Models and Experts: The Contribution of Expertise to Epidemic and Pandemic Modelling

Carlo Martini

Università Vita-Salute San Raffaele

Abstract

Modelling is a precious source of information in science. With models, we can simplify an otherwise messy reality in order to understand the fundamental driving forces of a system, like an epidemic, and we can try to predict the course of events in complex scenarios where there is a great degree of uncertainty. In short, models can be used to explain and predict phenomena. Yet models interact with expert opinions in two fundamental ways. They are sometimes in competition with expert opinion, and they are sometimes heavily dependent, for their proper working, on expert opinion.

In this paper I will illustrate the different ways in which a model interacts with expert opinion. I will focus on epidemiological models. I will explain how, in epidemic modelling, getting the expertise right is as important as getting the model right. I will briefly present epidemiological models with a focus on the specific contribution of expert judgment to the choice and use of these models. I will compare expert judgment with statistical judgment, highlighting the limits of the former. I will analyse the interconnectedness of modelling and expert judgment in epidemic simulations based on a case report and, finally, I will suggest some strategies for ameliorating the interaction between modelling and expert judgment.

Keywords: Epidemiological modeling, Expertise, Expert judgment, Statistical judgment.

1. Introduction

Modelling is a precious source of information in science. With models, we can simplify an otherwise messy reality in order to understand the fundamental driving forces of a system, like an epidemic. How different would the spread of a virus look like, if the driving mode of transmission was through airborne particles, or droplets, or fomites? With models, we can try to predict the course of events in complex scenarios where there is a great degree of uncertainty. Mill compared studying social phenomena to studying the laws of tides; we can only aim for approximation and inexactness when the course of a natural or social phenomenon is determined not

Argumenta 7, 1 (2021): 147-160
ISSN 2465-2334

DOI 10.14275/2465-2334/202113.mar

by a few generally well-known factors, but by a complex interaction of many causal factors:

> circumstances of a local or casual nature, such as the configuration of the bottom of the ocean, the degree of confinement from shores, the direction of the wind, &c., influence in many or in all places the height and time of the tide; and a portion of these circumstances being either not accurately knowable, not precisely measurable, or not capable of being certainly foreseen, the tide in known places commonly varies from the calculated result of general principles by some difference that we cannot explain, and in unknown ones may vary from it by a difference that we are not able to foresee or conjecture (Mill 1882: 587).

I will come back to this passage of Mill's work in due time, because it tells us something about the difficulties of modelling an epidemic. In short, models can be used to explain and predict phenomena. Allegedly, there are other purposes too, but for now I will stick to these two. Models work by isolation and idealization (Mäki 1992). For example, epidemiological models can give us at best a general idea of how a virus could spread if we make several simplifying assumptions. Most importantly, models often fail in performing their explanatory and predictive functions without expert judgment. The assumptions that go into a model, its parametrization, and its connection with a target system are all dependent in multiple ways on expert judgment.

In the rest of this paper, I will highlight the different phases in which expert judgment affects the development and application of a model. I will focus on epidemiological models.[1] While explanation is an important component of these types of models, epidemiological models are important because they are used to predict the spread of viruses in epidemic or pandemic situations given a range of pharmaceutical (e.g., antiviral drugs) and non-pharmaceutical (e.g., social distancing) policy measures that a society will usually put in place to respond to a pandemic or epidemic scenario. I will explain how, in epidemic modelling, getting the expertise right is as important as getting the model right. In the next section I will briefly present epidemiological models with a focus on the specific contribution of expert judgment to the choice and use of these models. In section 3 I will compare expert judgment with statistical judgment, highlighting the limits of the former. Section 4 will analyse the interconnectedness of modelling and expert judgment in epidemic simulations based on a report by the *National Academies of Sciences, Engineering, and Medicine* published (Institute of Medicine 2006). Section 5 will suggest some strategies for ameliorating the interaction between modelling and expert judgment and section 6 will conclude.

2. Expert Input in Epidemic Modelling: A Primer

Epidemic and pandemic models are usually systems of equations, often implemented in a computer programme. In Black's typology, they are mathematical

[1] In this paper I will use the term to specifically indicate mathematical modelling of infectious diseases—that is, that class of epidemic models that describe the spread of an infectious disease.

models (Black 1962), and they are theoretical models according to Achinstein's typology (1968):[2]

> The use of such a model characteristically involves the awareness and explicit acknowledgement that the real object is far more complex than its representation in the model: the theoretical model assumes away many complications while highlighting limited aspects of the object (Mäki 2001: 9932).

Indeed, epidemics and pandemics are complex phenomena: the spread of a virus depends on a very large number of factors, including biological aspects of the virus itself, of the range of hosts it can infect, and, as far as human health is concerned, it depends on human behaviour and available technology (e.g., availability of antiviral therapies and vaccines).

In this paper I focus on epidemiological models of two principal kinds: SIR models, and SIS models. The acronyms stand, respectively, for *Susceptible, Infectives, Removed*, and *Susceptible, Infectives, Susceptible*, the main difference being that in the latter model a virus can reinfect a host who has previously been infected. Depending on whether infection confers immunity in the hosts that survive the infection, the SIR and SIS models divide the total population in two or three macro subpopulations. At any given point in time a population will have three types of hosts: the susceptible are those who can be infected by the virus, the infective are those who are infected with the virus, and the removed are those who are either immune or dead, after infection.

Most viruses have a certain rate of reinfection, so model choice depends on knowledge about the virus and how it interacts with the host's immune system. Knowledge acquired in the early phases of a pandemic is very precious because it helps determine what kind of model we should be using, even before we think of the next steps, like parametrization and goodness of fit. For instance, in the early phases of the COVID-19 pandemic, sporadic cases of reinfection caused uncertainty about whether immunity from SARS-CoV-2 lasted more than a few weeks after recovery. As research progressed, scientists were able to determine relatively reliable rates of reinfection for different age groups, thus making model-choice easier.

Uncertainties about reinfection, and how it affects model choice, is just one example of how much modelling a pandemic needs reliable methods for collecting data. In a perfect world, we would be able to collect the information we need for model choice and parameterization in two ways: either (a) with instruments, for example, in the same way in which we measure the temperature, or pressure, in a patient, or (b) by widespread scientific consensus; for example, when we need to know the temperature of the sun we can rely on a reasonable level of scientific agreement and error rates. The world of pandemics is not perfect, and, instead, the input source for much of the knowledge that is needed is expert judgment, that is, the informed guesswork of specialists in certain areas of science. Accord-

[2] I have used mostly the term "epidemic modelling" in this paper, even though I may sometimes use the term "pandemic modelling" instead. The main difference is that epidemiological models usually contain primarily epidemiological elements, while a pandemic model may include social, economic, and other variables. The usage is not always consistent in the literature and for the scope of this paper any differences between the two can be ignored.

ing to Cooke (1991: 30) "musings, brainstorms, guesses, and speculations of experts can be significant input in a structured decision process". In epidemiological modelling, musing, guesses, etc. are often the main source of modelling input we have.

There are two possible interactions between modelling and expert judgment: Expert judgment can be a constitutive part of modelling—for example when using expert judgment as input for the parameters of the model—or expert judgment can be an alternative to model-based judgment—for example when the model is inadequate to represent the problem at hand. The following sections will deal with both of these scenarios in which there is interaction, and sometimes a conflict, between a model and an expert judgment.

Next, I list the main contribution of expert judgment to pandemic and epidemic modelling.

A. **Model Choice**: I have already illustrated this point above. It's important to note that the choice of the modelling framework is heavily dependent on epidemiological knowledge of the interaction between the organism and the infected host. For example, reinfection possibility (and rates) can significantly change the dynamics of the model. Other sources of model uncertainty are the choice of variables, the degree of detail, and the endemic dynamics that the model is designed to capture. Much of the knowledge needed to reduce model uncertainty can only be obtained by consulting experts, even though reinfection rates can be derived statistically, provided we have enough data. Whether this information can be obtained by reliable methods depends on the type of infectious disease we are considering. The older and the more data we have, the more likely it will be that a scientific consensus has formed around key assumptions. With pandemics as recent as COVID-19, especially in the early waves, relevant knowledge comes from the informed speculations of experts, and by comparison with data from similar viruses (e.g., other coronaviruses) and pandemics (e.g., SARS, MERS). Analogical reasoning is highly fallible and dependent on human reasoning, that is, expert judgment: "It is a fact about human cognition that we very commonly make a judgment that one case is similar to another in drawing conclusions about what to do in daily life" (Walton et al. 2008: 55).

B. **Parameter Selection and Parametrization**: One of the fundamental sources of uncertainty in models is parameter uncertainty. The list of parameters that are theoretically relevant to an epidemiological model is virtually endless. Even though parsimonious models containing only a few variables are often considered to be more valuable to isolate key features of a pandemic (Bertozzi et al. 2020), we still need experts to make a judgment of relevance in the first place. For example, the connectedness of a network structure is fundamental for understanding how many steps a virus needs to spread through an entire network, given the same number of nodes (hosts). Social habits, geographical distribution of hosts, and many other factors affect a network's connectedness. The same is true about the clustering of a network (in lay terms, the layout of a network and how its nodes are distributed), and its degree of centrality (whether there are nodes in the network that are connected to all or most of the other nodes). Other important parameters are, for example, the rate of infectiousness—the R-number that for COVID-19 made headlines time and again (Adam 2020)—and the mode of transmission, namely through aerosol, droplets or fomites. Parameterization of a model often needs significant expert input. Some of the parameters can be obtained by statistical calculation; for instance, the Basic Reproduction Number can be relatively straightforwardly obtained via a variety of estimation methods. Other parameters, however, are harder to estimate, like compliance with behavioural interventions. Other the parameters can only be estimated with significant uncertainty: Morse et al. (2006) highlight the significant amount of uncertainty for very important parameters (e.g., effectiveness of non-pharmacological practices) and state that there are no science-based compendia of best practices.

3. Expert Judgment and Statistical Judgment

At this point I must clarify the difference between expert judgment and statistical (also mechanical, or actuarial) judgment. Expert judgment refers to the judgment of a human expert. I cannot provide an account of expertise in this article, and I will generally refer to experts as those who have attained sufficient experience and competence in a relatively narrow field of human knowledge (Martini 2019). Experts are said to possess tacit knowledge, through the application of which they are able to perform tasks (know-how), or give answers to well-formulated problems (know-what). For example, expert forecasters can answer questions like "what is the probability of a 48-hour clear-weather window for the next seven days on Everest".[3] For obvious reasons, in this paper I am considering only know-what experts.

Statistical judgment, on the other hand, is judgment delivered by calculation. We can think of expert systems as an example of statistical judgment. An expert system is a system of rules that can be implemented into a computer to aid or substitute human decision-making (Dreyfus 1987). Medical diagnoses are examples of decision problems: what is the most likely diagnosis for patient X, given symptoms Y_1, Y_2, …, Y_n, and patient characteristics Z_1, Z_2, …, Z_n? To mention a few concrete cases, Seixas et al. (2014) develop an expert system to support the diagnosis of a range of neurological disorders, Samuel and Omisore (2013) give a Web-Based Decision Support System for typhoid fever, and the list could become very long. Medical expert systems are fed data, either automatically (e.g., patient temperature) or through an operator (is the patient experiencing shivers?) and use algorithms to reach a conclusion on a diagnostic query.

Since the 1950s, Paul Meehl and, later, his collaborators, have undertaken an extensive research programme to show the superiority of statistical judgment over expert judgment. In the field of psychiatric diagnostics, Meehl showed that simple algorithmic tools were often able to outperform clinical evaluation in predicting human behaviour. A simple example will illustrate: Let us imagine that our task is to distinguish between psychotic and neurotic patients. We have two options: a) a clinical evaluation where a physician examines the patient; b) the Goldberg Rule. In 1965 Goldberg's study showed that a simple actuarial rule was

[3] The example is not random: The world of alpinism regards Karl Gabel and Vitor Baía as two of the best weather forecasters for high-altitude mountaineering. They have provided forecasts to top-class alpinists during their expeditions around the world. Their forecasts rely on models and data but also on significant experience and knowledge about mountaineering and the behaviour of weather patterns around high mountains. See Benavides 2018.

performing better than clinicians in diagnosing patients as either psychotic or neurotic. The rule takes a number of inputs from validity and clinical scales and gives a diagnosis based on the outcome of a simple calculation.

$$X = (L + Pa + Sc) - (Hy + Pt)$$

[L is a validity scale and Pa, Sc, Hy, and Pt are clinical scales of the MMPI: in order Paranoia, Schizophrenia, Hypomania, Psychasthenia]

If $x < 45$, diagnose patient as neurotic. If $x > 45$, diagnose patient as psychotic.

Goldberg Rule (from Bishop and Trout 2005: 14)

Goldberg's rule doesn't need significant judgment input. A short training, or even a user-friendly interface, will be enough for asking the patient the right questions and collecting the right data, and the rest is left to the algorithm. According to Bishop and Trout, "Sometimes, it would be better for the experts to hand their caseload over to a simple formula that a smart 8-year-old could solve" (Bishop and Trout 2005: 25).

By looking at much of the literature on the benefits of statistical over expert judgment, we may be tempted to think that, for most or possibly all decision problems, computers and algorithms perform better than human judgment. The claim is probably true, but it comes with a very important caveat: it is true for most phenomena for which we either have a rather precise understanding of the mechanisms involved, or enough statistical data. Unfortunately, not all phenomena are of this kind.

Aspinall and Cooke provide a fitting example of a decision problem for which expert judgment is needed; namely, predicting volcano behaviour:

> When a potentially deadly volcano becomes restless, civil authorities invariably turn to scientific specialists to seek to anticipate what the volcano will do next, and to help them judge the danger. Although it is usually possible to discern the earliest signs of unrest, *forecasting the course and precise timing of eruptions* still remains awkwardly inexact (Aspinall and Cooke 1998: 2113, italics added).

Aspinall and Cooke's article explains the application of the Cooke Method of expert elicitation to the Soufrière Hills volcano, a previously dormant volcano that became active in 1995 in the populated island of Montserrat and has been claiming land and lives ever since. They state that, despite the extensive monitoring equipment the Montserrat Volcano Observatory has set up all around the active volcano, expert judgment is still the predominant source of actionable information on evacuation decisions. "Even though armed with arrays of sophisticated monitoring equipment, the scientists working on the problem have a wide range of opinions about what the volcano might or might not do next" (Aspinall and Cooke 1998: 2114). A major problem with using expert judgment, however, is that it tends to produce a lot of noise. Experts tend to disagree a lot, and, while harnessing the power of their tacit knowledge, they also suffer a wide range of biases that affect human judgment (Faust 1984). In the Montserrat case, the attempt was to use Cooke's methodology of expert judgment elicitation and aggregation (Cooke 1991) to reduce the noise from experts.

In the case of the Montserrat Volcano, expert judgment filters and fills in the knowledge gaps from models about volcano behaviour and data collected on the ground. Is it possible that one day we might be able to develop a rule for deciding whether to evacuate an area on the basis of a number of suitable inputs and an underlying algorithmic rule that can process the data? Most likely so, however, as of today, we need judgment based on expert's experience and tacit knowledge.

To conclude this section, I shall return to the topic of epidemic modelling, after the digression on expert judgment and statistical judgment. Does predicting the course of an epidemic, and the effectiveness of different containment measures, look more like predicting the behaviour of a volcano, or like diagnosing psychotic and neurotic patients? In the next sections, we will try to understand how much of epidemic modelling is dependent on hard data, and how much is dependent on expert judgment. I will illustrate the contribution of expert judgment in epidemiological modelling focusing on a 2006 report of the *National Academies of Sciences, Engineering, and Medicine*.

4. Using Expert Judgment to Forecast the Next Pandemic

The 2006 report *Modeling Community Containment for Pandemic Influenza* (Institute of Medicine 2006) illustrates very well the interaction between statistical and expert judgment in epidemiological modelling. One of the report's tasks was to assess "the quality of existing models about a potential influenza pandemic and their utility for predicting the effects of various community containment policies on disease mitigation" (Institute of Medicine 2006: 1). The motivation for the report was that in 2006 experts were aware that a major pandemic was to be expected, and the report is particularly interested in evaluating the ability of models to predict the effects of nonpharmaceutical interventions (NPIs, like social distancing and face masks) in mitigating future expected outbreaks. The report stresses that the models "should be viewed as aids to decision-making, rather than substitutes for decision-making" (Institute of Medicine 2006: 4).

In modelling an epidemic there is much model uncertainty to begin with. Model uncertainty refers to the intrinsic uncertainty about the architecture of the model we choose, and the report indicates that epidemiological models need to rely "heavily upon expert judgment as to the inherent reasonableness of the model as a representation of reality" (Institute of Medicine 2006: 4). In general, reducing parameter uncertainty is less dependent on expert judgment when it is possible to run robustness analyses or collect more data. Even then, however, robustness analysis and data-gathering have costs in terms of resources and time. Good expert judgment can then reduce the need of collecting additional data and testing for robustness.

Conscious modelers are well-aware of the dependency of their data-crunching machines on expert input, and of the limits that that implies for the quality of knowledge obtained through models. One way to reduce the input from experts is to set up automatic data feeding mechanisms:

> One way to improve predictive ability is to adapt or construct decision-aid models that can incorporate surveillance data in real time and adapt to the actual experiences of an outbreak as it occurs. *Current models are based on educated guesses for a range of plausible values based on information from previous pandemics*. As a result, they are not able to predict with any certainty the future course of a pandemic and the

> effectiveness of interventions to reduce transmission (Institute of Medicine 2006: 13, italics added).

One way of reducing the dependency of models on expert input and educated guesses is then to set up mechanism for feeding surveillance data directly to the model. Nonetheless, this set up has limitations.

As in the case of volcano behaviour, it is possible and desirable that in the future much of the data the models need could be obtained through automated or semi-automated means. "Automated" means that the model can access large databases maintained by governments or private institutions, in this context it is important that data be transparent and, as much as possible, open (Molloy 2011). "Semi-automated" means that the model cannot directly access the data, but that data is nonetheless straightforwardly available to the modelers in the same way as national GDP, employment, or mortality rates are readily available. The objectivity of data that can be tapped into by a model does not need to be uncontroversial, there can still be some uncertainty; but the fundamentals of the methodology with which experts arrive at the data are shared by the large majority of the established experts.

Even once we establish a methodology for gathering and collecting better data, as Recommendation 7 of the Institute of Medicine (2006) report suggests, we are left with the problem that modelling a pandemic is an iterative process. The course of a pandemic is highly dependent on geography, human behaviour, and of course the evolution of diseases (e.g., due to virus mutation). It is unlikely that there will be, in the short run, a highly accurate predictive model like a descriptive model of celestial mechanics, because different equilibria between infective-agents and infected-hosts are likely to arise in response to changes in all the macro areas listed above: the environment, social behaviour, and the evolution of biological organisms. For that and other reasons models need to be regularly updated with new knowledge from a range of disciplines: biological and social ones. Recommendation 6 of the report reads as follows:

> The committee recommends that policymakers regularly convene forums for public dialogue on pandemic influenza modeling and analyses, and recommends the development of a *standing expert panel* to provide ongoing advice regarding models of pandemic influenza (Institute of Medicine 2006: 13).

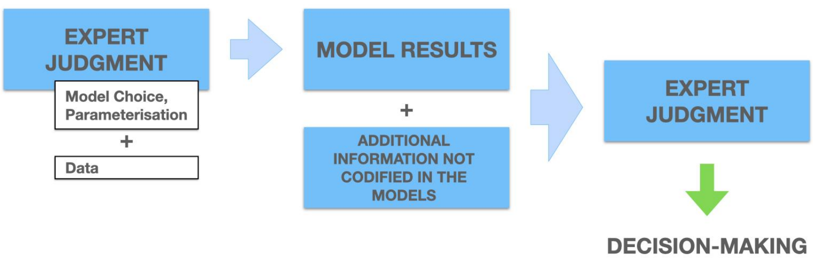
Ultimately, current epidemiological modelling, and especially the modelling of pandemic situations, relies heavily on the use of expert judgment at various phases of the modelling process. The last point I will discuss in this section, in relation to expert judgment, is the comparison between model-based evidence and expert-based evidence. The idea of expert-based evidence is that we can use experts and their judgments as estimators. Instead of relying on algorithm-driven estimates, we can use subjective probabilistic judgment (Cooke 1991). Once subjective judgments are elicited, possibly in a systematic way, they can be compared to model-based results for independent validation. This was the idea behind the work by Bankes, Aledort and colleagues, from the RAND corporation (see Institute of Medicine 2006: 9, RAND model). Aledort and colleagues ran an elicitation exercise to evaluate a package of non-pharmaceutical interventions, among which respiratory etiquette, hand hygiene, N95 respirators, etc. (see Aledort et al. 2007). The goal was to evaluate a package of non-pharmaceutical interventions during a seasonal influenza epidemic. The same package of choices was then run with two models—namely epidemiological and policy effectiveness models—and the package of preferred expert choices was compared to the results from 1000 runs of the linked models (Institute of Medicine 2006: 9).

The reason for resorting to subjective expert opinion was clear:

> In light of the evident lack of scientific evidence about specific non-pharmaceutical interventions in the context of seasonal or pandemic influenza, there was limited directly useable information from the majority of the studies identified in the formal Medline search. For this reason, *we turned to expert opinion to inform and categorize the findings* (Aledort et al. 2007: 3, italics added).

In this section I have explained how subjective expert judgment and modelling exercises for forecasting epidemiological phenomena and related containment measures are interlinked with one another. Given the complexity of the problem, the relative lack of standardized and agreed upon modelling framework, the lack of standardized procedures for collecting data, and the developing nature of the interaction between infectious pathogens and hosts, the starting point of understanding a specific pandemic situation is expert judgment. The role of epidemiological models is therefore twofold: one the one hand, simple models can help isolate and highlight key features of a phenomenon. In this respect, a model cuts out a piece of reality and allows modelers to analyse some of the chosen factors in isolation from the "messiness" of the real world. On the other hand, a computational model serves as a computational aid to otherwise intractable systems of equations, thereby making possible simulations of likely scenarios and outcomes, given varying initial conditions.

Ultimately, the results of a modelling exercise are dependent on the accuracy of the initial assumptions (model choice, and parameterization), and on the degree of uncertainty of each choice of variables and associated values. This type of complex post-hoc assessment is again in the hands of subjective expert judgment. Policy-making based on modelling results is neither straightforward nor independent of human judgment. In short, we can confidently claim that the starting and end points of modelling a pandemic are subjective expert judgments. The figure below illustrates this point.



The report on pandemic modelling reflects well Mill's point on studying complex phenomena like tides. We can see Mill's point in relation to a contrast class: the study of the movement of celestial bodies: astronomy. In astronomy, we can isolate a few causal factors, of similar nature to one another, and all acting in ways that can be relatively easily aggregated and taken into account. A small variation in parametrization is likely to make a difference if our instruments are good enough to detect the difference. Uncertainty and inexactness in astronomy is a matter of degree and precision of our instrumentation. Not so in the study of tides and epidemiology: tidal levels and the spread of diseases are dependent on causally interconnected factors that are very different in nature from one another, and small and large uncertainties for each of these factors can produce significantly different observations.

5. How to Handle Expert Judgment: Strategies

So far, I have assumed that the interaction between model-based and expert-based judgment is unproblematic. In this section I will explain why it is not. The main problem with expert judgment is that it tends to be biased in several ways. This is not to say that statistical judgment cannot be biased as well, but in this section I will focus on the specific ways in which biases in expert judgment reduce the epistemic quality of knowledge dependent on it. I will suggest established methodologies for reducing that bias, in particular: 1) the progressive substitution of subjective expert judgment with appropriate statistical methods, when possible, and 2) the reduction of subjective bias by means of aggregation techniques and structured expert judgment elicitation methods.

Since the work that Paul Meehl undertook in the 1950s, and the subsequent research that his *Disturbing Little Book* spurred (Meehl 1986), many scholars have highlighted the fact that human judgment is particularly ill suited at handling complex information and statistical data. Meehl, and later his collaborators, showed that those we call experts are often not so good at giving us good judgment on a number of practical questions. Trout and Bishop (2005) give the example of parole boards judging whether criminals should be eligible for parole. The heuristics and biases programme (Tversky and Kahneman 1974) has shown how human judgment fails to produce good estimates in a wide range of common tasks. The heuristics that we use to solve common daily problems, while useful with relatively menial tasks, can lead us into traps when the number of variables increases, and probabilistic interactions substitute deterministic causal pathways. As I explained above, pandemics and, in general, epidemiological phenomena, have all the characteristics of complex phenomena. The spread and impact of a virus on a host population depends on the interrelation of biological, ecological, and behavioural factors. Reliance on expert judgment, then, while it is inevitable because of our current state of knowledge, is also problematic.

The first important point here is that I have so far assumed that interrelation of model-based and expert-based judgment is straightforward. It is not. Significant literature on the interrelation has argued that whenever statistical and actuarial judgment is available, expert judgment will tend to ignore important insights from it and diminish the epistemic value of a combined judgment. In other words, when given the chance, the expert will tend to favour their own judgment, rather than the model's judgment, and often this will lead to an inferior overall assessment (Leli and Filskov 1984).

Trout and Bishop call the strategy of deviating from statistical judgment, when given the possibility, *epistemic exceptionalism* (Trout and Bishop 2005: 43). The reasoning goes as follows: let us suppose we have an algorithmic decision-making rule that works under most conditions. Under condition X, the rule tells us to choose Y. Clearly, there will be circumstances (sometimes called *broken leg* *scenarios*) under which the rule gives us the wrong answer, and we should deviate from its results. How do we know whether the rule is working or not, in a particular case? The judgment of an expert will have to provide that kind of information and, not surprisingly, experts tend to misjudge how often a case is an example of the class of cases in which the rule does not apply. Experts, that is, overestimate the number of cases in which a rule fails to provide the correct answer. Epistemic exceptionalism, therefore, is often the wrong strategy, when mixing actuarial and subjective decision-making. Bishop and Trout are clear on this: when a statistical method (an expert system) is available, experts should never stray from it except under very exceptional circumstances.

Suggestion 1. The first strategy for ameliorating our modelling practices is to improve our models in order to avoid excessive interactions between actuarial and subjective decision making. We should identify as clearly as possible those subdomains in which model-based judgment can be shown to be conclusively outperforming expert estimations. That will leave out those parts of the problems where, instead, we must rely on subjective judgment. Avoiding epistemic exceptionalism in key aspects of pandemic modelling should help clarify when subjective expert judgment ought to leave room to statistical judgment, a sort of "expert humility". The important point here is to understand that models are better at doing some things, and when we can identify the domains in which they are better we should let them handle the work.

The previous suggestion leaves implicit something I stated in the previous sections: there are domains in which models are either worse at producing knowledge, or not available at all. In sections 3 and 4 I argued we cannot predict the course of a pandemic with and without containment measures with models alone, so we must resort to established expertise. The question then is what we can do to reduce bias. For example, in the case of the COVID-19 pandemic, from the very beginning the main type of expertise that policy makers and the public listened to was health-centred. In the urgency of the initial steps that made sense because hospitals were being overrun and knowledge about the virus and possible treatments was scarce. Much thinking that went into the formulation of policies was also health-centred. This is not a note of criticism, but rather something to take at face value: Doctors, epidemiologists and health officials were in the public spotlight; their words were being analysed, criticised or otherwise glorified in the media. In that context health-centred thinking rightly influenced initial containment policies: "Up until late April, the Finnish government followed a script written predominantly by THL. THL is, by definition, a health utilitarian agency" (Häyry 2021: 117) The same is true for most governments around the world. In the medium and long run, it is possible to argue that health-centeredness can act as an epistemic bias and lead to groupthink. Groupthink is recognized as a cognitive bias that affects the quality of decision-making (Cleary et al. 2019). It is reasonable to argue that in the long run diversity of relevant expertise is important in modelling an epidemic.

Groupthink is an example of the possible biases that affect expert judgment, also in interaction with modelling efforts. The question is then whether and how expert judgment can be debiased. There is extensive literature on the subject, so in these final sections I will only be able to mention a few general points.

Suggestions 2. The second strategy for ameliorating our modelling practices is to reduce bias in expertise. There are a few ways to reduce biases:

A) **Prefer groups rather than individuals**. Committees have been shown to avoid some of the biases that would otherwise affect judgment, for example, when it comes to estimation, groups of experts tend to outperform individual experts. Cooke (1991) and his continuous work with various other collaborators, has developed a methodology of expert elicitation based on teams and judgment aggregation that is being used in volcanology (Aspinall 2010), by the Intergovernmental Panel for Climate Change (Kunreuther et al. 2014), and in several other applications of science. From an epistemological viewpoint, we should prefer judgment aggregation to singular thinking, even though aggregation has limitations and can be problematic (Martini and Sprenger 2017).

B) **Consider diversity in groups**. Diversity plays an important role in avoiding some of the biases that affect group decision-making and group-deliberation. Diversity ought to be limited when it affects the quality of the expertise base. To explain, there are two important elements that need to be considered when using expert judgment: the level of expertise and the diversity of the group. If the group lacks diversity, especially if the type of problems it deals with are complex and open-ended, then it is advisable to add diversity (Page 2008). But the diversity imperative cannot be absolute: If in order to add diversity we are adding group members that affect the level of expertise of the group, then we must be careful not to decrease the epistemic worth of the collective. In short, expertise and diversity must be balanced with one another.

6. Conclusion

In this article I have reviewed the role of expert judgment in epidemiological and pandemic modelling. I have highlighted how epidemiological and pandemic modelling are highly dependent on expert judgment, so much that getting the expertise right is as important as getting the model right. It is unlikely that there will be, in the short run, a highly accurate predictive model for pandemics; something like a descriptive model of celestial mechanics. If possible, that would be a welcome improvement, but for the time being we need to focus on expertise just as much as, if not more than, on technical issues about modelling.

Moreover, the starting and final points of modelling a pandemic are the same: expert judgment. That means that expert judgment is the first step in producing and parametrizing a model, and also that the product of a model of a pandemic is an input into the judgment of decision-making experts: "Above all, models should be viewed as aids to decision-making, rather than substitutes for decision-making" (Institute of Medicine 2006).

References

Achinstein, P. 1968, *Concepts of Science*, Baltimore: Johns Hopkins University Press.

Adam, D. 2020, "A Guide to R—The Pandemic's Misunderstood Metric: What the Reproduction Number Can and Can't Tell Us About Managing COVID-19", *Nature* 583, 346-48, doi: 10.1038/d41586-020-02009-w.

Aledort, J.E., Lurie, N., Wasserman, J., and Bozzette, S.A. 2007, "Non-Pharmaceutical Public Health Interventions for Pandemic Influenza: An Evaluation of the Evidence Base", *BMC Public Health*, 7, 208, doi: 10.1186/1471-2458-7-208.

Aspinall, W. 2010, "A Route to More Tractable Expert Advice", *Nature*, 463, 7279, 294-95.

Aspinall, W. and Cooke, R.M. 1998, "Expert Judgement and the Montserrat Volcano Eruption", in Mosleh, A. and Bari, R.A. (eds.) *Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management PSAM4*, September 13th -18th 1998, New York City, Vol. 3, London: Springer, 2113-18.

Benavides, A. 2018, "High-Altitude Forecasting: The Weather Wizards of the Greater Ranges", https://explorersweb.com/2018/11/20/high-altitude-forecasting-the-weather-wizards-of-the-greater-ranges/ (accessed: September 2021).

Bertozzi, A.L., Franco, E., Mohler, G., Short, M.B., and Sledge, D. 2020, "The Challenges of Modeling and Forecasting the Spread of COVID-19", *Proceedings of the National Academy of Sciences*, 117, 29, 16732-738.

Bishop, M.A. and Trout, J.D. 2005, *Epistemology and the Psychology of Human Judgment*, Oxford: Oxford University Press.

Black, M. 1962, *Models and Metaphors*, Ithaca: Cornell University Press.

Cleary, M., Lees, D., and Sayers, J. 2019, "Leadership, Thought Diversity, and the Influence of Groupthink", *Issues in Mental Health Nursing*, 40, 8, 731-33, doi: 10.1080/01612840.2019.1604050.

Cooke, R. 1991, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford: Oxford University Press.

Dreyfus, H.L. 1987, "From Socrates to Expert Systems: The Limits of Calculative Rationality", *Bulletin of the American Academy of Arts and Sciences*, 40, 4, 15-31.

Faust, D. 1984, *The Limits of Scientific Reasoning*, Minneapolis: University of Minnesota Press.

Goldberg, L.R. 1965, "Diagnosticians vs. Diagnostic Signs: The Diagnosis of Psychosis vs. Neurosis from the MMPI", *Psychological Monographs*, 79, 9, 1-28, doi: 10.1037/h0093885.

Häyry, M. 2021, "The COVID-19 Pandemic: A Month of Bioethics in Finland", *Cambridge Quarterly of Healthcare Ethics*, 30, 1, 114-22.

Kunreuther, H., Gupta, S., Bosetti, V., Cooke, R., Dutt, V., Ha-Duong, M., Held, H., Llanes-Regueiro, J., Patt, A., Shittu, E., and Weber, E. 2014, "Integrated Risk and Uncertainty Assessment of Climate Change Response Policies", in Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., and Minx, J.C. (eds.), *Climate Change 2014: Mitigation of Climate Change: Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge: Cambridge University Press.

Institute of Medicine 2006, "Modeling Community Containment for Pandemic Influenza: A Letter Report", Washington, DC: The National Academies Press, doi: 10.17226/11800.

Leli, D.A., and Filskov, S.B.F. 1984, "Clinical Detection of Intellectual Deterioration Associated with Brain Damage", *Journal of Clinical Psychology*, 40, 1435-41.

Mäki, U. 1992, "On the Method of Isolation in Economics", *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26, 4, 317-51.

Mäki, U. 2001, "Models, Metaphors, Narrative, and Rhetoric: Philosophical aspects", in Smelser, N.J. and Baltes, B. (eds.), *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam: Elsevier, 9931-37.

Martini, C. 2019, "The Epistemology of Expertise", in Fricker, M., Graham, P.J., Henderson, D., and Pedersen, N.J. (eds.), *The Routledge Handbook of Social Epistemology*, New York: Routledge, 115-22.

Martini, C. and Sprenger, J. 2017, "Opinion Aggregation and Individual Expertise", in Boyer-Kassem, T., Mayo-Wilson, C., and Weisberg, M. (eds.), *Scientific Collaboration and Collective Knowledge*, Oxford: Oxford University Press, 180-201.

Meehl, P.E. 1986, "Causes and Effects of my Disturbing Little Book", *Journal of Personality Assessment,* 50, 3, 370-75.

Mill, J.S. 1882 [1843], *A System of Logic, Ratiocinative and Inductive* (Eight Edition), New York: Harper & Brothers.

Molloy, J.C. 2011, "The Open Knowledge Foundation: Open Data Means Better Science", *PLoS Biol*, 9, 12, e1001195, doi: 10.1371/journal.pbio.1001195.

Morse, S.S., Garwin, R.L., and Olsiewski, P.J. 2006, "Next Flu Pandemic: What to Do Until the Vaccine Arrives?" *Science*, Nov. 10, 314, 5801, 929, doi: 10.1126/science.1135823.

Page, S. 2008, *The Difference*, Princeton: Princeton University Press.

Seixas, F.L., Zadrozny, B., Laks, J., Conci, A., and Saade, D.C.M. 2014, "A Bayesian Network Decision Model for Supporting the Diagnosis of Dementia, Alzheimer's Disease and Mild Cognitive Impairment", *Computers in Biology and Medicine*, 51, 140-58.

Samuel, O.W. and Omisore, M.O. 2013, "Hybrid Intelligent System for the Diagnosis of Typhoid Fever", *Journal of Computer Engineering and Information Technology*, 2, 2, doi: 10.4172/2324 9307.

Tversky, A. and Kahneman, D. 1974, "Judgment Under Uncertainty: Heuristics and Biases", *Science*, 185, 4157, 1124-31.

Walton, D., Reed, C., and Macagno, F. 2008, *Argumentation Schemes*, Cambridge: Cambridge University Press.