic forecasting: Why model imperfection is a poison pill. In: H. Andersen, D. Dieks, W. J. Gonzalez, Th. Uebel, & G. Wheeler (Eds.), *New challenges to philosophy of science*. Dordrecht: Springer.

Frigg, R., Smith, L., & Stainforth, D. (2013b). The myopia of imperfect climate models. *Philosophy of Science*, *80*, 886–897.

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

Frigg, R., Bradley, S., Du, H., & Smith, L. A. (2014b). Model error and ensemble forecasting: A cautionary tale. Guichun, C., Guo, Chuang, Liu (Eds.). *Scientific explanation and methodology of science* (pp. 58–66). Singapore: World Scientific.

Kirtman, B., Power, S. B., Adedoyin, J. A., Boer, G. J., Bojariu, R., Camilloni I. . . Wang, H. J. (2013). Near-term climate change: Projections and predictability. *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Stocker, T. F., Qin, D., Plattner, G. -K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., & Midgley, P. M. (Eds.). Cambridge, United Kingdom, New York, NY, USA: Cambridge University Press.

Knutti, R. (2016). Climate model confirmation: From philosophy to predicting climate in the real world (In press).

Lorenz, E. N. (1975). *Climate predictability. The physical basis of climate modelling*. Geneva: World Meteorological Organisation.

Parker, Wendy (2009). Confirmation and adequacy-for-purpose in climate modeling. *Aristotelian Society Supplementary Volume*, *83*(1), 233–249.

Smith, L. (2002). *Proceeding of the National Academy of Science of the United States of America*, *99*(1), 2487–2492.

Stewart, I. (2011). Sources of uncertainty in deterministic dynamics: An informal overview. *Philosophical Transactions of the Royal Society A*, *369*, 4705–4729.

Werndl, C. (2016). On defining climate and climate change. *The British Journal for the Philosophy of Science*. (In press)