ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Thompson, Erica L.; Smith, Leonard A.

Working Paper

Escape from model-land

Economics Discussion Papers, No. 2019-23

Provided in Cooperation with:
Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Thompson, Erica L.; Smith, Leonard A. (2019) : Escape from model-land, Economics Discussion Papers, No. 2019-23, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:
https://hdl.handle.net/10419/194875

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

CC BY https://creativecommons.org/licenses/by/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2019-23 | March 08, 2019 | http://www.economics-ejournal.org/economics/discussionpapers/2019-23

Escape from model-land

Erica L. Thompson and Leonard A. Smith

Abstract
Both mathematical modelling and simulation methods in general have contributed greatly to understanding, insight and forecasting in many fields including macroeconomics. Nevertheless, we must remain careful to distinguish model-land and model-land quantities from the real world. Decisions taken in the real world are more robust when informed by our best estimate of real-world quantities, than when "optimal" model-land quantities obtained from imperfect simulations are employed. The authors present a short guide to some of the temptations and pitfalls of model-land, some directions towards the exit, and two ways to escape.

(Published in Special Issue Bio-psycho-social foundations of macroeconomics)

JEL C52 C53 C6 D8 D81
Keywords Modelling and simulation; decision-making; model evaluation; uncertainty; structural model error; dynamical systems; radical uncertainty

Authors
Erica L. Thompson, London School of Economics and Political Science, Centre for the , Analysis of Time Serie (CATS), UK, e.thompson@lse.ac.u
Leonard A. Smith, Mathematical Institute, University of Oxford, UK, lenny@maths.ox.ac.uk

Citation Erica L. Thompson and Leonard A. Smith (2019). Escape from model-land. Economics Discussion Papers, No 2019-23, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2019-23

Received March 5, 2019 **Accepted as Economics Discussion Paper** March 5, 2019 **Published** March 8, 2019
© Author(s) 2019. Licensed under the Creative Commons License - Attribution 4.0 International (CC BY 4.0)

Computational simulations and associated graphical visualisations have become much more sophisticated in recent decades due to the availability of ever-greater computational resources. The qualitative visual appeal of these simulations has led to an explosion of simulation-based, often probabilistic forecasting in support of decision-making in everything from weather forecasting and American Football, to nuclear stewardship and climate adaptation. We argue that the utility and decision-relevance of these model simulations must be judged based on consistency with the past, and out-of-sample predictive performance and expert judgement, never based solely on the plausibility of their underlying principles or on the visual "realism" of outputs.

Model-land is a hypothetical world in which our simulations are perfect, an attractive fairy-tale state of mind in which optimising a simulation invariably reflects desirable pathways in the real world. Decision-support in model-land implies taking the output of model simulations at face value (perhaps using some form of statistical post-processing to account for blatant inconsistencies), and then interpreting frequencies in model-land to represent probabilities in the real-world. Elegant though these systems may be, something is lost in the move back to reality; very low probability events and model-inconceivable "Big Surprises" are much too frequent in applied meteorology, geology, and economics. We have found remarkably similar challenges to good model-based decision support in energy demand, fluid dynamics, hurricane formation, life boat operations, nuclear stewardship, weather forecasting, climate calculators, and sustainable governance of reindeer hunting.

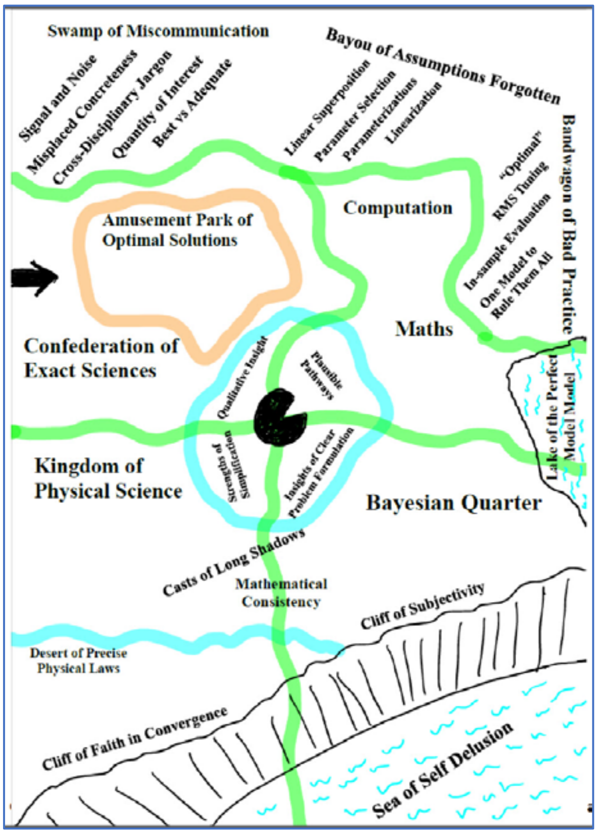

Figure 1: A map of Model-Land. The black hole in the middle is a way out.

One can justifiably aim to transform simulations in model-land into information regarding the real world, but only where such information exists. Our aim is a decision-making process that remains acceptable to all involved regardless of the outcome; ideally a process retained without modification and used again under similar conditions in the future regardless of the outcome, unless a deeper understanding of the system has been obtained. This cannot be accomplished in model-land. Uncomfortable departures from model-land are required for (good) decision support.

Big Surprises, for example, arise when something our simulation models cannot mimic turns out to have important implications for us. Big Surprises invalidate (not update) model-based probability forecasts: the I in $P(x|I)$ changes. In "weather-like" tasks, where there are many opportunities to test the outcome of our model against a real observed outcome, we can see when/how our models become silly. In "climate-like" tasks, where the forecasts are made truly out-of-sample, there is no such opportunity and we rely on judgements about the quality of the model given the degree to which it performs well under different conditions and expert judgement on the shortcomings of the model.

Simulations and Model-lands: the map is not the territory

As the simulation of complex systems becomes routine in many areas of research, the distinction between simulated variables and their real-world counterparts can become unclear. As a trivial example, when writing about forecasts of household consumption, energy prices, or global average surface temperature, many authors will use the same name and the same phrasing to refer to effects seen in the simulation as those used for the real world. It may not be the case that these authors are actually confused about which is which, the point is that readers of conclusions would benefit from a clear distinction being made, especially where such results are presented as if they have relevance to real-world phenomena and decision-making.

Why are we concerned about this? It is not just a philosophical worry about semantics but real implications we have observed when the consumers of this material realise just how different the model-variables are from the real-world phenomena they face. Something seen on the map may not correspond to what is in the territory; worse, something **not** seen on the map may be encountered when we explore the territory. Within model-land, we cannot even enunciate the possibility of a "Big Surprise", let alone think about the probability of such an event occurring. Yet the possibilities remain of economic surprises, previously-unseen weather events, energy price spikes, or worse-than-expected climate impacts, even where these are not simulated by today's models. Such events, in fact, happen disturbingly often. Can we escape model-land by targeting exclusively the less comfortable, but better-informed and much more relevant real-world entities in decision-making?

It is comfortable for researchers to remain in model-land as far as possible, since within model-land everything is well-defined, our statistical methods are all valid, and we can prove and utilise theorems. Exploring the furthest reaches of model-land in fact is a very productive career strategy, since it is limited only by the available computational resource. While pure mathematicians can, of course, thrive in model-land, applied mathematicians have a harder row to hoe, inasmuch as, for large classes of problems, the pure mathematicians have proven that **no** solution to the problem will hold in the real world.

Reasons for Staying in Model-land

Well-posed problems have exact solutions.
The use of which in the real world leads to disaster.

Optimal solutions exist, and might be found.
But they fail to perform as expected back in the real world.

I can be a Bayesian, with a principled path through at all stages of a project.
The Bayesian Way is invaluable for describing the challenge, but need never lead to a complete solution outside model land.

Arguably, one can construct actionable probabilities and use them optimally.
All Probabilities are conditioned on falsehood, and the Prob(Big Surprise) is never zero.

Nonlinear systems and Chaos pose a host of interesting mathematical problems to be solved!
Mathematicians thrive in model-land!
But for large classes of problems, we can prove those interesting solutions will not hold in the real world.

For what we term "climate-like" models, the realms of sophisticated statistical processing which variously "identify the best model", "calibrate the parameters of the model", "form a probability distribution from the ensemble", "calculate the size of the discrepancy" etc., are castles in the air built on a single assumption which is known to be incorrect: that the model is perfect. These mathematical "phantastic objects" are great works of logic but their outcomes are relevant only in model-land until a direct assertion is made that their underlying assumptions hold "well enough"; that they are shown to be adequate for purpose, not merely today's best available model. Of course, many assumptions are false in principle but negligible in practice and it is reasonable to ask, as we now do, whether this may not be the case here.

Structural model error and its implications: the Hawkmoth Effect

To understand the depth of the problem, it is helpful to unpick the mathematics further. Chaos is no longer as fashionable as it was a few decades ago, but most readers will be familiar with the so-called Butterfly Effect – the concept that a small difference in initial conditions (perhaps stepping on a butterfly) can result in a large difference in the outcome of a complex dynamical system over some timescale. This was first noticed by Edward Lorenz, coming to his attention due to slight numerical truncation error in a simple mathematical system.

In the 21st century, the Butterfly Effect is a solved problem. To account for the possibility of error in the initial conditions, instead of taking a single best-estimate of the system state, we instead use an ensemble (multiple initial conditions) to represent a probability distribution over the initial conditions consistent with both the observations and the mathematical model. This ensemble of model points is then interpreted as a probability distribution in the real world which encompasses all possible outcomes given the uncertainty in the initial conditions, parameter values, and other numbers.

This assumes that the equations of the dynamical system are known perfectly, as was the case for Lorenz's three-dimensional mathematical model. Where our complex model is not an end in itself, however, but a stand-in for a complex system such as the Earth's atmosphere, the economy, or the energy system, then we can say with confidence that our model is not perfect. We are then in the realm not of initial condition error but of structural model error: in a chaotic system if our model is only slightly mathematically mis-specified then a very large difference in outcome will evolve over time even with a “perfect” initial condition. We term this the Hawkmoth Effect.

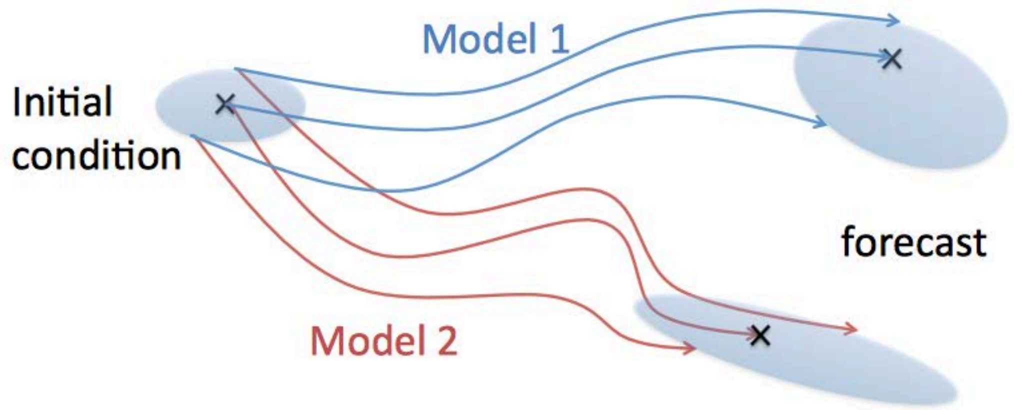

Figure 2: The Hawkmoth Effect. With a perfect estimation of the uncertainty around the initial condition, only a perfect model can result in a perfect specification of the uncertainty in the forecast. An imperfect model may be arbitrarily wrong even when initialised ideally. Note that identical initial conditions have finite separation immediately under Model 1 and Model 2; chaotic divergence identical initial conditions remain identical for all time.

If we have (somehow) perfectly specified our initial condition uncertainty, but have a structurally imperfect model, then the probability distribution that we arrive at by using multiple initial conditions will grow more and more misleading – misleadingly precise, misleadingly diverse, or just plain wrong in general. The natural response to this is then, by analogy with the solution to the Butterfly problem, to make an ensemble of multiple model structures, perhaps derived by systematic perturbations of the original model. Unfortunately, the strategy is no longer adequate. In initial condition space (a vector space), there a finite number of variables and a finite space of perturbations in which there are ensemble members consistent with both the observations and the model’s dynamics. Models lie in a function space where, by contrast, there are uncountably many possible structures. It is not clear why multi-model ensembles are taken to represent a probability distribution at all; the distributions from each imperfect model in the ensemble will differ from the desired perfect model probability distribution (if such a thing exists); it is not clear how combining them might lead to a relevant, much less precise, distribution for the real-world target of interest.

It is sometimes thought that if a model is only slightly wrong, then its outputs will correspondingly be only slightly wrong. The Butterfly Effect revealed that in deterministic nonlinear dynamical systems, a “slightly wrong” initial condition can yield wildly wrong outputs. The Hawkmoth Effect implies that when the mathematical structure of the model is only “slightly wrong” then one loses topological conjugacy (with probability one), and even the best formulated probability forecasts will be wildly wrong. This result, due to Smale in the early 1960’s holds consequences not only for the aims of prediction but also for model development and calibration, and of course for the formation of initial condition ensembles. Naïvely, we might hope that by making incremental improvements to the “realism” of a model (more accurate representations, greater details of processes, finer spatial or temporal resolution, etc.) we would also see incremental improvement in the outputs (either qualitative realism or according to some quantitative performance metric). Regarding the realism of short term trajectories, this may well be true! It is not expected to be true in terms of probability forecasts. And it is not always true in terms of short term trajectories; we note that fields of research where models have become dramatically more complex are experiencing exactly this problem: the nonlinear compound effects of any given small tweak or addition to the model structure are so great that calibration becomes a very computationally-intensive task and the marginal performance benefits of additional subroutines or processes may be zero or even negative. In plainer terms, adding detail to the model can make it less accurate.

The observation that complex models may be less informative than simple models (or comparatively informative but much more costly in terms of computational resource, human resource and cold hard cash) may, paradoxically, assist decision-making by providing a stopping-point to what is otherwise a potentially endless quest for “more research”, “better information” or “less uncertainty” before a decision is made. How good is a model before it is *good enough* to support a particular decision – i.e., adequate for the intended purpose? This of course depends on the decision as well as on the model. And ideally one would start with the decision and consider potential models in light of their ability (or not) in informing this decision. Starting in model-land, one can continue forever improving our model and exploring the implications of introducing new complexity: evaluating in model-land will no doubt show some manner of “improvement.” When the justification of the research is to inform some real-world time-sensitive decision, merely employing the best available model can undermine (and has undermined) the notion of the science-based support of decision making, when limitations like those above are not spelt out clearly.

Challenges for real-world decision-making

We have various illustrations of how to extract information from (ensembles of) simulations which out-perform naïve statistical model forecasts, and avoid some of the misleading assumptions that are unavoidable if one stays in model-land. These illustrations include the 2018 Pakistan heatwave, pricing and trading in the energy market in week out to two (when constrained by regulation), and experiments designed to explore model error in practice for nuclear stewardship.

In our work with the START Network, a group of humanitarian NGOs, we are looking at ways to streamline the use of information from weather (and other) forecasts to anticipate humanitarian crises. Following an alert, a 72-hour process decides whether to activate the release of funds and then how to allocate money to projects. In principle, for many situations it is possible to determine a timescale of applicability for the forecast. This can help both when it shows that information is available, as it allows confident use of a set of operating procedures based on the forecast, and also when it shows that relevant information is not available and the decision should be made based on other inputs. In the case of heatwave in Pakistan, it was made clear by one of us that a reasonably confident forecast can be made with sufficient lead time (several days) to follow START procedures and take actions which help to reduce the likely impact on potentially vulnerable groups. As we develop and extend this framework to other regions and hazards, such as tropical cyclones and droughts, we expect that in some cases the forecast information will be negligible and will then advise that the rapid-turnaround decision should focus more on other factors such as social, economic and practical bases for action. Taking one set of papers off the table is a valuable contribution to the decision process – using an uninformative forecast simply because it is the “best available” would at best result in a slower decision, where time is of the essence.

Working in Model-land

You may be living in model-land if you…

- try to **optimise** anything regarding the future;
- believe that **decision-relevant probabilities** can be extracted from models;
- believe that there are **precise parameter values** to be found;
- refuse to believe in anything that has not been seen in the model;
- think that **learning more will reduce the uncertainty** in a forecast;
- explicitly or implicitly **set the Probability of a Big Surprise to zero**; that there is nothing your model cannot simulate;
- want "**one model to rule them all**".

You may be near the exit if you…

- are now concerned about your forecast systems Probability of a Big Surprise;
- are uncomfortable to realise that given 100x the resources, your team would build a better model rather than refine the probability estimates of your current model;
- are careful to distinguish $\mathbf{x}$ from its model-land namesake $\mathbf{x}_{\text{model}}$;
- have forsaken the quest for a model that tells you everything, to search instead for models that tell you something about your Quantity of Interest;
- are intrigued by the idea of finding your model's Relevant Dominant Uncertainty – the aspect of the model you should improve first.

There are some critical distinctions that it is helpful to recognise in order to look for ways out of model-land and back to reality. Is the model used simply the "best available" at the present time, or is it truly adequate for the specific purpose of interest? How would adequacy for purpose be assessed, and what would it look like? Are you working with a weather-like task, where adequacy for purpose can more or less be quantified, or a climate-like task, where the forecasts are not verifiable on short timescales? Is the system reflexive; does it respond to the forecasts themselves? How do we evaluate models: against real-world variables, or against a contrived index, or against other models? Or are they primarily evaluated by means of their epistemic or physical foundations? Or, one step further, are they primarily explanatory models for insight and understanding rather than quantitative forecast machines? Does the model in fact assist with human understanding of the system, or is it so complex that it becomes a prosthesis of understanding in itself?

Escaping from Model-land

There are at least two ways to escape from model-land. The first, and most effective, is by repeatedly challenging your model to make out-of-sample predictions and seeing how well it performs. This, of course, can only be done in weather-like tasks, so named because tomorrow's weather forecast is an excellent example where it can be done. We can not only understand in broad terms the usefulness of the forecast on different timescales, but we can also quantify the degree of confidence in its success for different weather types and in different seasons. This is precisely the information needed for high-quality decision support: a forecast, *accompanied by a statement of its own limitations*. One would not attempt to use a detailed weather forecast for this day next year, even though it is in principle perfectly possible to extend the simulation indefinitely. In addition, these calculations will give us some understanding of where and how the model is performing poorly, which may assist us to improve the model itself.

In other cases, where the problem in hand does not allow out-of-sample evaluation, or at least prevents it on the time scales of the decision; there is simply not enough data available to construct this detailed understanding of the limitations of a model. We call these climate-like situations. Although one can never be as confident is such cases, there is an alternative way to escape from model-land: using expert judgement and the realism of simulations of the past to define the expected relationship of model with reality. An example: the most recent IPCC climate change assessment uses an expert judgement that there is only approximately a 2/3 chance that the actual outcome of global average temperatures in 2100 will fall into the 90% confidence interval generated by climate models (footnotes c and d to table SPM.2 on page 23 of the Summary for Policymakers). Again, this is precisely the information needed for high-quality decision support: a forecast, *accompanied by a statement of its own limitations* (the Probability of a “Big Surprise”). Structured procedures for generating and formalising expert judgements are available. It is worth noting here that presenting model output at face value as a prediction, or interpreting simulation frequencies as real-world probabilities, is equivalent to making an implicit expert judgement that the model structure is perfect; this claim can often be shown to be false. Indeed in-sample tests have limited power when claiming a model is adequate, but they might easily establish a high level of confidence (evidence) that the model is inadequate.

The example referred to in the previous paragraph is no small correction, but a first order change (of the order of tens of percentage points of probability mass) from one set of outcomes to another range of outcomes. This is likely to have a nontrivial impact on decision based on the probability of outcomes the simulations suggest when taken at face value. This is an expert judgement arrived at by the expert lead authors of the IPCC chapter, themselves scientists who worked on the underlying models and simulations which went into the generation of the 90% confidence interval. But if many thousands of work-hours have gone into refining the dynamical models and statistical techniques which produced the first interval, quantifying and reducing the model-land uncertainty, their valiant efforts are *known to be* swamped by the uncertainty which exists in the abyss between model-land and the real world. A bridge to escape is sorely needed here.

In neither route to escape from model-land do we work to indefinitely increase the complexity of models. Where out-of-sample predictions can be tested, they will reveal whether each model development does or does not lead to more informative outputs. Where we rely on expert judgement, it is likely that models intermediate in complexity will be the most intuitive to understand and assess. Even where we cannot test long range model-based predictions, we can test how well our model can reflect (shadow) the past, and learn the phenomena with which they cope most poorly. This informs judgement as to how far in the future a given model is likely to be relevant to the evolution of the real world.

In neither route to escape from model-land do we discard models completely: rather, we aim to use them more effectively. The choice is not between model-land or nothing. Instead, models and simulations are used to the furthest extent that confidence in their utility can be established, either by quantitative out-of-sample performance assessment or by well-founded critical expert judgement. A wide literature treats the use and calibration of expert judgement in such situations although there is certainly more to say about the interplay of model development and the parallel development of one’s own expert judgement.

More generally, letting go of the phantastic mathematical objects and achievables of model-land can lead to more relevant information on the real world and thus better-informed decision-making. Escaping from model-land may not always be comfortable, but it is necessary if we are to make better decisions.

Acknowledgements
This research was funded by ECOPOTENTIAL (H2020 grant number 641762) and by the LSE KEI Fund (Better Uncertainty Guidance). We gratefully acknowledge discussions with colleagues in the CRUISSE Network and at the ESRC-funded Rebuilding Macroeconomics Conference in October 2018.

References
[In review we would appreciate further suggestions of relevant references from across the economic or social science literature.]

Berger, J.O. and Smith, L.A. (2018) On the statistical formalism of uncertainty quantification, Annual Reviews of Statistics and its Application, 6.3.1-3.28. ISSN 2326-8298.
Beven, K., Buytaert, W. and Smith, L.A. (2012) On virtual observatories and modelled realities (or why discharge must be treated as a virtual variable), Hydrol. Process., 26 (12): 1905-1908.
Frigg, R., Smith, L.A. and Stainforth, D.A. (2015) 'An assessment of the foundational assumptions in high-resolution climate projections: the case of UKCP09', *Synthese.*
Judd, K. Reynolds, C.A. Rosmand, T.E. and Smith, L.A. (2008) The Geometry of Model Error. J AtmosSci 65 (6): 1749-1772.
Judd, K. and Smith, L.A. (2004) Indistinguishable states II: the imperfect model scenario, Physica D, 196: 224-242.
Parker, W. (2009). Confirmation and Adequacy-for-Purpose in Climate Modelling. *Aristotelian Society Supplementary Volume* **83**(1): 233-249.
Smith, L.A. (2016) 'Integrating information, misinformation and desire: improved weather-risk management for the energy sector', in Aston et al (ed.) *UK Success Stories in Industrial Mathematics*, 289-296. Springer.
Smith, L.A. (2002) What Might We Learn from Climate Forecasts? Proc Nat Acad Sci USA 4(99): 2487-2492.
Smith, L.A. (2007) “Chaos: A Very Short Introduction” OUP, Oxford
Smith, L.A. (2000) 'Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems' in Nonlinear Dynamics and Statistics, ed. Alistair I Mees, Boston: Birkhauser, 31-64.
Smith, L.A. (2006) Predictability past predictability present. Chapter 9 of Predictability of Weather and Climate (eds T. Palmer and R. Hagedorn). Cambridge, UK. Cambridge University Press.
Smith, L.A. and Petersen, A.C. (2014) 'Variations on reliability: connecting climate predictions to climate policy', in Boumans et al (ed.) *Error and Uncertainty in Scientific Practice*, London: Pickering & Chatto.
Smith, L.A. and Stern, N. (2011) 'Uncertainty in science and its role in climate policy', *Phil. Trans. R. Soc. A*,369, 1-24.
Thompson, E. (2013) “Modelling North Atlantic Storms in a Changing Climate,” PhD thesis, Imperial College, London.
Thompson, E., Frigg, R. and Helgeson, C. (2016) “Expert Judgment for Climate Change Adaptation,” *Philosophy of Science* 83(5):1110-1121.
Tuckett, D.,and Nikolic, M. (2017) The role of conviction and narrative in decision-making under radical uncertainty. *Theory & psychology*, *27*(4), 501-523.

Economics
The Open-Access, Open-Assessment E-Journal

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

http://www.economics-ejournal.org/economics/discussionpapers/2019-23

The Editor

© Author(s) 2019. Licensed under the Creative Commons License - Attribution 4.0 International (CC BY 4.0).