



# Eliciting Model Structures for Multivariate Probabilistic Risk Analysis

Mark Burgman<sup>1\*</sup>, Hannah Layman<sup>2</sup> and Simon French<sup>3</sup>

<sup>1</sup>Centre for Environmental Policy, Imperial College London, London, United Kingdom, <sup>2</sup>FTSE Russell, London Stock Exchange Group, London, United Kingdom, <sup>3</sup>Alliance Manchester Business School, Manchester, United Kingdom

Notionally objective probabilistic risk models, built around ideas of cause and effect, are used to predict impacts and evaluate trade-offs. In this paper, we focus on the use of expert judgement to fill gaps left by insufficient data and understanding. Psychological and contextual phenomena such as anchoring, availability bias, confirmation bias and overconfidence are pervasive and have powerful effects on individual judgements. Research across a range of fields has found that groups have access to more diverse information and ways of thinking about problems, and routinely outperform credentialed individuals on judgement and prediction tasks. In structured group elicitation, individuals make initial independent judgements, opinions are respected, participants consider the judgements made by others, and they may have the opportunity to reconsider and revise their initial estimates. Estimates may be aggregated using behavioural, mathematical or combined approaches. In contrast, mathematical modelers have been slower to accept that the host of psychological frailties and contextual biases that afflict judgements about parameters and events may also influence model assumptions and structures. Few, if any, quantitative risk analyses embrace sources of uncertainty comprehensively. However, several recent innovations aim to anticipate behavioural and social biases in model construction and to mitigate their effects. In this paper, we outline approaches to eliciting and combining alternative ideas of cause and effect. We discuss the translation of ideas into equations and assumptions, assessing the potential for psychological and social factors to affect the construction of models. We outline the strengths and weaknesses of recent advances in structured, group-based model construction that may accommodate a variety of understandings about cause and effect.

**Keywords:** parameters, model structures, uncertainty, elicitation, cognitive bias

## INTRODUCTION

Quantitative models often drive the risk analyses that estimate the probability of adverse events and assess their impacts on stakeholders. If these models do not capture all important perspectives on features that affect the probability and scale of the risk, then appropriate risk mitigation actions may be missed or not given sufficient importance, potentially amplifying adverse outcomes. For example, in the COV-19 pandemic, modellers omitted an important feature of the management of care homes, which may have impacted decision making about

## OPEN ACCESS

### Edited by:

Tina Nane,  
Delft University of Technology,  
Netherlands

### Reviewed by:

Rachel Wilkerson,  
Baylor University, United States  
Bruce G. Marcot,  
USDA Forest Service, United States

### \*Correspondence:

Mark Burgman  
mburgman@ic.ac.uk

### Specialty section:

This article was submitted to  
Mathematics of Computation and  
Data Science,  
a section of the journal  
Frontiers in Applied Mathematics and  
Statistics

**Received:** 15 February 2021

**Accepted:** 26 May 2021

**Published:** 09 June 2021

### Citation:

Burgman M, Layman H and French S  
(2021) Eliciting Model Structures for  
Multivariate Probabilistic Risk Analysis.  
*Front. Appl. Math. Stat.* 7:668037.  
doi: 10.3389/fams.2021.668037

responses and led to the significantly higher death rates among United Kingdom residents than expected. The BBC News reported<sup>1</sup>:

In the United Kingdom, modellers warned government that the virus could kill tens of thousands, and advised “cocooning” would reduce deaths. But Dr Ian Hall, of SPI-M, admits models did not reflect how care homes actually work, or identify the serious risk posed by agency staff working in different homes.

“The failure of those models, I guess, was that we didn’t know how connected the social care settings were with the community,” he says. “As modellers we didn’t know—I’m sure there are lots of academics and policy-makers out there, that could have told us this, if we’d asked them”.

Coronavirus would go on to kill more than 20,000 people in care homes.

There is a heavy responsibility on risk analysts (also sometimes called model-builders or knowledge engineers) to use every available tool to build appropriate models. Yet quantitative model building is still almost an amateurish art, seldom taught in university courses, its acquisition left to the vagaries of early career mentoring and experience. French [1] surveyed the elicitation processes that analysts use to build appropriate quantitative models, making the case that much further research and sharing of experiences is needed. Scientific judgement itself is value-laden, and bias and context are unavoidable when data are collected and interpreted. With this perspective in mind, we focus on building structures of cause and effect rather than building value models of the consequences, which involves the task of eliciting stakeholder preferences. Although we do not discuss these directly, many of our points apply equally to those.

Risk analysis deals with uncertainty. So, the models designed to assist in exploring the potential consequences are probabilistic, built around mathematical structures that represent ideas of cause and effect. Typically, model builders translate ideas about physical and biological processes into systems of equations, and then develop ways to represent epistemic and aleatory uncertainty. In doing so, they use their own judgement to decide, among other things, which probability density functions to sample, how to estimate the moments of these distributions and the dependencies between them. These are of course long-standing and well-known problems. When data are abundant and understanding of processes is more or less complete, a suite of mathematical tools are available to estimate functions and parameters and explore uncertainties, including statistical fitting, stochastic modelling and scenario analysis [2], and various forms of sensitivity analysis [3]. Mostly, however, the data for such models are scarce or unavailable and understanding is equivocal. In these

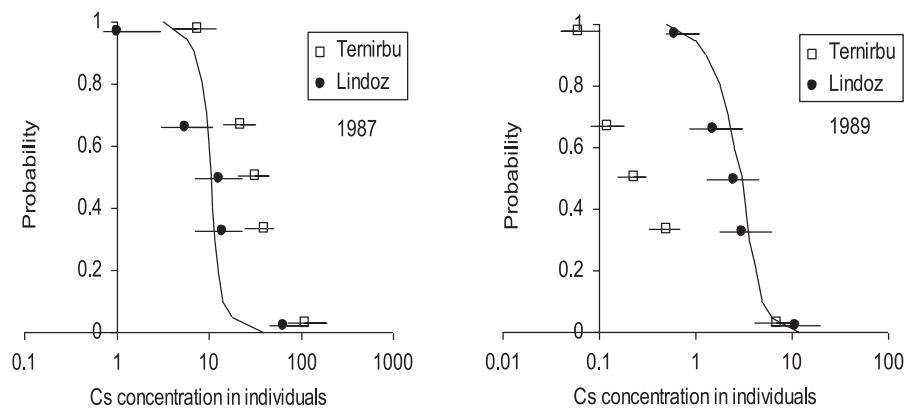
circumstances, analysts turn to expert judgement to fill knowledge gaps.

In this paper, we view “experts” not as those defined conventionally by their credentials and experience, but rather by broader criteria that include anyone who is interested in a problem, understands the context and the questions being asked and has substantive knowledge that may contribute to a solution [4]. There is well-established literature on the subjective, contextual and motivational biases to which individuals are susceptible. Phenomena such as self-interest, anchoring, availability bias, confirmation bias and overconfidence have pervasive and powerful effects on individual judgements of probabilities [5, 6]. Groups have access to more diverse information and ways of thinking about problems, and when managed appropriately, routinely provide more accurate judgements than do well-credentialled individual experts. This has been explored on a wide range of judgement and prediction tasks including political judgements, medical decisions, weather predictions, stock market forecasts and estimates for natural resource management [7–10]. Research has also led to the identification of specific limitations; groups may be misguided in situations in which one or a few opinions dominate, dissenting opinions are marginalized and motivational biases flourish [11], and the design of tools to avoid them.

Tools for structured group judgement are designed to anticipate and mitigate many psychological and contextual frailties that emerge in informal elicitation from individuals and groups [10, 12, 13]. The tools have many common features; interactions are facilitated, individuals make initial independent judgements, individual opinions are respected, participants consider the judgements made by others, and they may have the opportunity to reconsider and revise their initial estimates [14, 15]. Once a set of individual estimates has been acquired, they may be aggregated using behavioural, mathematical or combined approaches [15, 16]. These tools are designed to estimate the central tendencies and higher moments of single parameters.

In contrast, model structures and the assumptions behind them are usually chosen by individual analysts without recourse to any structured process to mitigate individual biases. The effects of well-established psychological and contextual frailties on the estimation of model structures remains a largely open question. In this paper, we outline approaches to eliciting and combining alternative judgments of cause and effect. We discuss the challenges in the transition from concepts to equations and parameters, including validating model structures, and we assess the potential for the psychological and social problems that afflict judgements of parameters to affect the construction of models. We describe novel approaches to solving these challenges that use collaborative elicitation and aggregation of conceptual and mathematical model structures that emulate the systems developed for structured elicitation of model parameters. While this paper attempts to cover current methods at each stage in the model development, given the lack of well documented case studies it consciously stops short of providing a road map for best practice. Instead it hopes to stimulate discussion and research to that end.

<sup>1</sup><https://www.bbc.co.uk/news/health-54976192>. Visited 19 November 2020.



**FIGURE 1 |** The predictions by two of the thirteen research groups (Ternirbu and Lindoz) together with the actual outcomes for a site (continuous curve) in central Bohemia in 1987 and 1989 for Cs concentrations resulting from Chernobyl fallout. The horizontal bars are 95% confidence intervals. The curves represent the probability (the proportion of individuals) in whom the concentration exceeded the concentration indicated on the x-axis (after [26], in [27]).

## EXPERT MODELS

Analysts typically choose how they will build a model. Yet, there are no formal guidelines for the process of model building itself [1]. In most cases the prior experience and skills of the modelers dominate the selection of methods [17]. Franco and Montibeller [18] provide a useful framework for discussing the contexts in which modelers operate. The expert mode of model construction is one in which analysts use established model structures and assumptions to solve a problem. They do so in relative isolation from other analysts, although they may consult other experts or stakeholders from time to time. This approach is commonplace especially when the problem and its solutions are within the known and knowable spaces of uncertainty (implying that at least some subsets of world are knowable; see Cynefin [19]).

However, there is a continuum between narrowly defined and well-established problems and those that have deeper uncertainties and wider implications. Outside a few very narrow problems, the so-called expert mode has deep and often unacknowledged problems. For example, French and Niculae [20] describe how in many emergencies such as extreme weather events, industrial accidents and epidemics, non-scientists and, sadly, many scientists put too much faith in model predictions, under-estimating inherent uncertainties and accounting poorly for social and economic impacts. Saltelli et al. [21] note that the choice of a particular modelling approach influences, and may determine, the outcome of an analysis. Our view is that the expert mode creates an environment in which the subjective preferences and values of the analysts themselves are inevitably embedded in the models and their assumptions, limiting the scope of alternative options and influencing the direction of solutions [21, 22, 23]. The hidden assumptions that drive value-laden outcomes can be difficult to challenge. Structures of privilege and status between institutions, disciplines and individuals within a group can emphasize particular perspectives and suppress others, privileging some model constructions and assumptions over others [21, 22, 24]. Instead of providing a vehicle to explore uncertainty,

quantitative analyses sometimes mask uncertainty and exclude alternative, equally valid options [25].

It is our experience that many people believe that if similarly-skilled and trained scientific specialists were provided with the same context, data and resources, they would reach more or less the same qualitative conclusions or recommend the same actions. Hoffman and Thiessen [26] tested the expert mode by asking a number of groups to build a model for the same problem, independently. They asked thirteen research groups in Europe to estimate (retrospectively) whole-body concentrations of  $^{137}\text{Cs}$  in people exposed to fallout from the Chernobyl nuclear accident. Analysts were provided details of the plume of radioactive waste together with data on radionucleotides in air, soil and water, local diets, and  $^{137}\text{Cs}$  concentrations in foods. The groups developed models and predictions without interacting with one another. **Figure 1** shows the results from Lindoz and Ternirbu, two of the modelling groups.

The geometric mean of the (approximately) lognormal distribution of concentrations predicted by Lindoz for 1987 matched the outcome quite well. However, the Lindoz group over-estimated the geometric standard deviation, resulting in underestimates at the low end and overestimates at the high end. For 1989, the predictions by the Lindoz group were nearly perfect. The Ternirbu group overestimated individual  $^{137}\text{Cs}$  concentrations in 1987 due to an overestimate of the amount of the contaminant in people's diets. They underestimated  $^{137}\text{Cs}$  concentrations in 1989 based on an incorrect assumption that soils would be tilled, which would have decreased bioavailability. Their agreement at the high end in 1989 was due to the fact that they also overestimated variability (an example of compensating errors).

While the comparisons above say nothing about the importance of differences, nor how the inaccuracies may have affected decision making, they serve to illustrate that equivalent training and common data alone are not sufficient to guarantee consistent predictions. Thus, single models that are built in relative isolation and that generate outputs that are difficult to verify, should be regarded with appropriate skepticism.

## FACILITATED CONCEPTUAL MODELS

As noted above, when analysts work in relative isolation without input from colleagues or stakeholders, they limit their access to the diverse perspectives and knowledge available to groups. Franco and Montibeller [18] outline an approach termed facilitated modelling. In this, analysts convene with stakeholders and subject matter experts to discuss a problem and formulate models. Model construction is iterative. Analysts identify factors of importance, draft models, outline assumptions and display model outputs. Stakeholders review the drafts and discuss ideas of cause and effect. The facilitator challenges participants to examine counterfactuals and alternative explanations, helping groups to avoid dysfunctional behaviours, ensuring they examine a problem in a structured fashion, and that group members share knowledge and perspectives comprehensively and equitably. Analysts refine the list of factors and their descriptions, revising their models and finding a consensus or capturing disparate ideas in alternative model assumptions [28].

Several approaches to facilitated modelling assimilate broad perspectives and access to information (see [1, 18]). One of the most popular is to represent ideas in causal maps ([29], closely related to mental models [30] and cognitive maps [31]), usually involving free-hand drawings of relationships between drivers (e.g., [32]). As the name suggests, causal maps are graphical mental models that show the direction of causation between variables. Participants represent variables diagrammatically as nodes (boxes) and their beliefs about the network of causal relationships between them as edges (arrows that represent the direction of causality). Typically, participants are free to arrange variables in any way (see examples in [28, 33]).

For example, Cawson et al. [34] convened a workshop to identify and rank the importance of factors driving fire in Australian forests. Workshop participants were recruited and grouped purposefully to include people with diverse experiences and perspectives. The convenors used the first stage of the workshop to task each group to independently create a list of factors influencing fire occurrence and fire intensity. They then used structured facilitated processes [10] to determine the relative importance of factors, identify common themes, discuss differences between individual models and develop a single (shared) conceptual model within each group. This approach assumes implicitly that there is no single “right” way to model the system. Two of their conceptual models are shown in **Figure 2**; one group decided to focus on “moisture” and another on “microclimate”. The convenors [34] noted different understandings of fire arising from systematically biased data collection (driven often by individual backgrounds), uncertainty about how things interact and subjective judgement resulting from unique experiences and the way different people process information.

Another approach to facilitate the creation of causal maps semantic webs. Semantic webs may be used to elicit ideas about cause and effect, in which participant interviews result in diagrams with a set of noun concepts represented as nodes in a network. Directional arrows labelled with relationship terms

(mostly verbs) show relatedness between concept nodes (see [35]). Alternatively, analysts may generate a list of factors (ideally using a group), and then present pairs of factors to participants, who are asked to comment on the potential for causal interactions between them. Participants consider all pairwise possibilities. This approach is comprehensive but time-consuming. Participants find it burdensome and the number of pairwise comparisons increases rapidly with the number of variables considered [29]. Causal maps constructed in this way tend to be much larger and include many more links than those based on free-hand, informal constructions [29; see also 17].

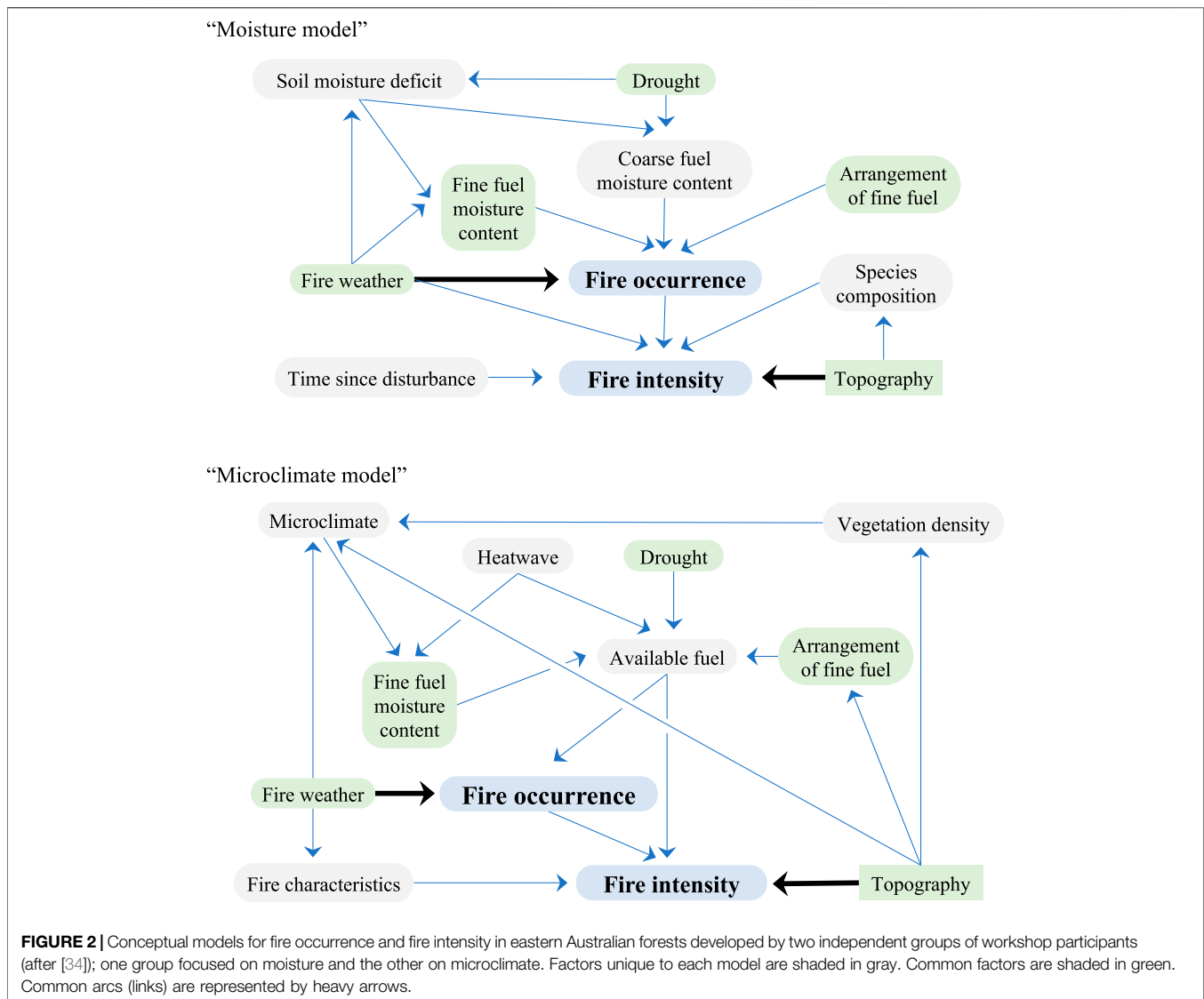
Conceptual model building faces a number of additional challenges that may be exacerbated by context or framing. Hysteresis or lag effects may cloud causal relationships between proximate and more ultimate causal factors (e.g., are catastrophic wildfires caused by local woody fuel build-up, or by decades of inappropriate forest management)? These issues reflect the need to establish the objective of the model at the outset, to clarify the context in which it is set and the purposes for which it will be used, recognizing that data will seldom be available for any substantial validation of the model, and that a judgement will need be made on whether it is fit for purpose [36, 37].

Cawson et al. [34] concluded that one of the principal benefits of facilitating alternative conceptual models is the opportunity to compare and contrast them, rather than reaching an explicit consensus. That is, it is useful to make alternative perspectives and experiences explicit, helping remove linguistic ambiguity and forming a basis for discussions about shared and divergent views. Saltelli et al. [21] likewise emphasized that the principal benefit of mathematical models is to explore ideas about cause and effect. Nevertheless, where alternative models change decision outcomes, decision-makers may need model outputs that provide some kind of consensus.

## CONSENSUS MENTAL MODELS

Moon et al. [38] describe two types of consensus mental models, termed “shared” and “team” mental models. “Shared mental models” are elicited as a facilitated, “collective task”, constructed when individuals interact together in a team setting [38]. The models created by each of the groups in Cawson et al.’s experiment ([34]; **Figure 2**) and the facilitated models envisaged by Franco and Montibeller [18] are examples of shared mental models. The process for generating them is based on facilitated, iterative group discussion and behavioural consensus (see also [38, 39]).

In team mental models, initially, different models are developed independently. A consensus view may be generated by “compiling the relationships among a group of elicited individual mental models into one model,” usually achieved mathematically. For example, Moon and Adams [33] constructed team mental models by amalgamating matrices (data adjacency matrices) representing connections between

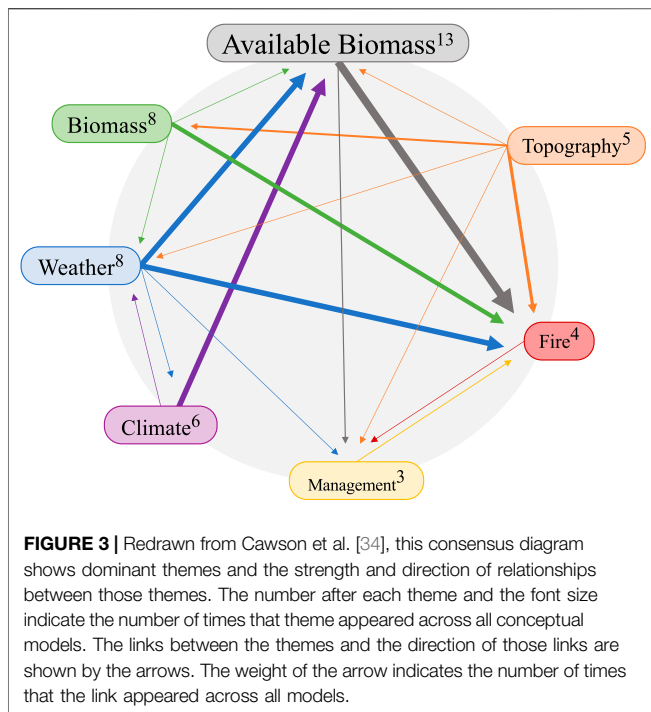


nodes; the edges between nodes are added, element by element, an approach used by, for example, [40, 41]. Markíczy and Goldberg [42] outline generalized measures of distances (dissimilarities) between pairs of causal maps, applicable if all groups use the same variables (see [39]). Different mental models that do not use the same variables may be assimilated into a consensus diagram by reducing them into a set of common elements. For example, using the shared models from each group, Cawson et al. [34] created a consensus diagram for the 13 different fire models to illustrate dominant themes, focusing on nodes and edges that most often occurred in causal maps created individually (**Figure 3**). We note that **Figure 3** makes a complete graph (in which all nodes are connected) except for the arc between climate and fire, which arguably is mediated by weather. Adding more arcs diminishes the prospect of evaluating what is non-causal in the graph, and impairing opportunities for validating model structures.

Scavarda et al. [43] outline an approach to creating causal maps that operates when individuals are dispersed and interact

asynchronously. They found initial causal maps, constructed independently had too many disconnected nodes and unmeasured arcs to be useful. In their approach to a team mental model, the analyst clusters nodes mathematically to generate a final “consensus” causal map. Moon and Browne [39] also developed a method to elicit and combine mental models.

There are current examples where modellers have access to consensus mental modelling, when their context creates the right incentives. For example, in finance, competition means that “sell side” modellers typically build valuation models in isolation from one another [44]. Sell side modellers can often see one another’s final predictions and so can adjust their forecasts relative to the consensus position. Modellers on the “buy side” not only have access to the individually developed “sell side” models but also opportunity to discuss ideas of cause/effect with sell side modellers and draw on their model structures to create their own models [45]. Despite the system mimicking some elements



of a structured elicitation processes, valuation modelers lack group diversity and typically don't seek broad input [46].

## TRANSLATING MENTAL MODELS INTO MATHEMATICAL MODELS

Box [47] noted, “Essentially all models are wrong, but some are useful,” and “Since all models are wrong the scientist must be alert to what is importantly wrong.” In decision analysis, the analyst has no basis for discerning important uncertainties from unimportant ones without knowing the subjective preferences of the decision-maker and the options under consideration. While uncertainty and differing conceptualisations are inevitable, models can also be logically wrong if, for instance, the units in the equations don't balance. In addition, equations or assumptions may clearly conflict with empirical evidence. These considerations matter because invalid or false models can lead to actions with severe, avoidable, detrimental consequences (e.g. [20, 48]). However, even when models are constructed carefully and accord with data and accepted theory, opinions about cause and effect may differ, and theory and data may be equivocal regarding assumptions that may affect important subjective preferences or options. So, what makes a model useful, remembering that there are many valid model constructions that reflect differing perspectives and contexts?

Mental models and causal maps may be translated into mathematical models to make predictions or test ideas. The translation step involves a raft of decisions about which variables to include, the kind of model, its equations and assumptions, among others [17]. The validity of the move from qualitative to quantitative analysis depends on the

narrative or visual representation of cause and effect and the mathematical expressions being closely aligned [29, 49]. Kemp-Benedict [50] suggested that analysts should faithfully represent the narrative in their mathematical models, as well as any fundamental constraints (e.g., energy or mass balances, economic limits), dependencies, and the spatial and temporal scales of key processes. These conditions place considerable demands on analysts to correctly interpret the narrative or causal map in a formal mathematical model. The analyst must yield “a large measure of control to the narrative team” [51].

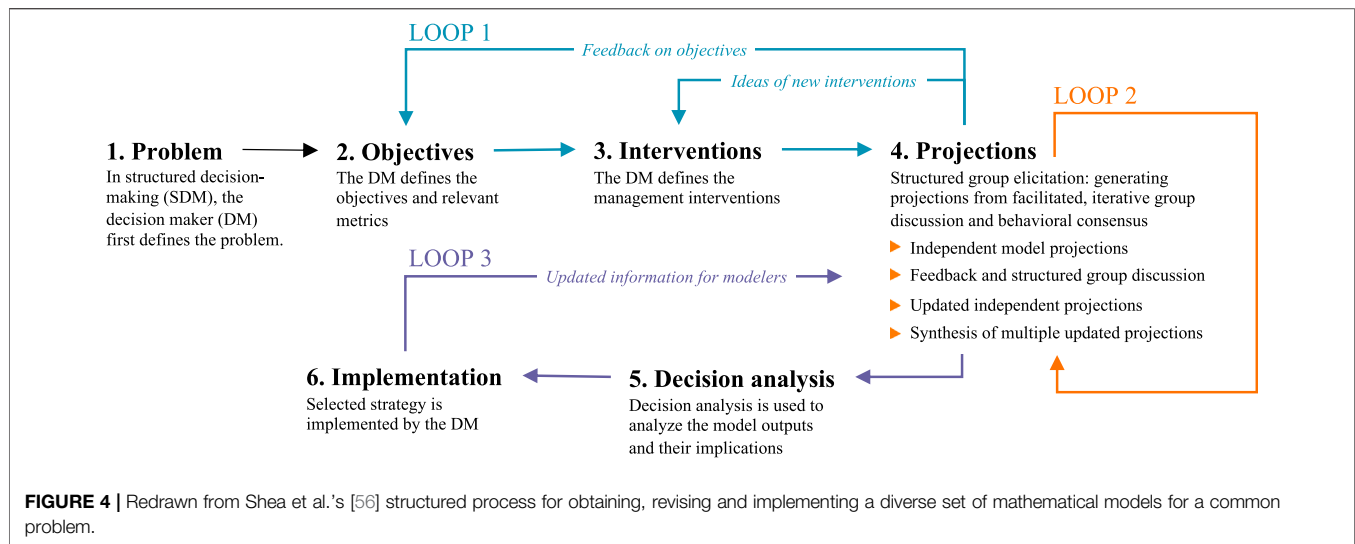
The model should reflect what the subject matter experts believe to be true, and it should allow the group to explore a neighbourhood of possibilities consistent with its consensus position and any areas of disagreement. The main role of the quantitative model is to take care of complications, keeping track of calculations, constraints and correlations. The complexity of the system—arizing from the mutual interactions between its constituent parts—is addressed principally by the narrative team. The point at which it is “worthwhile” to translate a qualitative model into a quantitative one is a question of the costs and benefits of the alternative actions in each circumstance.

Causal maps extend quite naturally into fuzzy cognitive maps ([52], e.g. [28, 39]) and Bayes nets ([53]; e.g., [28, 54]). Of course, it's not essential to begin with a causal map. Voinov and Bousquet [54] reviewed “companion modelling,” which uses agent-based models and role-playing games to build shared stakeholder models and “group model building,” which translates stakeholders' mental models into dynamic simulations by using system dynamic methods to identify feedback structures and cause–effect relationships).

Our conclusion is that model development should involve facilitated discussions between analysts and stakeholders and/or subject matter experts, illustrating the consequences of what they believe to be true, and testing the interpretations made by the analysts. Such iterative discussions may lead the analyst to revise their assumptions and model equations, or the stakeholders and subject matter experts to revise their thinking about cause and effect. The best we can hope for is that a model faithfully reflects the beliefs of subject matter specialists, that these beliefs accord with empirical evidence, and that the subsequent observations are a valid test of the model. Ideally, each model structural alternative will make unique and testable predictions.

## ELICITING AND COMBINING MULTIPLE MATHEMATICAL MODELS

A single analyst or analytical team can, of course, explore alternative model structural assumptions through sensitivity analysis, most straight-forwardly by varying the structure of the model, and its input parameter values and assessing changes in the response variables. This approach is limited implicitly by the scope of the analyst's perspectives on the problem, as noted above. The simplest way to elicit models independently is to provide context and background and ask analysts to develop models in relative isolation. There are a number of approaches for dealing with multiple models that



focus on mathematical averaging of model outputs (e.g., see [55–58] for examples).

As noted above, Cawson et al. [34] noted the benefits of a suite of models developed independently include the opportunities to compare ideas and discuss alternative perspectives. These conclusions resonate with others who have argued for independent model development as a way of exploring alternative beliefs about cause and effect ([21, 28, 36, 56]. Nevertheless, analysts may wish to assess the emergent judgements of a group of models, the consensus position, if there is one.

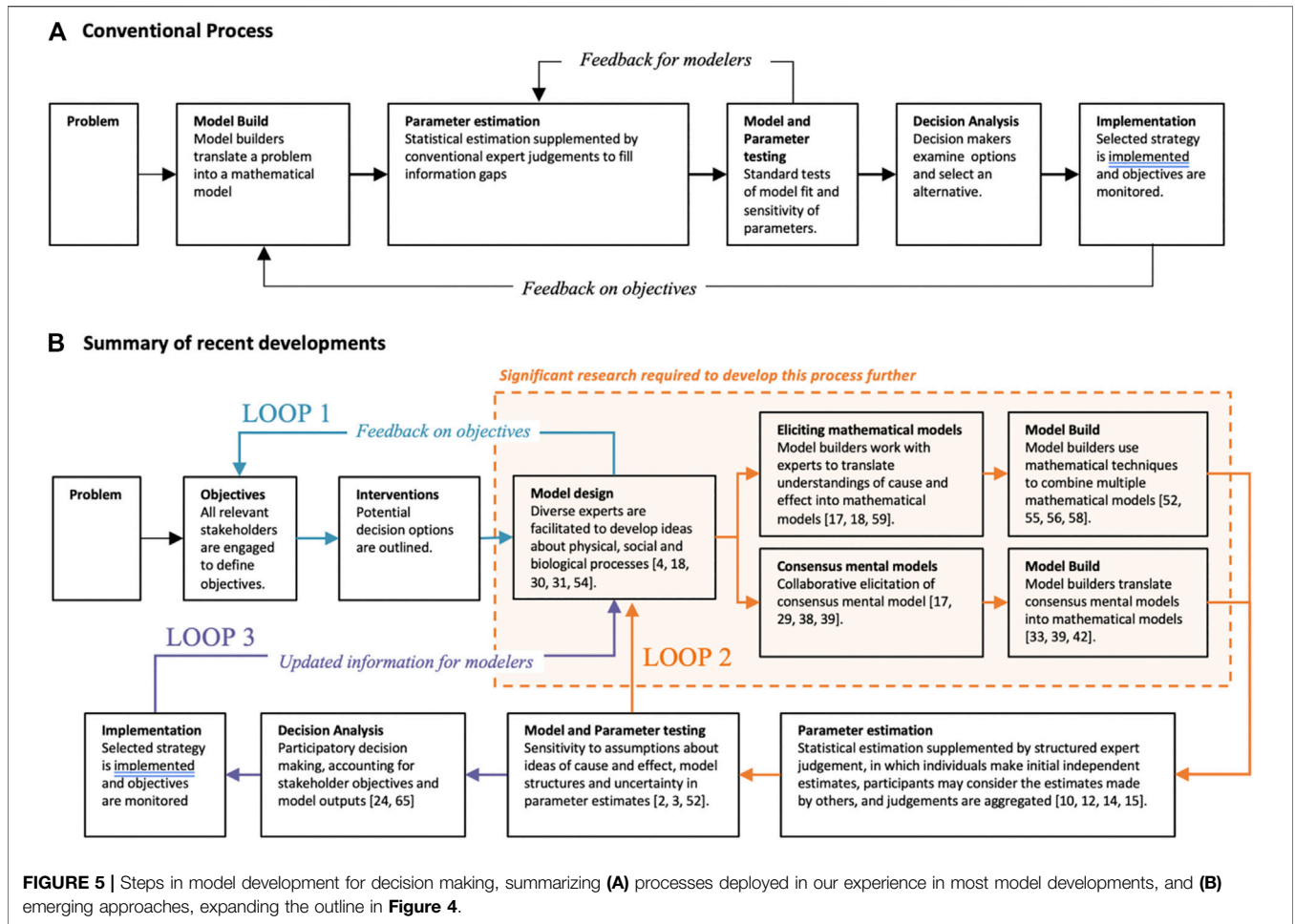
Shea et al. [56] recommended a structured, facilitated approach to developing multiple models (Figure 4). Initially, individual analysts or modelling groups build models and predict the outcomes of interventions. Of course, these groups could first explore ideas of cause and effect iteratively using mental models, before building mathematical models, as outlined above. In this approach however, after building a model, analysts then participate in formal, facilitated, Delphi-like structured discussions [9] to compare results, assess common features, discuss differences, and share information, thereby generating insights into the problem and its solutions. After discussion, analysts independently update their models and projections. Emergent properties can lead to robust and effective management actions. The process provides policy-makers with a sense of the central tendency of the projections across models, and a relatively comprehensive understanding of the underlying uncertainties.

In this scheme, “Implementation” includes the practicalities of interpretation of the model and associated strategies, often a challenging step in the process. It may lead to more detailed tactical problems, and the examination of additional trade-offs and social priorities (Loop 3).

Nicholson et al. [59] developed an approach to multi-model Bayes net development that also used Delphi-like interactions among multiple participants to elicit model structures, facilitate

discussion, mitigate cognitive and contextual biases, and aggregate outcomes. Initially, participants build causal maps independently. The system maintains participant anonymity to avoid dominance and halo effects. A facilitator moderates discussion. Participants agree on model variables (nodes) and structure (edges). Then, users specify the conditional probabilities for each child variable given each joint state of its parent variable. Group members use the resulting Bayes net to reason and explore the consequences of their thinking about variables and structure. Finally, participants cooperate in writing a report, organizing and explaining their analysis in detail, and including various key elements of good reasoning. This represents a significant advance in capturing domain expert ideas of cause and effect. However, eliciting an appropriate level of detail that is faithful to expert understandings and that is easily digestible to stakeholders remains a significant challenge. The arrows that domain experts may prescribe in conceptual models may not translate into the relationships that appear in a Bayes net. Diagnostics such as those in Cowell [60] may be used to check whether the conditional independences are consistent with data, even incomplete data. Barons et al. [61] offer an example of this in the context of eliciting structure.

In Figure 5A, we summarize the steps currently followed in most model-building and in Figure 5B, we expand Figure 4 to show some of the detail surrounding recent advances that current developments and discussions would suggest are more appropriate. However, the process outlined in Figure 5B is only a suggestion. Many decisions can be made without completing all, or even most, of these stages [62]. As noted above and in French [1], we are not proposing this process be adopted without substantial further research to establish good practice. The research needs to define best practice, ideally comparing and contrasting different processes and recommending approaches to model development for decision making that perform best and that are fit for purpose.



**FIGURE 5 |** Steps in model development for decision making, summarizing (A) processes deployed in our experience in most model developments, and (B) emerging approaches, expanding the outline in Figure 4.

## DISCUSSION

For the last 50 years, thinking about the psychology and social context of expert judgement has led to a revolution in the ways in which expert judgements are obtained and aggregated [5, 10, 12, 13, 60]. However, these advances have been restricted largely to the estimation of parameters and the outcomes of uniquely defined future events. Mathematical modelers have been slow to accept that the host of psychological frailties and contextual biases that afflict judgements about parameters and events may also influence the development of model structures and assumptions. Few, if any, quantitative risk analyses embrace sources of uncertainty comprehensively [27].

However, recent innovations [38, 53, 63] account for the potential for such influences. Similarly, there have been calls in operations research for a more inclusive and transparent approach to model development [21, 29]. These methods anticipate arbitrary behavioural and social biases and mitigate their effects, thereby generating model outcomes that are less influenced by the personal values and proclivities of individual modelers or decision-makers. They take advantage of the lessons learned from cognitive psychology about risk perception [63, 64]

and promise to make the art of model building more rigorous, accountable and faithful to the underlying narrative.

Emerging standards for model development demand that processes involve stakeholders, accommodate multiple views and promote transparency. Analysis of sensitivity and uncertainty are essential components of such systems [21] because they provide a framework for inclusive consideration of risks and equitable sharing of the burden of adverse consequences. Conceptual models (such as Cawson et al.'s [34] composite model) need to be designed in a way that leads to intuitive mathematical modelling. There is an urgent need for research and development of tools for the step between a mental model or suite of mental models and a combined mathematical model [1]. Bercht and Wijermans [65] describe how unstated and unreconciled mental models among researchers may hamper successful collaboration. As with judgments, the development of ideas of cause and effect, assumptions and modelling structure/tools should be done independently and iteratively, to pit diversity of thought against individual bias. This view is predicated on interdisciplinary diversity in modelling teams to mitigate the tunnel vision created by similar educational and disciplinary backgrounds. Such transparency enhances the public



acceptance of decisions, especially when embedded in participatory decision making processes [66].

Various methods are available to combine group judgments, outlined above. We conclude that it is less important “how” you translate to a mathematical model but rather that the process of translating is iterative, and that significant effort is dedicated to ensuring it obeys the consensus of ideas of cause and effect. Thus, the focus of them all should be on incorporating divergent views into an investigation of a model’s limitations. For example, in the context of the COV-19 pandemic, involving a diverse group of experts, and using independent and iterative process would have been more likely to identify the missing ideas of cause and effect in the model structures and led to better-informed decisions. Such procedures will inevitably add cost and time to the development of specific models, but with the considerable benefit of reducing the substantial costs that arise when models overlook critical factors and lead to poor decisions. As the example of the financial modelers illustrates, creating the right incentives is a critical step on the path toward the adoption of cooperative model building more broadly in public policy and decision-making. There is also a critical need to explore more fully whether and how concepts of cause and effect can be

translated faithfully into sets of equations and parameters, and for assessing the pervasiveness and importance of psychological and social influences on model development.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## ACKNOWLEDGMENTS

The authors thank Terry Walshe, Katie Moon and two reviewers for many insightful and helpful comments.

## REFERENCES

- French S. From Soft to Hard Elicitation. *J Oper Res Soc* (2021). 1–17 doi:10.1080/01605682.2021.1907244
- Lindgren M, and Bandhold H. *Scenario Planning: The Link between Future and Strategy*. London: MacMillan (2009) doi:10.1057/9780230233584
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global Sensitivity Analysis: The Primer*. Chichester: Chichester John Wiley (2008).
- Burgman M, Carr A, Godden L, Gregory R, McBride M, Flander L, et al. Redefining Expertise and Improving Ecological Judgment. *Conservation Lett* (2011) 4:81–7. doi:10.1111/j.1755-263x.2011.00165.x
- Slovic P. Trust, Emotion, Sex, Politics, and Science: Surveying the Risk-Assessment Battlefield. *Risk Anal* (1999) 19:689–701. doi:10.1111/j.1539-6924.1999.tb00439.x
- Burgman MA. *Trusting Judgements: How to Get the Best Out of Experts*. Cambridge: Cambridge University Press (2015) doi:10.1017/cbo9781316282472
- Hueffer K, Fonseca MA, Leiserowitz A, and Taylor KM. The Wisdom of Crowds: Predicting a Weather and Climate-Related Event. *Judgment Decis Making* (2013) 8:91–105.
- Tetlock PE, Mellers BA, Rohrbaugh N, and Chen E. Forecasting Tournaments. *Curr Dir Psychol Sci* (2014) 23:290–295. doi:10.1177/0963721414534257
- Kämmer JE, Hautz WE, Herzog SM, Kunina-Habenicht O, and Kurvers RHJM. The Potential of Collective Intelligence in Emergency Medicine: Pooling Medical Students’ Independent Decisions Improves Diagnostic Performance. *Med Decis Making* (2017) 37:715–724. doi:10.1177/0272989x17696998
- Hemming V, Walshe TV, Hanea AM, Fidler FM, and Burgman MA. Eliciting Improved Quantitative Judgements Using the IDEA Protocol: a Case Study in Natural Resource Management. *PLoS One* (2018) 13(6):e0198468. doi:10.1371/journal.pone.0198468
- Bénabou R. Groupthink: Collective Delusions in Organizations and Markets. *Rev Econ Stud* (2013) 80:429–62. doi:10.1093/restud/rds030
- Cooke RM. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford: Oxford University Press (1991).
- Tetlock PE, and Gardner D. *Superforecasting: The Art and Science of Prediction*. New York: Random House (2016).
- Hanea AM, McBride MF, Burgman MA, and Wintle BC. Classical Meets Modern in the IDEA Protocol for Structured Expert Judgement. *J Risk Res* (2018) 21:417–33. doi:10.1080/13669877.2016.1215346
- O’Hagan A. Expert Knowledge Elicitation: Subjective but Scientific. *The Am Statistician* (2019) 73(Suppl. 1):L69–81.
- Cooke RM, and Goossens LHJ. Procedures Guide for Structural Expert Judgement in Accident Consequence Modelling. *Radiat Prot Dosimetry* (2000) 90:303–9. doi:10.1093/oxfordjournals.rpd.a033152
- Voinov A, Jenni K, Gray S, Kolagani N, Glynn PD, Bommel P, et al. Tools and Methods in Participatory Modeling: Selecting the Right Tool for the Job. *Environ Model Softw* (2018) 109:232–55. doi:10.1016/j.envsoft.2018.08.028
- Franco LA, and Montibeller G. Facilitated Modelling in Operational Research. *Eur J Oper Res* (2010) 205:489–500. doi:10.1016/j.ejor.2009.09.030
- French S. Cynefin, Statistics and Decision Analysis. *J Oper Res Soc* (2013) 64: 547–61. doi:10.1057/jors.2012.23
- French S, and Nicolae C. Believe in the Model: Mishandle the Emergency. *J Homeland Security Emerg Manag* (2005) 2(1). Available from: <https://EconPapers.repec.org/RePEc:bpj:johsem:v:2:y:2005:i:1:p:18:n:5> (Accessed May 31, 2021).
- Saltelli A, Bammer G, Bruno I, Charters E, Di Fiore M, Didier E, et al. Five Ways to Ensure that Models Serve Society: a Manifesto. *Nature* (2020) 582: 482–4. doi:10.1038/d41586-020-01812-9
- O’Brien M. *Making Better Environmental Decisions: An Alternative to Risk Assessment*. Cambridge Mass: MIT Press (2000).
- Stirling A, and Mitchell C. Evaluate Power and Bias in Synthesizing Evidence for Policy. *Nature* (2018) 561:33. doi:10.1038/d41586-018-06128-3
- Stirling A. Opening up the Politics of Knowledge and Power in Bioscience. *Plos Biol* (2012) 10:e1001233. doi:10.1371/journal.pbio.1001233
- Stirling A. Precaution in the Governance of Technology. In: Brownsword R, Scotford E, and Yeung K, editors. *Oxford Handbook on the Law and Regulation of Technology*. Oxford: Oxford University Press (2017) 645–69. Chapter 26.
- Hoffman FO, and Thiessen KM. The Use of Chernobyl Data to Test Model Predictions for Interindividual Variability of <sup>137</sup>Cs Concentrations in Humans. *Reliability Eng Syst Saf* (1996) 54:197–202. doi:10.1016/s0951-8320(96)00075-0
- Burgman MA. *Risks and Decisions for Conservation and Environmental Management*. Cambridge: Cambridge University Press (2005). doi:10.1017/cbo9780511614279
- Walshe T, and Burgman M. A Framework for Assessing and Managing Risks Posed by Emerging Diseases. *Risk Anal* (2010) 30:236–49. doi:10.1111/j.1539-6924.2009.01305.x

29. Hodgkinson GP, Maule AJ, and Bown NJ. Causal Cognitive Mapping in the Organizational Strategy Field: a Comparison of Alternative Elicitation Procedures. *Organizational Res Methods* (2004) 7:3–26. doi:10.1177/1094428103259556
30. Morgan GM, Fischhoff B, Bostrom A, and Atman CJ. *Risk Communication: A Mental Models Approach*. Cambridge: Cambridge University Press (2002).
31. Eden C. Cognitive Mapping. *Eur J Oper Res* (1988) 36:1–13. doi:10.1016/0377-2217(88)90002-1
32. Green DW, and McManus IC. Cognitive Structural Models: The Perception of Risk and Prevention in Coronary Heart Disease. *Br J Psychol* (1995) 86:321–36. doi:10.1111/j.2044-8295.1995.tb02755.x
33. Moon K, and Adams VM. Using Quantitative Influence Diagrams to Map Natural Resource Managers' Mental Models of Invasive Species Management. *Land Use Policy* (2016) 50:341–51. doi:10.1016/j.landusepol.2015.10.013
34. Cawson JG, Hemming V, Ackland A, Anderson W, Bowman D, Bradstock R, et al. Exploring the Key Drivers of forest Flammability in Wet Eucalypt Forests Using Expert-Derived Conceptual Models. *Landscape Ecol* (2020) 35:1775–98. doi:10.1007/s10980-020-01055-z
35. Wood MD, Bostrom A, Bridges T, and Linkov I. Cognitive Mapping Tools: Review and Risk Management Needs. *Risk Anal* (2012) 32(8):1333–48. doi:10.1111/j.1539-6924.2011.01767.x
36. Phillips LD. A Theory of Requisite Decision Models. *Acta Psychologica* (1984) 56:29–48. doi:10.1016/0001-6918(84)90005-2
37. French S, Maule J, and Papamichail N. *Decision Behaviour, Analysis and Support*. Cambridge: Cambridge University Press (2009) doi:10.1017/cbo9780511609947
38. Moon K, Guerrero AM, Adams VM, Biggs D, Blackman DA, Craven L, et al. Mental Models for Conservation Research and Practice. *Conservation Lett* (2019) 12(3):e12642. doi:10.1111/conl.12642
39. Moon K, and Browne N. A Method to Develop a Shared Qualitative Model of a Complex System. *Conservation Biol* (2020) 35:1039–1050. doi:10.1111/cobi.13632
40. Langan-Fox J, Wirth A, Code S, Langfield-Smith K, and Wirth A. Analyzing Shared and Team Mental Models. *Int J Ind Ergon* (2008) 28(2):99–112.
41. Vassilides JM, and Jensen OP. Fuzzy Cognitive Mapping in Support of Integrated Ecosystem Assessments: Developing a Shared Conceptual Model Among Stakeholders. *J Environ Manage* (2016) 166:348–56. doi:10.1016/j.jenvman.2015.10.038
42. Markkiczy L, and Goldberg J. A Method for Eliciting and Comparing Causal Maps. *J Manag* (1995) 21:305–33. doi:10.1177/014920639502100207
43. Scavarda AJ, Bouzdine-Chameeva T, Goldstein SM, Hays JM, and Hill AV. A Methodology for Constructing Collective Causal Maps\*. *Decis Sci* (2006) 37(2): 263–83. doi:10.1111/j.1540-5915.2006.00124.x
44. Brown LD, Call AC, Clement MB, and Sharp NY. Inside the "Black Box" of Sell-Side Financial Analysts. *J Account Res* (2015) 53(1):1–47. doi:10.1111/1475-679x.12067
45. Hobbs J, and Singh V. A Comparison of Buy-Side and Sell-Side Analysts. *Rev Financial Econ* (2015) 24:42–51. doi:10.1016/j.rfe.2014.12.004
46. Merkley K, Michaely R, and Pacelli J. Cultural Diversity on Wall Street: Evidence from Consensus Earnings Forecasts. *J Account Econ* (2020) 70(1): 101330. doi:10.1016/j.jacceco.2020.101330
47. Box GEP. Science and Statistics. *J Am Stat Assoc* (1976) 71:791–9. doi:10.1080/01621459.1976.10480949
48. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for Investigating Causality of Sequence Variants in Human Disease. *Nature* (2014) 508:469–76. doi:10.1038/nature13127
49. Yarkoni T. *The Generalizability Crisis*. PsyArXiv. Advance online publication (2019) [Preprint].
50. Kemp-Benedict E. From Narrative to Number: a Role for Quantitative Models in Scenario Analysis. *Int Congress Environ Model Softw* (2004) 22, Available from: <https://scholarsarchive.byu.edu/iemssconference/2004/all/22> (Accessed May 31, 2021).
51. Kosko B Fuzzy Cognitive Maps. *Int J Man-Machine Stud* (1986) 24:65–75. doi:10.1016/s0020-7373(86)80040-2
52. Korb KB, and Nicholson AE. *Bayesian Artificial Intelligence*. CRC Press (2010) doi:10.1201/b10391
53. Forio MAE, Landuyt D, Bennetsen E, Lock K, Nguyen THT, Ambarita MND, et al. Bayesian Belief Network Models to Analyse and Predict Ecological Water Quality in Rivers. *Ecol Model* (2015) 312:222–38. doi:10.1016/j.ecolmodel.2015.05.025
54. Voinov A, and Bousquet F. Modelling with Stakeholders☆. *Environ Model Softw* (2010) 25:1268–81. doi:10.1016/j.envsoft.2010.03.007
55. Meinherz F, and Videira N. Integrating Qualitative and Quantitative Methods in Participatory Modeling to Elicit Behavioral Drivers in Environmental Dilemmas: the Case of Air Pollution in Talca, Chile. *Environ Manage* (2018) 62:260–76. doi:10.1007/s00267-018-1034-5
56. Shea K, Runge MC, Pannell D, Probert WJM, Li S-L, Tildesley M, et al. Harnessing Multiple Models for Outbreak Management. *Science* (2020) 368: 577–9. doi:10.1126/science.abb9934
57. Carlson CJ, Dougherty E, Boots M, Getz W, and Ryan SJ. Consensus and Conflict Among Ecological Forecasts of Zika Virus Outbreaks in the United States. *Sci Rep* (2018) 8:4921. doi:10.1038/s41598-018-22989-0
58. Hinne M, Gronau QF, van den Bergh D, and Wagenmakers E-J. A Conceptual Introduction to Bayesian Model Averaging. *Adv Methods Practices Psychol Sci* (2020) 3:200–15. doi:10.1177/2515245919898657
59. Nicholson AE, Korb KB, Nyberg EP, Wybrow M, Zukerman I, Mascaro S, et al. BARD: A Structured Technique for Group Elicitation of Bayesian Networks to Support Analytic Reasoning. arXiv preprint arXiv:2003.01207. 2020 Mar 2 (2020).
60. Cowell RG, Dawid P, Lauritzen SL, and Spiegelhalter DJ. *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. New York: Springer Science & Business Media (2006).
61. Barons MJ, Wright SK, and Smith JQ. Eliciting Probabilistic Judgements for Integrating Decision Support Systems. *Elicitation* (2018) 445–78. doi:10.1007/978-3-319-65052-4\_17
62. Keeney RL Making Better Decision Makers. *Decis Anal* (2004) 1:193–204. doi:10.1287/deca.1040.0009
63. Kahneman D. *Thinking, Fast and Slow*. Harmondsworth: Penguin (2012).
64. Morgan MG Use (And Abuse) of Expert Elicitation in Support of Decision Making for Public Policy. *Proc Natl Acad Sci* (2014) 111:7176–84. doi:10.1073/pnas.1319946111
65. Bercht AL, and Wijermans N. Mind the Mind: How to Effectively Communicate about Cognition in Social-Ecological Systems Research. *Ambio* (2109) 48:590–604. doi:10.1007/s13280-018-1099-7
66. Gregory R, Failing L, Harstone M, Long G, McDaniels T, and Ohlson D. *Structured Decision Making: A Practical Guide to Environmental Management Choices*. Chichester: John Wiley & Sons (2012). doi:10.1002/9781444398557

**Conflict of Interest:** HL was employed by FTSE Russell, London Stock Exchange Group.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past collaboration with one of the authors SF.

Copyright © 2021 Burgman, Layman and French. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.