Supplementary information

To accompany a Comment published in *Nature* **582**, 482–484 (2020)

https://doi.org/10.1038/d41586-020-01812-9

Five ways to ensure that models

serve society: a manifesto

Andrea Saltelli, Gabriele Bammer, Isabelle Bruno, Erica Charters, Monica Di Fiore, Emmanuel Didier, Wendy Nelson Espeland, John Kay, Samuele Lo Piano, Deborah Mayo, Roger Pielke Jr, Tommaso Portaluri, Theodore M. Porter, Arnald Puy, Ismael Rafols, Jerome R. Ravetz, Erik Reinert, Daniel Sarewitz, Philip B. Stark, Andrew Stirling, Jeroen van der Sluijs & Paolo Vineis

Contents

[Introductory section 4](#_Toc43992775)

[A.1 “Statisticians were debating” 4](#_Toc43992776)

[A.2 “Now computer modelling has come into the media and policy limelight” 5](#_Toc43992777)

[A.3 “Precise, reliable numbers” for COVID-19? 5](#_Toc43992778)

[A.4 “Models serve society extremely well”: weather forecasting models 6](#_Toc43992779)

[A.5 “Best practices for responsible mathematical modelling” 6](#_Toc43992780)

[1. Mind the assumptions 8](#_Toc43992781)

[1.1 “Models are often imported from one application to another”: seismic risk 8](#_Toc43992782)

[1.2 “A model used in the UK to guide transport policy” 8](#_Toc43992783)

[1.3 “Global uncertainty and sensitivity analysis” 8](#_Toc43992784)

[2. Mind the hubris 10](#_Toc43992785)

[2.1 Why talking of a modelling hubris? 10](#_Toc43992786)

[2.2 “Trade-off between the usefulness of a model and the breadth it tries to capture” 10](#_Toc43992787)

[2.3 Model accuracy should be measured prospectively 12](#_Toc43992788)

[2.4 “A simpler model” for HIV forecasting 13](#_Toc43992789)

[2.5 “Superspreading events” 13](#_Toc43992790)

[2.6 “Yucca Mountain repository”: a million years of certainty 14](#_Toc43992791)

[2.7 “Complexity as an end in itself”? 14](#_Toc43992792)

[2.8 “Models tend to be developed with large teams” 14](#_Toc43992793)

[3. Mind the framing 14](#_Toc43992794)

[3.1 “No one model can serve all purposes” 14](#_Toc43992795)

[3.2 “The technique is never neutral” 15](#_Toc43992796)

[3.3 “GENESIS model of shoreline erosion” 15](#_Toc43992797)

[3.4 “Value of a statistical life” 16](#_Toc43992798)

[3.5 “Involving stakeholders” 16](#_Toc43992799)

[3.6 “Participatory forecasting of flooding risk, and in the management of fisheries” 17](#_Toc43992800)

[4. Mind the consequences 19](#_Toc43992801)

[4.1 “Models contributed to crippling the global economy” 19](#_Toc43992802)

[4.2 “Once a number takes the center-stage” 19](#_Toc43992803)

[4.3 “Spurious precision adds to a false sense of certainty” 19](#_Toc43992804)

[4.4 “An analysis by the WHO Africa”, and the Imperial College’s “510,000 deaths” 20](#_Toc43992805)

[4.5 Opacity about uncertainty damages trust 20](#_Toc43992806)

[4.6 “Sociology of quantification” 20](#_Toc43992807)

[5. Mind the unknowns 21](#_Toc43992808)

[5.1 “Docta ignorantia” 21](#_Toc43992809)

[5.2 “Opening the door to surprises”; 21](#_Toc43992810)

[5.3 “Anthony Fauci”, and the use of deep uncertainty methods 22](#_Toc43992811)

[Questions not answers 22](#_Toc43992812)

[B.1 “Ritualistic use of quantification” 22](#_Toc43992813)

[B.2 Models’ assumptions and limitations must be appraised openly and honestly 23](#_Toc43992814)

[B.3 “Statactivistes” and other forms of engagement 24](#_Toc43992815)

[B.4 “The blog of Tomas Pueyo” 25](#_Toc43992816)

[B.5 “Not an end to quantification” 26](#_Toc43992817)

[Additional examples mentioned in Table 1. 26](#_Toc43992818)

[C.1 Species extinct 26](#_Toc43992819)

[C.2 Ubiquitous ranking 26](#_Toc43992820)

[C.3 An example of poor context choice 27](#_Toc43992821)

[C.4 Fraudulent quantifications 27](#_Toc43992822)

[C.5 Why bank fails 27](#_Toc43992823)

[C.6 Austerity policies and spreadsheet errors 27](#_Toc43992824)

[C.7 About consequentialism: the case of the ecological footprint 28](#_Toc43992825)

[C.8 Clinton’s precise victory 28](#_Toc43992826)

[C.9 NASA’s MESSENGER 28](#_Toc43992827)

[Table 1, list of examples 29](#_Toc43992828)

[References 32](#_Toc43992829)

**Supplementary material for “Five ways to make models serve society: a manifesto”**

We provide in the following those references and additional information which could not be accommodated in the body of the article, following the order of presentation of the article itself.

##

## Introductory section

### A.1 “Statisticians were debating”

For a number of years now1, statistical practitioners have been concerned about the misinterpretation and abuse of statistical significance testing. Recently2, some have argued that the concepts of statistical significance and P-value thresholds should be abandoned altogether. Others argue that while fallacies of testing should be avoided3, abandoning statistical significance and P-value thresholds is tantamount to precluding testing and falsification4.

Contributions to the debate from the authors of the present work are3–7.

In the US, the debates are found in editorials in journals of the American Statistical Association8–10 and are discussed in the blogs of philosophers of statistics such as Deborah Mayo (https://errorstatistics.com/) and the statistician Andrew Gelman (https://statmodeling.stat.columbia.edu/).

In COVID-19, a comment by Anthony Fauci about the drug Remdevisir had statisticians up in arms, also against the company developing the drug, for lack of transparency11.

One motivation of the present Comment is that while statisticians perceive the existence of a problem, this is less the case in modelling, notwithstanding the checklists and guidelines mentioned later in this text. Deborah Mayo discusses statisticians’ predicaments in this way:

*You might aver that we are too weak to fight off the lures of retaining the status quo – the carrots are too enticing, given that the sticks aren’t usually too painful. I’ve heard some people say that evoking traditional mantras for promoting reliability, now that science has become so crooked, only makes things worse. Really? Yes, there is gaming, but if we are not to become utter skeptics of good science, we should understand how the protections can work. In either case, I’d rather have rules to hold the “experts” accountable than live in a lawless wild west*3*,* p. 23.

If we are not to become utter sceptics of the science of modelling, ‘protections’ need to be developed there as well.

### A.2 “Now computer modelling has come into the media and policy limelight”

The irruption of modelling jargon, such as “flattening the curve”12 into public life has led to remarks about the pandemic operating a “domestication of modelling”13. Thus, “COVID-19 is coming to be known in maths and models”14:

*With COVID-19, we see that maths and models have agency as drivers of social action, translating models into citizen science and advocacy. #FlattenTheCurve entangles science into social practices, calculations into materialisations, abstracts into affects, and models into society.*14

This new kind of bonding between modelling and a pandemic is among the reasons for our reflection. This has also led to reactions, such as the expression “Wild-Ass Covid numbers”, coined by Rush Limbaugh, who goes on to add “The minute I hear anybody start talking about models and modeling, I blanch”. In many countries, the scientific models used to project the future course of viral spread have become deeply politicized15.

It is not unusual for science that its methods and its difficulties become a currency in political disputes16–18. There might well be frictions between experts and politicians on which numbers should be used. Governments might not find the method they dream of, or, conversely, experts might err to meet the political movement they aim at. During the New Deal in the US, Government officials expressed dissatisfaction with 1920s (pre-random) economic models proposed by statisticians, seeing them as useless to fight against the Depression. When statisticians came up with the new model of random sampling to help the Government fight the Depression, the Government did not understand its functioning and did not adopt it19.

Coming back to COVID-19, several voices have lamented abuse and misuse of modelling. For example, while John P. Ioannidis and Nassim N. Taleb disagree on everything which concerns the pandemic, they both see a failure in model-based forecasting20. A more damning judgment is offered by Caduff21, for whom what he perceives as excess reaction and fear is fuelled by neoliberal policies (systematic disinvestments in public health), nervous media reporting, authoritarian longings and

*mathematical disease modelling – a flexible and highly adaptable tool for prediction, mixing calculations with speculations, often based on codes that are kept secret and assumptions that are difficult to scrutinize from the outside.*

Caduff21 identifies the disease model report released in the United Kingdom22 (see section B2 below) a few days after Italy’s surprising (to Caduff) national lockdown as a topical moment, where “shock and surprise triggered a chain reaction in the pandemic response”.

### A.3 “Precise, reliable numbers” for COVID-19?

According to Waltner-Toews et al.23:

*Known unknowns [in the present pandemic] include: the real prevalence of the virus in the population; the role of asymptomatic cases in the rapid spread of the virus; the degree to which humans develop immunity; the dominant exposure pathways; the disease seasonal behaviour; the time to global availability of an effective vaccine or cure; and the nonlinear response of individuals and collectives to the social distancing interventions in the complex system of communities interconnected across multiple scales, with many tipping points, and hysteresis loops (implying that society cannot re-bounce to its pre-corona interventions state).*

At the moment of writing the range of available numbers for key parameters fluctuates24–27.

For example, as far as the role of asymptomatic patients is considered, the Centre for Evidence-Based Medicine indicates that “between 5% and 80% of people testing positive for SARS-CoV-2 may be asymptomatic”26 (April 2020), while an estimation of the asymptomatic cases on board the Diamond Princess cruise ship in Yokohama, Japan, completed on February 2020 arrived at a range of 15.5%-20.2%28.

Rarely has the agitated affair of science in the making – so distinct from the Olympic certainty of science textbooks – been so evidently on display. Society can perhaps gain a better understanding of science by accepting an image of it closer to its social reality.

Of course, it is harder, but not impossible, to identify at least some unknown unknowns. Unknown unknowns result from false convictions, as well as unknowns we aren’t aware of29. Involving a diverse group of stakeholders can help identify false convictions, as well as sometimes identifying unknowns many are not aware of. Humility is another important antidote. Further, being nimble in recognising and responding to surprises is also key to managing unknown unknowns.

As an example of a recent surprise, important studies about the potential of chloroquine and hydroxychloroquine published in the Lancet and NEJM were retracted by their own authors, who admitted limited control of the original, privately owned, data sources30.

### A.4 “Models serve society extremely well”: weather forecasting models

Myriad invisible models make our life better and safer. Models for weather forecast are an example in which a mutual process of domestication has taken place between models and society. By producing useful short-term predictions constantly updated by new information, and by communicating uncertainty carefully, these models make it normal for us to read on our mobile that tomorrow will be mostly sunny with a 20 per cent chance of rain. From agriculture to transport to energy, virtually all sectors of the economy benefit from these models31. Weather forecast becomes controversial only under conditions of extreme political interference, as shown by the recent story of hurricane Dorian in the US, or when high stakes events such as storms or flooding complexify the transmission from the technical knowledge of meteorologists to the takers of momentous political decisions, such as evacuation for coastlines or cities, situations for which apparently better societal arrangements are still needed31. Paradoxically, as the models become more visible, their use requires better political and societal coping strategies.

### A.5 “Best practices for responsible mathematical modelling”

There is no lack of offering in terms of existing guidelines and guidance for good practices, as every discipline goes about it in its own separate way in the multiverse of mathematical modelling32. This happens since – unlike statistics – modelling is not a discipline of its own, though the challenges met in the two domains are similar; modelling experiences difficulties in endowing itself with agreed quality standards for validation and verification. Relevant sources are32–34. It has taken more than twenty years (in the 1920s and 1930s) to make the practice of random sampling and the calculation of sampling errors understandable by and legitimate for government agencies (instead of “purposive selection” or “quotas selection” as it came to be called at the end of the 1930s)19. In the field of mathematical modelling it took two decades to achieve a modicum of agreed principles and language in the restricted community of practitioners of sensitivity analysis35.

There are good books about modelling36, about the design of numerical experiments37, about modelling practices and problems in different fields31,38,39, and good general prescriptions from impact assessment guidelines40–43.



Figure 1 Robert Rosen’s modelling relationship. For Rosen the act of encoding is not driven by causality, which would determine the model specification, but by the craft of the modeller. In other words, given the same data and extant theories, different modellers would produce different models and inferences.

Modelling hardly obeys a universal theory, though theoretical biologists would probably contend that Robert Rosen offers one in his “Life Itself”44,45. Rosen’s idea is that of modelling as a craft, and of the indeterminacy of the model given the evidence. An example of how different modelling teams can reach different conclusions given the same evidence is in46,47.

In the field of environmental sciences, modelling articles of increasing ambition tackle iterative model evaluation33, holistic system for model validation48, system of system modelling framework49, contextual practices specific to the different phases of the modelling process50 and grand challenges in socio-environmental systems modelling51.

Several of these articles correctly identify the need for a social-sciences input to the practice of modelling, and the need for reflexive approaches; see 52 and 32 for a discussion. Some authors34,52 also ran a survey among modellers to elicit their preferences.

In the field of research on the impact of infectious diseases interventions, in particular, several efforts have been made leading to quite similar conclusions about (i) minding context and purpose, (ii) revalidating the model any time it is revised, (iii) not compressing the uncertainty, (iv) revealing the conditionality (dependence upon assumptions) of model predictions, (v) cautious communication of the results, (vi) strengths and limitations of the analysis being equally important53. As another example, in the field of modelling rare infectious disease we find a checklist including stakeholder engagement and structured sensitivity analysis54. Poor practices bear a cost, as recent retractions, resignations, and cancelation of clinical drug trials at Duke University involved unreleased and unreproducible code55.

In the field of health care models, the need for better model transparency was also situated in the context of science credibility or reproducibility crisis, e.g. in56:

*Recommendations [for model validation] are particularly important in light of high-profile examples of scientific misconduct and fraudulent research published in leading scientific journals, leading to increasing emphasis on transparency and “shining a light on black boxes”*55.

Several of the authors of the present Comment have contributed work on science’s crisis and new ways forward3,6,7,17,46,57–61. Thus, we consider this an important aspect to consider, especially now that the difficulties in biomedical research62 cast a shadow over the fight for a rapid vaccine.

For some authors, laboratories that adopt questionable practices and cut corners in order to publish are more likely to reproduce than the virtuous ones, a phenomenon known as the “Darwinian fitness” of bad science63. Statisticians complain of similar causes for abuses of their discipline via P-hacking or HARKing64. We suggest that the same type of malpractice is tempting in the case of mathematical modelling, due to the absence of disciplinary oversight and the malleability of model assumptions32,65.

As noted also in the medical field, guidelines and best practices are rarely implemented54. This was also shown by our discussion in the Comment, which included the COVID-19 pandemic as a special case. For this reason, we believe that change will not be brought about by guidelines alone, but by the new societal sensibility and awareness of models that COVID-19 has imposed.

## Mind the assumptions

### 1.1 “Models are often imported from one application to another”: seismic risk

Poisson processes work well for radioactive decay but have no scientific justification in seismology, where they are widely applied66,67. Similarly, fault trees, which may be reasonable for engineered systems (e.g., spacecraft or nuclear power plants), are not justified for assessing seismic risk, where they are also widely used68. More examples from AIDS to fisheries can be found in38.

### 1.2 “A model used in the UK to guide transport policy”

*In the UK, investment decisions on transport projects are evaluated using a model known as WebTAG. In order to obtain state funding for a transport project, an appraisal must be made in accordance with highly detailed official guidance contained in WebTAG and the Treasury ‘Green Book’ guidance on project appraisal. In the world of WebTAG, time is given a monetary value depending on which of thirteen modes of transport an individual uses. The time of a taxi passenger is worth £13.57 per hour (as of 2018 and measured in 2002 prices) but the taxi driver’s time is less valuable, at £9.94 per hour. Hedge fund managers walking to work and journalists cycling to their offices share a time value of £7.69 per hour, but any delay to the Deliveroo courier on her motorbike is priced, like the taxi passenger’s, at £13.57 (less if she uses a pedal bike). The model demands that this level of precision continues into the future. Growth projections yield predictions of how valuable the time of each group will be in 2052, to the penny. If you would also like to know how many people will be travelling in a car on a weekday evening in 2036, the WebTAG spreadsheets will provide an answer. This exercise in fantasy ensures that every cell in the spreadsheet can be filled, and that at the end of the process some numbers will be provided.* 69

### 1.3 “Global uncertainty and sensitivity analysis”

Uncertainty analysis (UA) or uncertainty quantification looks at how the uncertainty in the input variables affect the output of the model, via e.g. an error propagation analysis. Sensitivity analysis (SA) investigates the relative importance of the input variables in determining the uncertainty in the output70. The appeal of SA is that most often a minority of the input variables accounts for a majority of the uncertainty in the output. Unfortunately, most sensitivity analyses seen in the literature are of poor quality, i.e. performed by moving one factor at a time (one-at-a-time SA), an approach which is only valid for linear models35.

Referring to Figure 2, UA, SA can help to find the soft spot in which model error is minimal. There are indeed other lessons that can be drawn by exposure to SA. One of these suggests that running a model just once, whatever the context and the phase of development of a model, may not be the best course of action. Since many inputs are uncertain, and today’s computers powerful, models should possibly be run a certain – even small – number of times, to get a feeling of what is happening with the model output/inference. Additionally, moving the model away from its comfort zone usually uncovers new errors or inadequacies – following what modelers jokingly call Lubarsky's Law of Cybernetic Entomology: ‘There's always one more bug’. Uncertainty and sensitivity analysis can also be used to explore alternatives in the model structure. The analyst can do this using triggers71 whose value activates one or another among a set of possible model structures or assumptions. If properly executed, this exercise may allow a ‘modelling of the modelling process’72, i.e. at the end of the analysis the uncertainty in the output will include contributions from model structures and assumptions together with an assessment of how much each of these ‘weighs’ globally. For example, one might discover that alternative ways of normalizing the variables before aggregating them in a particular exercise do not influence the result73. Another lesson from a probabilistic or global SA is the one treated in this Comment about context and purpose: SA is not performed on some unspecified model output, but on the model output that matters. Thus, SA has to be redone anew any time the model is asked to answer a different question, as argued in74. Moreover, it may happen that uncertainty and sensitivity analysis tell the analyst that the answer is not there – most often because it is too fuzzy/uncertain to be of use, see an example in75.

Practitioners76,77 have noted that discovering too much uncertainty in a model may lead the analyst to artificially compress the uncertainty in the input, so as to keep the uncertainty in the model output (the prediction) within more acceptable bounds. One rule of sensitivity auditing (Section 3.5) is to check against artificial compression of the uncertainties.

The analyst would be best advised to be candid under these circumstances, not let the love for one’s model prevail, and possibly change either the model or the question asked from it.

In fact, one may discover that this can indeed provide useful clues. UA and SA can also be used *via negativa*, for example to disprove that a certain conclusion can be reached at all, see an example in75. Finally, analysts would be well advised to renounce performing a pseudo-probabilistic SA where all uncertain input –magically – have the same 5% or 10% uncertainty range. The most delicate part of a UA, SA is to identify plausible ranges for each uncertain input, and if an analyst does not wish to go through this laborious step, there is no point in performing a SA. An example of this practice is in a recent modelling of SARS-CoV-2 infection in the WHO African Region78 (see Section 4.4).

A recent application of UA, SA containing some of the recommendations above, shows that the predicted world-wide demand on irrigated land by 2050 is seriously underestimated71.

As far as COVID-19 is concerned, good examples of global sensitivity analysis are available, where all uncertain factors are changed simultaneously to explore properly the space of the inputs79,80. In other studies this is not done81,82, and sensitivity is studied moving one factor at a time. Since the models studied are in all likelihood non-linear, this approach may be inappropriate.

Sensitivity analysis can be usefully complemented by sensitivity auditing (Section 3.5).

## Mind the hubris

### 2.1 Why talking of a modelling hubris?

Modelling appears to share some of the features of what Sheila Jasanoff calls “Technologies of hubris”83. Developed to reassure the public, and “to keep the wheels of science and industry turning”, these technologies include:

*a series of predictive methods (e.g., risk assessment, cost benefit analysis, climate modelling) that are designed, on the whole, to facilitate management and control, even in areas of high uncertainty. These methods achieve their power through claims of objectivity and a disciplined approach to analysis, but they suffer from three significant limitations.*

According to Jasanoff these limitations are

* Generating overconfidence thanks to the appearance of exhaustivity;
* Pre-empting political discussion;
* Remaining limited in the methods’ capacity to internalize challenges that arise outside their framing assumptions.

Jasanoff contrasts technologies of hubris with “technologies of humility”84 which invite researchers

*to reflect on the sources of ambiguity, indeterminacy and complexity …*

*to reframe problems so that their ethical dimensions are brought to light …*

*to alleviate known causes of people’s vulnerability …*

*to pay attention to the distribution of risks and benefits, and*

*to reflect on the social factors that promote or discourage learning.*

In relation to the COVID-19 discussion, humility in the development and use of mathematical models to understand the pandemic and cope with its consequences could be helpful in the long run85.

### 2.2 “Trade-off between the usefulness of a model and the breadth it tries to capture”

The conjecture of O’Neill86 (p. 70) starts by suggesting that a simple model may fail to describe the system, and thus lead to bias or systematic error. In general, the more complex a model is, the better it can fit the “training data”. But improved fit to the training data may come at the cost of “generalization error” when the model is applied to new data. In statistics a similar trade-off is that between under- and over-fitting. If a model fits the existing points too well, it will do badly when new points are added to the pool87. 

Figure 2 Too simple a model may be a poor description of the system (model inadequacy curve). Increasing complexity comes at the cost of adding model parameters, whose uncertainty propagates to the model output (propagation error). The model error results from the superposition of the two effects. Redrawn from Saltelli, Nature Communications **10**, 1–3, 2019.

In system analysis this is called Zadeh’s principle of incompatibility: when complexity increases “precision and significance (or relevance) become almost mutually exclusive characteristics”88. Model hubris, e.g. finding oneself on the right side of the error curve in Figure 2, is a known problem of modelling32. An old quote from the field of hydrology is89:

*[…] most simulation models will be complex, with many parameters, state-variables and non linear relations. Under the best circumstances, such models have many degrees of freedom and, with judicious fiddling, can be built to produce virtually any desired behaviour, often with both plausible structure and parameter values.*

Modellers use the term hyper-parametrized when a model has more parameters than can be reasonably calibrated given the existing data. Of course, a model can be hyper-parametrized, on the hubris side of the error curve, and miss important aspects of the phenomenon being modelled.

The trade-off between complexity and error (or bias and variance) is a common topic in books about modelling86,39 and statistics87,90. In modelling, it takes a different flavour depending on the scope of modelling – e.g., predictive/prognostic vs. causal/diagnostic. In predictive modelling a useful concept is that of uncertainty cascade39 (p. 86) due to the increasing numerosity and importance of the uncertainties as more elements are added to the model. The tendency to couple several models in Integrated Environmental Models (IEMs) or Integrated Assessment Models (IAMs) in order to realize socio-environmental systems modelling51 is prone to these problems.

In addition to the uncertainty of the parameters, also the flexibility of the model should be considered. The more flexible a model is, the greater its ability to model noise (i.e., accidental features) in the data, rather than just signal. The additional parameters may improve fit to the training data, making its predictions less reliable, because the noise in the new data is unlikely to be the same as the noise in the training data. Regularization methods allow control for the complexity of a model and monitoring its prediction risk, thus estimating the model more accurately91. In this respect, it is important to distinguish

* *error*, i.e. the difference between the fitted model and the truth, or between the predictions of the model and what actually happens;
* *formal uncertainty*, the error bars on parameters in the model or predictions from the model, calculated on the assumption that the model parametrization is correct – i.e., what statisticians and modelers usually report;
* *actual uncertainty*, which takes into account the fact that parametrization itself imposes an approximation, that the observational error might not be Gaussian, etc. – which is what sensitivity analysis and some approaches to inverse problems address.

Most contemporary statistics concerns itself with formal uncertainty and ignores actual uncertainty, with predictable consequences for the reliability of inferences. The formal uncertainties for the coefficients in a parametrization that is not rich enough to include the “true” model are generally poor. For instance, fitting a constant to a large set of data will have small formal errors, even if the true model is nothing close to constant. The result is that including additional terms in the model increasingly means fitting the noise rather than the signal, because the signal at high frequencies is greatly attenuated.

Merrick and Weyant discuss this problem as the trade-off between the accuracy of representation enabled by the available data and model parsimony, and suggest tackling it using established information theoretic ideas92, with the objective to achieve model simplification while minimizing the loss of information.

Finally, it is worth mentioning that models are often the result of trial and error, formal and informal “model selection” – sometimes referred to as the *garden of forking* path93,94 – which introduces uncertainties and biases that are rarely taken into account in predictions. Part of responsible quantification is to explain what else you tried and abandoned in arriving at the reported model. In this respect some sort of “modelling of the modelling process”, which ideally retraces a modeller’s steps, could be a helpful exercise; see72 for an application to sensitivity analysis.

### 2.3 Model accuracy should be measured prospectively

In general, predictive models should be evaluated by holding them up to future data. Instead, what we often see in economic modelling, COVID modelling, earthquake modelling – to name just a few – is re-jiggering the model as new data come in and never measuring or reporting actual predictive accuracy. The model becomes a moving target, continually adjusted to fit the past (as of the present moment), with no evidence of how it really works prospectively. This is sometimes called “Texas sharpshooting”, where you shoot first, then draw the bullseye where the bullet hits, and is also very close to the “judicious fiddling” mentioned above89.

Another perspective on possible model misuse it offered by Paul Pfleiderer95 with his description of chameleon models in economics. A chameleon model:

*asserts that it has implications for policy, but when challenged about the reasonableness of its assumptions and its connection with the real world, it changes its colour and retreats to being a simply a theoretical (bookshelf) model that has diplomatic immunity when it comes to questioning its assumptions.*

The diplomatic immunity alluded to by Paul Pfleiderer is that related to a very famous and contested statement of the economist Milton Friedman, for whom “the more significant the theory, the more unrealistic the assumptions”96. In Economics this is a fascinating and never-ending discussion, touching on the level of abstraction in economic theory (see Section B1). At the most fundamental level, economic theories can be based on metaphors from physics (physics envy) or on biology, with great differences in implications for action. The former metaphor, with its concept of equilibrium, is more amenable to mathematical treatment then the latter, based on disequilibrium97. For this, some economists can be heard saying that their discipline “follows the line of least mathematical resistance”97.

### 2.4 “A simpler model” for HIV forecasting

*Tasked with creating a model to guide policymakers as to how the disease would spread and the level of intervention necessary, the WHO designed a complex model informed by the latest country-by-country demographic data. A far simpler model was developed by mathematicians Robert May and Roy Anderson*98*, who came up with more pessimistic projections for the spread of HIV. Unfortunately, their projections proved much closer to the eventual outturns. AIDS infections accelerated across the world, causing particular harm in Southern Africa: in 1990 there were estimated to be 120,000 people living with AIDS, a number which had grown to 3.4 million by 2000. The number of new HIV infections had risen ninefold. The world was, it seemed, a much less stable place than the WHO model had assumed.*

*Why did the (apparently) more sophisticated WHO model fail compared to May and Anderson’s simple one? The key factors governing the spread of disease included the probability of an infected person transferring the infection to another person. May and Anderson realised that the probability of infecting another person had two components: the probability that any sexual act would transfer infection; and how many sexual partners infected people had. It was crucial to distinguish between the two. An HIV-positive sex worker who sleeps with 10 different people is more likely to spread the disease than someone who sleeps with the same person ten times. But the WHO model did not make this distinction, and that is why its predictions of the spread of AIDS were, tragically, inaccurate.*69

### 2.5 “Superspreading events”

The present discussion of COVID-19 “superspreading events” as being responsible for a large portion of the infections has similarities to the AIDS test case just discussed. Thus, in COVID-19 the infection would seem to be driven by a feature of the event, e.g. a setting where people sing or shout in a closed setting99. A recent study performed in Israel100 points to the existence of “superspreaders”:

*Genetic analysis of more than 200 SARS-CoV-2 genomes from people across Israel show that only 1–10% of infected people caused 80% of the next wave of cases. The results illustrate the power of ‘superspreaders’ in viral transmission.*

### 2.6 “Yucca Mountain repository”: a million years of certainty

The Total System Performance Assessment by the US Department of Energy for the risk assessment of the Yucca Mountain repository for radioactive waste disposal (TSPA) is composed of 286 sub-models with thousands of parameters, tasked by regulators to predict one million years of safety; see chapter 3, “Yucca Mountain: A Million Years of Safety”,38 pages 45-65. The case is also described elsewhere101. The discovery of 36Cl at the level of the repository, 300 m underground, surprised the investigator involved in modelling the repository system. Since the isotope comes from nuclear weapons tests before 1963, its presence implied that water could travel 300 meters in less than fifty years, while, according to the percolation rate used in the model (less than 10 mm per year), this time should have been of the order of tens of thousands of years. Modellers may be interested in knowing that the error committed at Yucca Mountain (assuming a homogeneous medium without preferential pathways for migration of contaminants) was the same committed by the state authority modellers on the occasion of the Love Canal affair102, a serious case of contamination with many innocent victims. As a local activist, a housewife103 helped identify the problem, the episode gave birth to the so-called movement of popular104 or housewife102,103 epidemiology.

It would be perfect to close the Yucca Mountain text case here, but history decided otherwise. New determinations105,106 have severely questioned the presence of ‘pulse isotopes’ at the repository site, thus leaving the entire matter still hanging on political decisions107.

### 2.7 “Complexity as an end in itself”?

At times, complexity is sought as an end in itself, if not as a marketing strategy. The field of algorithms offers the example of a software for recidivism, COMPAS, commercialized with 137 features, whose prediction can be reproduced with just two features108. Similar cases and the perverse implications of these algorithms are discussed in “Weapons of Math Destruction” by Cathy O’Neil109. An analysis of algorithms used in settings such as criminal justice and child-protective services revealed a consistent inability of existing algorithms to predict life outcomes110: using an approach known as common-task method, 160 different teams built predictive models for six life outcomes using data from the Fragile Families and Child Wellbeing Study. The results were disappointing in that the best predictions “were not very accurate and were only slightly better than those from a simple benchmark model”110. While machine learning methods are extremely popular at the time of the present work, their performance has been questioned in relation to simpler and faster statistical methods111.

### 2.8 “Models tend to be developed with large teams”

Attributing liability for software-driven decisions is an issue fraught with several levels of difficulties, among which one of the most prominent is the well-known problem of many hands. In structured organisations, where software development is the result of a collective work, attributing individual responsibility may be challenging112. See the example of ‘defeat devices’ in Section C4.

## Mind the framing

### 3.1 “No one model can serve all purposes”

*… a model built for one purpose but then used for another needs to be re-justified for the new purpose and this will probably mean it also has to be rechecked, re-validated and maybe even re-built in a different way.*74

### 3.2 “The technique is never neutral”

The analytical tool used to investigate a practical problem to a large extent determines the nature of the conclusions. The German sociologist Ulrich Beck made the point in his 1986 work Risk Society113:

*It is not uncommon for political programs to be decided in advance simply by the choice of what expert representatives are included in the circle of advisers.*

The topic of non-neutrality of technique, and of how methodological choices condition the generation of narratives, is often examined in sustainability science114.

The non-neutrality of technique is also relevant due to the knowledge asymmetry between producers and users of mathematical modelling. Modellers know vastly more about the conditionality of the predictions of their models than is usually communicated to the users/customers of the analysis33. This is not necessarily the result of bad faith but a consequence of the often-large number of assumptions used in modelling. One rule of sensitivity auditing (Section 3.5) is the hunt for hidden or implicit assumptions.

### 3.3 “GENESIS model of shoreline erosion”

The model GENESIS used for behaviour of shorelines in response to coastal engineering and/or beach replenishment activities115 has been criticized for not tackling the system being studied, the shorelines, but an engineered vision of the same system. Beach and coastal erosion models used to justify coastal engineering projects include many idealized assumptions that have been empirically falsified. According to Orrin H. Pilkey116, one source of the problem in GENESIS is an engineering mindset as applied to ecological problems:

*… it is difficult to transfer the mathematical modelling approach from predicting the behavior of steel and concrete structures to predicting the course of earth surface processes,*116p. 163.

*All of this uncertainty makes GENESIS, at best, a qualitative, not quantitative model, and at worst a model that, after a certain amount of assuming and adjusting input parameters, produces a result that the coastal “expert” employing its services expected - a way of backing up one's judgment with what appear to be real numbers,*116 p. 172.

GENESIS and its limitations have been discussed in other works38,115. A common theme of these critiques is the use, in GENESIS, of averages for natural phenomena, such as storms, which do not lend themselves to such a mathematical simplification: beaches are not eroded by average storms. In other words, as discussed above, a model may be complex and still miss important details:

*Perhaps most common among abuses, and not always easy to recognize, are situations where mathematical models are constructed with an excruciating abundance of detail in some aspects, whilst other important facets of the problem are misty or a vital parameter is uncertain to within, at best, an order of magnitude. It makes no sense to convey a beguiling sense of “reality” with irrelevant detail, when other equally important factors can only be guessed at.*98

### 3.4 “Value of a statistical life”

Policy prescription may need recourse to cost-benefit analysis and related concepts, such as the value of a statistical life (VSL) used in117 to conclude that social distancing in the US will lead to a net benefit of about $5.2 trillion. Please note that here as elsewhere in the present work we are not contesting the benefits of a policy, in this case social distancing measures; i.e., we do not criticize the policies per-se, however different the opinions about a given policy might be, but the use of mathematical models to justify the policy, when such a justification appears doubtful or instrumental.

Concepts such as VSL – when moved from actuarial science to complex environmental, societal or global health issues – lose their appearance of rigour, and may appear to disguise political problems as technical ones114, while offering a false sense of precision and control118. Concepts of VSL are already controversial when applied where they should belong, e.g., in the setting of compensatory damages119, as happens when an airplane falls and the families of the victims are given different compensations depending on the nationality of the deceased or the jurisdiction in which the case is filed120. Within the US different regulatory agencies use different values of VSL119. Porter describes the concept of VSL as a “matter of continuing controversy”121, in his text “Objectivity as Standardization: The Rhetoric of Impersonality in Measurement, Statistics, and Cost-Benefit Analysis”. Here we learn that the value of a human life was already the subject of satire in the XVII century, when it was suggested by the influential William Petty that Irish – save for a few cowherds – should be forcibly moved to England on the ground of their higher value there compensated for the cost of the shipping121.

For most of the twentieth century, the value of life in cost-benefit analysis for water and road projects was based on a life insurance number, the discounted value of lost earnings, even though it was widely recognized as the wrong thing to measure. The reason for this choice was that in this way the calculation was more readily standardized.

VSL is a possible ingredient of cost-benefit analyses, which are themselves defined, even by their advocates, as economic solutions to political problems121. These analyses are normally loved or hated through the lenses of the desired end – improve efficiency, increase profit, promote hygiene and sanitation, achieve better education, uphold the care of the environment, and so on. It is not surprising that all that is technical in a cost-benefit analysis is the subject of conflict and manipulations. Discount rates, for example, are wanted low by those who favour the implementation of a project, so that the benefits will keep accruing with time, and high by those who oppose the same project121. Since these analyses are unlikely to run out of fashion, some degree of societal circumspection in use might be advisable.

### 3.5 “Involving stakeholders”

Model validation and verification approaches can be distinguished between the purely technical – such as e.g., in using sensitivity analysis, or sharing the code in repositories – and those which we could call reflexive52,122,123, looking at a richer set of attributes of the modelling process, including framings124, expectations and motivations of the developers, tacit assumptions, implicit biases in the use of a particular modelling approach, care in not over-interpreting the results, and so on.

As an example of a purely technical prescription for model quality for the case of COVID-19 see125, where transparency is achieved by sharing the code. While this is surely a useful recommendation, its scope does not match the scale of the problem.

Coming to the more reflexive approaches, NUSAP77,126 is a notational system for the analysis and communication of uncertainty in science for policy. It is based on five categories: Numeral, Unit, Spread, Assessment and Pedigree. While the first three correspond to the usual scientific practices, the last two inform about the characteristics of the number production process and of the involved actors. NUSAP is thus designed for a participatory approach to the construction and evaluation of models, one where stakeholders’ engagement is considered essential. Examples of stakeholder involvement in the modelling process can be found in environmental sciences127, in medicine128, and in mathematical modelling proper47 and has led to the coining of the term “participatory modelling”. NUSAP is recommended in several impact assessment guidelines, such as those published recently from SAPEA (Science Advice for Policy by European Academies)43. See also <https://en.wikipedia.org/wiki/NUSAP>

Similar to NUSAP, sensitivity auditing addresses specifically the quality of mathematical or statistical models. Its seven rules129,130 are:

1. Check against the rhetorical use of the model;
2. Adopt an ‘assumption hunting’ attitude;
3. Detect artificial deflation or inflation of uncertainty;
4. Find sensitive assumptions before they find you;
5. Allow interested parties to make sense of, and possibly replicate, your results;
6. Check the framing against alternative worldviews:
7. Perform a thorough sensitivity analysis.

These points loosely cover the same ground as the five recommendations of the present paper. Sensitivity auditing is recommended in several impact assessment guidelines, e.g. SAPEA 201943 above and European Commission, 200940. Recent examples of sensitivity auditing are in sustainable food production131 and nutrition and public health132. See also <https://en.wikipedia.org/wiki/Sensitivity_auditing>

Both NUSAP and sensitivity auditing are inspired by post-normal science, a set of heuristics for the use of science on issues where “facts [are] uncertain, values in dispute, stakes high and decisions urgent”122,133. It is clearly evident that these elements are all at play in the pandemic. NUSAP and sensitivity auditing borrow from PNS problem-solving strategies in which different kinds of knowledge are brought to bear by recourse to extended peer communities134. These are conceived to include all disciplines and all actors who have a stake or an interest in the issue being faced. See also <https://en.wikipedia.org/wiki/Post-normal_science>.

Two recent articles on the pandemic written with a PNS orientation23,135 touch upon some of the themes of the present Comment, e.g. in terms of scientific humility, of a possible new covenant between science and society, and remark that with COVID-19 the entire world becomes an extended peer community. More perspectives on PNS published in the journal Nature are136–141.

### 3.6 “Participatory forecasting of flooding risk, and in the management of fisheries”

Lane et al.127 offer an instructive example of how modelling work can be co-produced in a way that helps both modellers and the community concerned with the modelling study.

The objective was to alleviate periodic flooding in a small market town, Ryedale, North Yorkshire, UK. The residents were involved.

*[…] knowledge regarding flooding was co-produced. This illustrates a way of working with experts, both certified (academic natural and social scientists) and noncertified (local people affected by flooding), […] We reveal a deep and distributed understanding of flood hydrology across all experts, certified and uncertified, …*127

“Distributed understanding” above means that that knowledge was not all on the side of the experts. Years of modelling stream flows and cost/benefit ratios for flood protection structures had failed to consider an alternative intervention – upstream storage of flood waters – until local stakeholders were brought into the modelling process. According to Lane and colleagues, upstream storage was neglected in the models because of the “use of a pit-filling algorithm that made sure that all water flows downhill”. Interestingly, in order to complete the analysis, the experts had to construct a brand-new model to properly account for upstream water storage processes.

As per the management of fisheries, participatory modelling142 was among the objectives of a project (JAKFISH, Judgement and Knowledge in Fisheries Involving Stakeholders) to address models’ quantitative and qualitative uncertainties in fisheries management.

Interaction with stakeholders advanced the understanding of the potential risks for stocks and fisheries associated with biological, socio-economic and management aspects of the issue. The project was avowedly inspired by post-normal science, as fisheries management was considered to be a typical high-stakes, high uncertainty setting where

…*one cannot rely on textbook knowledge, and trust that scientists alone will be able to give the answers–because there is not one single answer due to the uncertainties and decision stakes involved.*142

Among the main conclusion of the study, based on four test cases, were the acknowledgment of the need to confront diverging agendas early on in the process:

*Stakeholders have an agenda, and at the same time, scientists have scientific agendas or at least personal scientific ambitions*

In these situations, reaching a common agreement and compromise on the purpose of the process is an important preliminary step. The cases also revealed that early stakeholder involvement is a prerequisite for success; this realization is based on empirical observations, and not on value judgments. The trade-off between model complexity and model transparency – discussed in this Comment – was also an issue, and, in two of the cases, ad-hoc models had to be developed. Unsurprisingly, once uncertainty was better acknowledged,

*lower fishing mortality targets were required to maintain pre-agreed stock levels with a certain probability than if no uncertainty was considered.*142

With commitment, time and resources, participatory modelling with model users can be combined with model complexity through the integration of different kinds of models, as demonstrated in an additional fisheries example143. Uncertainty analysis was a strong feature. Based on the modelling, the Australian Fisheries Management Authority rewrote the fisheries management plan for the Australian southern and eastern scalefish and shark fishery, and ultimately restructured this fishery along with Australia’s other federal fisheries.

All of these examples demonstrate benefits of modelling beyond prediction, especially their ability to a) represent and help understand systems, including complex systems, b) provide tools for understanding and managing unknowns, c) provide tools for decision support, d) incorporate participatory processes that expose diversity (e.g., differences in mental models), e) provide tools for integration (e.g., of different knowledge and perspectives) and f) communicate key issues to critical audiences.

## Mind the consequences

Bad modelling damages health and the environment31,38,144. In relation to COVID-19, Nassim Taleb has written that “[I]f we base our pandemic response plans on flawed academic models, people die”145. Discussed among sociologists and other social scientists is the question whether statistics “perform” the world they, at the same time, describe146. Mathematical models are likewise performative147; they influence society and are by this influenced in their becoming, as shown by the many examples discussed in this work, from economics to ecology, as well in the present pandemic. Discussing COVID-19 models’ adaptability and ductility the authors in65 speak of “mathematical models as public troubles”, but conclude on a positive note that these troubles “can be treated as generative events for energising new forms of expertise, as well as for remaking experts and publics, through participation.”

For Caduff21 (Section A2) indiscriminate use of mathematical disease modelling in place of actual monitoring has led to the adoption of what he perceives as unsustainable generalized lockdown measures:

*The lack of testing created a void that was filled by the flexible evidence of disease modeling. In the absence of robust data, disease modeling emerged as the presumably best and only available science to inform policy.*

### 4.1 “Models contributed to crippling the global economy”

The role of numbers in the financialization of the economy and their role in the latest recession has been discussed by philosophers148, historians149, sociologists150, jurists151 and traders152. A highlight of this debate is the so-called Li formula153 for copula calculation on the compound risk of bundled mortgages. The ‘formula that killed Wall Street’154 was calibrated on a period of market upswing, and became totally unrealistic when the market went into downturn. ‘Faith’ in these formulae was possibly instrumental to profit maximisation, with scarce concern for global economic consequences154.

### 4.2 “Once a number takes the center-stage”

Are we looking at all the numbers or to a selective subset which determine our choices to the exclusion of others which would also be possible? The depth of the issue is illustrated in the Santa Clara debate155 and related controversies145. The numbers of deaths and infections have taken center-stage in the narrative of COVID-19. For Emmanuel Didier156, when a set of numbers establishes itself at center-stage, other possible numbers and stories may be neglected, including the potential loss of civil liberties associated to the fight against the virus. Similarly, for Caduff21 the Imperial College model (Section A2 and B2), possibly the most influential reports to emerge in the COVID-19 pandemic – by focusing exclusively on “non-pharmaceutical interventions” – sidelined a possible discussion of a public health strategy known to be effective: systematic testing and contact tracing. For the same Caduff this led to neglect of the social, political and economic implications of lockdowns.

Similar questions are pondered by philosophers157. And these problems are exacerbated by confirmation and other biases158,159.

### 4.3 “Spurious precision adds to a false sense of certainty”

For the mathematician Gauss:

*Lack of mathematical culture is revealed nowhere so conspicuously, as in meaningless precision in numerical computations.* (Carl Friedrich Gauss160)

### 4.4 “An analysis by the WHO Africa”, and the Imperial College’s “510,000 deaths”

An interesting case is a recent modelling of SARS-CoV-2 infection in the WHO African Region78. The model includes twelve parameters for transmission dynamics in a Markov-chain model, ten of which are varied by plus/minus 10%. One should assume that all these parameters are equally uncertain, which is unlikely. If no valid information went into the input of such an uncertainty analysis, how can the output be of use? Surprisingly, the same WHO announces the results as “New WHO estimates: Up to 190 000 people could die of COVID-19 in Africa if not controlled”161. 190,000 deaths were in fact the worst-case scenario, corresponding to all parameters set to their most unfavourable value. In this way, a totally speculative 10% shift in parameters value has become the substance of news about possible deaths.

Something similar happened in the case of the Imperial College model which played an important role in the adoption of policies both in the UK and US162. The predictions of 510,000 COVID-19-related deaths in Britain and 2.2 million in the US are too precise given the uncertainty in model inputs, and the authors’ own limited assessment of the uncertainties, (a range between 410,000 and 550,000, p. 13) achieved by looking at the dependence of the results from the uncertain value of infection rate22.

### 4.5 Opacity about uncertainty damages trust

The classic work here is “Trust in Numbers”163 already mentioned in relation to sociology of numbers. The book notes how trust may depend not just on the solidity of the numbers, but on the needs of expert and client communities.

*The appeal of numbers is especially compelling to bureaucratic officials who lack the mandate of a popular election, or divine right […] Quantification is a way of making decisions without seeming to decide. Objectivity lends authority to officials who have very little of their own.* (p. 8)

Porter’s book guides readers through the different histories and styles of quantification in public decisions in the UK, France and the US. Porter describes the different dynamics of trust in numbers when the quantification is ‘mandated’, i.e. prescribed by regulation in a possibly adversarial setting (the case of US), versus more flexible arrangements (the case of the French engineers of the *Corps des Ponts et Chaussées*). As noted by Ravetz in a review of this work164, the historical perspective of “Trust in numbers” illustrates how even in the hardest of disciplines – high energy physics – a community of trust links the research community and permits it to function163 (pp. 222-223). The last three chapters of Porter’s book – on the rise of cost-benefit analysis, on objectivity in different disciplines, and on the importance of communities in science – could usefully be part of a syllabus in modelling.

### 4.6 “Sociology of quantification”

An introduction to the topic of sociology of quantification is a 1995 work by Espeland and Stevens165. Popp Berman and Hirschman offer a recent, readable review and a pointer to existing literature on the same topic166. Classic works on quantification are from Hacking167, Norton Wise168, Porter163, and Desrosières169, and more recently Didier19. The topic has gained urgency recently due to the perceived downsides of all sorts of ratings and quantifications driven by algorithms109,151,170,171. More than just sociology, these works include history, philosophy, and ethics of quantification, the latter being already advocated by Espeland and Stevens165. Recent works and initiatives on ethics of quantification from the authors are171–174.

## Mind the unknowns

### 5.1 “Docta ignorantia”

*Docta Ignorantia* is an essay by polymath Nicholas of Cusa. In more recent times another polymath, the Danish thinker Piet Hein, noted “Knowing what thou knowest not is in a sense omniscience”. Do we need knowing of not knowing? According to Jerome Ravetz:

*…the art of choosing research problems can be described as sensing where that border [between knowledge and ignorance] can be penetrated and to what depth. Similarly, the art of monitoring for possible accidents or realized hazards, be they in industrial plant or environmental disruption, consists in having a border with ignorance that is permeable to signals coming from the other side, signs of incipient harmful processes or events that should be identified and controlled.*175

The same author also notes:

*… if ignorance is recognized to be severe, then no amount of sophisticated calculation with uncertainties in a decision algorithm can be adequate for a decision.*

In other words, an appraisal of ignorance176 can allow the detection of ‘trans-scientific’ problems, i.e. problems that can be formulated in the language of science but to which science cannot offer a solution177. Quantification in these cases is futile or counterproductive. On transforming a political problem into a technical one, see chapters 12 and 13 in160.

An effective use of models that embraces ignorance is *via negativa*. *Via negativa* is an expression borrowed from theology by Nassim Taleb to indicate knowledge about what is not the case178. Models can be effectively used to show that a hypothesis is false, rather than to prove it179: “a model may confirm our biases and support incorrect intuitions. Therefore, models are most useful when they are used to challenge existing formulations, rather than to validate or verify them. Any scientist who is asked to use a model to verify or validate a predetermined result should be suspicious”. In a later work, Naomi Oreskes180 offers a discussion of the differences between a crisp physical law and a mathematical model: the former can be more easily refuted (falsified) than the latter.

Five points which to some extent overlap with those of the present Comment are given in the text “Wrong but Useful - What Covid-19 Epidemiologic Models Can and Cannot Tell Us”181.

### 5.2 “Opening the door to surprises”;

Both the UK’s Queen Elisabeth II and the US senate were befuddled by the incapacity of their respective economists to justify why their big models could not see the last recession coming. The media reported with titles such as “Queen Elisabeth questions the predictive capacity of British economists”182, and Philip Mirowski describes the issue at length in one of his books183, p.275-286. Mirowski’s critique of the model used in economics (specifically of DSGE, dynamic stochastic general equilibrium models) focuses on the economists’ standard ‘caeteris paribus’ hypothesis (=all the rest being equal) and on the key modelling assumptions of agents’ rationality and markets’ efficiency183.

### 5.3 “Anthony Fauci”, and the use of deep uncertainty methods

A sign of the tension between the demand for quantification and the difficulty of quantifying responsibly is the reply given by Anthony Fauci, member of the White House Coronavirus Task Force and director of the National Institute of Allergy and Infectious Diseases, to a politician insisting on a numerical estimate of casualties: “There is no number-answer to your question”184.

The same point is made in185, titled “Don’t try to predict COVID-19. If you must, use Deep Uncertainty methods”. The three methods suggested are exploratory modelling186, the already discussed global sensitivity analysis70, and robust decision making187 – an iterative process that compares candidate strategies, possible future states of the world and related vulnerabilities, and seeks for appropriate hedging for the trade-offs among these vulnerabilities. These deep uncertainty strategies are similar to the ‘quantitative storytelling’ which stress-test policy options against criteria of feasibility (e.g., do we have the biophysical resources?), viability (can it be done in the present social arrangement?) and desirability (is it what society wants?); see for discussion and references <https://en.wikipedia.org/wiki/Quantitative_storytelling>.

## Questions not answers

### B.1 “Ritualistic use of quantification”

Excessive or ritualistic use of mathematics is a recurring theme in economics. For Reinert188 the mathematization of economics goes in historical cycles, and – by eliminating qualitative differences – overly abstract models may even create economic rent-seeking between nations189,190. According to Romer191 ‘mathiness’, i.e., instrumental use of mathematics, serves to hide normative dimensions. Critiques of models used in macroeconomics for their normative bias can be found in Mirowski183, Romer191, and Stiglitz192. Thompson and Smith193 offer a refreshing picture of the problem of ‘escaping from model-land’. The discussion about mathematical, ritualistic excesses is present in several disciplines:

*… the widespread misappropriation of the language of mathematics in the social and biological sciences has to be one of the great tragedies of our time. Nothing can be sadder than the sight of equations crawling down a page of literary theory, nothing more raucous than the invasion of simple rules of cause and effect into the language of psychoanalysis. Far less obvious in its poverty of reasoning is the inappropriate application of mathematical methods to the analysis of certain scientific problems for which we have no obvious solutions. These projects are usually driven by our inability to cope with the unpredictable — stock-market crashes, hurricanes, earthquakes and epidemics*.194

Statistical rituals are discussed by Gigerenzer195. Models can also be used ritualistically, because their use is prescribed by law or by a bureaucracy163. A cost-benefit analysis may be mandated even when there are no evident ways to assign costs122,163. Nobel laureate Kenneth Arrow, employed during World War II as a weather officer in the US Army Air Corps, recalls when they were requested to produce monthly weather forecasts:

*The statisticians among us subjected these forecasts to verification and they differed in no way from chance. The forecasters themselves were convinced and requested that the forecasts be discontinued. The reply read approximately like this: “The commanding general is well aware that the forecasts are no good. However, he needs them for planning purposes.”*196

### B.2 Models’ assumptions and limitations must be appraised openly and honestly

The limits of modelling should be made evident; for example a quantitative evaluation197 of the coronavirus model of the Institute for Health Metrics and Evaluation (IHME198, also referenced in White House press conferences) finds observations in the real-world fall into the 5% probability zone more than 70% of the time – pointing thus to poor model performance. The model assumes that cases change exponentially, where the exponent is quadratic198. Among other things, that imposes symmetry on the rise and fall. For a broader critique see199. In May 2020 the IHME model was modified to include a disease-transmission component, which took into consideration the relaxation of stay-at-home orders in different US states. According to a comparative study of different models200, this improved the IHME performance.

As for the model of the Imperial College201,202, which warned about the possibility of 510,000 deaths in Britain and 2.2 million in the US and which purportedly convinced UK and US to adopt social distancing measures22, its track record is in our opinion relevant to the present discussion for its pivotal role in the making of policies. Imperial College modellers’ prediction of the impact of the foot and mouth disease made in 2001203 resulted in the controversial204,205 mass culling of animals, as well as predicting up to 50 thousand UK deaths in the same year from mad cow disease (BSE)206 – there were actually less than 200 deaths. Then, there are the 150 thousand UK deaths from avian flu predicted for 2005207 – there were less than 300 globally. Finally, the dire prediction of tens of thousands of UK deaths from swine flu in 2009, after Imperial College estimated 4% mortality rate208 based on209 – there were not even 500.

The topic of these off-mark predictions has been intensely debated in the media210,211.

As per the foot and mouth disease case, and the six (or ten) million or more animals culled due to the possibly wrong prediction203–205, some authors note – albeit with an excess of significant digits204:

*the nightly appearance on television of apparently healthy cattle and sheep being sacrificed to bring FMD under control horrified both rural and urban communities, and involved government at the highest level. In Britain, four million FMD-susceptible animals on 10,157 premises were slaughtered: 2,026 premises were declared infected, 4,762 premises were considered dangerous contacts of infected premises and 3,369 premises were located near to infected premises. A further 2.5 million animals were slaughtered for reasons of welfare, such as overcrowding and compromised nutrition. The official figure for the number of animals slaughtered was approximately 6.5 million, but when the total number of still-suckling lambs, calves and pigs that were slaughtered is included, the total could be as high as ten million.*204

Note that the discussion about the UK response to the foot and mouth disease has been going on for several years204,205,212. In a response to the criticism213, Neil Ferguson noted that a more targeted approach would have been impossible at the time of the outbreak:

*The biggest problem in 2001 was that by the time we realised what was happening, there were something between 30 and 50 infected farms … It took a huge amount of effort to deal with that, and so very intensive surveillance of infected areas proved impossible from the outset.*

Ferguson had also motivated the culling measures in a modelling study published in 2007214.

Coming back to COVID-19, social media discussions have highlighted the issue of the non-easy accessibility/readability of the Imperial College model. According to the available sources, including the same Ferguson215, experts from GitHub and Microsoft supported Imperial College to review the code; the resulting revised version of the original code was then the one that was made publicly available. The history of changes to the original code had been deleted, for the sake of an easier download of the repository, as explained by one of Imperial College co-authors on GitHub. The same person assured however that “the code here is essentially the same functionally as that used for Report 9 and can be used to reproduce the results”. In our opinion, however, it looks evident that Microsoft and GitHub intervention has not been just layout and documentation:

Results have been claimed to be non-reproducible. Some researchers from Edinburgh University, for instance, found a discrepancy of almost 80.000 deaths, even when using the same inputs and pseudo-random numbers. This was firstly dismissed as due to the stochastic nature of the model, but it was later admitted that it was due to a bug in the code.

The discussion is available on GitHub, and the relevant links are available at <https://bit.ly/2TknWR7>. Nevertheless, the journal Nature reported that the revised version of the Imperial College model has received a ‘thumb-up’ from Codecheck, a project which awards “certificate of executable computation”216,217. The positive message of this certificate includes the reproducibility of the code’s results. The piece in Nature reports that other scientists have already privately verified that the code is reproducible216. We felt it useful to report all these details to underscore the urgency of the points discussed in the Comment, from the models becoming the object of controversy to the need for modellers to do their homework promptly and transparently.

A review of 31 different models used in the COVID-19 context led the authors of the editorial in BMJ218 to conclude that the quality of these models was still poor and that models developed in such a short time could well do more harm than good. For this they conclude: “As no COVID-19 clinical prediction models can currently be recommended, clinicians will have to rely on their clinical acumen and shared experiences of best practice for now”.

Panovska-Griffiths219 – in relation to the UK discussion of the epidemic (namely, Ferguson’s model201 vs. Gupta’s model220) – acknowledges explicitly that different models (stochastic individual-based model the former, SIR model the latter) can lead to very different results, especially if they are calibrated on different data.

A discussion of the need for COVID-19 models to explain where predictions differ or change over time is given in199. Interestingly, some modellers now acknowledge the need to inform their users on the performance record of their forecasts, e.g. for different US states200; see Section 2.2.

### B.3 “Statactivistes” and other forms of engagement

The orientation of the present Comment is that principles or criteria of modelling are not separate from those of statistics. The mix of politics, economics, statistics and much else is woven into both. This is why more communication between the two fields would appear natural. Next, we consider that modelling and statistical work do not need to be neutral in order to be good171. What is being objected to in the present Comment is so-called stealth advocacy221, whereby the normative implications of a model are hidden, out of interest or disciplinary bias191. There are examples of modelling and statistical works that do not eschew a normative or political dimension.

The agenda of the French movement of Statactivism222,223 is to ‘fight against’ as well as ‘fight with’ numbers. Its repertoire of tactics against perceived statistical abuse include:

* Self-defence or ‘statistical judo’ – i.e. gaming the metrics, a strategic use of the Goodhart law;
* Exposing the faults of existing measures, e.g. by denouncing the middle-class bias of the existing French consumer price index (PPI);
* Developing new measures, e.g., in relation to the above, a new PPI in defence of the purchasing power of the poor;
* Identifying areas of exclusion and neglect of existing official statistics.

For example, a new barometer for inequality and poverty (BIP40)224 was implemented between 2001 and 2007 by a French collective of activists (*Réseau d'alerte sur les inégalités*) in dissent with official statistics. The French newspaper Le Monde picked up the story, and this led to a fruitful engagement of the activists, supported by unions, with the French statistical authorities, to develop better measures of poverty and inequality. Statactivism takes inspiration from the French tradition of sociology of numbers19,169,222. In the analysis of Alain Desrosières225, who quotes Porter163, statistics can be paradoxically taken as “tools of weakness” to be used by the oppressed. In this vision, which contrasts that of statistics as a tool in the hands of elites, statistical argumentation can be “put forward by dominated groups to break the old order and render injustice visible” 225.

In the US and globally, statactivists have especially fought against management numbers used by the police226. Cathy O’Neil109 (p.91-92) describes how data activists use ‘Hackathons’ to open up software used by the police, in order to check if the algorithms embed racist practices. At the university of Cardiff (UK), a Data Justice Lab examines “the intricate relationship between datafication and social justice, highlighting the politics and impacts of data-driven processes and big data”227. Modellers and statisticians fight gerrymandering in the US,228 and defend in the wider world the integrity of the voting process in elections229. Important normative and philosophical elements also characterize the discussions in the present movement for statistical reform3.

Perhaps a last word on statistics and values can be left to the philosopher Richard Rudner230. He wrote in 1953 that it is impossible to use a test of significance without knowing to what it is being applied, i.e. without making a value judgment on the implications of the inference231. Rudner made this example about statisticians in order to make a more general point: that scientists do need to make value judgments.

### B.4 “The blog of Tomas Pueyo”

Tomas Pueyo uses plain-language explanations of the implications of model uncertainties for policy options in his blog postings on COVID-19 epidemiological models. His Medium article “Coronavirus: why you must act now232” has been read by c. 40 million people as of June 2020 and has received extensive international press coverage (i.e. the New York Times, The Guardian, CNN). According to an interview he gave to *Agencia EFE*233, government members from Peru, Germany and Bulgaria reached out asking for advice on how to manage the epidemic once his op-ed became viral. He is also the author of the article “The Hammer and the Dance”, a metaphor to describe the nuanced responses and the fluid period that will follow the lockdown promoted to crush the COVID-19 disease. As a sign of the present extraordinary state, the New York Times columnist Thomas L. Friedman asks president Trump to read precisely this piece234:

*If you [President Trump] have not seen them, check out the widely referenced graphs in an analysis on medium.com titled “Coronavirus: The Hammer and the Dance” by the engineer-entrepreneur Thomas Pueyo.*

Also as a follow up of our discussion of the controversy surrounding the Imperial College model (Section B2), it is instructive to contrast Pueyo, the ‘engineer-entrepreneur’ with an academic modeller, Professor Neil Ferguson, who candidly admits in an interview235:

*“For me the code is not a mess, but it’s all in my head, completely undocumented. Nobody would be able to use it . . . and I don’t have the bandwidth to support individual users.”*

The same Ferguson notes elsewhere216 that lack of time was a determining factor why his simulation didn’t use current best-practice coding methods.

What’s refreshing about Pueyo’s analysis is that it is responsive not to any single model or prediction but to the wide range of models and model outputs encompassing the radical uncertainties of the pandemic. The approach focuses attention on the uncertainty space within which decisions must be made, and supports strategies that are reasonable under a variety of possible futures while also hedging against especially dire consequences. This approach broadly and implicitly adapts the principles we articulate in our Commentary. It is also reminiscent of the “robust decision making” modelling approach of Lempert and colleagues236.

We felt it useful to align Pueyo and Ferguson to stress the heightened importance of transparency and openness when research is deployed in the service of society, and to reinforce the thesis of the present Comment that modellers cannot be left alone in this enterprise.

### B.5 “Not an end to quantification”

Books have been written on the misuse of statistics62, that of ratings and algorithms109,170, and that of predictive modelling31,38,69. The proximity of these instances of quantification, and the common elements in these disparate fields, are among the themes of the present Comment166. We need reliable statistics as much as we need responsible algorithms and models.

*Modelling is essential to the scientific enterprise. When Steven Shapin, a scholar studying science and technology, talks about “invisible science”*237*—meaning scientific and technological products which improve our life—one chapter could be devoted to “invisible models” underpinning these technologies.*32

## Additional examples mentioned in Table 1.

### C.1 Species extinct

The finding by Urban238 that 7.9% of species would become extinct as a result of climate change is puzzling46. Among the reasons for the impossibility to predict at two digits precision the fraction of species that will be extinct is the present uncertainty about the total number of existing species.

### C.2 Ubiquitous ranking

“The Tyranny of Metric”170 of Jerry Muller has an interesting set of cases. The output and quality of research is intensively modelled and ranked109,170,239, as well as that of higher education170. Rating is believed to have conditioned the higher education sector for the worse, creating a global market for education which has sent prices up and created a distorted system of incentives240. Yet these university rankings can be shown to be arbitrary and volatile.241

### C.3 An example of poor context choice

As discussed in242, the European Commission met serious obstacles trying to support legislation to cut carbon emissions by using a computer code which was proprietary and hence not open to inspection. “The economic model, known as “PRIMES”, is owned by the National Technical University of Athens and is designed to show how using different mixes of energy sources affect the wider economy. The European Commission has used it for many years to help guide the bloc’s energy policies but industry critics complain that its assumptions are impossible to question because the model is privately owned. One trade group, Business Europe, has called for the Commission to use other, more transparent models.”

### C.4 Fraudulent quantifications

Though this is not the focus of the present work, it might be mentioned that models embedded in software are a convenient mantle to hide fraudulent purposes, such as in the well-known case of the automobile industry coding the software of the car so that it would ‘recognize’ testing conditions and hence decrease the car’s harmful nitrogen oxides emissions243. As in modelling, statistics can provide the occasion for fraudulent behaviour. In 1905, an agricultural statistician at the USDA used his position which involved estimating the quantity of crops to be produced during the season to speculate on the crop futures market. He reaped huge benefits and created a scandal. In response, the USDA established in the same year the first “Crop Reporting Board” where statisticians were locked up while performing their task, and Congress voted sanctions in case of violations19.

### C.5 Why bank fails

*Banks do not fail usually for the reasons described by value at risk models (VaR). The case of Northern Rock. In the spring of 2007, the UK bank Northern Rock announced at its AGM that it was the best capitalised bank in the United Kingdom, and would be returning ’surplus’ capital to shareholders. And according to the internationally agreed risk calculations embodied in the Basel regulations that had come into force at the beginning of the year, it was indeed the best capitalised bank in Britain. The risk weights mandated by those new regulations assumed that mortgages were among the safest assets in which a bank could invest; however if you stripped out the risk weighting, the liabilities of Northern Rock were eighty times its equity capital.*

*Northern Rock had moved far from the traditional building society it had been only ten years earlier, which took deposits from retail savers and lent them to home buyers. The bank now financed much of its lending from day to day borrowing in money markets before selling packages of securitised mortgages to other financial institutions. In August 2007 both the market for short term borrowing and the market for re-selling packages of loans dried up and the bank simply ran out of money. Queues formed outside the company’s branches as savers scrambled to get what was left in the tills. The panic subsided after the government guaranteed deposits and the Bank of England provided financial support. In February 2008, beyond rescue, Northern Rock was nationalised.*69

### C.6 Austerity policies and spreadsheet errors

A model was deployed to justify austerity policy in Europe based on the so-called 90% rule by Reinhart and Rogoff 244 on the relationship between growth and debt. The work was found (by a grad student) to have error in the spreadsheet calculations245,246, invalidating the model findings supporting the policies.

### C.7 About consequentialism: the case of the ecological footprint

The success of the Ecological Footprint is to a large extent due to the simplicity of its message and to the desirability of the message it conveys. It offers a crisp, single number capturing the impact of humans on the planet. The mathematical protocol developed by the Global Footprint Network (GFN) aims to assess humanity’s impact on the planet by aggregating across scales and compartments. Thus, the result of the analysis can be communicated as an overall measure of human impact, such as the ‘Earth overshot day’; e.g., “July 29 is Earth Overshoot Day 2019” (www.overshootday.org) signifies that in less than seven months, humanity exhausts Earth's budget for the year. This number is problematic. To give an example, the Ecological Footprint protocol suggests that we are overshooting the capacity of the planet by about 50%, but the quantity of nitrogen used in agriculture alone would already require 2.5 planets to be fixed by natural processes – not just 1.6. The 1.6 number could become 16 or 160 or infinity, depending on what impact of humans is considered – infinity corresponding to irreversible processes, e.g. humans extinguishing a species by overfishing or by pesticides. Several practitioners – including a commission to evaluate measures of progress led by economists Joseph Stiglitz, Amartya Sen, and Jean-Paul Fitoussi247 – have objected to the EF. See248 for a discussion and249 for a reply of the proponents.

### C.8 Clinton’s precise victory

A much-discussed case of spurious accuracy is Nate Silver’s election forecast models250 that predicted Clinton victory in 2016 to decimal point accuracy251.

### C.9 NASA’s MESSENGER

Thanks to clever modelling, NASA could position around Mercury the probe MESSENGER, launched in 2004, after five billions miles and six ½ years69. The trajectory of the probe involved flyby manoeuvres at Earth, Venus, and Mercury to reduce the speed relative to Mercury, for final insertion into the planet’s orbit.

## Table 1, list of examples

|  |  |  |
| --- | --- | --- |
|  Section | **Positive and negative examples (see SM for more details)** | **Reference** |
| 1. Mind the assumptions
 | **Negative**: Beach models used to justify coastal engineering projects, which include many idealized assumptions that have been empirically falsified. See chapter 8 from 31, chapter 5,6 in 38. For the model GENESIS see also 115. | 252 |
| **Negative:** Transport policy in the UK using WebTAG, which needs as an input the average number of passengers sitting in a car decades from now 69. | 253 |
| **Negative**: IHME model on COVID-19 assumes the cumulative death rate to be exponential, where the exponent is quadratic. Among other things, that imposes symmetry on the rise and fall. See 199 for other critical aspects of the model. | 254 |
| **Negative**: Early COVID-19 models were not subject to adequate sensitivity analysis. Global sensitivity analysis should have been applied. | 81 |
| **Negative**: One-at-a-time sensitivity analysis on a model for COVID-19 in Italy. Global sensitivity analysis should have been applied. | 82 |
| **Positive**: Two examples of accurate global sensitivity analysis on a simple model for COVID-19. | 79,80 |
| 1. Mind the hubris
 | **Negative:** Transport policy in the UK using WebTAG, a model whose complexity is not matched by the quality of the data used in its calibration according to 69 | 253 |
| **Negative**: The Total System Performance Assessment (TSPA) by the US Department of Energy for the risk assessment of the Yucca Mountain repository for radioactive waste disposal is composed of 286 sub-models with thousands of parameters, described in chapter 3 of O. H. Pilkey and L. Pilkey-Jarvis38, who note that the model is designed to predict “A million years of Certainty”. | 255 |
| **Positive:** Sridhar and Majumder (2020) recommend humility in the development and use of mathematical models to understand COVID-19 and cope with its consequences. | 85 |
| 1. Mind the framing
 | **Negative**: PRIMES, a proprietary model for EU energy policy, was controversially used for policy negotiation although it was a proprietary, non/accessible software242. | 256 |
| **Negative:** Ferguson et al.’s COVID-19 model was presented as if it offered a basis for policies far beyond those it addressed257. | 201,202 |
| **Negative:** Cost-benefit models of COVID-19 policies are highly sensitive to assumptions (e.g., on the value of a statistical life) which are not neutral. For a general discussion see163. | 117  |
| **Positive**: In the UK discussion on the start of the epidemic (namely, Ferguson’s model vs. Gupta’s model) the work in219 acknowledges explicitly that different models (stochastic individual-based model the former, SIR model the latter) can lead to very different results, especially if they are calibrated on different data.  | 219220 |
| **Positive:** Lane et al.127 offer an instructive example of how modelling work in flood management can be co-produced in a way that helps both modellers and the involved community. | 127 |
| **Positive:** Participatory modelling in the project (JAKFISH142, Judgement and Knowledge in Fisheries Involving Stakeholders). Likewise in143, by the Australian Fisheries Management Authority. | 142,143 |
| 1. Mind the consequences
 | **Negative**: The Northern Rock case shows banks don’t normally fail for reasons described in value at risk models69.  | Northern Rock case  |
| **Negative**: The model and theorem of Reinhart and Rogoff on the relationship between growth and debt was used to justify austerity policies. The theorem was found to be the result of a coding error. See 245 and 246 for more details. | 244 |
| **Negative**: Copula calculation (Li formula) on the compound risk of mortgages was calibrated on a period of market upswing. Financial modelling of pricing derivatives played a major role in the last recession 148,149,152,154. | 153 |
| **Negative**: About consequentialism: ecological footprint models seek to account for the state of the entire planet with a single number248,249. | 258,259 |
| **Negative**: The election forecast models that predicted Clinton victory in 2016 to decimal point accuracy251. | 250 |
| **Negative**: Numbers running wild – modelling of species extinction is subject to many kinds of serious unknowns but was predicted with two-digit precision46. | 238 |
| **Negative**: Transport policy in the UK using WebTAG, an assumption-rich model often reported with undue precision, with growth projections yielding predictions of how valuable the time in different transport sectors will be in 2052, to the penny69. | 253 |
| **Negative:** International rating and ranking 170 including that of the output and quality of research109,170,239, and of higher education170. For example higher education ratings create a perverse system of incentives and prices for students and their families240. The ranking can be shown to be arbitrary and volatile241.  | 260,261 |
| 1. Mind the unknown
 | **Positive**:Success of weather forecasting262,263. | National Weather Forecast |
| **Negative**: A quantitative evaluation of the Institute for Health Metrics and Evaluation (IHME) coronavirus model (also referenced in a White House press conference) finds observations in the real-world fall into the 5% probability zone more than 70% of the time197. | 254 |
| **Negative:** Modelling of species extinction is subject to many kinds of serious unknowns but uncertainty was largely underestimated46. | 238 |
| **Positive**: Blog postings on COVID-19 epidemiological models by Tomas Pueyo, who uses plain-language explanations of the implications of model uncertainties for policy options. | 232 |
| **Positive**: Sperrin, Grant, and Peek (2020) note that machine learning models that compose prediction models (e.g., for likely patient outcomes) have all been found to be biased and suggest: “As no covid-19 clinical prediction models can currently be recommended, clinicians will have to rely on their clinical acumen and shared experiences of best practices for now”218. | 218 |
| **Positive**: “There is no number-answer to your question”, by Anthony Fauci. | 184 |
| **Positive:** Clever modelling by NASA: MESSENGER mission to Mercury.  | 264 |

## References

1. Wasserstein, R. L. & Lazar, N. A. The ASA’s Statement on p-Values: Context, Process, and Purpose. *Am. Stat.* **70**, 129–133 (2016).

2. Amrhein, V., Greenland, S. & McShane, B. Scientists rise up against statistical significance. *Nature* **567**, 305–307 (2019).

3. Mayo, D. G. *Statistical Inference as Severe Testing. How to Get Beyond the Statistics Wars*. (Cambridge University Press, 2018). doi:10.1017/9781107286184

4. Mayo, D. P-Values on Trial: Selective Reporting of (Best Practice Guides Against) Selective Reporting. *Harvard Data Sci. Rev.* **2**, (2020).

5. Saltelli, A. & Stark, P. B. Fixing statistics is more than a technical issue. *Nature* **553**, 281–281 (2018).

6. Stark, P. B. & Saltelli, A. Cargo-cult statistics and scientific crisis. *Significance* **15**, 40–43 (2018).

7. Stark, P. B. Before reproducibility must come preproducibility. *Nature* **557**, 613–613 (2018).

8. Wasserstein, R. L. & Lazar, N. A. The ASA’s Statement on *p* -Values: Context, Process, and Purpose. *Am. Stat.* **70**, 129–133 (2016).

9. Wasserstein, R. L., Schirm, A. L. & Lazar, N. A. Moving to a World Beyond “p < 0.05”. *Am. Stat.* **73**, (2019).

10. Kafadar, K. The Year in Review … And More to Come. *Amstat News* (2019).

11. Silverman, E. Where’s the data? Now is no time to sit on Covid-19 trial results. *Statnews* (2020).

12. Gross, J. & Padilla, M. Coronavirus Glossary: Flattening the Curve, Pandemic, Covid-19 and More - The New York Times. *New York Times* (2020).

13. Montgomery, C. & Engelmann, L. Epidemiological Publics? On the Domestication of Modelling in the era of COVID-19. *http://somatosphere.net/* (2020). Available at: http://somatosphere.net/2020/epidemiological-publics-on-the-domestication-of-modelling-in-the-era-of-covid-19.html/. (Accessed: 15th May 2020)

14. Rhodes, T., Lancaster, K. & Rosengarten, M. A model society: maths, models and expertise in viral outbreaks. *Crit. Public Health* **30**, 253–256 (2020).

15. Pielke, R. J. The Mudfight Over ‘Wild-Ass’ Covid Numbers Is Pathological. *Wired* (2020).

16. Sarewitz, D. How science makes environmental controversies worse. *Environ. Sci. Policy* **7**, 385–403 (2004).

17. Saltelli, A. Why science’s crisis should not become a political battling ground. *Futures* **104**, 85–90 (2018).

18. Oreskes, N. Beware: transparency rule is a Trojan Horse. *Nature* **557**, 469–469 (2018).

19. Didier, E. *America by the numbers : quantification, democracy, and the birth of national statistics*. (MIT press, 2020).

20. Pinson, P. & Makridakis, S. COVID-19: Ioannidis vs. Taleb. *International Institute of Forecasters* (2020). Available at: https://forecasters.org/blog/2020/06/14/covid-19-ioannidis-vs-taleb/. (Accessed: 17th June 2020)

21. Caduff, C. What Went Wrong: Corona and the World after the Full Stop. *Med. Anthropol. Q.* **accepted**, (2020).

22. Ferguson, N. M. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College London* (2020). Available at: https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf.

23. Waltner-Toews, D. *et al.* Post-normal pandemics: Why COVID-19 requires a new approach to science. *STEPS Centre Blog* (2020). Available at: https://steps-centre.org/blog/postnormal-pandemics-why-covid-19-requires-a-new-approach-to-science/.

24. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report –46. *WHO* (2020). Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf\_4. (Accessed: 6th May 2020)

25. Baud, D. *et al.* Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* **0**, (2020).

26. Heneghan, C., Brassey, J. & Jefferson, T. COVID-19: What proportion are asymptomatic? - CEBM. *Centre for Evidence-Based Medicine* (2020). Available at: https://www.cebm.net/covid-19/covid-19-what-proportion-are-asymptomatic/. (Accessed: 29th April 2020)

27. Lai, C. C. *et al.* Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. *J. Microbiol. Immunol. Infect.* (2020). doi:10.1016/j.jmii.2020.02.012

28. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *European Communicable Disease Bullettin* **25**, (2020).

29. Smithson, M. How can we know unknown unknowns? *Integration and Implementation Insights* (2019). Available at: https://i2insights.org/2019/09/10/how-can-we-know-unknown-unknowns/. (Accessed: 6th June 2020)

30. Joseph, A. Lancet, NEJM retract Covid-19 studies that sparked backlash. *Statnews* (2020).

31. Sarewitz, D., Pielke, R. A. & Byerly, R. *Prediction: Science, Decision Making, and the Future of Nature*. (Island Press, 2000).

32. Saltelli, A. Statistical versus mathematical modelling: a short comment. *Nat. Commun.* **10**, 1–3 (2019).

33. Jakeman, A. J., Letcher, R. A. & Norton, J. P. Ten iterative steps in development and evaluation of environmental models,. *Environ. Model. Softw.* **21**, 602–614 (2006).

34. Padilla, J. J., Diallo, S. Y., Lynch, C. J. & Gore, R. Observations on the practice and profession of modeling and simulation: A survey approach. *Simulation* **94**, 493–506 (2018).

35. Saltelli, A. *et al.* Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environ. Model. Softw.* **114**, 29–39 (2019).

36. Morrison, F. *The art of modeling dynamic systems : forecasting for chaos, randomness, and determinism*. (Dover Publications, 1991).

37. Santner, T. J., Williams, B. J. & Notz, W. I. *The Design and Analysis of Computer Experiments*. (Springer-Verlag, 2003).

38. Pilkey, O. H. & Pilkey-Jarvis, L. *Useless Arithmetic: Why Environmental Scientists Can’t Predict the Future*. (Columbia University Press, 2009).

39. Christie, M., Cliffe, A., Dawid, P. & Senn, S. S. *Simplicity, Complexity and Modelling*. (Wiley, 2011).

40. European Commission. *European Commission IMPACT ASSESSMENT GUIDELINES*. (2009).

41. Office of Management and Budget. *Proposed Risk Assessment Bulletin*. (2006).

42. U.S. Environmental Protection Agency (EPA). *Guidance on the Development, Evaluation, and Application of Environmental Models*. (2009).

43. Science Advice for Policy by European Academies. *Making sense of science for policy under conditions of complexity and uncertainty*. (2019).

44. Rosen, R. *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*. (Columbia University Press, 1991).

45. Louie, A. H. Robert Rosen’s anticipatory systems. *Foresight* **12**, 18–29 (2010).

46. van der Sluijs, J. P. Numbers running wild. in *Science on the Verge* (eds. Benessia, A. et al.) 151–188 (Arizona State University, 2016).

47. Refsgaard, J. C., van der Sluijs, J. P., Brown, J. & van der Keur, P. A framework for dealing with uncertainty due to model structure error. *Adv. Water Resour.* **29**, 1586–1597 (2006).

48. Hamilton, S. H. *et al.* A framework for characterising and evaluating the effectiveness of environmental modelling. *Environ. Model. Softw.* **118**, 83–98 (2019).

49. Little, J. C. *et al.* A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. *Environ. Model. Softw.* **112**, 82–94 (2019).

50. Badham, J. *et al.* Effective modeling for Integrated Water Resource Management: A guide to contextual practices by phases and steps and future opportunities. *Environ. Model. Softw.* **116**, 40–56 (2019).

51. Elsawah, S. *et al.* Eight grand challenges in socio-environmental systems modeling. *Socio-Environmental Syst. Model.* **2**, 16226 (2020).

52. Eker, S., Rovenskaya, E., Obersteiner, M. & Langan, S. Practice and perspectives in the validation of resource management models. *Nat. Commun.* **9**, 5359 (2018).

53. Den Boon, S. *et al.* Guidelines for multi-model comparisons of the impact of infectious disease interventions. *BMC Med.* **17**, 163 (2019).

54. Behrend, M. R. *et al.* Modelling for policy: The five principles of the Neglected Tropical Diseases Modelling Consortium. *PLoS Negl. Trop. Dis.* **14**, e0008033 (2020).

55. Morin, A. *et al.* Shining light into black boxes. *Science (80-. ).* **336**, 159–160 (2012).

56. Eddy, D. M. *et al.* Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices task force-7. *Med. Decis. Mak.* **32**, 733–743 (2012).

57. Benessia, A. *et al.* *Science on the Verge*. (Arizona State University, 2016).

58. Saltelli, A. & Funtowicz, S. What is science’s crisis really about? *Futures* **91**, 5–11 (2017).

59. Sarewitz, D. Saving Science. *New Atl.* **49**, 4–40 (2016).

60. Saltelli, A. & Boulanger, P.-M. Technoscience, policy and the new media. Nexus or vortex? *Futures* 102491 (2019). doi:10.1016/J.FUTURES.2019.102491

61. Bammer, G. *et al.* Expertise in research integration and implementation for tackling complex problems: when is it needed, where can it be found and how can it be strengthened? *Palgrave Commun.* **6**, 1–16 (2020).

62. Harris, R. F. *Rigor mortis : how sloppy science creates worthless cures, crushes hope, and wastes billions*. (Basic Books, 2017).

63. Smaldino, P. E. & McElreath, R. The natural selection of bad science. *R. Soc. Open Sci.* **3**, (2016).

64. Kerr, N. L. HARKing: Hypothesizing After the Results are Known. *Personal. Soc. Psychol. Rev.* **2**, 196–217 (1998).

65. Rhodes, T. & Lancaster, K. Mathematical models as public troubles in COVID-19 infection control: following the numbers. *Heal. Sociol. Rev.* 1–18 (2020). doi:10.1080/14461242.2020.1764376

66. Mulargia, F., Stark, P. B. & Geller, R. J. Why is Probabilistic Seismic Hazard Analysis (PSHA) still used? *Phys. Earth Planet. Inter.* **264**, 63–75 (2017).

67. Geller, R. J., Mulargia, F. & Stark, P. B. Why we need a new paradigm for earthquake occurrence. in *Subduction Dynamics: From Mantle Flow to Mega Disasters, Geophysical Monograph 211* (eds. Morra, G., Yuen, D. A., King, S. D., Lee, S.-M. & Stein, S.) (Wiley, 2016).

68. Stark, P. B. & Freedman, D. A. What is the Chance of an Earthquake? in *Earthquake Science and Seismic Risk Reduction, NATO Science Series IV: Earth and Environmental Sciences* (eds. Mulargia, F. & Geller, R. J.) (Kluwer, 2003).

69. Kay, J. A. & King, M. A. *Radical uncertainty : decision-making beyond the numbers*. (W. W. Norton & Company, 2020).

70. Saltelli, A. *et al.* *Global sensitivity analysis : the primer*. (John Wiley, 2008).

71. Puy, A., Piano, S. Lo & Saltelli, A. Current models underestimate future irrigated areas. *Geophys. Res. Lett.* (2020). doi:10.1029/2020GL087360

72. Puy, A., Lo Piano, S. & Saltelli, A. A sensitivity analysis of the PAWN sensitivity index. *Environ. Model. Softw.* **127**, (2020).

73. Saisana, M., Saltelli, A. & Tarantola, S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J. R. Stat. Soc. Ser. A (Statistics Soc.* **168**, 307–323 (2005).

74. Edmonds, B. *et al.* Different Modelling Purposes. *J. Artif. Soc. Soc. Simul.* **22**, 6 (2019).

75. Saltelli, A. & d’Hombres, B. Sensitivity analysis didn’t help. A practitioner’s critique of the Stern review. *Glob. Environ. Chang.* **20**, 298–302 (2010).

76. Leamer, E. E. Sensitivity Analyses Would Help. *Am. Econ. Rev.* **75**, 308–313 (1985).

77. Funtowicz, S. & Ravetz, J. R. *Uncertainty and Quality in Science for Policy*. (Kluwer, 1990). doi:10.1007/978-94-009-0621-1\_3

78. Cabore, J. W. *et al.* The potential effects of widespread community transmission of SARS-CoV-2 infection in the WHO African Region: a predictive model. *Br. Med. J. Glob. Heal.* **accepted**, (2020).

79. Arino, J. & Portet, S. A simple model for COVID-19. *Infect. Dis. Model.* **5**, 309–315 (2020).

80. Borgonovo, E. & Lu, X. Is Time to Intervention in the COVID-19 Outbreak Really Important? A Global Sensitivity Analysis Approach. *arXiv.org* **2005.01833**, (2020).

81. Ndaïrou, F., Area, I., Nieto, J. J. & Torres, D. F. M. Mathematical Modeling of COVID-19 Transmission Dynamics with a Case Study of Wuhan. *Chaos, Solitons & Fractals* 109846 (2020). doi:10.1016/j.chaos.2020.109846

82. Giordano, G. *et al.* Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* 1–6 (2020). doi:10.1038/s41591-020-0883-7

83. Jasanoff, S. Technologies of humility: Citizen participation in governing science. *Minerva* **41**, 223–244 (2003).

84. Jasanoff, S. Technologies of humility. *Nature* **450**, 33 (2007).

85. Sridhar, D. & Majumder, M. S. Modelling the pandemic. *BMJ* **369**, m1567 (2020).

86. Turner, M. G. & Gardner, R. H. Introduction to Models. in *Landscape Ecology in Theory and Practice* 63–95 (Springer New York, 2015). doi:10.1007/978-1-4939-2794-4\_3

87. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference, and prediction*. (Springer, 2009). doi:10.1007/978-0-387-84858-7\_2

88. Zadeh, L. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Trans. Syst. Man. Cybern.* **3**, 28–44 (1973).

89. Hornberger, G. M. & Spear, R. C. An approach to the preliminary analysis of environmental systems. *J. Environ. Manage.* **12**, (1981).

90. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An introduction to statistical learning : with applications in R*. (Springer, 2017).

91. Hoerl, A. E. & Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **42**, 80 (2000).

92. Merrick, J. H. & Weyant, J. P. On choosing the resolution of normative models. *Eur. J. Oper. Res.* **279**, 511–523 (2019).

93. Borges, J. *El Jardin de los Senderos que se Bifurcan*. (SUR, 1941).

94. Gelman, A. & Loken, E. *The garden of forking paths*. *Working Paper Department of Statistics, Columbia University* (2013).

95. Pfleiderer, P. Chameleons: The Misuse of Theoretical Models in Finance and Economics. *Economica* **87**, 81–107 (2020).

96. Friedman, M. *Essays in Positive Economics*. (University of Chicago Press, 1953).

97. Reinert, E. S. & Daastøl, A. M. Exploring the genesis of economic innovations: The religious gestalt-switch and the duty to invent as preconditions for economic growth. *Eur. J. Law Econ.* **4**, 233–283 (1997).

98. May, R. M. Uses and Abuses of Mathematics in Biology. *Science (80-. ).* **303**, 790–793 (2004).

99. Kupferschmidt, K. Why do some COVID-19 patients infect many others, whereas most don’t spread the virus at all? *Science (80-. ).* (2020). doi:10.1126/science.abc8931

100. Miller, D. *et al.* Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *medRxiv* (2020).

101. Metlay, D. From Tin Roof to Torn Wet Blanket: Predicting and Observing Ground Water Movement at a Proposed Nuclear Waste Site. in *Prediction: Science, Decision Making, and the Future of Nature* (eds. Sarewitz, D., Pielke, R. A. J. & Byerly, R.) (Island Press, 2000).

102. Fjelland, R. When Laypeople are Right and Experts are Wrong: Lessons from Love Canal. *Int. J. Philos. Chem.* **22**, 105–125 (2016).

103. Gibbs, L. M. & Levine, M. *Love Canal. My story*. (State University of New York Press, Albany, 1982).

104. Brown, P. Popular Epidemiology Revisited. *Curr. Sociol.* **45**, 137–156 (1997).

105. Cizdziel, J. & Smiecinski, A. *Bomb-pulse Chlorine-36 at the proposed Yucca Mountain repository horizon: An investigation of previous conflicting results and collection of new data*. *Nevada System of Higher Education* (2006).

106. Stothoff, S. A. & Walter, G. R. Average infiltration at Yucca Mountain over the next million years. *Water Resour. Res.* **49**, 7528–7545 (2013).

107. Sarewitz, D. Of Cold Mice and Isotopes Or Should We Do Less Science. *Summer School in Higher Education Research and Science Studies Universität Bonn* (2018). Available at: https://sfis.asu.edu/sites/default/files/should\_we\_do\_less\_science-revised\_distrib.pdf.

108. Dressel, J. & Farid, H. The accuracy, fairness, and limits of predicting recidivism. *Sci. Adv.* **4**, eaao5580 (2018).

109. O’Neil, C. *Weapons of math destruction : how big data increases inequality and threatens democracy*. (Random House Publishing Group, 2016).

110. Salganik, M. J. *et al.* Measuring the predictability of life outcomes with a scientific mass collaboration. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8398–8403 (2020).

111. Makridakis, S., Spiliotis, E. & Assimakopoulos, V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS One* **13**, e0194889 (2018).

112. Nissenbaum, H. Accountability in a computerized society. *Sci. Eng. Ethics* **2**, 25–42 (1996).

113. Beck, P. U. *Risk Society: Towards a New Modernity*. (Sage Publications (CA), 1992).

114. Saltelli, A. *et al.* The technique is never neutral. How methodological choices condition the generation of narratives for sustainability. *Environ. Sci. Policy* **106**, 87–98 (2020).

115. Young, R. S., Pilkey, O. H., Bush, D. M. & Thieler, E. R. A Discussion of the Generalized Model for Simulating Shoreline Change (GENESIS). *J. Coast. Res.* **11**, 875–886 (1995).

116. Pilkey, O. H. What you know can hurt you: predicting the behavior of nourished beaches. in *Prediction: Science, Decision Making, and the Future of Nature* (eds. Sarewitz, D., Pielke, R. A. J. & Byerly, R.) 159–184 (2000).

117. Thunstrom, L., Newbold, S., Finnoff, D., Ashworth, M. & Shogren, J. F. The Benefits and Costs of Flattening the Curve for COVID-19. *SSRN Electron. J. (to Appear J. Cost Benefit Anal.* (2020). doi:10.2139/ssrn.3561934

118. Stirling, A. How politics closes down uncertainty - STEPS Centre. *STEPS Centre* Feb. 20 (2019). Available at: https://steps-centre.org/blog/how-politics-closes-down-uncertainty/. (Accessed: 31st March 2019)

119. Viscusi, W. K. The Flawed Hedonic Damages Measure of Compensation for Wrongful Death and Personal Injury. *J. Forensic Econ.* **20**, 113–135 (2008).

120. Linshi, J. Germanwings Plane Crash: How Much Compensation for Victims’ Families? | Time. *Time* (2015).

121. Porter, T. M. Objectivity as Standardization: The Rhetoric of Impersonality in Measurement, Statistics, and Cost-Benefit Analysis. in *Rethinking objectivity* (ed. Megill, A.) 197–237 (Duke University Press, 1994).

122. Funtowicz, S. & Ravetz, J. R. The worth of a songbird: ecological economics as a post-normal science. *Ecol. Econ.* **10**, 197–207 (1994).

123. Ravetz, J. R. Models as metaphors. in *Public participation in sustainability science : a handbook* (ed. B. Kasemir, J. Jäger, C. Jaeger, Gardner Matthew T., Clark William C., and W. A.) (Cambridge University Press, 2003).

124. Stirling, A. Keep it complex. *Nature* **468**, 1029–1031 (2010).

125. Barton, C. M. *et al.* Call for transparency of COVID-19 models. *Science (80-. ).* **368**, 482.2-483 (2020).

126. van der Sluijs, J. P. *et al.* Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Anal.* **25**, 481–492 (2005).

127. Lane, S. N. *et al.* Doing flood risk science differently: an experiment in radical scientific method. *Trans. Inst. Br. Geogr.* **36**, 15–36 (2011).

128. Miedema, F. To confront 21st century challenges, science must rethink its reward system | Science | The Guardian. *The Guardian* (2016).

129. Saltelli, A., Guimaraes Pereira, Â., van der Sluijs, J. P. & Funtowicz, S. What do I make of your latinorumc Sensitivity auditing of mathematical modelling. *Int. J. Foresight Innov. Policy* **9**, 213–234 (2013).

130. Saltelli, A. & Funtowicz, S. When All Models Are Wrong. *Issues Sci. Technol.* **30**, 79–85 (2014).

131. Saltelli, A. & Lo Piano, S. Problematic quantifications: a critical appraisal of scenario making for a global ‘sustainable’ food production. *Food Ethics* **1**, 173–179 (2017).

132. Lo Piano, S. & Robinson, M. Nutrition and public health economic evaluations under the lenses of post normal science. *Futures* **112**, 102436 (2019).

133. Funtowicz, S. & Ravetz, J. R. Science for the post-normal age. *Futures* **25**, 739–755 (1993).

134. van der Sluijs, J. P. A way out of the credibility crisis around model-use in Integrated Environmental Assessment. *Futures* **34**, (2002).

135. Ravetz, J. R. Pandemic Recovery Requires Post-Normal Science. *Issues Sci. Technol.* **June 19**, (2020).

136. Editorial. Value judgements. *Nature* **473**, 123–124 (2011).

137. Gluckman, P. Policy: The art of science advice to government. *Nature* **507**, 163–165 (2014).

138. Grinnell, F. Rethink our approach to assessing risk. *Nature* **522**, 257 (2015).

139. Dietl, G. P. Ecology: Different worlds. *Nature* **529**, 29–30 (2016).

140. Editorial. Future present. *Nature* **531**, 7–8 (2016).

141. Ravetz, J. Stop the science training that demands ‘don’t ask’. *Nature* **575**, 417 (2019).

142. Röckmann, C. *et al.* The added value of participatory modelling in fisheries management - what has been learnt? *Mar. Policy* **36**, 1072–1085 (2012).

143. Fulton, E. A., Smith, A. D. M., Smith, D. C. & Johnson, P. An Integrated Approach Is Needed for Ecosystem Based Fisheries Management: Insights from Ecosystem-Level Management Strategy Evaluation. *PLoS One* **9**, e84242 (2014).

144. Pielke Jr., R. The role of models in prediction for decision. in *Models in ecosystem science* (eds. Canham, C. D., Cole, J. J. & Lauenroth, W. K.) 111–135 (Princeton University Press, 2003).

145. Taleb, N. N. & Bar-Yam, Y. The UK’s coronavirus policy may sound scientific. It isn’t. *The Guardian* (2020).

146. Didier, E. Do Statistics ‘Perform’ the Economy ? in *Do Economists Make Markets? On the Performativity of Economics* (eds. MacKenzie, D., Muniesa, F. & Liucia Sue) (Princeton University Press, 2007).

147. Egmond, S. van & Zeiss, R. Modeling for Policy Science-based models as performative boundary objects for Dutch policy making. *Sci. Technol. Stud.* **23**, 58–78 (2010).

148. Ravetz, J. R. Faith and Reason in the Mathematics of the Credit Crunch. *Oxford Mag. Eight Week, Michaelmas term 14–16* (2008).

149. Porter, T. M. Funny Numbers. *Cult. Unbound* **4**, 585–598 (2012).

150. MacKenzie, D. Is economics performative? Option theory and the construction of derivatives markets. *J. Hist. Econ. Thought* **28**, 29–55 (2006).

151. Supiot, A. *Governance by Numbers: The Making of a Legal Model of Allegiance*. (Oxford University Press, 2007).

152. Wilmott, P. & Orrell, D. *The Money Formula*. (Wiley & Sons, 2017).

153. Li, D. X. On Default Correlation. *J. Fixed Income* **9**, 43–54 (2000).

154. Salmon, F. Recipe for Disaster: The Formula That Killed Wall Street. *Wired*  (2009).

155. Bendavid, E. *et al.* COVID-19 Antibody Seroprevalence in Santa Clara County, California. *medRxiv* 2020.04.14.20062463 (2020). doi:10.1101/2020.04.14.20062463

156. Didier, E. Politique du nombre de morts. *AOC, Analyse Opinion Critique* (2020).

157. Foucault, M. *et al.* Coronavirus and philosophers. *European Journal of Psychoanalysis* (2020). Available at: https://www.journal-psychoanalysis.eu/coronavirus-and-philosophers/. (Accessed: 15th April 2020)

158. Welsh, M. You are biased! *Integration and Implementation Insights* (2018). Available at: https://i2insights.org/2018/09/18/confirmation-bias/. (Accessed: 6th June 2020)

159. Welsh, M. *Bias in Science and CommunicationA field guide*. *Bias in Science and CommunicationA field guide* (IOP Publishing, 2018). doi:10.1088/978-0-7503-1311-7

160. Ravetz, J. R. *Scientific knowledge and its social problems*. (Oxford University Press, 1971).

161. WHO, R. O. for A. New WHO estimates: Up to 190 000 people could die of COVID-19 in Africa if not controlled. *Newsroom* (2020). Available at: https://www.afro.who.int/news/new-who-estimates-190-000-people-could-die-covid-19-africa-if-not-controlled. (Accessed: 24th May 2020)

162. Landler, M. & Castle, S. Behind the Virus Report That Jarred the U.S. and the U.K. to Action - The New York Times. *The New York Times* (2020).

163. Porter, T. M. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. (Princeton University Press, 1995).

164. Ravetz, J. R. In Numbers We Trust | Issues in Science and Technology. *Issues Sci. Technol.* **13**, (1997).

165. Espeland, W. N. & Stevens, M. L. A sociology of quantification. *Eur. J. Sociol.* **49**, 401–436 (2008).

166. Popp Berman, E. & Hirschman, D. The Sociology of Quantification: Where Are We Now? *Contemp. Sociol.* **47**, 257–266 (2018).

167. Hacking, I. *The taming of chance*. (Cambridge University Press, 1990).

168. Norton Wise, M. *The values of precision*. (Princeton University Press, 1995).

169. Desrosières, A. *The politics of large numbers : a history of statistical reasoning*. (Harvard University Press, 1998).

170. Muller, J. Z. *The tyranny of metrics*. (Princeton University Press , 2018).

171. Saltelli, A. Ethics of quantification or quantification of ethics? *Futures* **116**, (2020).

172. Centre for the Study of the Sciences and the Humanities, U. of B. Ethics of quantification. *University of Bergen*  (2019). Available at: https://www.uib.no/en/svt/127044/ethics-quantification. (Accessed: 12th August 2019)

173. Sareen, S., Rommetveit, K. & Saltelli, A. Ethics of Quantification: Illumination, Obfuscation and Performative Legitimation. *Palgrave Commun.* **Accepted**, (2020).

174. Saltelli, A. & Di Fiore, M. From sociology of quantification to ethics of quantification. *Palgrave Commun.* **revised**, (2020).

175. Ravetz, J. R. Usable Knowledge, Usable Ignorance. *Knowledge* **9**, 87–116 (1987).

176. Ravetz, J. R. Preface: Descartes and the rediscovery of ignorance. in *The end of the Cartesian dream* (eds. Guimarães Pereira, Â. & Funtowicz, S.) XV–XVIII (Routledge, 2015).

177. Weinberg, A. Science and trans-science. *Minerva* **10**, 209–222 (1972).

178. Taleb, N. N. *Antifragile: Things That Gain from Disorder*. (Random House Publishing Group, 2012).

179. Oreskes, N., Shrader-Frechette, K. & Belitz, K. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science (80-. ).* **263**, (1994).

180. Oreskes, N. Why Predict? Historical Perspectives on Prediction in Earth Science. in *Prediction: Science, Decision Making, and the Future of Nature* 23–40 (2000).

181. Holmdahl, I. & Buckee, C. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *N. Engl. J. Med.* NEJMp2016822 (2020). doi:10.1056/NEJMp2016822

182. Pierce, A. The Queen asks why no one saw the credit crunch coming. *The Telegraph* (2008).

183. Mirowski, P. *Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Meltdown*. (Verso, 2013).

184. Fauci, A. There is no number-answer. *Twitter* (2020). Available at: https://twitter.com/marioricciard18/status/1237778247011663872.

185. Steinmann, P., Wang, J. R., Voorn, G. A. K. van & Kwakkel, J. H. Don’t try to predict COVID-19. If you must, use Deep Uncertainty methods. *Rev. Artif. Soc. Soc. Simul.* (2020).

186. Bankes, S. Exploratory Modeling for Policy Analysis. *Oper. Res.* **41**, 435–449 (1993).

187. Lempert, R. J., Groves, D. G., Popper, S. W. & Bankes, S. C. A general, analytic method for generating robust strategies and narrative scenarios. *Manage. Sci.* **52**, 514–528 (2006).

188. Reinert, E. S. Full circle: economics from scholasticism through innovation and back into mathematical scholasticism. *J. Econ. Stud.* **27**, 364–376 (2000).

189. Reinert, E. S. *The Terrible Simplifers: Common Origins of Financial Crises and PersistentPoverty in Economic Theory and the new ‘1848 Moment’, DESA working paper Paper No. 8, 8ST/ESA/2009/DWP/88*. (2009).

190. Reinert, E. S. The Inequalities that could not happen: What the Cold War did to Economics. in *The Inequality Crisis, Bristol, World Economics Association.* (eds. Fullbrook, E. & Morgan, J.) Forthcoming (World Economics Association, 2020).

191. Romer, P. Mathiness in the Theory of Economic Growth. *Am. Econ. Rev.* **105**, 89–93 (2015).

192. Stiglitz, J. E. Rethinking Macroeconomics: What Failed, and How to Repair It. *J. Eur. Econ. Assoc.* **9**, 591–645 (2011).

193. Thompson, E. L. . & Smith, L. A. Escape from model-land. *Econ. Open access, Open Assess. J.* **23**, (2019).

194. Gupta, S. Avoiding ambiguity: Scientist sometimes use mathematics to give the illusion of certainty. *Nature* **412**, 589 (2001).

195. Gigerenzer, G. Statistical Rituals: The Replication Delusion and How We Got There. *Adv. Methods Pract. Psychol. Sci.* **1**, 198–218 (2018).

196. Szenberg, M. *Eminent economists : their life philosophies*. (Cambridge University Press, 1992).

197. Marchant, R., Samia, N. I., Rosen, O., Tanner, M. A. & Cripps, S. Learning as We Go: An Examination of the Statistical Accuracy of COVID19 Daily Death Count Predictions. *medRxiv* 2020.04.11.20062257 (2020). doi:10.1101/2020.04.11.20062257

198. IHME COVID-19 health service utilization forecasting team. *Forecasting the impact of the first wave of the COVID-19 1 pandemic on hospital demand and deaths for the USA and 2 European Economic Area countries, on MedRxiv, https://doi.org/10.1101/2020.04.21.20074732*. (2020). doi:https://doi.org/10.1101/2020.04.21.20074732

199. Jewell, N. P., Lewnard, J. A. & Jewell, B. L. Predictive Mathematical Models of the COVID-19 Pandemic. *JAMA* (2020). doi:10.1001/jama.2020.6585

200. Kramer, D. COVID-19 pandemic modeling is fraught with uncertainties. *Phys. Today* **73**, 25–27 (2020).

201. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452 (2006).

202. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic - SUPPLEMENTARY INFORMATION. *Nature* (2006).

203. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science (80-. ).* **292**, 1155–1160 (2001).

204. Kitching, R. P., Thrusfield, M. V. & Taylor, N. M. Use and abuse of mathematical models: an illustration from the 2001 foot and mouthdisease epidemic in the United Kingdom. *Sci. Tech. Rev. Off. Int. des Epizoot.* **25**, (2006).

205. Mansley, L. M., Donaldson, A. I., Thrusfield, M. V. & Honhold, N. Destructive tension: Mathematics versus experience - The progress and control of the 2001 foot and mouth disease epidemic in Great Britain. *OIE Rev. Sci. Tech.* **30**, 483–498 (2011).

206. Ferguson, N. M., Ghani, A. C., Donnelly, C. A., Hagenaars, T. J. & Anderson, R. M. Estimating the human health riskfrom possible BSE infection ofthe British sheep Flock. *Nature* **415**, 420–424 (2002).

207. Sturcke, J. Bird flu pandemic ‘could kill 150m’. *The Guardian* (2005).

208. Imperial College London. Swine flu: early findings about pandemic potential reported in new study. *Imperial News* (2009). Available at: https://www.imperial.ac.uk/news/66374/swine-early-findings-about-pandemic-potential/. (Accessed: 13th May 2020)

209. Fraser, C. *et al.* Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science (80-. ).* **324**, 1557–1561 (2009).

210. Steerpike. Six questions that Neil Ferguson should be asked. *The Spectator* (2020).

211. Fund, J. Neil Ferguson’s Coronavirus Imperial College Model -- ‘Professor Lockdown’ Resigns in Disgrace. *National Review* (2020).

212. Charleston, B. *et al.* Relationship between clinical signs and transmission of an infectious disease and the implications for control. *Science (80-. ).* **332**, 726–729 (2011).

213. Ghosh, P. Mass culling for foot-and-mouth ‘may be unnecessary’. *BBC News* (2011).

214. Chis Ster, I. & Ferguson, N. M. Transmission Parameters of the 2001 Foot and Mouth Epidemic in Great Britain. *PLoS One* **2**, e502 (2007).

215. Ferguson, N. M. Microsoft and GitHub working with the Imperial College. *Twitter* (2020). Available at: https://twitter.com/neil\_ferguson/status/1241835456947519492.

216. Singh Chawla, D. Critiqued coronavirus simulation gets thumbs up from code-checking efforts. *Nature* (2020). doi:10.1038/d41586-020-01685-y

217. Eglen, S. J. CODECHECK Certificate 2020-010. *Zenodo* (2020). Available at: https://doi.org/10.5281/zenodo.386 5491.

218. Sperrin, M., Grant, S. W. & Peek, N. Prediction models for diagnosis and prognosis in Covid-19. *BMJ* **369**, (2020).

219. Panovska-Griffiths, J. Can mathematical modelling solve the current Covid-19 crisis? *BMC Public Health* **20**, 551 (2020).

220. Lourenco, J. *et al.* Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *medRxiv* 2020.03.24.20042291 (2020). doi:10.1101/2020.03.24.20042291

221. Pielke, Jr, R. A. *The Honest Broker*. (Cambridge University Press, 2007).

222. Bruno, I., Didier, E. & Prévieux, J. *Statactivisme. Comment lutter avec des nombres*. (Zones, La Découverte, 2014).

223. Bruno, I., Didier, E. & Vitale, T. Special Issue. Statistics and Activism. *Particip. Conflict. Open J. Sociopolitical Stud.* **7**, (2014).

224. Concialdi, P. Le BIP40: alerte sur la pauvreté. in *Statactivisme. Comment lutter avec des nombres* (eds. Bruno, I., Didier, E. & Prévieux, J.) 199–211 (Zones, La Découverte, 2014).

225. Desrosières, A. Statistics and social critique. *Open J. Sociopolitical Stud.* **7**, 348–359 (2014).

226. Didier, E. Globalization of Quantitative Policing: Between Management and Statactivism. *Annu. Rev. Sociol.* **44**, 515–534 (2018).

227. Cardiff University. Data Justice Lab. *School of Journalism, Media and Culture* (2020). Available at: https://datajusticelab.org/.

228. Arnold, C. The mathematicians who want to save democracy. *Nature* **546**, 200–202 (2017).

229. Lindeman, M. & Stark, P. B. A Gentle Introduction to Risk-Limiting Audits. *IEEE Secur. Priv.* **10**, 42–49 (2012).

230. Rudner, R. The Scientist Qua Scientist Makes Value Judgments. *Philosophy of Science* **20**, 1–6 (1953).

231. Saltelli, A. What can we learn from the debate on statistical significance? *Blog: Error Statistics Philosophy* (2019). Available at: https://errorstatistics.com/2019/11/22/a-saltelli-guest-post-what-can-we-learn-from-the-debate-on-statistical-significance/. (Accessed: 8th June 2020)

232. Pueyo, T. Coronavirus: Why You Must Act Now. *Medium* (2020). Available at: https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca. (Accessed: 7th May 2020)

233. Agencia EFE. Go-to expert on Covid-19 fight hails Asia’s response, criticizes Europe, US. *Interview* (2020). Available at: https://www.efe.com/efe/english/world/go-to-expert-on-covid-19-fight-hails-asia-s-response-criticizes-europe-us/50000262-4220303. (Accessed: 5th June 2020)

234. Friedman, T. L. An Open Letter to President Trump. *The New York Times* (2020).

235. Leake, J. Neil Ferguson interview: No 10’s infection guru recruits game developers to build coronavirus pandemic model. *Sunday Times* (2020).

236. Lempert, R. J., Popper, S. W. & Bankes, S. C. Robust Decision Making: Coping with Uncertainty. *Futurist* **44**, 47 (2010).

237. Shapin, S. Invisible Science. *Hedgehog Rev.* **18**, (2016).

238. Urban, M. C. Accelerating extinction risk from climate change. *Science (80-. ).* **348**, 571–573 (2015).

239. Hicks, D., Wouters, P., Waltman, L., De Rijcke, S. & Rafols, I. Bibliometrics: The Leiden Manifesto for research metrics. *Nature* **520**, 429–431 (2015).

240. Espeland, W. N. & Sauder, M. *Engines of anxiety : academic rankings, reputation, and accountability*. (Russell Sage Foundation, 2016).

241. Saisana, M., d’Hombres, B. & Saltelli, A. Rickety numbers: Volatility of university rankings and policy implications. *Res. Policy* **40**, 165–177 (2011).

242. Clark, P. Credibility of EU energy review questioned. *Financial Times* (2011).

243. Gates, G., Ewing, J., Russell, K. & Watkins, D. How Volkswagen’s ‘Defeat Devices’ Worked. *The New York Times* (2017).

244. Reinhart, C. M. & Rogoff, K. S. *Growth In a Time of Debt, NBER Working Paper Series*. (2010).

245. Munchau W. Perils of placing faith in a thin theory. *Financ. Times* **April 21**, (2013).

246. Alexander, R. Reinhart, Rogoff... and Herndon: The student who caught out the profs - BBC News. *BBC News* (2013).

247. Stiglitz, J. E., Sen, A. & Fitoussi, J.-P. Report by the Commission on the Measurement of Economic Performance and Social Progress. *Sustain. Dev.* **12**, 292 (2009).

248. Giampietro, M. & Saltelli, A. Footprints to nowhere. *Ecol. Indic.* **46**, 610–621 (2014).

249. Galli, A. *et al.* Questioning the Ecological Footprint. *Ecol. Indic.* **69**, 224–232 (2016).

250. Silver, N. 2016 Election Forecast. *FiveThirtyEight* (2016). Available at: https://projects.fivethirtyeight.com/2016-election-forecast/. (Accessed: 10th May 2020)

251. Mutalik, P. The Devil in the Polling Data. *Quanta Magazine* (2016).

252. U.S. Army Corps of Engineers. *Shoreline Protection and Beach Erosion Control Study*. (1994).

253. HM Treasury. *The Green Book, Central Government Guidance on Appraisaland Evaluation*. (2018).

254. Murray, C. J. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries (IHME). *medRxiv* 2020.04.21.20074732 (2020). doi:10.1101/2020.04.21.20074732

255. US Nuclear Regulatory Commission. NRC: Package ML010300375 - Total System Performance Assessment (TSPA) Model for Site Recommendation. *Package description* (2001). Available at: https://www.nrc.gov/docs/ML0103/ML010300375.html. (Accessed: 10th May 2020)

256. E3MLab/ICCS at National Technical University of Athens. *PRIMES, Detailed model description*. (2013).

257. Magness, P. W. How Wrong Were the Models and Why? – AIER. *Am. Inst. Econ. Res.* (2020).

258. Global Footprint Network. Ecological Footprint. (2020). Available at: https://www.footprintnetwork.org/our-work/ecological-footprint/. (Accessed: 10th May 2020)

259. Global Footprint Network. Earth Overshoot Day. (2019). Available at: https://www.overshootday.org/. (Accessed: 13th August 2019)

260. ARWU, Academic Ranking of World Universities. *World University Rankings 2019 |* (2019). Available at: http://www.shanghairanking.com/. (Accessed: 4th June 2020)

261. Times Higher Education (THE). *Academic & University News |* (2020). Available at: https://www.timeshighereducation.com/. (Accessed: 4th June 2020)

262. Lazo, J. K., Morss, R. E. & Demuth, J. L. 300 billion served. *Bull. Am. Meteorol. Soc.* **90**, 785–798 (2009).

263. Pielke, R. J., Carbone, R. E., Jr., R. P. & Carbone, R. E. WEATHER IMPACTS, FORECASTS, AND POLICY. *Bull. Am. Meteorol. Soc.* **83**, 293–403 (2002).

264. Talbert, T. MESSENGER. *NASA* (2019). Available at: https://www.nasa.gov/mission\_pages/messenger/main/index.html. (Accessed: 7th June 2020)