# An ineffective antidote for hawkmoths

Forthcoming in *European Journal for Philosophy of Science*.

**Roman Frigg**

Department of Philosophy, Logic and Scientific Method
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
UK
r.p.frigg@lse.ac.uk

**Leonard A. Smith**

Bradley Department of Electrical and Computer Engineering
1185 Perry Street
453 Whittemore (0111)
Virginia Tech
Blacksburg, VA 24061
USA
lennys@vt.edu

**Abstract**

In recent publications we have drawn attention to the fact that if the dynamics of a model is structurally unstable, then the presence of structural model error places in-principle limits on the model's ability to generate decision-relevant probability forecasts. Writing with a varying array of co-authors, Eric Winsberg has now produced at least four publications in which he dismisses our points as unfounded; the most recent of these appeared in this journal. In this paper we respond to the arguments of Winsberg and his co-workers, and we point out that their criticisms fail. We take this as an opportunity to restate and explain our arguments, and to point to fruitful directions for future research.

# 1. Introduction

In the sciences, mathematical models are frequently used to predict the future. In geophysics, computations are often too complicated to do by hand and computers are employed. These mathematical models target the dynamics of physical systems of interest. If one is willing to assume that physical systems themselves are governed by mathematical equations,[1] then structural model error corresponds to a difference in the mathematical structure of the model and the mathematical structure of the system. In the absence of structural model error, uncertainties in the values of parameters and the precise state of the system can be approached successfully with a number of methods (Judd and Smith 2001). These methods can quantify this imprecision with a probability distribution because parameters, states and the like are well-defined; they are merely imprecisely known. The status of these entities is not similarly well-defined in nonlinear models with structural model error (Smith 2001; Du and Smith 2012; Berger and Smith 2019). These considerations should inevitably lead to the question "what confidence can we have that the future will look anything like the model predictions?". This is an important question because in scientific practice models always have some structural model error, and, indeed, both imprecision and structural error are longstanding concerns in the prediction of both weather and climate.

In recent publications we have drawn attention to the fact that if the dynamics of a model is structurally unstable, then the presence of structural model error places in-principle limits on the model's ability to generate decision-relevant probability forecasts. The IPCC shares these concerns explicitly.[2] We have argued that both in-practice issues (imprecision) and in-principle issues (structural model error) pose problems for probabilistic modelling projects like UKCP09, which produces high-resolution climate projections up to the end of the century. Writing with a varying array of co-authors, Eric Winsberg has now produced at least four publication in which he dismisses our points as unfounded, or, as the title of one of the papers puts it, as "[t]he adventures of climate science in the sweet land of idle arguments" (Winsberg and Goodwin 2016, 9).[3] In the most recent of these publications, which came out in this journal, he and his co-authors claim to present "[a]n antidote for hawkmoths". In this paper we respond to the arguments of Winsberg and his co-workers in all four publications. We take this as an opportunity to restate and explain our argument, and to point to fruitful directions for future research.

[1] That a system – a part of the physical world – is itself governed by a mathematical structure is not an unproblematic assumption. It remains useful, however, to speak as if a governing model structure with well-defined parameter values existed and we do so in what follows. But even if no such governing equations exist, the issues of model error discussed below are no less serious.

[2] This is documented in Section 5 below.

[3] Throughout the paper we use the following abbreviations: "WG" for (Winsberg and Goodwin 2016), "GW" for (Goodwin and Winsberg 2016), "W" for (Winsberg 2018), and "NNW" for (Nabergall et al. 2019).

The paper is organised as follows. In Section 2 we explain the problem of structural model error and restate our position. In Section 3 we discuss, and reject, the claim that our argument is based on an analogy between the logistic map and climate models. In Section 4 we show that a number of mathematical objections that have been levelled against us are mistaken. In Section 5 we discuss arguments to the effect that the manifestations of the hawkmoth effect are less dramatic than we suggest they may be, and we argue that Winsberg and his collaborators oversell their case. Section 6 concludes.

## 2. The Problem of Structural Model Error

In this section we discuss structural model error. We first introduce the notions of structural model error and structural instability, and then describe their consequences (Subsection 2.1). We then turn to the question of when, and under what circumstances, we should expect to encounter structural instability (Subsection 2.2).

### 2.1 Structural Model Error and Its Consequences

Consider a system that is defined by a flow in a Euclidean state space with *n* dimensions (where n > 2). The flow is generated by the differential equation

$$\dot{x} = V(x), \tag{1}$$

where $x$ is an $n$ dimensional state vector containing the variables that specify the state of the system, $\dot{x}$ is the first time derivative of that vector (where time is taken to be continuous and to range over the real numbers ), and $V$ is the vector field that drives the change of $x$.[4] As an example consider a pendulum that can move in two spatial dimensions. The $x$ then consists of four components – the position and momentum variables for both directions in which the pendulum can move – and $V$ encodes the forces acting on the pendulum bob. A *trajectory* is the motion of a point through the state space over time, and a *flow* is the totality of all such trajectories. Other examples include electric circuits, the planets of the solar system, convection between two plates each at a constant temperature, a rotating and thermally driven rotating annulus, and a global climate system.[5]

For simplicity, assume that such a system is represented by a model that has the mathematical structure of a flow in the *same* state space.[6] That is, the model that traces the time evolution of the system's variables is assumed to have the same variables as the system itself. It is a sufficient condition for a model to be perfect that the model flow and the target flow be identical (arguably they need only be identical in the regions of the state space that are relevant to a particular situation). If the two flows are not identical, they can nevertheless be *topologically equivalent*.[7] Consider a second differential equation

[4] A rigorous discussion of the definitions and results that we present in this section can be found in Chapter 2 of Pilyugin's (1991). The main points are also stated in Section 4 of our (2014).

[5] Climate models are build up from deterministic equations. They are then run on digital computers, which are deterministic. The IPCC requires that all model outputs are completely reproducible, and so these models cannot have a truly stochastic element.

[6] In some cases we may not know what a system's variables are, if such things exists at all. For the sake of argument, we set such worries aside.

[7] Topologically equivalent flows are sometimes also called *topologically conjugate*.

$$\dot{x} = W(x), \qquad (2)$$

where $W$ is vector field and $x$ is a state as above. A *homeomorphism* is a continuous mapping between two topological spaces that also has a continuous inverse. Now consider a homeomorphism $h$ mapping the Euclidean state space onto itself. Such a homeomorphism is called a *topological equivalence* of the two flows (generated by $V$ and $W$ respectively) iff it maps trajectories of one flow onto trajectories of the other flow. The existence of such a mapping imposes stringent conditions. For instance, a topological equivalence cannot map a flow with three fixed points onto a flow with two fixed points, or map a flow with only periodic orbits onto a flow with only non-periodic orbits. If two flows are "qualitatively different", there is no topological equivalence.

If there is a topological equivalence between the model and the system, then there is a sense in which the dynamical properties of the two flows coincide: their attractors (if any) can be mapped to each other and their dynamical invariants such as Lyapunov exponents are identical. In this case, an ensemble of initial system states (representing our uncertainty in the state of the system at an initial instant of time) can be mapped into a corresponding ensemble of model states which can be evolved forward in time under the model's dynamics and then mapped back onto the system's state space so that the mapped-back ensemble accurately reflects our uncertainty about a system's state at a later time. In short, if the flow of the model and the flow of the system are topologically equivalent, then evolving information about the uncertainty in an initial condition forward in time under the model's dynamics yields the same result as evolving uncertainty forward in time under the system's own dynamics. In this situation the model can reliably trace uncertainty about the system.

This ability is lost if the flows of the model and the system are not topologically equivalent. The attractor(s) of the system need not correspond to the attractor(s) of the model, and dynamical invariants need not be identical. In this case the model cannot be used to define, much less to trace precisely uncertainty about the system.[8] Initial condition ensemble distributions[9] are frequently used to express the uncertainty about system's state at certain time. Ideally, one would be able to evolve this initial condition ensemble distribution forward in time with the model to obtain a distribution reflecting the uncertainty in a future state of the system. In the absence of topological equivalence this cannot be done because evolving an initial condition ensemble distribution forward under the dynamics of the model

[8] This result is established in Judd and Smith (2001, 2004) and in Smith (2000). The proposition is the result of lengthy mathematical argument, the leading idea of which is follows. Assume, as is the case in any physical experiment, that our observations are noisy; that is, we can only make measurements with finite precision. Two states are indistinguishable at $t_0$ (now), iff any number of discrete observations on the system at any time before $t_0$ does not provide information sufficient to distinguish between them. The argument then proceeds in two steps. The first step is to show that to produce an accountable probabilistic forecast at $t_0$ one has to use a probability distribution that (a) is consistent with observations, which amounts to using a probability distribution over the indistinguishable states at $t_0$, and that (b) takes into account the long-term behaviour of the dynamics (i.e. its attractor). If one has a perfect model, a set of indistinguishable states exists and accountable forecasts can be produced. The second step of the argument consists in establishing that if a nonlinear dynamical simulation model is not perfect (i.e. if the model's and the system's dynamics are not topologically conjugate), then the set of indistinguishable states is empty with a probability equal to one. One cannot have probability distribution on an empty set, and so there is no distribution that produces an accountable forecast.

[9] Defined by the model given the observations.

may not provide a probability distribution reflecting the distribution of uncertainty in a future state of the system.[10] This inability is illustrated in Section 3 of our (2014). We call a probability distribution *accountable* iff the forecast probabilities reflect the relative frequencies with which outcomes are observed over future times.[11] For a probability distribution to be decision-relevant it has to be accountable. This is because if a forecast does not reflect the actual frequency of outcomes, the forecast can be misleading (for instance by regarding something that rarely happens to be a frequent occurrence), and decisions made on the basis of the forecast may not have the desired effects.[12] Our point can then be stated thus: in the absence of topological equivalence, it cannot be taken for granted that models provide accountable (and thereby decision-relevant) probabilities about the uncertainty of a future state of the system.[13]

Two remarks are in place. First, this point must not be paraphrased as "models are unfit for decision making", or as any other statement to the effect that models are useless wholesale. Our claim is restricted to probabilities about the uncertainty of a future state of the system; whether models provide other kinds of decision relevant information is, at this point, left as open problem (to which we return in Section 5). Second, there is an important question about how one should react to this point. If one knows that a model is not topologically equivalent to its target, and if one therefore knows that it cannot be taken for granted that model probabilities are accountable, should one renounce the use of model probabilities about the uncertainty of a future state of the system? Our cautionary position is that we should.

To capture structural model error, consider the set of all vector fields on the Euclidean state space that have a certain desirable property. In the current context, the standard choice for that property is that vector fields are continuously differentiable at least once. Then introduce a distance on this set. This distance says how far apart two vector fields are.[14] Intuitively, if $\rho$ is small, the two fields are similar; the values of $\rho$ increase if the fields become more dissimilar. We are now in a position to define structural stability. The flow generated by (1) is *structurally stable* iff, given a small real number $\varepsilon > 0$, there always exists a small real number $\delta > 0$ so that the following condition holds: for any vector field

[10] For an extensive discussion see Judd and Smith (2004). We note that a lack of topological conjugacy does not preclude that the model might have trajectories that shadow trajectories of the system for an arbitrarily long but *finite* duration. However, whether or not such trajectories exist, and, if so, the duration they will shadow, depends on the particulars of the system and of the model, and even the initial state of the system (Smith et al. 1999). Such trajectories cannot be exploited for predictions. Shadowing says that there is a trajectory that stays close to observations but it is usually not known which trajectory this is, and the existence of one trajectory does to warrant that claim that the ensemble behaves as the system does, which would be required for probability forecast. For an introductory discussion of shadowing see P. Smith (1998, Ch. 4).

[11] Accountability is defined in Smith's (1995). The property Smith calls "accountability" is now a way that forecast accuracy is measured. For a discussion see Jolliffe and Stephenson (2012).

[12] In our (2014) we put the point in terms on non-linear models and said that if a nonlinear model has structural model error, its ability to generate decision-relevant probabilities is compromised. The mathematical theorems we mention, both in our (2014) and in the current section, are general and apply to all systems that are governed by equation (1). The mention of non-linearity was owed to the pragmatics of modelling. The systems that we are interested in – the global climate, weather, the motion of planets, and so – are all best modelled as non-linear systems. We aimed so way what the implications of structural model error for *these* systems is. Since the result mentioned in this section are general, non-linear system are obviously within their scope. But we emphasise that the results are not in any way restricted to non-linear systems.

[13] *A fortiori*, it cannot be taken for granted that models provide accurate point predictions.

[14] Technically, the relevant class is $C^1$, the class of vector fields that are continuously differentiable at least once. For a formal definition of $\rho$ see Pilyugin (1991, 110).

in the chosen set with $\rho(V,W) < \delta$ there exists a topological equivalence $h$ of the flows generated by (1) and (2) so that $r(x,h(x)) < \varepsilon$ for all points $x$ in the state space, where $r$ is the metric on the state space. Intuitively structural stability means that if two vector fields are close (in the sense of not being further apart than $\delta$), then the two flows that they generate are also close (in the sense of not being further apart than $\varepsilon$). The flow generated by (1) is *structurally unstable* iff it is not structurally stable.[15]

If a system's trajectories for infinitesimally close initial conditions can diverge exponentially-on-average in the limit of infinite time, then the system exhibits what is known as sensitive dependence on initial conditions (SDIC). The manifestations of SDIC in chaotic systems are commonly referred to as the *butterfly effect* (Smith 2007). If a system is structurally unstable, the flow of an arbitrarily close vector field may diverge from the flow of the system (in sense specified in the previous paragraph). This was the motivation for Thompson (2013) to dub the implications of the lack of structural stability the *hawkmoth effect*.[16]

The issue of structural stability has important implications for scientific modelling.[17] Consider a system whose flow is generated by $V$ and let the system be represented by a model whose flow is generated by $W$. If the model's or the system's flow is structurally unstable, then even a small difference between the model's vector field and the system's vector field can result in the two flows not being topologically equivalent. In his book on chaotic dynamical systems, Devaney points out that such differences are to be expected because in modelling a system, "certain assumptions will have been made, and certain approximations and experimental errors will be present", and as result  the model will be "only an approximation to reality" (1989, 53). In other words, $V$ and $W$ will not be identical. This can lead to difficulties because

> if the dynamical system in question is not structurally stable, then the small errors and approximations made in the model have a chance of dramatically changing the structure of the real solution to the system. That is, our 'solution' could be radically wrong or unstable. If, on the other hand, the dynamical system in question is structurally stable, then the small errors introduced by approximations and experimental errors may not matter at all: the solution to the model system may be equivalent or topologically conjugate to the actual solution. (*ibid.*)

This problem is also clearly stated in Abraham and Marsden (1978, xix-xx) and in Katok and Hasselblatt (1995, 287). So, if a model is structurally unstable, then the system, even if it is close to the model, can have a qualitatively different dynamics.

Abraham and Marsden emphasise the importance of this point for scientific modelling and report that considerations of this kind led Duhem to list the "criterion of *stability*" alongside criteria like verifiability of predictions and agreement with data as a criterion of theory

[15] In passing we note that maps that can be defined from flows, for instance through Poincaré maps and time-discrete versions of a flow (stroboscopic maps). Maps inherit the flow's stability properties: if a flow is structurally unstable, then so are all maps that derive from the flow.

[16] This leaves open how exactly these effects should be characterised or defined. The definition of the butterfly effect and dynamical chaos has been the subject of a heated debate; see Devaney (1989), Ott (1993) and Werndl (2009) for a review and a proposed definition of chaos.

[17] In short, as uncertainty in initial condition decreases, in the case of the butterfly effect the forecast remains informative farther and farther into the future (and we can determine its evolution). This is not the case with the hawkmoth effect, which persists even when epsilon equals zero (in which case there is no forecast error due to SDIC).

choice, where the criterion requires the "stability or continuity of the predictions, or their adequacy, when the model is slightly perturbed" (1978, xix, original emphasis). They note that while this criterion is not often discussed explicitly, it "functions as a tacit assumption, which may be called the *dogma of stability*". If stability cannot be proven and empirical values for coefficients have to be used in a model, we are in an epistemically difficult position because "[p]robably the physicist must rely on faith at this point, analogous to the faith of a mathematician in the consistency of set theory" (1978, xx, original emphasis).

This brings home the point that structural model error matters. It is curious, though, that all these statements appear in mathematics books, while there does not seem to be any recognition in the philosophical literature on modelling that structural model error raises methodological issues that ought to be taken seriously.[18] An important aim of our papers was to bring this problem to the attention of philosophers. Where leading practitioners feel that there is a tacit assumption at work which, if absent, forces scientist to rely on faith when using models, there is a challenge that philosophers should rise to.

### 2.2 When Should We Expect to Encounter Structural Instability?

One might now try to deflect this challenge by pointing out that what has been established so far is the *conditional* claim that if models are unstable, then difficult issues arise. This will have to occupy us only if the models that we are interested indeed are structurally unstable. But when should we expect to be faced with structural instability, and how common an occurrence is it? At the end of Section 4 of our (2014) we have briefly mentioned three reasons – each individually sufficient – for thinking that structural instability is common, and we now want to discuss these in more detail.

The first reason is that the relevant models do not in general satisfy the conditions necessary for structural stability. It is a deep theorem in the theory of dynamical systems that a flow is structurally stable iff it satisfies *Axiom A* and the *Strong Transversality Condition*.[19] Axiom A essentially says that the system is hyperbolic, and the strong transversality condition says that stable and unstable manifolds must intersect transversely at every point. The flows of interest in atmospheric modelling (and this includes climate) do not, in general, satisfy these conditions.[20] These general mathematical results have wide-ranging implications, and the lack of structural stability disappointed Hirsch and Smale, who report that related insights "dismayed" Poincaré (Hirsch and Smale 1974, 321).[21]

The second reason is the results obtained in Judd and Smith (2004). Judd and Smith show that if the model is imperfect (i.e. has structural model error), then it is almost certain that no trajectory of the model is consistent with an infinite series of observations. This implies that it is not possible to estimate the projection of a system state using trajectories or form accountable ensembles.

[18] Indeed, Abraham and Marsden note that the "traditional mutuality of mechanics and philosophy has declined in recent years" (1978, xix), and little has changed since 1978, certainly as far as the study of structural stability is concerned.

[19] This result carries over to maps. See Pilyugin (1991) for details.

[20] Systems need not be large or complex to face these issues. Devaney notes that there are "simple examples of systems such as the Lorenz system from meteorology that are 'far' from being structurally stable. These systems cannot even be approximated […] by stable systems." (1989, 53-54).

[21] Poincaré's own pronouncements are also documented in Barrow-Green (1996) and Smith (2001, 2002).

The third reason is a series of theorems regarding structural stability that have been proven in the dynamical systems literature since the 1960s.[22] Smale (1966, 491) poses what he calls “the problem of structural stability”: “are the structurally stable differential equations dense in the $C^1$ topology in all (first order, ordinary, autonomous) differential equations?”. The equations Smale talks about are equations like (1) and (2), and the $C^1$ topology is a distance on the set of vector fields.[23] Smale then states that he gives a “negative answer” to this problem, namely that “structurally stable systems are not dense” (*ibid.*); i.e. they are not dense in the set of systems that are defined by equation (1).

To see the relevance of this theorem, we have to unpack some of the notions that appear in it. Doing so is important not only to grasp its mathematical content, but also to see its epistemic relevance.[24] Intuitively, a set is open if its boundary does not belong to it. Think, for instance, of the surface of circle without the periphery. Now consider a set $A$ that is a subset of a larger set $S$. Set $A$ is dense (in $S$) if $A$ intersects every nonempty open subset of $S$. Intuitively this means that for every element of $S$, the element is either in $A$ or is arbitrarily close to an element of $A$. As an example, think of the rational numbers, which are dense in the real numbers, meaning that every real number either is a rational number or has a rational number arbitrarily close to it. The fact that structurally stable systems are not dense then means that not every system is arbitrarily close to a stable system and that there are non-empty sets in the space of vector fields in which there are no stable systems at all. Coming back to Devaney’s characterisation of the process of modelling, the *best* we can do is to come up with a model that is close to the true mathematical representation of the target system. If the true mathematical representation is unstable, then a model close to it can exhibit substantially different dynamical behaviour. If the true mathematical representation is stable, then we cannot expect a nearby model to be stable because stable models are not dense, which, again, can result in the model behaving differently than the system.

Unqualified talk of “expectations” may strike some as unsatisfactory because unless we have a means to quantify expectations, ideally with probabilities, we don’t really know what to expect. One way to try to address this demand is though the notion of a property being “generic”. Consider a property $\Pi$ and let $P$ be the subset of elements of $S$ that have $\Pi$. The notion of set being “generic” tries to capture the intuitive idea that the elements of the set $P$ are “typical”, or that “most” elements have $\Pi$ (Katok and Hasselblatt 1995, 287-88). The technical definition of “generic” proceeds via the notion of a set being a residual: a set $A \subset S$ is a residual set iff $A$ is the intersection of a countable family of open dense subsets of $S$ (Abraham and Marsden 1978, 16). Then, a property $\Pi$ is generic if $P$ contains a residual set (*ibid.*, 532). Unfortunately, this definition is hardly intuitive. As Katok and Hasselblatt explain, the motivation for it is that in a complete metric space the countable intersection of open dense sets is dense (Katok and Hasselblatt 1995, 288). If so, a generic set is one that is dense, and we have seen above what that means. With this in place, it then

[22] For surveys see Abraham and Marden (1978), Katok and Hasselblatt (1995), and Pilyugin (1991). Several important papers on the issue are in the second volume of Smale’s collected works (Cucker and Wong 2000). These theorems have so far gone unnoticed in the philosophical literature on models. Yet, a systematic exploration of all these results and their consequences for scientific modelling is beyond the scope of this paper. In the remainder of this section we briefly discuss one of these theorems. We hope that this discussion shows how important these theorems are for the epistemology of modelling, and that the it will trigger interest in the subject matter.

[23] For a discussion see Abraham and Marsen (1978, 532).

[24] We give intuitive characterisations of the relevant concepts. Rigorous definition can be found in any textbook on topology, for instance Jänich (1984).

follows that structurally stable systems are not generic the set of systems that are defined by equation (1) (Abraham and Marsden 1978, 535).

Some may see the restatement of the claim in terms of “generic” as a step in the right direction, but they would prefer to have statement of the kind: if we pick a model at random from the set of models,[25] the probability of this model being stable is zero (or close to zero). A restatement of this kinds may seem within reach due to the fact that the complement of a generic set is of “first category” and that a collection of sets of first category “can be viewed as a topological analog of the collection of sets of measure zero” (Katok and Hasselblatt 1995, 288). So structurally stable sets would then be of measure zero, which, with a suitably chosen measure, could be translated into the desired claim concerning probabilities.

The extent to which and under under what conditions, this shift from the topological notion of genericity to the measure-theoretic notion of probability is possible is, to be best of our knowledge, an open question. Katok and Hasselblatt are careful to say that first category sets are a topological analogue of measure zero sets; they do not say that they *are* measure zero sets. And the passage from topology to measure theory is not without perils. Katok and Hasselblatt note two problems (*ibid.*). First, in finite-dimensional cases there are natural Lebesgue measures on the relevant spaces, but it turns out that there exist generic sets of measure zero. Second, in the case of systems with an infinite-dimensional state space, there are no natural measures at all. This raises interesting and important questions. Do we really need probabilities for decision making? If not, what sort of framework for decision making should we use when there are no probabilities? If we do need probabilities, where do we get them from? To what extent can we build on a topological analogy? These are questions for fruitful future research.[26]

In sum, we have introduced structural model error and pointed out that in the absence of topological equivalence models do not provide accountable probabilities about the uncertainty of a future state of the system. We have then given three reasons why structural model error should be taken seriously in the context of scientific modelling. We have also drawn attention to open questions and indicated some avenues for future research. We will now turn to the arguments of Winsberg and his co-authors, who dismiss our points as unfounded.

## 3. No Analogy with the Logistic Map

The general mathematical concepts discussed in the last section may be counterintuitive and the consequences of the theorems can be difficult to grasp. For this reason we followed the lead of May (1976) and used the logistic map to *illustrate* the relevant features with more easily visualized dynamical behaviours. We discuss a thought experiment involving Laplace’s Demon and two of his apprentices (2014, Sec. 3).[27] In this thought experiment the

[25] Ideally, from the set of available models that will run on today's hardware.

[26] We note that this is a point where the debate about modelling can fruitfully interact with discussions in formal epistemology. In the wake of Belot’s (2013), a discussion has ensued about the relationship between measure-theoretic and topological notions of size in Bayesian epistemology, with contributions from Cisewski *et al*. (2018), Elga (2016) and Nielsen and Stewart (2019), among others. Future discussion about the relevance of structural instability ought to take insights gained in this discussion on board. We are grateful to an anonymous referee for pointing this out to us.

[27] These charaters were introduced in Smith (2007).

demon knows the true dynamics of a system, which is specified by what we call the *quartic map*, a function of 4th order. The junior apprentice models the situation with the well-known *logistic map*, a 2nd order function. So the model exhibits structural model error. The parameters of the system (quartic map) and the model (logistic map) are such that the one-step-error of the model dynamics with respect the system dynamics is of the magnitude of one part in a thousand. We run a series of computer simulations showing that even though the model-dynamics and the system-dynamics are similar (in that the model has small one-step-error almost everywhere), the trajectories of a distribution of initial conditions moved forward in time under the system-dynamics differ markedly from the trajectories of the same distribution of initial conditions moved forward in time under the model-dynamics. It is common in computational contexts to employ an ensemble of states to gain information regarding the dynamics of the probability distribution from which they were drawn (evolving the probability distribution analytically is rarely possible). Adopting this convention, we considered an ensemble of 1024 initial conditions and showed that the distribution that these points reflect behaves very differently under the model dynamics than under the system dynamics. Since the state spaces of both the model and the system are one-dimensional, one can present the outcomes in the form of easily comprehensible graphs. The thought experiment thus *illustrates* some effects of structural model error in a simple and intuitive setting.

Winsberg and Goodwin describe our argument as follows:

> The form of their [Frigg and Smith's] 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. (WG 9)

They then expand on this and explain that we take the "logistic equation" as the "base case" from which we "generalize" to "other modelling projects", which we do "because of their similarity to the base case". We are reported to "contend that most time evolutions are relevantly similar to the logistic equation" from which we are said to draw the conclusion that the effects observed in the logistic map also apply in "modeling projects of climate scientists" (WG 11-12). This analysis is repeated in (Goodwin and Winsberg 2016, 1125-27). In Winserberg's recent book (2018) it reappears as the second route to the hawkmoth effect (W 232). Accordingly, Winsberg and Goodwin conceptualise the task of evaluating our argument as assessing "the strength of the argument by analogy that the authors offer" (WG 13), and later claim to have debunked it by showing that the analogy is weak.

This is wrong. Our argument is not an argument by analogy with the logistic map, and none of our conclusions rest on similarity claims between the logistic map and other models. Winsberg seems to acknowledge this in passing in the appendix of his book (W 235), but throughout his paper with Goodwin (WG and GW), the authors conflate a *pedagogical illustration* of a theorem with the *scope* of the theorem. The mathematical results we described in Section 2 are general.[28] They apply to any system that has the mathematical structure described in the last section, and any such system is structurally stable iff it satisfies both Axiom A and the Strong Transversality Condition. This is a general mathematical result that applies to systems because they have the requisite mathematical

[28] To be painfully clear that our use of a one-dimensional system-model pair was merely illustrative of the effects of model error, we note that the relevant theorems on structural stability apply only in higher dimensional systems even though structural model error can, of course, arise in models of any dimension.

structure and not because they are in any way similar to the logistic map. Those who set out to show that models of a particular system do not face the effects of structural model error have three options: (a) they can show that the results we cite in Section 2 are mathematically flawed; (b) they can show that the mathematical results do not apply to the system of interest; or (c) they can show that the system of interest satisfies Axiom A and the Strong Transversality Condition and is therefore structurally stable.

Demonstrating any of these would provide great insight. Winsberg and Goodwin have not, however, attempted to so. Instead they go to great lengths to assess the strength of the (wrongly) alleged analogy between the logistic map and modelling projects in climate science. This is tilting at windmills: no such analogy has been invoked to prove anything anywhere. We use the model-system pair consisting of the logistic map and the quartic map to *illustrate* what can happen when small structural errors occur in nonlinear dynamical systems. This one-dimensional example allows for visual illustration because pictures are easier to draw in such cases, which allows us to present the mathematical points in an intuitively accessible way. But the example is merely an illustration. Our general claim in no way rests on the example and can be made without ever mentioning either the logistic or the quartic map, or the other model-system pairs mentioned in (Judd and Smith 2004).

Finally, in our example we calculated the relative entropy of 2048 probability distributions that are evolved both under the system's and under the model's dynamics. We found that after eight time steps model probabilities and system probabilities have low relative entropy in only about a quarter of the cases; the relative entropy increases in the other three quarters of cases. The specifics of the example don't generalise: we cannot infer that in other cases one quarter of model distributions will remain close to system distributions or that the concrete values of the relative entropy will be seen in the example, nor can we infer anything about the timescales on which a growth of relative entropy can be observed. These details will depend on the specifics of the model-system pair under investigation. What we do claim is that the *qualitative* overall pattern generalises: one can expect a majority model-system distribution pairs to drift apart and hence have increasing relative entropy. For this not to happen, it would have to be the case that the two distributions either never separate or that they re-converge, which is something that generally does not occur if the model and the system have a dynamics that is not topologically conjugate.

## 4. Mathematical Considerations

In two recent publications Winsberg (W), and Nabergall, Navas, and Winsberg (2019) take a different line. Rather than attributing to us an argument by analogy between the logistic map and climate models, they construe the analogy as being between sensitive dependence on initial conditions and structural model error. Thus, we are reported to hold that "when structural stability is absent […], errors in output depend on errors in model structure in a way that is tightly analogous to the phenomenon of sensitive dependence on initial conditions in chaotic systems" (NNW 3-4). In his book Winsberg summarises what he takes to be our view by saying that "the hawkmoth effect is supposed to be a model-structure analog of the butterfly effect" (W 232, cf. 71). Their project then is to assess the strength of that analogy. We welcome the focus on the formal aspects of the problem, which has so far gone largely unnoticed, and we agree that progress will most likely be made through paying careful attention to the relevant mathematical results. Unfortunately, NNW's discussion suffers from a number of inadequacies that undermine their conclusions.

NNW begin by framing the problem as the question whether "model error destroys forecast skill faster than the ordinary or 'classical' chaos", and they attribute to us an affirmative answer to this question.[29] This is a misattribution. We never said anything about how fast the hawkmoth effect becomes manifest; we did not engage in comparisons between the speeds of the butterfly effect and the hawkmoth effect; and we never claimed that the hawkmoth effect was "faster" than "ordinary chaos". Indeed, there are good reasons not to engage in any such a comparison because how fast effects manifest themselves vary with the particulars of the system. If the aim is to understand the relation between the two effects in general, then focussing on which effect is faster is a red herring.[30]

Setting aside matters of speed, we can identify two claims in NNW. First, they argue that the implications of the hawkmoth effect for forecasting skill are much more benign than the implications of the butterfly effect, and that the two are therefore not "tightly analogous". Second, they question whether we have provided valid evidence for thinking that the hawkmoth effect is relevant in the context of climate modelling. We address these claims one at a time and argue that they are mistaken.

NNW argue that there is a profound mismatch between the butterfly effect and the hawkmoth effect. For structural model error to be the analogue of SDIC, it would have to be the case that "for almost any $\phi \in \Phi$", where $\Phi$ is "a space of time-evolution functions", a small change in $\phi$ would result in a significantly different time evolution (NNW 11). This, so the argument continues, is not the case because structural instability only warrants the much weaker claim that there exists "one trajectory" that can be deformed by more than $\varepsilon$ (*ibid.*). In more detail: "Rather than requiring that *most* trajectories […] go *far* away, it [the absence of structural stability] only requires that *one* trajectory go *more than a very small epsilon away*" (NNW 12, original emphasis). The absence of structural stability is therefore reported to be much weaker than SDIC, and they summarise their findings as follows: "*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*" (NNW 14, original emphasis). So, according to NNW, the contrast between SDIC and the absence of structural stability is that SDIC establishes that *almost all* initial conditions diverge, while the absence of structural stability only guarantees that *there exists one* time evolution function for which the trajectories differ.

How do NNW reach this conclusion? The standard definition of SDIC is that a topological dynamical system consisting of phase space *X*, a metric *d*, and a time evolution $\phi_t$ exhibits SDIC at $x \in X$ iff there exists an $\varepsilon > 0$ such for every $\delta > 0$ it is the case that there exists a $y \in X$ with $d(x, y) < \delta$ and $d(\phi_t(x), \phi_t(y)) > \varepsilon$ for some time *t*. In other words, there is a distance $\varepsilon > 0$ such that no matter how small a region around *x* we consider, that region will always contain at least one state *y* that lies on a trajectory that will eventually move more

[29] The question whether structural model error destroys forecast skill "faster" than ordinary chaos is presented as a crucial issue throughout the paper, and the discussion makes it clear that NNW attribute an affirmative answer to us (see NNW 1, 3, 4, 5, 6, 10, 13, 15).

[30] There are of course specific cases in which structural model error accumulates faster than initial condition error. This can happen, for instance, in weather forecasting where, in a perfect model scenario, the system can remain predictable for a couple of weeks, while real forecasts do not exhibit that much skill due to structural model errors (we are grateful to an anonymous referee for point this out). Our point is that this is not a general feature of structural model error.

than $\varepsilon$ away from $x$. If this holds for all $x \in X$ (or subset of $X$ that maps onto itself under t dynamcis) then the system exhibits SDIC on $X$. This definition is standard both in the physics literature and in the philosophy literature, and when a system is said to exhibit SDIC then *this* definition is appealed to.[31]

NNW state this definition (on pp. 6-7), but immediately after the definition they add that "[w]e could also strengthen the definition of sensitivity to initial conditions to require that *almost all* such states have this property" (NNW 7, emphasis added). This is not just a casual throw-away remark. In fact, this then becomes their go-to definition of SDIC on which they base their argument. Accordingly, their Definition 2 states that a system exhibits SDIC if "almost all elements $y \in X$" lead to divergent trajectories (NNW 7). This definition is also repeated in Winsberg's book (W 233).

But moving from "there exists a $y \in X$" to "almost all $y \in X$" is huge departure from the standard definition of SDIC in the physics and the philosophy literature, and one that is problematic for two reasons. First, NNW give no justification for this change. While it is true that *some* systems are such that the "strengthened" condition holds true in them, the mathematical results of nonlinear dynamics are established with the standard definition (based only on an existential claim). But one cannot base a general characterisation the difference between SDIC and structural model error on a version of SDIC that is true only of some but not all systems that have SDIC. Second, NNW's "strengthened" definition is the crucial ingredient that drives their entire argument. As we have seen, they claim that structural instability is weaker than SDIC because it only establishes that there exits *one* trajectory that will go *more than a very small epsilon away*, while SDIC shows that *almost all* trajectories go *far* away. To reach this conclusion they have to appeal to their "strengthened" definition. But this difference does not exist in the standard mathematical definitions of the relevant notions; NNW have artificially created this difference by altering the standard mathematical definition of SDIC. The entire massive difference they claim to have uncovered between SDIC and structural instability is an artefact supported by a tweaked definition!

A further problem with NNW's argument is that it is unclear what the qualification "almost all" could even *mean* in the context of a topological dynamical system. "Almost all" is a measure theoretic concept that is *undefined* in a topological system. NNW admit this in a footnote and say that they state their definition "without there being specific mention of a measure", but they immediately assure the reader that "[t]his 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" (NNW 7). Neither do they say what they mean by a measure being "conditional on a specified metric"; nor do they give a reason for the claim that taking the Lebesgue measure is standard. In fact, measures have to be invariant under the system's dynamics for statements about "almost all" initial conditions to be meaningful, and the Lebesgue measure is *not* invariant under many dynamical laws. And even if one were to restrict attention to invariant measures, the measure of sets can vary depending on what measure is used, and whether a set of initial conditions qualifies as "almost all" will depend on the choice of the particular measure.

[31] As regards the definition in the physics literature, see, for instance, the classic text by Devaney (1989, 49), Hirsch, Smale and Devaney (2004, 338) and Wiggins (2003, 574); as regards the philosophy literature see, for instance, P. Smith (1998, 15) and Werndl (2009, 205).

Their claim that the definitions of SDIC and structural stability imply that SDIC is maximally and structural instability is maximally good is odd because both definitions make existential claims and because SDIC allows for informative forecasts on arbitrarily long timescales when decreasing uncertainty in initial conditions, while SME does not. But NNW seem to take it for granted that there will always be such an asymmetry when they ask the rhetoric question: "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 should such a weirdly concocted example shift any burdens of proof of the kind the LSE group demand of us?" Those who decide to discard the principled arguments we presented may still want to pay attention to the practice of modelling, where examples of "small model errors of the kind we would expect" are not hard to come by. Gettelman at al. (2019), for instance, report that the equilibrium climate sensitivity in the most recent version of the Community Earth System Model (CESM2) is significantly influenced by how aerosol cloud interactions are modelled.[32]

NNW accuse us of conflating two claims: "failure to stay arbitrarily close is not the same thing as being guaranteed to go an arbitrary (bounded) distance away. But in analogizing absence of structural stability to SDIC, the LSE group are engaging in exactly this conflation" (NNW 10, cf. 11). We make no such conflation, and NNW misrepresent our argument. We are not saying that the lack of structural stability guarantees trajectories to "go an arbitrary (bounded) distance away". What we are saying is that in the absence of structural stability it is always *possible* that trajectories do so (and we cannot see it coming), and that it is therefore always *possible* that forecasts are misleading; our illustrations show what *can* happen in situations of structural model error, not what we are somehow guaranteed to see. The point of contention between NNW and what they call "the LSE group" seems to be how to react to this situation. Our position is that models should be regarded as unreliable in such situation unless there is additional evidence to support their predictive capability in the specific context and on the relevant timescale. NNW seem to take issues with this cautionary attitude when reject our "shift in burden of proof" as "overstated" (NNW 4). This is a difference in the attitude to risk. How much risk one is willing or able tolerate may well depend on the context. We are concerned with forecasts (or projections, more about this later) at a local scale. This is also the scale at which businesses like insurances operate, for instance when they sell policies against storm damage. In the UK, the national regulators oblige insurance companies to provide evidence on whether or not there is no more than a 1/200 chance that an event occurs. If you know that you face structural model error, making this case will involve more than just saying that nothing has undermined forecasts so far.

What should one do in such a situation? NNW conclude "*that we are forced to examine the empirical evidence on a case-by-case basis*" and that we "*need to look carefully at similar evidence, and at the decision maker's context, to decide what models are decision relevant and can produce decision relevant probabilities even when the likelihood of very small model error is high*" (NNW 4, original italics). We agree with that. But we insist that if the evidence has not been examined in this way, then the cautionary note for the decision-maker remains.

Let us now turn to NNW's second argument, that we have provided no valid evidence for thinking that the hawkmoth effect is relevant in the current context. We are reported to

[32] Thanks to an anonymous referee for suggesting this example.

present a "very weak argument and at worst a confusion" (NNW 13). NNW's reasons for thinking so are, first, that "Smale himself certainly never uses the term 'generic,' nor any term that we would regard as being a close cousin" and, second, that "density is a topological notion with no obvious measure-theoretic implications" (*ibid*.). This is false. Readers of Volume 2 Smale's of collected papers (Cucker and Wong 2000) will find ample mention of the term, and they will encounter publications with titles like "Stability and Genericity in Dynamical Systems". Smale notes that "attempts at solving the problem of structural stability, guide one toward the study of the generic or general dynamical systems in contrast to the exceptional ones" (*ibid*., 616), and Smale proves the "nongenericity" of certain kinds of stability (see, e.g., *ibid*., 735). As regards NNW's second point, they are correct in pointing out that Smale's results are topological, but they fail to elaborate on why this is supposed to be an argument against them, and why an argument that is not a "confusion" would have to be couched in measure-theoretic terms.

NNW then go on to claim that Smale's result that we cite in our (2014) "would show that structural stability is generic in certain settings". We have discussed this result in some detail in Section 2. If NNW think that this should be interpreted as establishing that structural *stability* is generic, then the reader would need an explanation of how they reach this conclusion. This said, we agree that Smale's results, as well as other mathematical results concerning structural stability, deserve more attention from philosophers than they have hitherto received, and much could be learned about the epistemic issues that arise in connection with models by studying the implications of these results for predictive tasks.

## 5. Matters of Significance and Scope

There is a question of how significant the points we made in Section 2 are. A number of arguments in the papers by WG try show that the manifestations of model error are more benign than they take us to be saying they are.

In a first move WG try to establish that structural stability as defined in Section 2 is irrelevant in practical applications. They note that 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" (WG 13). They then claim that the reference class we are looking at is too large:

> 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 modelled by second order equations, it is not at all clear that the analogy between the pond of fish and the earth's climate holds […]" (WG 14)

So the argument is that we consider structural variations in a class of models that is too large because systems that are governed by "physical laws" obey "second order equations" (in passing we note that the function we use is of 4th and not of 5th order).[33] A demonstration that a function is unstable in that large class is therefore irrelevant because one knows that the real system is located in a much smaller class.

[33] A similar point is made in (NNW 24).

Winsberg and Goodwin are right in pointing out that structural stability is a property that a dynamics has with respect to a certain reference class, and it would constitute momentous progress if one could (a) narrow down the reference class that is relevant for the Earth's climate (or physics in general) and (b) show that the time evolution of the Earth's climate is in fact stable in that relevant reference class.

WG have achieved neither one nor the other. We do not know what WG have in mind when they say that physical laws dictate that the Earth's climate is best modelled by a second order equation. Many laws of physics do not have the mathematical form of a second order equation (think of the law of gravity, the law of electrostatic attraction, and the Stefan-Boltzman law, which plays a fundamental role in atmospheric dynamics). Furthermore, later in the paper they note that "there are far too many open questions about climate modelling. 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" (WG 14). If so, there is no way to support their claim that relevant physical system are governed by second order equations.

The next argument is that structural instability becomes irrelevant once one focuses on the correct dynamical properties. WG submit that our focus is too narrowly on point predictions and as soon as one looks at a larger class of predictions the problem goes away. This is because "[i]t 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" (WG 15). They cite Fillion and Corless (2014) as showing that "almost every model that has any degree of empirical confirmation will have some statistics that are relevantly stable under any class of perturbations" (*ibid*.).

There can be predictions or features of distributions, as well as qualitative aspects of the finite time dynamics, that are stable under small perturbations either in model parameters or in model structure. We have never claimed that structural model error makes models useless wholesale, and we have drawn attention to this ourselves (see, for instance, our 2014, 48-50): models can, for instance, provide physical understanding of large-scale processes. There are, however, no blanket answers to how understanding is gained and an analysis of what we can learn from models will have to proceed case-by-case. We are, however, unable to decipher WG's position on this issue with precision. They point out that it is "perfectly possible" that an unstable system allows for "kinds of predictive tasks" to be stable, and that models with "any degree of empirical confirmation" will have "some" statistics that remains unchanged under perturbations. WG's wording here is too unspecific to address the specific concerns about stability which they criticise. To put a specific predictive task of concern to us – probability forecasting of, for instance, the temperature on the hottest day in central London in 2080 – on a firm epistemic footing it is not enough to know that it is *possible* that there is *some* statistic that is invariant under some unspecified perturbation on some unspecified time scale. Rather, one must establish that the specific statistic that is *relevant to that particular task* is stable. Indeed, for each particular predictive task one would first have to identify the relevant dynamic properties that would have to be stable for the task to be carried out successfully, and then one would have to show that the relevant properties are indeed stable on desired time scales as required. This is a task that has to be carried out case-by-case, and focussing on such properties is the way forward to put model results on firm epistemic grounds. WG do nothing of that kind.

WG suggest that dimensionality is key to any response to this challenge and claim that "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" (WG 15). Neither reasons nor citations are given for this "almost always" claim. As we have seen in Section 2, the general theorems concerning structural stability apply to any system with a state space of more than two dimensions. It would therefore be of value to know the scope of WG's claim that large systems are less likely to manifest instabilities in a nontrivial way.[34]

WG then draw attention to time scales and point out that "it is entirely unclear what the relevant time scales would be for such instabilities [i.e. ones due to the hawkmoth effect] to manifest themselves" and that even if one could establish a priori that models are unstable "such an a priori argument would not provide us with anything like the resources for determining what the relevant notion of 'short run' is" (WG 14). We agree that questions of time scale matter for decision makers and have drawn attention to this ourselves (2014, 48). That said, the lack of topological conjugacy implies that the model attractor differs from the system attractor, and this holds consequences at all lead times, as discussed in Smith (2001), Judd et al (2008) and in Berger and Smith (2019).

We also note that the lack of a priori arguments cuts both ways. There are no a priori arguments regarding the lead times on which structural model errors manifest themselves. As in the case of initial condition uncertainty in a structurally perfect model, whether such uncertainties have an impact on a certain predictive task is matter that has to be decided on a case-by-case basis; the conclusion will depend both on the system and the task at hand. WG provide no argument for the conclusion that time scales in the case we mention (UKCP09's local climate projections) are such that structural model errors do not manifest themselves; nor do they offer arguments that relevant lead times are unproblematic in other cases. The issue of lead times on which climate models are informative is an interesting open issue that should receive more attention than it has received so far. That said, the systematic differences between today's best climate models (even on global scales) and the nonlinearity of the relevant physics in those models suggest that probability predictions for the wettest day of 2099 at a 25 square km resolution are untenable. Lower bounds on current structural model error suggest in-practice limits which swamp a priori timescales for in-principle limits due to structural stability. This is a reason why climate modellers today are concerned about structural model error.

The next issue concerns the scope of the argument. According to WG we have "provocative moments" when our argument apparently shows that "it is safe to say that much of what climate scientists claim to know would have to be regarded as untrustworthy" (WG 10). This implies that "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 modelling endeavors" (WG 9). They then turn to the IPCC report and point out that "according to the IPCC, establishing the reality of anthropogenic climate change requires, both detecting and attributing climate change" and claim that we cast doubt on the reality and risk of climate change, and that our argument

[34] In passing we note that Winsberg and Goodwin's discussion of dynamical systems not only suffers from omissions; it is also misleading. When discussing the uncertainties in models they say that dynamical systems "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" (WG 16). By definition an attractor is set a set of states that neighbouring states in a given basin of attraction asymptotically approach in the course of dynamic evolution and that is invariant under the dynamics. A trajectory on one attractor simply cannot "jump to another attractor." Transients can be long, basins of attractions can be riddled, and "appears to be" might have been intended to reflect the rich variety of mathematical behaviours nonlinear systems can display; nevertheless it is mathematical nonsense to say that a trajectory can jump from one attractor to another.

therefore "undermines the most basic conclusions of contemporary climate science" (WG 10). In his recent book Winsberg describes the stance of "the LSE group" as "highly skeptical" (W 71) and adds that our position "would have devastating consequences for attribution claims, and for policy making recommendations based on climate science if any of their [i.e. our] view were *true*." (W 72, original italics). Finally, NNW accuse us of violating the maxim "do no harm" because we argue for "wildly sceptical scenarios" (NNW 20-21).

While we maintain with Richard Feynman that all scientific knowledge is uncertain, we reject any and all anti-science sceptical positions with no ifs, ands, or buts, and repeated attempts to attribute an anti-science sceptical position to us are groundless. In none of the papers Winsberg and co-workers cite do we reject the basic conclusions of contemporary climate science. We call into doubt detailed predictions like the high-resolution local climate projections produced by UKCP09.[35] And we hold this to be well within the bounds of healthy scientific criticism, even where we may later learn we were wrong. Furthermore, Winsberg (W 71-72) and Winsberg and Goodwin (WG 10-11) take us to task for our views on attribution and detection. To our knowledge there is *no mention of these two topics* in any of the papers of ours they cite. Whether these arguments have any implication for detection and attribution, and if so, what the implications, are, is at best an open question, and, as far as we can see, it is one that is not discussed in their papers.

While we do not discuss the IPCC in the papers WC cite, the IPCC does cite our papers on structural model error. Specifically, we note that the IPCC does not disagree with us regarding the significant role of structural model error, and it quotes Smith (2002) explicitly in its 4th assessment report: "Such limitations imply that distributions of future climate responses from ensemble simulations are themselves subject to uncertainty Smith (2002) and would be wider were uncertainty due to structural model errors accounted for" (Solomon et al. 2007, 797). The IPCC rejects a literal interpretation of the fidelity of model-based distributions. The most recent IPCC Assessment Report also supports this view and attempts to address it quantitatively, making it clear that the probability placed in the (outer 10%) tails of model-distribution for future global mean temperature must be increased to 34%, the additional 24% taken from the central (90%) model-range. These numbers are based on their expert judgement, as are the models themselves, of course (more about this case below).

WG appear to claim that even if we don't state an anti-science climate-change denier's position explicitly, such a position is implied by what we say (WG 10). They provide no argument for this conclusion that we can identify as such. It is unclear to us whether WG believe that the role of structural model errors in today's climate models have negligible impact in the applications we criticise, or if they hold that issues of topological conjugacy and structural stability play no role in the class of large fluid dynamical models. It would advance the discussion if they made their position on this matter explicit.

WG are correct in claiming that in practice climate probabilities are computed differently than was done in our example used to define "the default position" (WG 12). We are happy to consider renaming this position, if the name causes confusion. In our illustrations we

[35] Winsberg and Goodwin keep repeating that UKCP09 produces "projections" and not "predictions", which they see as a further reason why our arguments fail (WG 12). This is what we say in the detailed discussion of UKCP09 in our (2015), where we also explain the relation between predictions and projections. For reasons unbeknownst to us, they appear to have ignored the detailed discussion of UKCP09 in our (2015) completely.

made many simplifications and considered an ensemble of 1024 initial conditions, and we use a histogram based on 32 bins of equal size on the unit interval as a proxy for a continuous probability distribution (2014, 37-39). In an early paper we called this the "default position" (2013a, 481). Winsberg and Goodwin object that this position is irrelevant because probabilities are not calculated in this way (WG 12). The projections produced by UKCP09 are the result of complicated statistical techniques, which we describe in some detail in our (2015), which, curiously, Winsberg and his co-workers systematically ignore (judged by the absence a reference in any of their papers). WG are right to point out that the determination of probabilities by UKCP09 is more complex than the one in our example. We said as much both in our (2013b, 890) and (2015, 3989-90, 3998-4000), and those complicated methods are also discussed in (Berger and Smith 2019). But this does not render structural model error irrelevant to the determination of probabilities.

It is true that probabilistic predictions are not always generated using our so-called "default position"; real applications are more complex than our simple illustration, as one would expect. Consider the case of the IPCC's assessment of global mean temperature projections for 2100 in the most recent assessment report.[36] The projection is arrived at by using the CMIP5 ensemble, a set of state-of-the-art climate models that have each been run under the RCP8.5 scenario to supply a value for the projected temperature change in 2100. It turns out that [2.67C, 4.87C] is the interval that symmetrically spans 90% of the ensemble distribution, which the IPCC classifies as "very likely". The IPCC authors then downgrade the model-derived "very likely" range to "likely", which means that the probability that the real climate system will produce this outcome is only above 66%. Hence, the IPCC reckons that there an additional 24% probability that we would experience something outside the 90% central range of model predictions in this projection. That is, the probability in the tails (10%) is increased to 34%. This judgment has been made by the IPCC authors to account structural model errors, for instance due to shared systematic biases, similar resolutions, common algorithms (and sometimes code), computational constraints, and similar parameterizations. So the authors of the IPCC report regard structural model error as an important factor, explicitly even for coarse variables like global mean temperature.

Finally, Winsberg and Goodwin make a methodological recommendation. To find out whether a quantity of interest is stable, "the best we can do is to play around with our models and see which predictions are stable under the perturbations that we think ought to concern us" (WG 15). While there is merit in such explorations, there are doubts that the question can be settled in this way. First, as noted above, it is unclear which perturbations to consider. Second, carrying out simulations is extremely costly and it takes a long time to produce a model run. "Playing around" with state-of-the-art models is arguably impossible by definition of a "state-of-the-art model" (Smith 2002). While we would hesitate to deem climateprediction.net (Stainforth et al. 2005) "playing around" with any model, it is interesting to examine the response of the climate community to these results. In general, viable trial and error approaches to determine which properties are stable under perturbations and which aren't have not yet been employed in the study we criticised.[37] Principled mathematics, and statistics, and physical science would each play a role in such an endeavour.

[36] For a discussion of this case see Thompson *et al.*'s (2016).

[37] The UKCO09 project had only 17 runs with full model (HadCM3) and 280 runs with a reduced model with a slab ocean (i.e. an ocean with no currents and a uniform effective depth of 50 m). For details see our (2015).

## 6. Conclusion

In their papers, Winsberg and co-workers appear to attempt a wholesale rejection of everything we say about structural model error, yet they do not come out as saying that models' structural error is ignorable and that UKCP09's high resolution probability statements are trustworthy. They emphasise – as we do – that the particulars of the case such as lead times and choice variables matter, and they conclude that "the question is highly complex, and depends on details of the definition of structural stability being employed and on the timeframe of the prediction" (WG 15). As regards UKCP09, they say that "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" (WG 16). This *is* structural model error. We agree, and our (2015, 3994) contains a graph showing the range of the simulations for global mean annual temperature over the twentieth century.[38] These very by several degrees. Current models, as high resolution prediction engines, are far from perfect. But poorly understood features of a system and poor parameterizations are a source of structural model error, and so Winsberg and co-workers in fact attribute the models' shortcomings to structural model error. We are in agreement with Winsberg and co-workers that in current climate science model errors are not epsilon-small, they are huge. It is nevertheless important to understand what effects even small errors can have, and a discussion of the hawkmoth effect gives us exactly that.

We feel the failure to acknowledge clearly and transparently the limits that either structural model error or the large systematic errors in our current models place on their real-world predictions puts the credibility of solid climate science at risk (one of us has regularly faced resistance to statements made in the IPCC reports). Indeed, lack of transparency in how science is presented in public debates places the credibility of all science in support of decision making at greater risk. We expect that these errors could and will be reduced, noting that climate scientists themselves deem this a difficult task (Palmer and Weisheimer 2011). Even if these errors were reduced to well below the magnitude of current weather models, we note that structural stability – and the recognition that models aren't structurally stable – implies limits on models, which, unlike many current shortcomings, it appears we cannot overcome. This, however, leaves open the question where these limits lie and how they affect particular modelling endeavours. As noted in Section 5, there may exist project-specific goals which lessen the impact of model error; accountable probability forecasts however appear to be out of reach.

WG proffer Parker's (2018) recommendation to assess adequacy for purpose on a case-by-case basis as the solution to the problem (WG 16). We agree and add that assessing adequacy for purpose will involve evaluating stability properties of relevant variables and considering appropriate lead-times. That said, once one recognises that climate prediction operate on a grander scale than weather prediction and accepts that it is simply not possible to show adequacy on climate targets in as useful a way as might be done in weather forecasting, the purposes we have considered (probability forecasts at high resolution long lead (2090) time scales) simply cannot be deemed adequate today. We also hold that while one can never prove any forecast system to be adequate for this purpose, one can nevertheless, show a forecast system to be inadequate. Theoretical considerations, and consistency between both model and simulations with the present conditions will have to

[38] See also Figure 9.9a on p. 768 in the IPCC's AR5 (Stocker et al. 2013).

play an important role in developing an appropriate level of confidence in climate modelling simulation, and acknowledging the limitations of our current models plays a critical role in maintaining the credibility of science in application. The aim of our project was to draw attention to the importance of these issues.

## Acknowledgments

We would like to thank Jim Berger, Vincent Lam, Ana Lopez, Wendy Parker, David Stainforth, Erica Thompson, and Charlotte Werndl for helpful discussions, and four referees for the journal for useful suggestions.

## References

Abraham, R., & Marsden, J. E. (1978). *Foundations of Mechanics* (2nd ed.). Redwood City: Addison-Wesley.

Barrow-Green, J. (1996). *Poincaré and the Three Body Problem* Washington: American Mathematical Society.

Belot, G. (2013). Bayesian Orgulity. *Philosophy of Science, 80*(4), 483-503.

Berger, J., & Smith, L. A. (2019). Uncertainty Quantification. *Annual Review of Statistics and Its Application, 6*, 433-460.

Cisewski, J., Kadane, J. B., Schervish, M. J., Seidenfeld, T., & Stern, R. (2018). Standards for Modest Bayesian Credences. *Philosophy of Science, 85*(1), 53-78.

Cucker, F., & Wong, R. (Eds.). (2000). *The Collected Papers of Stephen Smale. Volume 2.* Singapore: Singapore University Press and World Scientific.

Devaney, R. L. (1989). *An Introduction to Chaotic Dynamical Systems* (2nd ed.). Boulder, Colorado: Westview Press.

Du, H., & Smith, L. A. (2012). Parameter estimation through ignorance. *Physical Review E, 86*(1), 016213.

Elga, A. (2016). Bayesian Humility. *Philosophy of Science, 83*(3), 305-323.

Fillion, N., & Corless, R. M. (2014). On the Epistemological Analysis of Modeling and Computational Error in the Mathematical Sciences. *Synthese, 191*, 1451-1467.

Frigg, R., Bradley, S., Du, H., & Smith, L. A. (2014). The adventures of Laplace's demon and his apprentices. *Philosophy of Science, 81*(1), 31-59.

Frigg, R., Bradley, S., Machete, R. L., & Smith, L. A. (2013a). Probabilistic Forecasting: Why Model Imperfection Is a Poison Pill. In H. Anderson, D. Dieks, G. Wheeler, W. Gonzalez, & T. Uebel (Eds.), *New Challenges to Philosophy of Science* (pp. 479-491). Berlin and New York: Springer.

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

Frigg, R., Smith, L. A., & Stainforth, D. A. (2015). An Assessment of the Foundational Assumptions in High-Resolution Climate Projections: The Case of UKCP09. *Synthese, 192*, 3979-4008.

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters, 46*(14), 8329-8337.

Goodwin, W. M., & Winsberg, E. (2016). Missing the Forest for the Fish: How Much Does the 'Hawkmoth Effect' Threaten the Viability of Climate Projections? *Philosophy of Science, 83*(5), 1122-1132.

Hirsch, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*. San Diego: Academic Press.

Hirsch, M. W., Smale, S., & Devaney, R. L. (2004). *Differential equations, dynamical systems, and an introduction to chaos* (2nd ed.). San Diego: Elsevier.

Jänich, K. (1984). *Topology*. New York: Springer.

Jolliffe, I. T., & Stephenson, D. B. (2012). *Forecast verification: a practitioner's guide in atmospheric science* (2nd ed.): Wiley-Blackwell.

Judd, K., Reynolds, C. A., Smith, L. A., & Rosmond, T. E. (2008). The Geometry of Model Error. *Journal of the Atmospheric Sciences, 65*(6), 1749-1772.

Judd, K., & Smith, L. A. (2001). Indistinguishable States I: perfect model scenario. *Physica D, 151*, 125-141.

Judd, K., & Smith, L. A. (2004). Indistinguishable states II: The imperfect model scenario. *Physica D, 196*, 224-242.

Katok, A., & Hasselblatt, B. (1995). *Introduction to the Modern Theory of Dynamical Systems*. Cambridge: Cambridge University Press.

May, R. (1976). A Simple Mathematical Equation with very Complicated Dynamics. *Nature, 261*, 459-469.

Nabergall, L., Navas, A., & Winsberg, E. (2019). An antidote for hawkmoths: on the prevalence of structural chaos in non-linear modeling. *European Journal for Philosophy of Science, 9*(21), 1-28.

Nielsen, M., & Stewart, R. T. (2019). Obligation, Permission, and Bayesian Orgulity. *Ergo, 6*(3), doi:https://doi.org/10.3998/ergo.12405314.0006.003.

Ott, E. (1993). *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press.

Palmer, T. N., & Weisheimer, A. (2011). Diagnosing the causes of bias in climate models – why is it so hard. *Geophysical & Astrophysical Fluid Dynamics, 105*, 351-365.

Parker, W. S. (2018). The Significance of Robust Climate Projections. In E. A. Lloyd, & E. Winsberg (Eds.), *Climate Modelling. Philosophical and Conceptual Issues* (pp. 273-296). Cham: Palgrave Macmillan.

Pilyugin, S. Y. (1991). *Shadowing in Dynamical Systems*. Berlin, Heidelberg and New York: Springer.

Smale, S. (1966). Structurally Stable Systems Are Not Dense. *American Journal of Mathematics, 88*, 491-496.

Smith, L. A. (1995). Accountability and error in ensemble forecasting. In *Proceedings of the ECMWF Seminar on Predictability, Vol. 1* (pp. 351-369). Shinfield Park, Reading: ECMWF.

Smith, L. A. (2001). Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems. In A. I. Mees (Ed.), *Nonlinear Dynamics and Statistics* (pp. 31-64). Boston: Birkhäuser.

Smith, L. A. (2002). What might we learn from climate forecasts? *Proceedings of the National Academy of Science, USA 4*(99), 2487-2492.

Smith, L. A. (2007). *Chaos. A Very Short Introduction*. Oxford: Oxford University Press.

Smith, L. A., Ziehmann, C., & Fraedrich, K. (1999). Uncertainty dynamics and predictability in chaotic systems. *Quarterly Journal of the Royal Meteorological Society, 125*, 2855-2886.

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

Solomon, S., Qin, D., & Manning, M. (Eds.). (2007). *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., et al. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature, 433*(7024), 403-406.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M. B., Allen , S. K., Boschung, J., et al. (Eds.). (2013). *Climate change 2013. The physical science basis. Working Group I contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Thompson, E., Frigg, R., & Helgeson, C. (2016). Expert Judgment for Climate Change Adaptation. *Philosophy of Science, 83*(6), 1110-1121.

Thompson, E. L. (2013). *Modelling North Atlantic Storms in a Changing Climate. Ph.D. Thesis*: Imperial College, London, UK.

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

Wiggins, S. (2003). *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (2nd ed.). New York: Springer.

Winsberg, E. (2018). *Philosophy and Climate Science*. Cambridge: Cambridge University Press.

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