When Experts or Models Disagree

Simon French
Manchester Business School,
The University of Manchester
simon.french@mbs.ac.uk

Emma Carter
Manchester Business School,
The University of Manchester

Carmen Niculae
Manchester Business School,
The University of Manchester

ABSTRACT

In managing crises, decision makers are confronted with a plethora of uncertainties. Many arise because the world is uncertain, particularly in the context of a crisis. But some arise because analyses based upon different, but seemingly equivalent models lead to different forecasts. Other times expert advisors differ in their explanations and predictions of the evolving situation. We argue that when handled correctly such conflict can alert the decision makers to the inherent complexity and uncertainty of the situation and improve their management of the crisis.

Keywords

conflict of expert opinion; highly reliable organisations (HRO); mixture of models; risk communication; shared mental models (SMM); the expert problem.

INTRODUCTION

In Harrisburg, less than 10 miles away, the state's new governor struggled with conflicting advice on whether to begin an evacuation that might affect more than 600,000 people.
Washington Post, March 27th, 1999 (*20th Anniversary of Three Mile Island*)

Conflicting advice is one of the banes of decision making. Decision makers (DMs) understandably wish that all the evidence and advice offered them to point unambiguously towards just one course of action: but this seldom is the case. In crisis management this is just as true as elsewhere, but there is less time to resolve conflicts and decide what to do. Moreover, we are all far from the calm, rational, reasoning DMs able to balance and resolve conflicting information that we would like to think ourselves (Bazerman, 2002, Kahneman and Tversky, 2000), particularly in stressful contexts (Flin et al., 1997). Intuitive decision making is guided by various heuristics and prone to many psychological biases. When groups of DMs are involved the dynamics of their interactions can lead to further dysfunctional behaviours including groupthink (Janis, 1982, Janis, 1989). In a crisis, therefore, it is important that the processes, systems and information are structured so that they guide the DMs towards developing a balanced response.

The aim of this paper is to explore the issue of dealing with conflicting advice within a crisis from a number of perspectives and reflect on possible ways forward. Although there are many forms that advice may take, we shall deal with just two broad categories: advice from experts and forecasts from models. This distinction is more convenient than conceptual since experts often base their advice on models and models encode current expert knowledge. The next section provides several examples of conflicting advice and information and the processes currently in place to resolve them. We then turn to three relevant literatures: decision theory and statistics, shared mental models (SMM) and highly reliable organisations (HRO). With this background we offer a number of recommendations, some of which do not sit comfortably with present practices in some emergency management organisations.

While we are concerned primarily with issue of conflicting advice within the emergency management process, there is a related important issue of the danger of providing conflicting advice to the public (Renn and Levine, 1991, Bennett and Calman, 1999, Powell and Leiss, 1997).

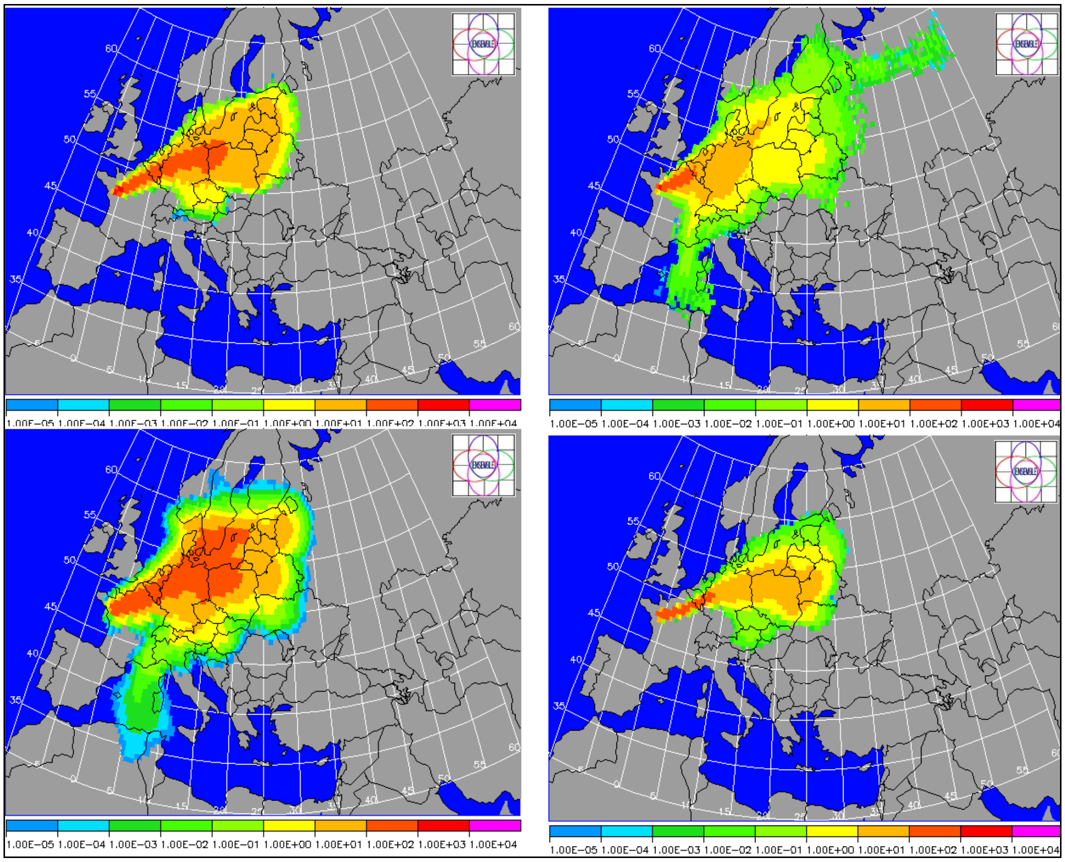

Figure 1: Plots of time integrated concentrations at about 60 hours after release for four different models in Exercise 4.

SOME EXAMPLES: REAL AND CONCEPTUAL

The opening quotation notes that conflicting advice was received by one key DM facing a significant decision about whether to evacuate a substantial population – and that he was not comfortable with this. Moreover, note that the DM was the state's governor, someone who would have little contextual knowledge of nuclear safety and the broad risks faced. More than ever he would need the help and advice of experts. This is typical of crisis management. While there are professionals involved in setting up the process and systems at the outset of an incident, many of the team will have completely different 'day jobs' and have only rehearsed their roles in crisis management infrequently. Moreover, the more senior decision makers have more significant day jobs and may have rehearsed least.

The EU funded ENSEMBLE project explored conflicts between long range atmospheric dispersion models during a nuclear emergency. The project produced a web-based tool in which the differences and similarities between different forecasts could be explored and evaluated (Mikkelsen et al., 2003, Galmarini et al., 2004). During the project several exercises were run in which national meteorological institutes were simultaneously sent details of a hypothetical release (location, source term, time, etc.) and asked to forecast in real-time the long range spread of the plume. Our role was to explore the conflicts between and uncertainties inherent in the forecasts. Whilst in the majority of the exercises there was much agreement between the predictions, in all cases there were differences; in some cases substantial differences (French and Bayley, 2003). Figure 1 illustrates the predicted plume spread 60 hours after a release was supposed to have occurred in France. There are clear differences in terms of easterly and southerly plume travel. In the earlier European Tracer Experiments similar conflicts between model predictions were obtained (Girardi et al., 1998). In a real emergency how would DMs react to such conflicts in forecast?

We have run or observed several workshops in various EU projects to simulate the handling of different phases of a nuclear accident. A common finding in that is that DMs are discomforted by and tend to assume away uncertainties in the advice offered them by experts and model predictions (French et al., 2000, Niculae, 2005). During the sequence of mini-crises that arose during the UK's handling of BSE, politicians continually mishandled or simply ignored conflicting advice (Lord Phillips, 2000).

Within the EU funded EVATECH project, one of our roles was to research the structure of nuclear emergency management processes in four countries: Belgium, Germany, Slovakia and the UK (Carter and French, 2005). Some countries, we discovered, encouraged a diversity of models and advice; others tended to base their analysis on a single approved model for any context and arrange for the expert advice to be drawn together into an unambiguous summary before taking it to the DMs. Clearly there are different cultures and methodologies at work here: but which supports crisis management better?

Suppose you ask an expert for advice on an issue: say a doctor on a significant health problem that you have. He advises a particular treatment. Because of the importance of the issue, you seek advice from a second doctor. She also advises the same treatment. Do you feel more comfortable in taking that advice? Would you, if you discovered that the second doctor trained under the first? Just because the experts agree does not mean that it is straightforward to assimilate their advice.

THEORETICAL APPROACHES

Decision Theory and Statistics

Within the decision theoretic and statistical literatures there are many discussions of how DMs should update their beliefs in the light of expert advice (see,e.g., French and Rios Insua, 2000, Goossens and Kelly, 2000, Clemen and Winkler, 1999) or in the light of model forecasts (see, e.g., West and Harrison, 1989, Pole et al., 1994, Craig et al., 2001). All show that the issue is far from transparent. Consider first the case of conflicting expert judgements.

Experts may disagree for a number of reasons (Mumpower and Stewart, 1996). Most fundamentally, they may differ in their worldviews, believing in fundamentally different theories (e.g. Keynesians versus monetarists or evolutionists versus creationists). They may have access to different data. There may be political or business pressures upon them. As the final example in the last section indicates, the advice from several experts may be correlated through common experiences or training, making it difficult to weight their advice appropriately. Moreover, experts are human and thus subject to various behavioural biases and, typically in the case of experts, overconfidence (Wright and Goodwin, 1998, Wright and Ayton, 1994). All this means that one should not necessary take their predictions and uncertainty bounds at face value. Their judgements should be assessed and calibrated before being used (Cooke, 1991, French, 1985). There are quantitative techniques for calibrating and combining expert judgements (Clemen and Lichtendahl, 2002, Goossens and Kelly, 2000, Clemen and Winkler, 1999), but they are time consuming and far from applicable in crisis management. Typically, they require that a calibration set of judgements is elicited from the experts in order to analyse the quality of their judgements. One might hope that this could be done in advance and kept on record, but the relative calibration of experts may vary over time. In crises, the aggregation of expert advice is usually undertaken behaviourally: i.e. they are encouraged to discuss the situation and reach a consensus. Within decision analysis there are a range of processes aimed at building shared understanding and commitment among groups, e.g., decision conferencing (French, 1988, Goodwin and Wright, 2003). However, in emergency management some balance is needed between pushing for a common view of the crisis and maintaining disagreements so that one monitors events to check which of the disagreeing views seems to be coming to pass.

If one is faced with several models predicting different outcomes then a fairly clear statistical finding is that one should average them in some sense: mixtures of models forecast better than individual models (Draper, 1995, West and Harrison, 1989). This is often perceived as a surprising result by scientists trained in hypothesis testing. Surely one should identify the correct or best performing model and use that? Such a suggestion flies not just in the face of many empirical results to the contrary; it confuses the task of forecasting in risk management with that of explanation in scientific investigation. Scientists are trained not to commit to a theory until accumulated evidence effectively contradicts all others. However when analysing risks in crisis management there is seldom a large body of evidence, thus a mixture of models which keeps open a wide range of possibilities provides more robust predictions: and just as this is true in terms of statistical methods, it is also true in discussing conflicts with experts. There are many methods proposed in the forecasting literature for weighing together a collection of models to provide a 'good' average for prediction that reflects in some sense the overall uncertainty in what may happen. Here we simply draw upon the key point that it is generally better to work with a range of models which span the possibilities, rather than focus down on the predictions of a single model.

Shared Mental Models

An SMM is a common world view, shared by a group of people or management team. Ody (1995) observed that DMs respond not to crisis itself but to their own mental representation of it, and saw that the accuracy of their mental models is dependent on how well DMs can gather information, maintain a flexible impression of the event and continually adapt. 'The identification of an unstructured event, and the management of that event as it creates itself can only be facilitated if the DMs have flexible procedures, and an effective communication system, enabling continual gathering of information, from initial incident conception through to the containment of that threat'. They need to build a family of SMMs, not just of what they believe may be happening, but also of the procedures that they are adopting, the organisational structures they will be working through, the pubic and other stakeholder responses, and so on. Many others consider the concept of an SMM (see, e.g., Flin et al., 1997, Ford and A, 2000, Smith and Dowell, 2000). An SMM enables team members to generate similar expectations about a dynamic situation (Smith and Dowell, 2000). Rouse et al (1992) suggest that SMMs have three functions: description, explanation and prediction. The description function relates knowledge of the system, what it is for, and what it looks like. The explanation function relates to how the system – i.e. here, crisis – works and also interpretations of what it is currently doing. The prediction function creates expectations of how it is likely to develop. When experts disagree, their mental models are, to some extent and often considerable extent, in conflict. The crisis management team must handle this conflict and, in particular, use it to recognise the breadth of uncertainty that is present. We argued above that the DMs should work with the full set of expert advice and model predictions: i.e. we recognise that that there will not be a single SMM and, hence, the DMs should work with a family of SMMs which capture the uncertainty and breadth of possibilities before them. Note that in arguing that a team should work with a family of SMMs, we are not arguing that different members should hold different mental models: rather that each member should appreciate all of the perspectives encapsulated in all of the models.

Highly Reliable Organisations

The literature on HROs (Weick, 1987, Weick and Sutcliffe, 2001) provides further evidence that diversity is important in emergency management and should be encouraged. HROs have *mindful* management of the unexpected, organising themselves so that they are better able to notice the unexpected in the making and halt its development. HROs are very aware of their SMMs and, more importantly, continually monitoring events to check that their perceptions, understandings and expectations cohere with what is actually happening. There do all in their power to minimise the risk of defensive avoidance (Janis and Mann, 1977). They adopt organic management structures, continually directing their attention to the risks that face them. Emergency management needs to embrace this concept of mindful management and monitor the situation for changes that are unexpected. Conflicts in expert advice and model predictions can indicate where such changes might occur. The literature agrees that diversity is an important part of team decision making as different people size problems up differently. An important part of this is that it is the existence of diversity that is important not the essence.

HANDLING CONFLICT IN THE EMERGENCY MANAGEMENT PROCESS

In (French and Niculae, 2005, Niculae et al., 2004) the Cynefin categorisation of decision contexts was applied to emergency management. There it was argued too often that emergency managers and their advisors perceive that the situation is evolving along known lines according to some well understood model. Their single SMM is too restricted and cannot anticipate some of the complexity of the event; they fail to monitor the situation widely enough; and do not adapt their policies quickly enough. They can end up mishandling the physical aspects of the situation and also losing the trust of the public and stakeholders.

Thus any conflict between experts or models or both could be an important signal that things are not as straightforward as the emergency managers imagine and that the uncertainty is greater than they have appreciated. If managers recognise this, they can use the spread of expert and model predictions as a rough guide to some of the uncertainty in the situation and choose countermeasure and remedial policies that are robust in the sense that they cope will with the future however it evolves. In their advice to and dealings with the public, they may be encouraged to become less dogmatic in their assertions about the likely course of events and the manner in which they will bring things under control.

Some of the quantitative methods of eliciting and combining expert judgement and of forming mixtures of models alluded to above may be of help in dealing with the details. Given the time required for such techniques, this may be difficult. We leave that for another paper. Here we argue that that the real benefits of recognising and acknowledging conflict are qualitative. They catalyse the DMs recognition of the inherent uncertainty and complexity of the situation and, hence, of the need to adopt sensitive, robust policies. How should emergency management processes and organisations be structured and supported to enable this response to conflicting advice? Management practices within HROs should be drawn into the planning of emergency management to help support this. Emergency management processes should maintain the breadth of possibilities, recognising that genuine conflict arises from genuine uncertainties, and not shut down possible models of the evolution of the event too early. Thus countries which have built their processes upon the use of a single model in any situation might pause and ask themselves if this is wise. Similarly, they would be unwise to listen to one expert only on any topic. And perhaps one should be concerned by crisis management processes in which a small central group of experts listen to the advice of all the other experts, summarise this into a single coherent advisory report for the DMs and then approach them with this. In doing this, they are essentially falling prey to the scientific imperative of trying to identify a true model rather than the risk management imperative of monitoring all possible hazards. Moreover, DMs and their advisors should be wary of giving too much credence to the agreement between several experts if they share a common background, experience and education. One would expect their advice to be correlated and so should not give it too much weight just because it is. Technically we believe that the decision support systems used should include collaborative tools to enable debate between all parties as models are analysed and their import understood (French and Niculae, 2005, French et al., 2005).

Above all we believe that there is a need for a change of culture in which conflicting advice is recognised for its value as a signal of inherent uncertainty and not as an unwelcome eventuality. Such a culture change will not be easy to achieve. It may require a different form of planning exercise in which events are not assumed to evolve according to some model, but a much more uncertain environment is simulated.

ACKNOWLEDGMENTS

We are grateful to many colleagues for discussions that have informed these reflections. Much of this work stems from my involvement in several projects on nuclear emergency management funded under the EU Framework Programmes. In particular, many of the ideas were developed during the EVATECH project (EU Contract: FIKR-CT-2001-00193).

REFERENCES

1. Bazerman, M. (2002) *Managerial Decision Making,* John Wiley and Sons, New York.

2. Bennett, P. G. and Calman, K. C. (Eds.) (1999) *Risk Communication and Public Health: Policy Science and Participation,* Oxford University Press, Oxford.

3. Carter, E. and French, S. (2005) Nuclear emergency management in Europe: a review of approaches to decision making. In *ISCRAM 2005: Information Systems for Crisis Response and Management*, (Eds, Van de Walle, B. and Carle, B.) Brussels.

4. Clemen, R. T. and Lichtendahl, K. C. (2002) Debiasing expert overconfidence: a Bayesian calibration model. In *PSAM6*, San Juan, Puerto Rico.

5. Clemen, R. T. and Winkler, R. L. (1999) Combining probability distributions from experts in risk analysis *Risk Analysis,* **19**.

6. Cooke, R. (1991) *Experts in Uncertainty,* Oxford University Press, Oxford.

7. Craig, P. S., Goldstein, M., Rougier, J. C. and Seheult, A. H. (2001) Bayesian forecasting for complex systems using computer simulators *Journal of the American Statistical Association,* **96,** 717-729.

8. Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion) *Journal of the Royal Statistical Society,* **B57,** 45-97.

9. Flin, R., Salas, E., Strub, M. and Martin, L. (1997) *Decision Making under Stress,* Ashgate, Aldershot.

10. Ford, K. and A, S. (2000) Emergency response training: strategies for enhancing real-world performance *Journal of Hazardous Materials,* **75,** 195-215.

11. French, S. (1985) Group consensus probability distributions: a critical survey. In *Bayesian Statistics 2* (Eds, Bernardo, J. M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M.) North-Holland, pp. 183-201.

12. French, S. (Ed.) (1988) *Readings in Decision Analysis,* Chapman and Hall, London.

13. French, S., Bartzis, J., Ehrhardt, J., Lochard, J., Morrey, M., Papamichail, K. N., Sinkko, K. and Sohier, A. (2000) RODOS: Decision support for nuclear emergencies. In *Recent Developments and Applications in Decision Making* (Eds, Zanakis, S. H., Doukidis, G. and Zopounidis, G.) Kluwer Academic Publishers, Dordrecht, pp. 379-394.

14. French, S. and Bayley, C. (2003) An analysis of the ENSEMBLE exercises from the perspective of decision making. In *ENSEMBLE Project Report*, Manchester Business School, Booth Street West, Manchester, M15 6PB.

15. French, S., Carter, E. and Niculae, C. (2005) Decision Support in Nuclear and Radiological Emergency Situations: Are we too focused on models and technology? *International Journal of Risk Assessment and Management,* **(in press)**.

16. French, S. and Niculae, C. (2005) Believe in the Model: Mishandle the Emergency *Journal of Homeland Security and Emergency Management,* **2**.

17. French, S. and Rios Insua, D. (2000) *Statistical Decision Theory,* Arnold, London.

18. Galmarini, S., Bianconi, R., Klug, W., Mikkelsen, T., Addis, R., Andronopoulos, S., Astrup, P., Baklanov, A., Bartniki, J., Bartzis, J., Bellasio, R., Bompay, F., Buckley, R., Bouzom, M., Champion, H., D'Amours, R., Davakis, E., Eleveld, H., Geertsema, G. T., Glaab, H., Kollax, M., Ilvonen, M., A, M., Pechinger, U., Persson, C., Polreich, E., Ptempski, S., Prodanova, M., Saltbones, J., Slaper, H., Sofiev, M. A., Syrakov, D., Sorenson, J. H., Van der Auwera, L., Valkama, I. and Zelazny, R. (2004) Can the confidence in long range atmospheric transport models be increased? The pan-European experience of ENSEMBLE. *Radiation Protection Dosimetry,* **109,** 19-24.

19. Girardi, F., Graziani, G., van Velzen, D., Mosca, S., Bianconi, R., Bellasio, R., Klug, W. and Fraser, G. (1998) *The European Tracer Experiment,* Office for Official Publications of the European Communities., Luxembourg.

20. Goodwin, P. and Wright, G. (2003) *Decision Analysis for Management Judgement,* John Wiley and Sons, Chichester.

21. Goossens, L. H. J. and Kelly, G. N. (2000) Special: Issue: Expert Judgement and Accident Consequence Uncertainty Analysis *Radiation Protection Dosimetry,* **90,** 293-381.

22. Janis, I. L. (1982) *Groupthink: Psychological Studies of Policy Decisions and Fiascos,* Houghton Mifflin, Boston.

23. Janis, I. L. (1989) *Crucial Decisions - Leadership in Policy Making and Crisis Management,* The Free Press.

24. Janis, I. L. and Mann, L. (1977) *Decision Making: a Psychological Analysis of Conflict, Choice and Commitment,* Free Press, New York.

25. Kahneman, D. and Tversky, A. (Eds.) (2000) *Choices, Values and Frames,* Cambridge University Press, Cambridge.

26. Lord Phillips (2000) The Phillips Report on the BSE Crisis. HMSO, London.

27. Mikkelsen, T., Galmarini, S., Bianconi, R. and French, S. (2003) ENSEMBLE: methods to reconcile disparate national long-range dispersion forecasts. RISO, Roskilde, Denmark.

28. Mumpower, J. L. and Stewart, T. R. (1996) Expert judgemen and expert disagreement *Thinking and Reasoning,* **2,** 191-211.

29. Niculae, C. (2005) A socio-technical perspective on the use of RODOS in nuclear emergency management. PhD Thesis, Manchester Business School, The University of Manchester

30. Niculae, C., French, S. and Carter, E. (2004) Emergency Management: Does it have a sufficiently comprehensive understanding of decision-making, process and context? *Radiation Protection Dosimetry,* **109,** 97-100.

31. Ody, K. (1995) Facilitating the "right" decision in crisis: Supporting the crisis decision maker through analysis of their needs *Safety Science,* **20,** 125-133.

32. Pole, A., West, M. and Harrison, J. (1994) *Applied Bayesian Forecasting and Time Series Analysis,* Chapman and Hall, London.

33. Powell, D. and Leiss, W. (1997) *Mad Cows and Mother's Milk: The Perils of Poor Risk Communication,* Queens University Press.

34. Renn, O. and Levine, D. (1991) Credibility and trust in risk communication. In *Communicating Risks to the Public The Netherlands* (Eds, Kasperson, R. E. and Stallen, P. J. M.) Kluwer Academic Publishers, The Hague, pp. 175-218.

35. Rouse, W. B., Cannon-Bowers, J. A. and Salas, E. (1992) The role of mental models in team performance in complex systems *IEEE Transactions on Systems Man and Cybernetics,* **SMC-22,** 1226.

36. Smith, W. and Dowell, J. (2000) A case study of co-ordinative decision making in disaster management *Ergonomics,* **43,** 1153-1166.

37. Weick, K. E. (1987) Organisational culture as a source of high reliability *California Management Review,* **29,** 112-127.

38. Weick, K. E. and Sutcliffe, K. (2001) *Managing the Unexpected: Assuring High Performance in an Age of Complexity,* Jossey Bass, San Francisco.

39. West, M. and Harrison, J. (1989) *Bayesian Forecasting and Dynamic Models,* Springer Verlag, New York.

40. Wright, G. and Ayton, P. (Eds.) (1994) *Subjective Probability,* John Wiley and Sons, Chichester.

41. Wright, G. and Goodwin, P. (Eds.) (1998) *Forecasting with Judgement,* John Wiley and Sons, Chichester.