# Combining Expert Judgement: A Review for Decision Makers

**Simon French**

Manchester Business School  
Manchester M15 6PB, UK  
simon.french@mbs.ac.uk

May 2009

Abstract

In a paper written over twenty years ago, I distinguished three contexts in which one might wish to combine expert judgements of uncertainty: the expert problem, the group decision problem and the text-book problem. Over the intervening years much has been written on the first two, which have the focus of a single decision context, but little on the third, though the closely related field of meta-analysis has developed considerably. With many developments in internet technology, particularly in relation to interactivity and communication, the text-book problem is gaining in importance since data and expert judgements can be made available over the web to be used by many different individuals to shape their own beliefs in many different contexts. Moreover, applications such as web-based decision support, *e*‑participation and *e*‑democracy are making algorithmic ‘solutions’ to the group decision problem attractive, but we know that such solutions are, at best, rare and, at worst, illusory.

In this paper I survey developments since my earlier paper, and then I turn to how expert judgement might be used within web-based group decision support, as well as in *e*‑participation and *e*‑democracy contexts. The latter points to a growing importance of the textbook problem and suggests that Cooke’s principles for scientific reporting of expert judgement studies may need enhancing for such studies to be used by a wider audience.

**Keywords:** *e*‑democracy; *e*‑participation; expertproblem; group decision problem; meta-analysis; text book problem; web-based group decision support systems (*w*GDSS).

## Introduction

Decision makers have long relied on expert judgement to inform their decision making, and there is a long literature relating to the incorporation of such advice into formal decision analyses. A key issue is not just how to assimilate the advice of one expert, but also how to draw together and learn from the advice of several, particularly when there is conflict in their views. At the second Valencia Conference on Bayesian Statistics, I reviewed and categorised Bayesian approaches to the combination of expert judgement (French 1985). There are many situations in which one might need to draw on the opinions of a group of experts. In that paper I considered three broad contexts:

*The expert problem.* In this the group of experts are asked for advice by a decision maker (dm) who faces a specific real decision problem. The dm is, or can be taken to be outside the group, and it is she[[1]](#footnote-1) who has the responsibility and accountability for the consequences of the decision. The experts are free from such responsibility and accountability. In this context the emphasis is on the dm learning from the experts.

*The group decision problem*. Here there is still the focus of a predefined problem; but it the group itself that is jointly responsible accountable for the decision. The dms are their own experts. They may, indeed almost certainly will, wish that to the outside world their decision appears rational and, possibly, also fair and democratic. Thus they may wish to combine their judgements in some formal structured way.

*The text-book problem*. The group may simply be required to give their judgements for others to use in the future in as yet undefined circumstances. Here there is no predefined decision problem, but many potential ones that are only vaguely perceived.

In all cases we assume that the experts are asked for their opinion on the *science* of the situation, i.e. how likely something is to happen. They are not asked for value judgements (French *et al.* 2009). We also assume that they provide their judgements as probabilities. A recent review of practical protocols for elicitation of expert probability judgements is provided by O’Hagan *et al* (2006).

It is possible, indeed likely that some combination of the three contexts may occur. For instance, a group of DMs might be informed by a group of experts before making their decision. We might view such problems as a two stage problem in which each DM listens the experts and updates her probabilities in the light of their opinions and then DMs act as a group coming to a decision (but see below).

The first two contexts have the focus of a specific decision problem; the third does not. This has the effect of making discussion of the expert and group decision problems more straightforward. There is structure against which possible approaches to their resolution can be judged – which does not mean that their resolution is similar. What may an appropriate approach to the expert problem may be much less suited to the group problem. I explored such matters in the 1985 paper; but I said little on the text-book problem, perhaps because its lack of structure made it somewhat difficult to address. Since 1985 there has been many developments in approaches to the expert and group decision problems (for reviews see, inter alia, Clemen and Winkler 1999; Cooke 1991; French and Rios Insua 2000; Garthwaite *et al.* 2005; Goossens and Kelly 2000; Jacobs 1995; O'Hagan *et al.* 2006). However, little further has been said on the text-book problem. Perhaps I gave a non-problem a name – certainly the discussion at Valencia raised this as a possibility – but perhaps not. Current developments in web-based public participation and deliberative democracy, particularly in the area of societal risk management, are giving the topic some importance. Moreover, there are strong connections with meta-analysis (Hartung *et al.* 2008; Morris and Normand 1992; Sutton and Abrams 2001), a family of methodologies that have grown in importance over the past quarter century.

In this paper I briefly survey developments since my earlier paper, and then turn to how expert judgement might be used within decision support, particularly web-based group decision support systems (*w*gdss), as well as in *e*‑participation and *e*‑democracy contexts. There are many issues related to the use of expert judgement in such contexts. The use of information and communication technology (ict) to provide the infrastructure for the interactions between participants favours algorithmic approaches to the combination of judgements in place of more behavioural approaches; yet we know that the former are fraught with difficulty, if not intrinsically flawed beyond redemption. Moreover, the web provides the potential for a much wider public to access expert reports and brings the need to address issues related to the text-book problem to the fore. I close by arguing for the need for more research at the interface between meta-analysis, formal theories of the use of expert judgement and public participation in democratic processes.

Throughout my discussion I lean strongly towards the Bayesian paradigm of decision analysis (French and Rios Insua 2000).

## The expert problem: a brief survey of developments

Expert opinions can be combined either by eliciting them individually and explicitly and then applying some mathematical or algorithmic aggregation rule or by allowing the group to discuss issues and through discussion agree a consensus view; the latter being known as behavioural aggregation. In my 1985 paper, I explored a number of approaches to mathematical aggregation for the expert problem, but only two strands have really stood the test of time: the Bayesian and opinion pooling. Bayesian approaches treat the experts’ judgements as *data* for the dm and then seek to develop appropriate likelihood functions to represent the dm’s relative confidence in the experts. The difficulty is in developing tractable probability models that capture the dm’s understanding of:

* the ability of the experts to encode their knowledge probabilistically (O'Hagan *et al.* 2006; Wright and Ayton 1994);
* the correlation between expert judgements that arises from their shared knowledge and common professional backgrounds (Booker and Meyer 1988; French 1981; Mumpower and Stewart 1996; Shanteau 1995);
* the correlation between the experts’ judgements and the dm’s own judgements (French 1980);
* the effects of other biasing pressures such as may arise from conflicts of interests, fear of being an ‘outlier’, concern about future accountabilities, and more general psychological ‘biases’ and emotional and cultural responses to context (French *et al.* 2009; Gigerenzer 2002; Gilovich *et al.* 2002; Hockey *et al.* 2000; Kahneman *et al.* 1982).

Providing experts with training in encoding uncertainty probabilistically can be beneficial and may reduce the need for complex likelihoods to allow for their ‘errors’ in the judgement process (Cooke and Goossens 2004), but it does not address the other issues. Moreover, when several uncertainties are elicited, there is a need to elicit the dependency structure between the uncertain quantities. Eliciting correlations is difficult and little is known about the sorts of errors to which experts are prone in providing such judgements. One obvious and oft used approach is to elicit belief nets, influence diagrams or other graphical structures which give form to the dependencies from the experts’ scientific understanding of the context (Clemen and Reilly 1999; Clemen and Winkler 1993; Kurowicka and Cooke 2006; Renooij 2001).

There have been several practical studies which use Bayesian methods to combine expert judgement (e.g. Szwed and van Dorp 2002; Wiper and French 1995), but despite indications of reasonable performance, few address all the issues of calibration and correlation mentioned above. One is left with a feeling that the underlying models omit many of the factors that need to be included. Thus the Bayesian approach, although extremely strong conceptually, remains impractical because of the difficulty in developing the likelihood function.

Opinion pools, on the other hand, take a much more pragmatic approach: they simply weight together the experts’ judgements, taking them at least intuitively, if not conceptually as probabilities in their own right. The process may use an arithmetic or geometrically weighted mean or perhaps something rather more generalised (French 1985; Genest and Zidek 1986). In their ‘vanilla’ form, the weights in an opinion pools are often simply given by the dm based on her judgement of the experts’ relative expertise, seldom with any operational meaning being offered for the concept of ‘relative expertise’. Alternatively, the weights may be taken as equal, perhaps on some Laplacian Principle of Indifference, or of equity, or, even, that all the experts are paid the same fee[[2]](#footnote-2). Recently, it has been suggested that the weights might be defined from the social network which surrounds the experts, one operationalisation being the relative frequency with which their work is cited in the literature (Cooke *et al.* 2007).

Since the mid 1980’s Cooke and his co-workers at TU Delft have developed the *Classical* approach. This defines the weights empirically on the basis of the experts’ relative performance on a calibration set of uncertain quantities, the resolution of which are unknown to the experts, but are or become known to the analyst and dm by the time they form the opinion pool on the quantities of interest (Bedford and Cooke 2001; Cooke 1991). Arguably, Cooke’s Classical approach is the most applied of all expert judgement methods: for case studies, see (Cooke 2007; Goossens and Kelly 2000). Recently, a database of their studies, some 45 in number, has been created (Cooke and Goossens 2007). This offers a real opportunity for careful empirical evaluation of their methods as well as comparative assessments of other approaches[[3]](#footnote-3).

The choice of quantities to include in the calibration set is far from easy. Behavioural studies have shown that an individual’s ability to encode judgments in probabilistic terms varies according his or her expertise in that domain. Typically, but far from universally, individuals become better at encoding their uncertainties the more feedback they have had on similar judgements in the past (Cooke 1991; Lichtenstein *et al.* 1982; Suantak *et al.* 1996). This tends to mean that they are better at encoding uncertainty in their domain of expertise, because they tend to receive more feedback in areas in which they work regularly. Thus the calibration set has to be chosen to reflect the domain that covers the variables which are important in the real problem. Unfortunately, the DM and analyst are more likely to seek expert advice on in precisely those areas in which there are few past data − otherwise, the analyst would analyse those data and thus construct statistical forecasts. Nonetheless, with care and some subtlety, experience has shown that Cooke’s method is applicable.

There have been suggestions that both Bayesian models and opinion should be constrained to satisfy various axiomatic principles which are supposed to encode rational behaviour in learning from others opinions. For instance, should they be externally Bayesian (Madansky 1964); i.e. should the result of assimilating the information in external data depend on whether the DM allows the experts to do this individually and then combines their judgements or whether the DM first combines their prior judgements then updates the combination using the external data via Bayes’ Theorem? However, the ‘rationality’ arguments which motivate such axioms are seldom entirely persuasive (French 1985; French and Rios Insua 2000; Genest and Zidek 1986) and their result can be to limit the distributional forms so much as to make some of the other modelling requirements if not intractable, then certainly uncomfortable (Faria and Smith 1997). One axiomatic principle which has enormous *prime facie* appeal, but which does not stand close inspection is that of independence preservation (French 1987; Genest and Wagner 1987). Suppose all the experts and the DM judge a series of random quantities to be probabilistically independent. Surely the combination of their judgements should preserve this independence? Yet a little thought shows that to assume this would be to deny the possibility of re-calibrating the experts’ judgements. Independence of random quantities implies that learning the values of any does not affect the DM’s beliefs about the remaining quantities. But this is precisely contrary to the concept of a calibration set. The DM and her analyst observe the performance of the experts on a ‘known set of unknowns’ to gather information on the experts’ calibration and then use this to modify their combination of the experts’ judgements on the quantities of real interest. The essence of this somewhat paradoxical point is that while the experts are considering a series of random quantities, the DM is considering a system comprising *both* those random quantities *and* the experts’ judgements. The context for the DM is much more complex than assumed in the simplistic statement of the independence preservation axiom. And that is the nub of the issue with all such axiomatic principles: they oversimplify a very complex context.

Both the Bayesian approach and opinion pools represent mathematical approaches to aggregation of expert judgement; there are also behavioural approaches. The analyst could gather the experts together and let them come to some consensus agreement about the numbers to put into the models. He might allow them to do this in totally free discussion, in a more facilitated and structured, but still free-flowing discussion, or in a more formal process. Among the more structured processes, perhaps the *Delphi Technique* is the most well-known and commonly used (Dalkey and Helmer 1963; Linstone and Turoff 1978), though it is not without its critics (e.g., Cooke 1991). Rowe and Wright (1999) provide a critical evaluation of the technique and its performance, noting that in practice, there are many versions of the Delphi process and their appropriateness and effectiveness will vary with the precise circumstances of their use.

There are many pros and cons to be balanced in choosing between mathematical and behavioural aggregation. Behavioural aggregation can be subject to many group dysfunctional behaviours (French *et al.* 2009). On the other hand, behavioural aggregation can build commitment to the eventual decision and, hence, encourage effective implementation. If decision making within an organisation is the prime objective of the process, this can be a powerful argument in its favour. For a planning or a regulatory decision with a wide variety of stakeholders, mathematical aggregation can be an advantage, both because it is explicit, auditable and, in a sense, objective; and also because it leaves all opinions and eventualities in the analysis. Moreover, behavioural aggregation can win people over and thus concerns about hazards held by a small minority of the group of experts may be lost in building a common group view. In terms of empirical performance there is some evidence that mathematical aggregation can outperform behavioural approaches (Clemen and Winkler 1999).

There is, of course, the question whether a group can perform better than its best performing member. Should the analyst combine expert judgements or simply seek to identify the most accurate? Cooke’s experiences suggest that combining judgements is the more effective, though his method does ‘discard’ the poorest performers (Cooke 2007). Moreover, there is much general evidence that combining forecasting models is more effective than selecting the best (Draper 1995). For behavioural methods, the results tend to be inconclusive. Reagan-Cirincoine (1994) reported an experiment in which facilitated groups generally outperformed their best members in forecasting tasks, but also summarise a literature in which the reverse finding seems to be true. An earlier survey of the literature was provided by Hastie (1986).

Discussions of protocols which implement elicitation and combination of expert judgements in practice may be found in, *inter alia*, Bedford et al (2006), Cooke and Probst (2006), Cooke and Goossens (2004; 2007), Goossens and Kelly (2000) and O’Hagan *et al* (2006).

## Group decision making

One of the most startling things about Arrow's Impossibility Theorem (Arrow 1963) is that it is so much ignored. It and the many variants proved subsequently in many respects deny the possibility that democracy can ever really exist (Bacharach 1975; Hodge and Klima 2005; Kelly 1978). Yet among the research communities working on supporting decision making, be it at the group or societal level, few address the very real issues that Arrow raised. Essentially, the conclusion of his and related work is that *any* algorithmic method of combining the judgements of a group of dms is subject to paradox and inconsistency (French 1985; 2006). Moreover, if one considers the possibility of agenda rigging, manipulability and straightforward, game-playing dishonesty, then one despairs of achieving a valid combination of judgements of uncertainty in the context of group decision making simply by algorithmic means. Rather one is led to the view that the idea of a group decision is a flawed concept. Much in the same sense that DeFinetti (1974) claimed that “Probability does not exist!”, so group decisions do not exist. Groups do not decide: individuals do. Beliefs, values, intentions all reside in individuals, not groups. Groups are better viewed as social processes which translate the individual decisions of the group members – their ‘votes’ – into a course of action (Dryzek and List 2003; French 1981; 1985; French *et al.* 2009).

Because of this, most decision analysts moved away from algorithmic approaches to group decision making during the 1980s. They did not seek to combine individual subjective probabilities and utilities to form group consensus probabilities and utilities and then rank alternatives through a group expected utility. Instead they adopted behavioural approaches based around something like a decision conference (Eden and Radford 1990; French *et al.* 2009; Phillips 1984). In such events they worked with the group discussing issues and moving towards an agreed decision. Certainly, they used models, but the models did not claim to represent consensus. Rather they were a reference analysis against which differences in opinion could be explored through sensitivity analysis (French 2003). Indeed, decision analysis was – and is – more often used to communicate than to structure a ‘rational’ decision. The process was behavioural and the final decision taken by agreement, often non-verbalised agreement in which pragmatism and politics might shade the outcome. Specifically, the process would not form some explicit agreed judgement of uncertainty represented by some consensus probability distribution. Instead the implications of different levels of uncertainty would have been explored through sensitivity analysis before the group agreed a course of action. French (2006) charts the development of use of the Bayesian paradigm to support first individual decision makers, then groups of decision makers and also makes the point that the same issues arise whatever paradigm of rationality, Bayesian or non-Bayesian, is adopted.

The tremendous growth and accessibility of the internet over the past decade has in many ways reversed this trend towards behavioural aggregation. There has been a resurgence in algorithmic approaches to guide group decisions (see, e.g., Karacapilidis and Pappis 1997; Koning 2003; Morton *et al.* 2003), perhaps because of:

* the attraction of being able to use *w*gdss to involve dms asynchronously in the deliberation without the need for a meeting;
* the use of computers to access the internet simultaneously provides the means to implement algorithms.

This resurgence is, of course, extremely attractive to software developers. The technical issues in building such *w*gdss are relatively small. Generally, optimisation algorithms work and the systems point to a strategy as the best, whether or not it would stand scrutiny from a decision theoretic perspective. Most dms are unfamiliar any of the conceptual problems of group decision making discussed in – to them – an esoteric literature. So the systems will sell and be used unless we find effective ways to communicate their potentially flawed working. The development of *e*‑participation and *e*‑democracy may give this further impetus as efforts to create greater engagement of the public and stakeholders in societal decision making deploy *w*gdss with deliberative democratic contexts (French *et al.* 2006). We discuss this further below.

Interesting questions can arise when one asked how a group of DMs should be advised by a group of experts. A version of external Bayesianity (Madansky 1964) might suggest that the process should achieve the same result whether each DM is advised by the group of experts separately and then interacts with the other DMs; or whether all the DMs are advised together by the experts. But too many other issues confound the matter to follow this through. Trivially, if one takes the import of Arrow’s theorem to heart, one cannot aggregate the DMs’ judgements mathematically; so there is no question of whether mathematical aggregation of expert opinion commutes with mathematical aggregation of the DMs’ judgements. From a behavioural perspective of facilitating the process, there is much to be said for gathering all the experts and DMs together and discussing the issues in plenary. It helps ensure that all speak a common language and understand the issues in the same way. But it also changes the group dynamics and even though the experts may be intended to be advisory on matters of science only, they can sway the debate on matters of value judgement through their presence. Logistics and perhaps legal governance issues may prohibit a gathering of all together. Perhaps the experts should be brought in one at a time to the meeting of DMs? Or perhaps one should just use mathematical aggregation and report the result to the DMs, and then leave them to decide? I am unaware of any comparative studies which seek to explore these issues.

## Textbook problem

In my 1985 paper I said little on the text-book problem. Of course, the whole discipline of statistics can be seen as the science and art of displaying evidence so that others may learn from it to inform their future activities. The foundational work of writers such as Jeffreys (1961) and Savage (1972) introduces debates on how a subjective view of statistics can be reconciled with the sharing of knowledge across Science and the development of consensus. Such work has developed into discussions of topics such as Bayesian conversations (DeGroot and Bayarri 1991; Kadane *et al.* 1999; Lehmann and Goodman 2000), evidence based medicine and policy (Ashby and Smith 2000), robustness and sensitivity analysis (French 2003; Leamer 1978; Rios Insua and Ruggeri 2000), summarization and data presentation (Cleveland 1994; O'Hagan and Forester 2004), and, above all, meta-analysis (Hartung *et al.* 2008; Morris and Normand 1992; Sutton and Abrams 2001). However, such statistical approaches have pretty much focused on evidence arising from data. There has been little work on how to draw together studies built upon expert judgement. Perhaps the solution to the text-book problem might be termed meta-analysis of expert judgement studies; but such methods would need to be developed.

To a large extent, the text book problem has remained unaddressed because it simply has not been a problem: at least not in the sense of combination of judgements. Consider. The context is that a group of experts are charge with writing a report on an issue and they offer their probabilities for the uncertainties in their conclusions. At the time they have no knowledge of any specific decision that their conclusion might inform. For instance, working parties were set up by the British and American governments a few years back to consider how likely the Earth was to be destroyed by an asteroid. Although one can imagine a plethora of decisions that the information might support from societal protection to the personal and religious, there was no single specific decision identified for the reports to support. For many a year such expert reports have been commissioned, although only recently has it been common to quantify the uncertainties in their conclusions. DMs have used such reports to inform the issue and problem formulation phase of their decision making (Franco *et al.* 2006; 2007; French *et al.* 2009). Mostly such decision making has taken place in small groups with the DMs being advised by further experts who read and summarise the reports forming their own probabilities[[4]](#footnote-4). Thus the problem has moved to an example of the expert problem discussed above. In so far as the text-book problem itself has been concerned, general good professional practice indicated that, along with their judgements of uncertainty, the experts should publish whatever objective evidence and data they did have so that others could form their own judgements independently.

Cooke has suggested four principles that should underpin the publication expert judgement data, originally for the expert problem, but applicable to the text-book problem too (Cooke 1991; Cooke and Goossens 2007).

Cooke’s principles

* *Scrutability/Accountability*: all data, including experts’ names and assessments, and all processing tools are open to peer review and results must be reproducible by competent reviewers.
* *Empirical Control*: quantitative expert assessments are subject to empirical quality controls.
* *Neutrality*: The method for combining and evaluating expert opinion should encourage experts to state their true opinions, and must not bias results.
* *Fairness*: Experts are not prejudged, prior to processing the results of their assessments.

Essentially, these principles seek to assure that the study is auditable, open to peer review and hence ‘scientific’. These principles are not uncontroversial. One may wonder about *Fairness*. Experts are by definition prejudged, if only by virtue of their being dubbed ‘experts’. So unless one accepts that society is divided into two homogenous groups of the lay public and the experts, respectively, there may be some reason to weight the experts unequally *a priori*. A Bayesian would surely believe this. Moreover, Cooke does so implicitly in his exploration of social networking as a basis for the weights (Cooke *et al.* 2007). *Neutrality* is persuasive, and Cooke uses it to justify his classical weighting scheme on the basis of scoring rules (Cooke 1991). In these days when there is a pleasing exhortation at all political and managerial levels to use evidence-based decision making, there can be few qualms about *Empirical Control*. However *Scrutability/Accountability* has always been an issue. In most circumstances experts insist on anonymity before they will take part in a study. In many legal systems they are concerned about the implicit liability they bear for any advice they give. At the very least, instead of full anonymity they expect *Chatham House Rules* to apply: what is said may be reported but not who said it. Nonetheless, to a large extent these principles have informed the production of reports of many expert judgement studies: see, e.g., the USNRC/EU accident consequence studies referred to in Goossens and Kelly (2000).

In general, however, there has been little discussion of the text-book problem *per se* over the years. Indeed, there has been less discussion of Cooke’s principles than one might expect. But the advent of greater public participation in societal decisions and risk governance may change that. Across Europe and indeed globally, there is a general agreement that societal issues would be more effectively and acceptably handled if there were a greater element of involvement and integration of the perspectives of *all* stakeholders (Beierle and Cayford 2002; French *et al.* 2007; Gregory *et al.* 2005; Renn 1998; 2008; Renn *et al.* 1993; Renn *et al.* 1995)[[5]](#footnote-5). Participation of citizens in the assessment, analysis and management of the challenges and risks that face society can seemingly avoid some of the issues that have arisen in the past handling of such matters. But the involvement of the lay public in such deliberations will involve conveying to them the import of past studies, both empirical and expert, and that brings to the fore the need to address the text-book problem.

Consider specifically societal risk governance and risk management. Wider citizen and stakeholder involvement brings several challenges.

There is a need to convey the quantitative advice of scientific experts alongside their more qualitative advice. There are many issues relating to the public communication of science and of presenting evidence on societal risks (Bennett and Calman 1999; Department of Health 1998; Gigerenzer 2002; Leach *et al.* 2005; Maule 2004; 2008; Slovic 2001). The majority of the work in this area relates to the communication of simple probabilities and risks: the more difficult, yet more relevant issue in most risk analyses of conveying distributions and interactions or correlations remains much less well addressed.

Scientists and engineers, drawing information from expert judgement studies, are familiar with scientific conflicts. They expect some of the experts to differ in their judgements. But it is known that the lay public tend to discomforted by disagreements and conflicts between experts (Department of Health 1998). Thus how an expert judgement study chooses to present differing expert opinions and the overall combination of these could either inform or discomfort the public.

If, as is likely, in a wide ranging public debate one has to draw together several studies, how should one do this? The development of meta-analysis methodologies has, by and large, focused on studies based on empirical data. Expert judgement studies have a very different character to published statistical ones. Most fundamentally, perhaps, empirically-based studies are seldom published unless the results are statistically significant, introducing a bias that is the bane of meta-analysis. Such is not the case for studies based on expert judgement. Generally they are published whatever the result. Some of the meta-analytic theory on learning form several opinion polls may be relevant, but broadly there is no meta-analysis methodology that is appropriate to this context; and much more work is needed.

A common finding of including the stakeholders and public in the full process of analysis and deliberation is that they often do not ask the same questions or show interest in the same technical issues as have traditionally concerned regulators and government departments. This may be particularly important in relation to expert judgement. Thus expert judgement studies may be framed in terms of uncertainties that are not immediately important to lay discussants but silent on ones which are.

Finally, Cooke’s four principles need much greater discussion. If expert judgement studies are to be used to inform wider debate, we need to be sure that they are conducted and reported to the highest possible professional standards. We have noted that he argues that the experts’ names should be published and their personal judgements identified in any report. On grounds of scientific audit, this seems more than reasonable though we have noted that issues of legal liability provide the experts with cogent reasons to demand anonymity. In the context of public participation, there is a further issue. Like it or not, we live in a ‘celebrity age’ in which the opinions and examples of individuals are valued not only for their expertise, but for who they are in the public’s eyes. Does this mean we should enforce anonymity all the harder? Also should the affiliation of an expert matter? There is much evidence that the public’s trust in expert advice can depend quite strongly on whether they work for a government department, regulator, industry, NGO, university, or a particular profession such as medicine (French *et al.* 2002; Langford *et al.* 1999; Slovic 1997). I am not sure where these arguments are leading; but I am sure that if expert judgement is to be used in participatory deliberation on societal issues, then we need more clarity on the ethical and professional issues in doing so.

## Conclusions

In the above, I have surveyed some of the developments in the use of expert judgement in risk and decision making my 1985 paper. Far from feeling that the field is mature, I believe that modern developments in society mean that we need to explore many more issues in its use. In particular:

* Developments in web technology are facilitating the use of distributed decision support systems; and sadly some of this is based upon very naive and simplistic models which run counter to much of what has been learnt in decision theory. We need to be voluble in our criticism of these and to promote more sophisticated ways of supporting groups of DMs.
* The text-book problem is gaining in importance and to address this we need to develop meta-analytic methodologies for drawing together expert judgement studies. Moreover, as part of these developments we need to consider how to convey the import of the meta-analysis to the lay public.
* Cooke’s principles need much greater discussion if expert judgement studies are to inform more public deliberation.

References

K. J. Arrow (1963). Social Choice and Individual Values. 2nd Edn. John Wiley and Sons, New York.

D. Ashby and A. F. M. Smith (2000). "Evidence-based medicine as Bayesian decision-making." *Statistics in Medicine* **19** pp. 3291-305.

M. Bacharach (1975). "Group decisions in the face of differences of opinion." *Management Science* **22**(2) pp. 182-91.

T. Bedford and R. Cooke (2001). Probabilistic Risk Analysis: Foundations and Methods. Edn. Cambridge, Cambridge University Press.

T. Bedford, J. Quigley and L. Walls (2006). "Expert elicitation for reliable system design." *Statistical Science* **21**(4) pp. 428-50.

T. Beierle and J. Cayford (2002). Democracy in Practice: Public Participation in Environmental Decisions, Resources for the Futurepp.

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

J. M. Booker and M. A. Meyer (1988). "Sources and Effects of Interexpert Correlation: An Empirical Study , 8, 135-142. 1988." *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-8** pp. 135-42.

R. F. Bordley (2009). "Combining the opinions of experts who partition events differently." *Decision Analysis* **6**(1) pp. 38-46.

R. T. Clemen and T. Reilly (1999). "Correlations and copulas for decision and risk analyses." *Management Science* **45**(2) pp. 208-24.

R. T. Clemen and R. Winkler (1993). "Aggregating point estimates: a flexible modelling approach." *Management Science* **39** pp. 501-15.

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

W. S. Cleveland (1994). The Elements of Graphing Data. Edn. Murray Hill, New Jersey.

R. M. Cooke (1991). Experts in Uncertainty. Edn. Oxford, Oxford University Press.

R. M. Cooke (2007). Expert Judgement Studies. *Reliability Engineering and System Safety*. **93** pp. 655-777.

R. M. Cooke, S. ElSaadany and H. Xinzheng (2007). "On the performance of social network and likelihood based expert weighting schemes." *Reliability Engineering and System Safety* **93**(5) pp. 745-56.

R. M. Cooke and L. H. J. Goossens (2004). "Expert judgement elicitation for risk assessments of critical infrastructures." *Journal of Risk Research* **7**(6) pp. 643-56.

R. M. Cooke and L. H. J. Goossens (2007). "TU Delft Expert Judgement Database." *Reliability Engineering and System Safety* **93**(5) pp. 657-74.

R. M. Cooke and K. N. Probst (2006). Highlights of the Expert Judgement Policy Symposium and Technical Workshop. Washington, D C, Reources for the Future, 1616 P St NW, Washington, DC 20036pp.

N. Dalkey and O. Helmer (1963). "An experimental application of the Delphi method to the use of experts." *Management Science* **9**(3) pp. 458-67.

B. De Finetti (1974). Theory of Probability. Edn. Chichester, John Wiley and Sons.

M. H. DeGroot and M. J. Bayarri (1991). "What Bayesians expect of each other." *Journal of the American Statistical Association* **86**(416) pp. 924-32.

Department of Health (1998). Communicating About Risks to Public Health: Pointers to Good Practice. Edn. London, HMSO.

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

J. S. Dryzek and C. List (2003). "Social choice theory and deliberative democracy: a reconciliation." *British Journal of Political Science* **33**(1) pp. 1-28.

C. Eden and J. Radford, eds. (1990). *Tackling Strategic Problems: the Role of Group Decision Support*. London, Sage.

A. Faria and J. Q. Smith (1997). "Conditionally externally Bayesian pooling operators in chain graphs." *Annals of Statistics* **25**(4) pp. 1740-61.

A. Franco, D. Shaw and M. Westcombe (2006). Problem Structuring Methods I. *Journal of the Operational Research Society*. **57** pp. 757-878.

A. Franco, D. Shaw and M. Westcombe (2007). Problem Structuring Methods II. *Journal of the Operational Research Society*. **58** pp. 545- 682.

S. French (1980). "Updating of belief in the light of someone else's opinion." *Journal of the Royal Statistical Society* **A143** pp. 43-8.

S. French (1981). "Consensus of opinion." *European Journal of Operational Research* **7** pp. 332-40.

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

S. French (1987). Conflict of belief: when advisers disagree. In *Analysing Conflict and its Resolution: Some Mathematical Contributions*. P. G. Bennett, ed. Oxford, Oxford University Presspp.93-111.

S. French (2003). "Modelling, making inferences and making decisions: the roles of sensitivity analysis." *TOP* **11**(2) pp. 229-52.

S. French (2006). "Web-enabled strategic GDSS, e-democracy and Arrow’s Theorem: a Bayesian perspective." *Decision Support Systems* **(in press)** pp.

S. French, A. J. Maule, G. Mythen and C. Wales (2002). Trust and Risk Communications. Manchester, Manchester Business Schoolpp.

S. French, A. J. Maule and K. N. Papamichail (2009). Decision Behaviour, Analysis and Support. Edn. Cambridge, Cambridge University Press.

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

S. French, D. Rios Insua and F. Ruggeri (2006). "e-participation and decision analysis." *Decision Analysis* pp. (under revision).

S. French, D. Rios Insua and F. Ruggeri (2007). "e-participation and decision analysis." *Decision Analysis* **4**(4) pp. 1-16.

P. H. Garthwaite, J. B. Kadane and A. O'Hagan (2005). "Statistical methods for eliciting probability distributions." *Journal of the American Statistical Association* **100**(470) pp. 680-701.

C. Genest and C. G. Wagner (1987). "Further evidence against independence preservation in expert judgement synthesis." *Aequationes Mathematicae* **32** pp. 74-86.

C. Genest and J. V. Zidek (1986). "Combining probability distributions: a critique and annotated bibliography." *Statistical Science* **1** pp. 114-48.

G. Gigerenzer (2002). Reckoning with Risk: Learning to live with Uncertainty. Edn. Harmondsworth, Penguin Books.

T. Gilovich, D. Griffin and D. Kahneman, eds. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, Cambridge University Press.

L. H. J. Goossens and G. N. Kelly (2000). "Special: Issue: Expert Judgement and Accident Consequence Uncertainty Analysis." *Radiation Protection Dosimetry* **90**(3) pp. 293-381.

R. S. Gregory, B. Fischhoff and T. McDaniels (2005). "Acceptable input: using decision analysis to guide public policy deliberations." *Decision Analysis* **2**(1) pp. 4-16.

J. Hartung, J. Knapp and B. K. Sinha (2008). Statistical Meta-Analysis with Applications. Edn. Hoboken, NJ, John Wiley and Sons.

R. Hastie (1986). Experimental evidence of group accuracy. In *Information Pooling and Group Decision Making*. B. Grofman and G. Owen, eds. Greenwich, Ct, JAI Presspp.129-57.

G. R. J. Hockey, A. J. Maule, P. J. Clough and L. Bdzola (2000). "Effects of negative mood on risk in everyday decision making." *Cognition and Emotion* **14** pp. 823-56.

J. K. Hodge and R. E. Klima (2005). The Mathematics of Voting and Elections: a Hands-On Approach. Edn. Rhode Island, American Mathematical Society.

R. A. Jacobs (1995). "Methods for combining experts' probability assessments." *Neural Computing* **7** pp. 867-88.

H. Jeffreys (1961). Theory of Probability. 3 Edn. Oxford, Oxford University Press.

J. B. Kadane, M. J. Schervish and T. Seidenfeld (1999). Rethinking the Foundations of Statistics. Edn. Cambridge, Cambridge University Press.

D. Kahneman, P. Slovic and A. Tversky, eds. (1982). *Judgement under Uncertainty*. Cambridge, Cambridge University Press.

N. I. Karacapilidis and C. P. Pappis (1997). "A framework for group decision support systems: combining AI tools and OR techniques." *European Journal of Operational Research* **103** pp. 373-88.

F. S. Kelly (1978). Arrow Impossibility Theorems. Edn. New York, Academic Press.

J.-L. Koning (2003). How electronic voting can escape Arrow's Impossibility Theorem. In *e-Government and e-Democracy: Progress and Challenges*. J. Padget, R. Neira and J. L. Diaz de Leon, eds. Zacatenco, Mexico, Instituo Politecnico Nacional, Centro de Investigacion en Computacion, Unidad Profesional "Aldofo Lopez Mateos"pp.138-46.

D. Kurowicka and R. M. Cooke (2006). Uncertainty Analysis with High Dimensional Modelling. Edn. Chichester, Wiley.

I. Langford, C. Marris and T. O'Riordan (1999). Public Reactions to Risk: Social Structures, Images of Science and the Role of Trust. In *Risk Communication and Public Health: Policy, Science and Participation*. P. G. Bennett and K. C. Calman, eds. Oxford, Oxford University Presspp.33-50.

M. Leach, I. Scoones and B. Wynne, eds. (2005). *Science and Citizens*. London, Zed Books.

E. E. Leamer (1978). Specification Searches. Edn. Chichester, John Wiley and Sons.

H. P. Lehmann and S. N. Goodman (2000). "Bayesian communication: a clinically significant paradigm for electronic publication." *Journal of the American Medical Informatics Association* **7**(3) pp. 254-66.

S. Lichtenstein, B. Fischhoff and L. D. Phillips (1982). Calibration of probabilities: the state of the art to 1980. In *Judgement under Uncertainty*. D. Kahneman, P. Slovic and A. Tversky, eds. Cambridge, Cambridge University Presspp.306-34.

H. A. Linstone and M. Turoff (1978). The Delphi Method: Techniques and Applications. Edn. London, Addison- Wesley.

A. Madansky (1964). Externally Bayesian groups. RM-4141-PR, RANDpp.

A. J. Maule (2004). "Translating risk management knowledge: the lessons to be learned from research on the perception and communication of risk." *Risk Management* **6**(2) pp. 15-27.

A. J. Maule (2008). Risk communication and organisations. In *The Oxford Handbook of Organizational Decision Making*. W. Starbuck and G. Hodgkinson, eds. Oxford, Oxford University. Presspp.

C. N. Morris and S. L. Normand (1992). Hierarchical models for combining information and meta-analyses. In *Bayesian Statistics 4*. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds. Oxford, Oxford University Presspp.321-44.

A. Morton, F. Ackermann and V. Belton (2003). "Technology-driven and model-driven approaches to group decision support: focus, research philosophy, and key concepts." *European Journal of Information Systems* **12**(2) pp. 110-26.

J. L. Mumpower and T. R. Stewart (1996). "Expert judgemen and expert disagreement." *Thinking and Reasoning* **2**(2-3) pp. 191-211.

A. O'Hagan, C. E. Buck, A. Daneshkhah, R. Eiser, P. H. Garthwaite, D. Jenkinson, J. E. Oakley and T. Rakow (2006). Uncertain Judgements: Eliciting Experts' Probabilities. Edn. Chichester, John Wiley and Sons.

A. O'Hagan and J. Forester (2004). Bayesian Statistics. Edn. London, Edward Arnold.

L. D. Phillips (1984). "A theory of requisite decision models." *Acta Psychologica* **56** pp. 29-48.

P. Reagan-Cirincione (1994). "Combining group facilitation, decision modelling and information technology to improve the accuracy of group judgement Improving the accuracy of group judgment: a process intervention combining group facilitation, social judgment analysis, and information technology." *Organisational Behaviour and Human Decision Process* **58** pp. 246-70.

O. Renn (1998). "The Role of Risk Communication and Public Dialogue for Improving Risk Management." *Risk, Decision and Policy* **3** pp. 3-50.

O. Renn (2008). Risk Governance. Edn. London, Earthscan.

O. Renn, T. Webler, H. Rakel, P. Dienel and B. Johnson (1993). "Public participation in decision making: a three step-procedure." *Policy Sciences* **26**(3) pp. 189-214.

O. Renn, T. Webler and P. Wiedermann, eds. (1995). *Fairness and Competence in Citizen Participation: Evaluating Models and Environmental Discourse*. Dordrecht, Kluwer.

S. Renooij (2001). "Probability elicitation for belief networks: issues to consider." *The Knowledge Engineering Review* **16**(3) pp. 255-69.

D. Rios Insua and F. Ruggeri (2000). Robust Bayesian Analysis. Edn. New York, Springer-Verlag.

G. Rowe and G. Wright (1999). "The Delphi technique as a forecasting tool: issues and analysis." *International Journal of Forecasting* **15** pp. 353-75.

L. J. Savage (1972). The Foundations of Statistics. 2nd Edn. New York, Dover.

J. Shanteau (1995). Expert Judgment and Financial Decision Making. In *Risky Business: Risk Behavior and Risk Management*. B. Green, ed. Stockholm, Stockholm Universitypp.

P. Slovic (1997). Trust, emotion, sex, politics and science: surveying the risk-assessment battlefield. In *Environment, Ethics and Behaviour*. M. Bazerman, D. Messick, A. Tenbrunsel and K. Wade-Benzoni, eds. San Francisco, New Lexington Presspp.

P. Slovic (2001). Perceptions of Risk. Edn. London, Earthscan Library.

L. Suantak, F. Bolger and W. R. Ferrell (1996). "The hard-easy effect in subjective probability calibration." *Organisational Behaviour and Human Decision Processes* **67** pp. 201-21.

A. J. Sutton and K. R. Abrams (2001). "Bayesian methods in meta-analysis and evidence synthesis." *Statistical Methods in Medical Research* **10**(4) pp. 277-303.

P. S. Szwed and J. R. van Dorp (2002). *A Bayesian Model for rare event risk assessment using expert judgement about paired scenario comparisons*. ASEM National Conference.

M. W. Wiper and S. French (1995). "Combining experts' opinions using a normal-Wishart model." *Journal of Forecasting* **14** pp. 25-34.

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

1. The dm is referred to in the feminine throughout, the analyst in the masculine, and experts usually in the plural. [↑](#footnote-ref-1)
2. Though I have never heard of unequal weighting based upon relative salaries. [↑](#footnote-ref-2)
3. For the record, one early comparison with a Bayesian method showed pretty much the same performance on one study, but Bayesian method was much more costly computationally (Wiper and French 1995). [↑](#footnote-ref-3)
4. One issue of relevance here is how to draw together expert judgements offered over different but related uncertainties – in mathematical terms, based upon different *σ*-fields of events. Bordley (2009) offers some early thoughts on this. [↑](#footnote-ref-4)
5. See also the European Science Foundation programme *Towards Electronic Democracy: Internet based complex decision support* (TED). (http://www.esf.org/ted and http://infodoc.escet.urjc.es/ted/) [↑](#footnote-ref-5)