

# Expert Judgement, Meta-Analysis and Participatory Risk Analysis

Simon French

Department of Statistics

University of Warwick

Coventry, CV4 7AL, UK

simon.french@warwick.ac.uk

## Abstract

There are three contexts in which one might wish to combine expert judgements of uncertainty: the expert problem, the group decision problem and the textbook problem. Much has been written on the first two, which have the focus of a single decision context, but little on the third. The textbook problem arises when one needs to draw together expert judgements into a decision analysis when their judgements were made originally in a context free manner or perhaps for other decision contexts. In many ways the textbook problem parallels that of performing a meta-analysis of empirical studies. However, there are differences. In this paper, we discuss those difficulties and then focus on two closely related issues: how should expert judgement studies be published so as to facilitate subsequent meta-analyses and how should such meta-analyses be performed?

**Keywords:** aggregation of expert judgement; meta-analysis; public participation; the textbook problem.

## 1 Introduction

As the complexity of modern society and its systems grows, it becomes more necessary to rely on expert judgement in many risk and decision analyses. Recently I surveyed the combination of experts' judgements for such analyses, considering three different contexts (French 2011):

- *The expert problem.* In this a group of experts are consulted by a decision maker who faces a specific decision and is not a member of the group. The decision maker alone

is responsible and accountable for the decision. Here, the emphasis is on how the decision maker should learn from the experts.

- *The group decision problem.* The group itself is jointly responsible and accountable for the decision; they are also their own experts. They wish that to the outside world their decision appears rational and, possibly, also fair and democratic. Thus, the emphasis is on how they may combine their judgements in some formal structured way.
- *The textbook problem.* The group is simply required to give their judgements for others to use in future, as yet undefined circumstances. Thus the emphasis here is on reporting their judgements in a manner that offers the greatest potential for future use.

The first two cases have the focus of a specific decision problem, but the third does not.

The textbook problem was so termed by French (1985), and it was moot then whether it existed other than in concept. Recently, however, it has gained importance as part of societal risk and decision analysis (French 2011). As yet there are *no* established methodologically sound approaches to its solution. In addressing its solution or, at least, exploring the issues that need be overcome in its solution, my arguments will follow two different, but closely strands: firstly, the need to report expert studies in a ‘scientific’ manner; secondly, the technical issues faced by meta-analyses specifically designed to draw together previously published expert studies. Further I shall argue that these meta-analytic techniques for expert studies should differ considerably from those developed for combining empirical studies, since expert judgement data and empirical data have very different qualities.

## **2 The Textbook Problem in Societal Risk and Decision Analysis**

How should a group of experts write a report containing their probabilistic forecasts of some potential events or uncertainty judgements of some quantities without the context of a specific decision or risk? Savage (1972) claimed that this is the role of scientific reporting in building consensus on our understanding of the natural world; but that is not what I mean. To

give an example: a few years back, the US and UK governments commissioned reports on the likelihood of an earth impact of a major asteroid. Although a scientific issue, it was not one that was driving current research; rather, it was a question that politicians asked, and one they asked without the context of a specific decision. How should the experts have framed their answer; and how might their answers be used?

Two developments have led to the textbook problem gaining reality. Firstly, paternalistic approaches of governments and their agencies to societal issues are being replaced by more participatory methods (see, e.g., Gregory *et al.* 2005; French *et al.* 2007; Renn 2008). Twenty years ago, most democracies were purely representative democracies. Citizens elected politicians who with their administrations made societal decisions. Once made, decisions were announced and defended against any public criticism. Public disillusion with politics along with greater adherence to deliberative democratic ideals have led to growing use of stakeholder involvement and public participation (Bayley 2008). In turn, this has led to a demand for wider access to information on risks, etc., which introduces the second development. The web has allowed citizens to find and access reports and other data much easier than ever before, and to deliberate upon these (Rios Insua and French 2010). The development of the semantic web and open linked data (Bizer *et al.* 2009) and of web-based analytic technologies for building model chains, e.g. those envisaged by the *UncertWeb Project* ([www.uncertweb.org](http://www.uncertweb.org)), are adding further impetus. Analyses may be conducted using a range of data, models and other material available across the web: and this may be done by lay, often inexperienced analysts. Expert studies and the judgemental data therein will be increasingly consulted for the purposes of risk and decision analyses. Reports and expert judgements from studies for earlier decisions may be reused in later analyses to support a different decision, a factor that may confound some of the issues in dealing with the expert and group decision problems with those relating to the textbook problem.

### 3 Publication, Cooke's Principles and Meta-Analysis

Little has been written on good practice in reporting of expert studies. One exception is found in the work of Cooke (1991), who suggested four underpinning principles.

#### *Cooke's Principles*

- *Scrutability/Accountability.* All data, including experts' names and assessments, and all processing tools should be open to peer review and results must be reproducible by competent reviewers.
- *Empirical Control.* Quantitative expert assessments should be 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 should not be prejudged, prior to processing the results of their assessments.

These principles seek to assure that any expert judgement study may be audited, open to peer review and hence 'scientific'. While there are issues with their framing and practicability (French 2011), these principles – surprisingly – remain the only significant guidance on how this might be done; and – even more surprisingly – have been little discussed (for some discussion, see Cooke 2007). In contrast, the research community and scientific journals have developed and enforced a wide range of principles to govern the peer review, publication and use of empirical studies, alongside which has grown a recognition of the importance of evidence-based decision making (Pfeffer and Sutton 2006; Shemilt *et al.* 2010). The latter developments began within medicine, but the imperatives of basing decisions on evidence are now changing thinking in many domains. Methodologies of meta-analysis have been established (Borenstein *et al.* 2009; Geyskens *et al.* 2009). To date, meta-analysis has focused on issues of drawing out inferences that may be drawn from several published empirical studies;

methods for expert judgement data have not been considered. Indeed, one might anticipate logical positivist arguments that evidence-based decision making should draw solely upon objective empirical data. It is often assumed that evidence-based decision making should be ‘objective’, but evidence need not be confined to empirical data (Briner *et al.* 2009). Although Glanville and Paisley (2010, p89) state: “... expert opinion would not be considered a reliable source of evidence on treatment effects”, they continue “the ‘concurrence of experts’ is a recognised form of model validation and strong anecdotal evidence suggests that expert advice is used in informing the development of the scope and structure of decision models in health care.” Cooper *et al.* (2007) suggest that expert input is common in model selection and in providing parameter values, though it is not considered acceptable for assessing clinical effectiveness<sup>1</sup>.

Following my Bayesian leanings, I would argue that expert judgement should inform a meta-analysis *providing that the relative quality of the expert input to that of empirical data is appropriately assessed*. And therein lays the rub: how do we assess that relative quality? We need meta-analytic techniques that draw together both empirical and expert judgemental data. French (2011) explored some of the differences that might be expected between meta-analyses of expert judgement and empirical studies: see Table 1. These differences, especially the probable correlation between expert judgement studies, are sufficient to make almost all of the meta-analytic techniques for empirical data of questionable value for expert judgement data. One exception might be the work on forest plots comparing the results of several studies (Anzures-Cabrera and Higgins 2010). Cooke’s *Excalibr* software for expert judgement already produces graphical output very similar to these.

---

<sup>1</sup> Given that the form of any model may well have more effect in determining the outcome of an analysis than the inclusion of other expert judgemental data, this seems somewhat perverse, but that is another matter.

	<b>Meta-analysis for Empirical Studies</b>	<b>Meta-Analysis for Expert Judgement Studies</b>
<b>Peer review of source studies</b>	In the greater majority of cases one knows whether published studies have been peer-reviewed. Often a selection criterion in meta-analysis is to limit the meta-analysis to peer-reviewed studies.	There are no regularised procedures of peer review of expert judgement studies. Such studies are usually published by organisations or projects and are seldom subject to rigours of scientific publication.
<b>Publication versus context bias</b>	It is well know that it is much easier to get significant results published than insignificant or negative ones, causing potential biases in meta-analyses.	Expert judgement studies are commissioned for contexts with significant, often potentially disastrous risks; and sometimes widely promoted by a pressure group. Publication relates more to political issues than significance of the results.
<b>Independent experiments versus correlated experts</b>	Generally the experiments reported in the studies are independent. Even if there is a reason to suspect dependence, there is usually enough structure to investigate and model this. Thus allowance for correlation may be made in the meta-analysis	Experts are correlated sources of information. Common education, exposure to similar literatures and the general consensus on scientific knowledge implies that they are not independent sources. Moreover, some experts may be common to several studies. But very little is known about how to model the correlations that arise.
<b>Experimental Design versus Elicitation Protocol</b>	Drawing together several empirical studies requires a careful comparison of their experimental designs to ensure that the data are commensurate and can sensibly be drawn into a common analysis.	To compare and draw together expert judgement studies one needs assess the similarities and differences in the elicitation protocol. Were the questions and assessments focused on the same uncertainties or unknown quantities?
<b>Outliers</b>	The treatment of outliers has long been moot in statistics and it is not less so within meta-analysis. One needs to allow for different outlier treatment in different studies.	An outlying expert judgement may be an indicator that the expert holds a different mental model from the other experts, perhaps a different scientific paradigm. In risk analyses it is important <i>not</i> to lose these as they may represent hazards that need to be risk managed.

**Table 1:** Comparison of meta-analyses of empirical studies and of expert judgement studies

One point about publication that is immediate from Table 1 is the importance of a full publication of the elicitation protocol of an expert study. No empirical study should be published without its experimental design, and a parallel principle should govern the publication of expert judgement studies. Whether this is consequence of Cooke's Scrutability principle or an additional one, I leave to the reader to decide.

We have discussed expert judgement studies as if they are *solely* based on expert judgement. This is not always the case. In some cases, experts consult their computer models or do some back of the envelope calculations before giving an opinion. Either way their probability statements may be the complex result of combining their personal opinions about parameters, the validity of models and modelling error. Moreover, the models are a complex expression of the experts' selection of human knowledge. Applications of Cooke's classical model (Cooke 1991; 2007) in the context of the expert problem have often recognised this, but few other discussions of the use of expert judgement have (Cooke *et al.* 1997; Goossens and Kelly 2000). The expert problem has different characteristics to the textbook problem (French 2011), so the development of Cooke's method to the latter context is not straightforward. Moreover, Cooke's method requires that experts are assessed against comparable calibration sets, a practicality that will seldom be satisfied in the textbook problem.

Finally, note that expert judgemental studies and meta-analyses thereof need to be reported in a way that is mindful the literature on risk communication and the public understanding of science. It is a non-trivial task to present uncertainty information in a manner which both lay publics and experts will understand in an unbiased way (for general discussion and advice, see, e.g., Morgan *et al.* 2002; Maule 2008; French *et al.* 2009; Bennett *et al.* 2010).

<i>Size</i>	<i>Number near Earth</i>	<i>Frequency of impacts</i>	<i>Last impact</i>	<i>Annual probability of impact</i>
10 - 50m	2 hundred million	1 in every 5 years	Siberia, 1908: area not populated.	0.2
100m	2 hundred thousand	1 in every thousand years	China, 1490: 10,000 deaths	0.001
1-2km	2 thousand	1 in every 100,000 to 1 million years	Argentina, 3 million years ago: local extinctions and global cooling	0.00001
15km	50	1 in every 65 million years	Mexico, 65 million years ago: dinosaur extinction	0.00000002

**Table 2:** Risk of asteroid impact  
 From [http://www.risk-ed.org/pages/risk/asteroid\\_prob.htm](http://www.risk-ed.org/pages/risk/asteroid_prob.htm) (visited 13/5/2011)  
[www.risk-ed.org](http://www.risk-ed.org) is established by the Chemical Industries Education Centre of the Institute of Materials.

#### 4 An example of the issues

To explore some of these issues, I have supposed that I need to assess the probability of a catastrophic asteroid<sup>2</sup> impact on the Earth. I chose this problem simply because I had no professional knowledge or experience of the science and could seek to learn as member of the lay public. Using Google with search terms: ‘risk’ ‘asteroid’ and ‘impact’, I quickly found: [http://www.risk-ed.org/pages/risk/asteroid\\_prob.htm](http://www.risk-ed.org/pages/risk/asteroid_prob.htm). This began with the advice that: “The largest asteroids will cause the most damage but there are less of them. Astronomers have used mathematical models and historical data to estimate the probability of an asteroid hitting the earth.” Table 2 was then presented.

Several points are immediate:

- There are missing ranges of asteroid size: e.g. 100m – 1km. Can we just interpolate such missing data?
- For 10m – 50m the annual probability of impact is 0.2; yet the last recorded impact was in 1908. This seems incongruous.
- The frequency of 1 in 65million years for impact of 15km asteroid and the last occurrence being 65million years ago is suggestive of a single data point.

---

<sup>2</sup> A better term would be ‘near earth object’ or NEO, but I adopt lay language.



- How were these estimates derived: statistical modelling of frequency data, some astrophysical modelling, or what? The introductory sentences suggest that expert judgement was not involved; but the data would seem too sparse to be interpreted without some expertise and judgement.

Would the lay reader notice these points? Indeed, given the ease with which probabilities are misinterpreted (Gilovich *et al.* 2002), might they be concerned that after 65 million years a 15km asteroid impact is imminent!

Google identified several other relevant sites<sup>3</sup>. Quoting from four of these:

- "The most dangerous asteroids, capable of a global disaster, are extremely rare. The threshold size is believed to be 1/2 to 1 km. These bodies impact the Earth only once every 1,000 centuries on average. Comets in this size range are thought to impact even less frequently, perhaps once every 5,000 centuries or so." and "The threshold for an impact that causes widespread global mortality and threatens civilization almost certainly lies between about 0.5 and 5 km diameter, perhaps near 2 km. Impacts of objects this large occur from one to several times per million years."

[http://imagine.gsfc.nasa.gov/docs/ask\\_astro/answers/danger.html](http://imagine.gsfc.nasa.gov/docs/ask_astro/answers/danger.html)

- "Such global environmental catastrophes, which place the entire population of the Earth at risk, are estimated to take place several times per million years on average."

<http://www2.jpl.nasa.gov/sl9/back2.html>

- "A 10 km diameter asteroid entering at 30 km/s would have the energy equivalent to 15 billion times that of the Hiroshima nuclear bomb ... we would, on average, expect one such dramatic impact every ten million years."

<http://miac.uqac.ca/MIAC/impactearth.htm>

---

<sup>3</sup> All sites were visited on the same day: 13/5/2011.

- A report for the European Space Agency (ESA) provides the summary in Table 3.

[http://www.esa.int/gsp/NEO/doc/NEOMAP\\_report\\_June23\\_wCover.pdf](http://www.esa.int/gsp/NEO/doc/NEOMAP_report_June23_wCover.pdf)

There are several points to note. The values from these sites are somewhat inconsistent with those in Table 2. Two of these sites have been established by NASA, one by the ESA, and one has (had?) some connection with the Canadian Space Agency; so one would think that they have some authority. No site gave the basis or evidence for their probabilities. There is no transparency and auditability of the form that Cooke's principles would demand. All the sites frame the risks negatively in the sense of talking about the risk of death and catastrophe. Such negative framing is likely to create a risk prone behaviour on the part of the reader, not something that one might wish in deliberating societal risks (see, e.g., Maule 1989). Finally, while I confess that I do not have a clear idea of the form a meta-analysis of expert judgement might take, the lack of any provenance of the probability estimates on these sites would make the task nigh on impossible.

Exploring the web a little further I came across several discussions of the Torino scale (e.g. <http://impact.arc.nasa.gov/torino.cfm>; for academic background see Binzel 2000). This is *not* related to the probability an impact on the earth *per se*, but classifies the seriousness of the threat of known asteroids, recognising that the effect of an impact depends on the velocity and angle at which the earth's and asteroid's orbits intersect. The scale runs from 0 to 10,

Impactor Size (m)	Mean Impact Interval (yr)	Possible effects/ comparable event
30	200	Fireball, shock-wave, minor damage
50	2500	Tunguska explosion or small crater
100	5000	Largest H-bomb detonation
200	47000	Destruction on a national scale
500	200,000	Destruction on a European Scale
1000	600,000	Many millions dead, global effects
5000	20 million	Billions dead, global climate change
10,000	100 million	Extinction of human civilisation

**Table 3:** Estimated frequency and effects of impactors (asteroids, etc.) from (NEOMAP 2004)

with scores of 0 or 1 implying virtually no significant threat. A NASA database, the *Sentry Risk Table* (<http://neo.jpl.nasa.gov/risk/>), catalogues all known asteroids along with their Torino scores. Comfortingly, there are none above 1, all but 2 scoring 0. But this raises the question of what data underpin Table 2, which claims that there are some 50 asteroids near earth of greater than 15km diameter. The *Sentry Risk Table* lists none.

I then used Google Scholar to search for more information, recognising that this is an academic rather than a lay approach. I immediately found Chapman and Morrison's 1994 paper in *Nature*, which contains much information, albeit, somewhat old on the risks of asteroid impact. It also contains an argument that runs roughly as follows. The frequency of impacts which would annihilate humanity is about 1 in 100 million years. The earth's population is about 6 billion. So the annualised death rate from this cause is about 60 lives per annum. Adding lesser impacts which might only kill millions or a few billion raises the annual death rate to several hundred: a figure that is commensurate with many other modern risks, such as deaths from air crashes. So the risk of major asteroid impact should be ascribed as much importance as air safety. I find this logic hard to follow. Chapman and Morrison's argument seems to be muddling event probabilities, individual and population risks: that it occurred in an article in *Nature* is of particular concern. Probably subject to a similar confusion, Lembit Opik, a British MP, remarked in 2000 that the risk of being killed by an asteroid was 750 times higher than winning the national lottery (BBC News report, 4/1/2000). Given that there have been *no* deaths from asteroid impact since the national lottery began, but quite a few lottery winners; empirically, one doubts the value of this comparison. These last two examples indicate the importance in publishing expert judgement studies of paying attention to good practice in risk communication (see remarks at the end of Section 3). The potential for lay misinterpretation and incorrect deductions about risks is great.

For those interested, Mellor (2010) explores the issue of asteroid impact from the perspective of the public understanding of science and some of the political background<sup>4</sup>. Her paper charts the reporting and development of estimates of the probabilities of asteroid impact, noting that the process may not have been a shining example of the ‘objectivity’ of Science.

I recognise that that asteroid impact is unlikely to be central to a real societal decision in which stakeholder and public participation were used. I chose it for investigation because of my ignorance the risks concerned. However, looking at other contexts via the web does not encourage me that the results were atypical. Judgemental probabilities and advice on uncertainties are offered to the public from all manner of websites without any semblance of their provenance or their overlap with other sites offering similar advice.

## **5 The Intergovernmental Panel on Climate Change**

The work of the Intergovernmental Panel on Climate Change (IPCC) is arguably the most complex example of the aggregation of vast amounts of data and expert opinion to date. Since 1988 it has been drawing together evidence and synthesising evidence relating to climate change and its mitigation. It has published four major reports and is embarked on developing a fifth due for publication in 2013/14: the latest synthesis report is IPCC (2008); all reports and contributing studies are on the website (<http://www.ipcc.ch>). The IPCC has broadly adopted Bayesian methodologies and embraced analyses in which probability models are based on both data and expert judgement.

In developing their syntheses the IPCC has been careful to provide and publish guidance material to all its subpanels on a range of issues including uncertainty modelling, elicitation and communication (Moss and Schneider 2000; Intergovernmental Panel on Climate Change Core Writing Team *et al.* 2008; Intergovernmental Panel on Climate Change Core Writing

---

<sup>4</sup> Note I read this paper *after* my explorations of the web described in this section: so it did not bias them.

Team *et al.* 2010). These documents emphasise the need to assess and communicate the uncertainty consistently, recognise the need to draw in expert judgements, and are particularly strong on applying understandings derived from psychological studies of uncertainly perception, both in relation to the elicitation and communication of uncertainty. The guidance also emphasises the need to provide ‘traceable’, i.e. auditable, analyses so that the results can be subject to effective peer review. But there is a passage in Moss and Schneider (Section 2, 2000) in which it is suggested that detailed data arising in elicitation of expert judgements or the production of a ‘group aggregate’ need not be reported; so the spirit of Cooke’s Scrutability Principle is not fully adopted.

The guidance is, however, almost entirely silent on aggregating expert judgements *per se*. Aggregation is achieved via ensembles (mixtures) of models. Expert judgement may used to provide probability distributions encoding uncertainties in each model; and then models that predict the same quantities are drawn into ensembles, which in this context may be viewed an approach to meta-analysis. While expert judgement is embraced, there is, nonetheless, a tone in the documents that derives from more positivist approaches to analysing empirical data. The ensemble methodology that IPCC use does not directly address the issues raised in Table 1. Remarkably, there is little discussion of correlation between models in ensembles and I could find no discussion of correlations that derive from the non-independence of experts. More positively, there is considerable discussion of how evidence should be presented qualitatively but in a consistent format so that conflicts between the models can be judged by reader with different levels of sophistication in uncertainty modelling.

A full critique of the many hundreds of pages of analysis in the IPCC syntheses, specifically the use and aggregation of expert judgement therein, would require much more space than available in this paper, but it does seem clear that some of the issues raised here are not entirely answered by the guidance documents. Although reporting standards are commenda-

bly high, the reporting of any expert judgements is not as detailed as perhaps it might be. The entire IPCC project may be thought of as an exercise in meta-analysis, but their approach relies on ensembles for quantitative aggregation supported by careful and standardised presentation of the qualitative evidence to report conflicts.

## 6 Discussion

So what might be done to address these issues? Nothing is a possibility. We might declare that people have always had opportunities to be misinformed; the web just makes the process easier. But I believe that it behoves us to think about how things *should* be done, even if practice falls well short of the ideal. Only when we understand what is ideal, can we help people improve. Only then can we explain how people may have been misled. It is clear to me that in the case of reporting expert judgement we do *not* know the ideal. Indeed, I might argue that the absence of any serious debate about the validity of Cooke's principles suggests that as a community of risk and decision analysts we do not care. Nor do we know how to learn from earlier several studies. Thus two strands of development are needed:

- reporting standards for expert judgement studies that allows them to be audited and evaluated;
- meta-analytic methodologies for expert judgement data.

Firstly, I recognise that we cannot change the exciting anarchy that is the Web. We have to accept that material that is dubious or just plain wrong will always be published. The same is true and accepted in the case of empirical studies: there are plenty of dubious ones out there. But in the case of empirical studies there are also a peer-reviewed, quality assured journals which are much more trustworthy, carefully indexed and backed up by archives of the underlying datasets. We need to create the same sort of archive for expert judgement studies. Cooke and Goossens (2007) have set up a database of expert judgement studies, but

with more of an intention of supporting research into expert judgement than as an archive for future risk analyses.

As community of risk and decision analysts we need define good principles for publishing expert judgement studies. Cooke's principles give us a good starting point, but we should recognise that they are only a starting point. His ideas need deep discussion, possible modification and extension, and then adoption by us all in our protocols for designing, running and reporting expert judgement studies. In establishing the principles we need recognise that it must be possible to audit studies against them so that peer-review methodologies can be defined and implemented. The principles must be operational. We also need to establish one or more archives in which studies can be deposited: Cooke and Goossens have provided a prototype, but it may need modification and there certainly need to some organisational ownership of such archives so that their future existence has some assurance.

Turning to the other strand of developments, we need to develop techniques that in a specific context allow us to select relevant expert studies and then produce a meta-analysis which establishes what may reasonably be learnt from these. Table 1 indicates that we cannot simply adopt standard meta-analytic techniques, but will need to develop some afresh.

Until we have such formal meta-analytic procedures, ones that have similar authority to the good practices promulgated by, e.g., the Cochrane Collaboration, we will need to proceed pragmatically. For most societal risk analyses, the need to draw in understandings from previous expert judgement studies will occur more in the formulation phase and primarily in defining prior distributions (French *et al.* 2009). There is a large literature on eliciting and constructing probability models to represent beliefs and uncertainties – any of the Bayesian modelling literature is relevant here, but see Morgan (2003) and O'Hagan *et al.* (2006). Furthermore, as computers get more powerful more complex sensitivity analyses can be applied to large multi-parameter probability models (Saltelli *et al.* 2000; French 2003); so the possible

effects of, say, the largely unknown correlations between different expert judgements within the models can be investigated. We might also look to current developments in combining scenario planning approaches with decision analysis (Wright and Goodwin 1999; Montibeller *et al.* 2006; Stewart *et al.* 2010; Stewart *et al.* 2011). In these several interesting scenarios are created which may form backdrops against which decision analyses may be constructed. There is often no intention that such scenarios are constructed to span the future in some balanced way, only an intention that they are all of interest in some sense, perhaps of particular concern to one or more stakeholder groups. We might construct scenarios such that each reflects the import of one of the selected expert judgement studies. This would avoid the complex issue of aggregating judgements across several studies, albeit at the cost of placing the onus of learning from the several scenarios on the intuition and understanding of the decision makers.

But I return to the main point of this paper: a plea that we address the issues that arise in the publication and subsequent use of expert judgement studies in some future context. Finally, the referees and others have queried the correctness of the term ‘textbook’ problem. I have used it simply because that is what I unthinkingly dubbed it in French (1985). Maybe the *archive problem* would be a better term? It would recognise that the issues that we are discussing relates to formal expert opinion and how it should be carefully documented, archived and subsequently used.

## **Acknowledgements**

I presented draft papers on this topic at *Bayesian Statistics 9* in Valencia in June 2010, to the *Real Academia de Ciencias: Exactas, Fisicas y Naturales* in Madrid in November 2010, at a Seminar on Policy Analytics at Paris Dauphine in December 2010, and at the second meeting on Risk, Reliability, Games and Decision at Belgirate in May 2011. I am grateful for the many helpful and thoughtful comments that I received on these occasions. I am also



very grateful for the constructive comments offered by the referees and editor of *Decision Analysis* and for their encouragement to consider the IPCC activities.

## References

- J. Anzures-Cabrera and J. P. T. Higgins (2010). "Graphical displays for meta-analysis; an overview with suggestions for practice." *Research Synthesis Methods* **1**(1).
- C. Bayley (2008). Public Participation. *Encyclopedia of Quantitative Risk Analysis and Assessment*. E. L. Melnick and B. S. Everitt. Chichester, John Wiley and Sons.
- P. G. Bennett, K. C. Calman, S. Curtis and D. Fischbacher-Smith, Eds. (2010). *Risk Communication and Public Health. 2nd Edition*. Oxford, Oxford University Press.
- R. P. Binzel (2000). "The Torino impact hazard scale." *Planetary and Space Science* **48**: 297-303.
- C. Bizer, T. Heath and T. Berners Lee (2009). "Linked Data - The Story So Far." *International Journal on Semantic Web and Information Systems* **5**(3): 1-22.
- M. Borenstein, L. V. Hedges, J. P. T. Higgins and H. R. Rothstein (2009). *Introduction to Meta-Analysis*. Chichester, John Wiley and Sons.
- R. B. Briner, D. Denyer and D. M. Rousseau (2009). "Evidence-Based Management: Concept Cleanup Time?" *Academy of Management Perspectives* **23**(4): 19-32.
- C. R. Chapman and D. Morrison (1994). "Impacts on the Earth by asteroids and comets: assessing the hazard." *Nature* **367**: 33-40.
- R. M. Cooke (1991). *Experts in Uncertainty*. Oxford, Oxford University Press.
- R. M. Cooke, Ed. (2007). *Expert Judgement Studies*. Reliability Engineering and System Safety.
- R. M. Cooke, T. Bedford and B. Kraan (1997). Report on the Benchmark Workshop on Uncertainty/ Sensitivity Analysis Codes. , European Safety and Reliability Association. Copies: Mathematics and Informatics, TU Delft, PO Box 5031, 2800GA Delft.

- R. M. Cooke and L. H. J. Goossens (2007). "TU Delft Expert Judgement Database." *Reliability Engineering and System Safety* **93**(5): 657-674.
- N. J. Cooper, A. J. Sutton, A. E. Ades, S. Paisley and D. R. Jones (2007). "Use of evidence in economic decision models: practical issues and methodological challenges." *Health Economics* **16**(12): 1277-1286.
- S. French (1985). Group consensus probability distributions: a critical survey (with discussion). *Bayesian Statistics 2*. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, North-Holland: 183-201.
- S. French (2003). "Modelling, making inferences and making decisions: the roles of sensitivity analysis." *TOP* **11**(2): 229-252.
- S. French (2011). "Aggregating Expert Judgement." *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales* **105**(1): 181–206.
- S. French, A. J. Maule and K. N. Papamichail (2009). *Decision Behaviour, Analysis and Support*. Cambridge, Cambridge University Press.
- S. French, D. Rios Insua and F. Ruggeri (2007). "e-participation and decision analysis." *Decision Analysis* **4**(4): 1-16.
- I. Geyskens, R. Krishnan, J.-B. E. M. Steenkamp and P. V. Cunha (2009). "A review and evaluation of meta-analysis practices in management research." *Journal of Management*(in press): Publihed online as doi:10.1177/0149206308328501.
- T. Gilovich, D. Griffin and D. Kahneman, Eds. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, Cambridge University Press.
- J. Glanville and S. Paisley (2010). Searching for evidence for cost-effectiveness decisions. *Evidence-Based Decisions and Economics*. I. Shemilt, M. Mugford, L. Vale, K. Marsh and C. Donaldson. Oxford, Oxford University Press: 79-92.

- L. H. J. Goossens and G. N. Kelly (2000). "Special: Issue: Expert Judgement and Accident Consequence Uncertainty Analysis." *Radiation Protection Dosimetry* **90**(3): 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): 4-16.
- Intergovernmental Panel on Climate Change Core Writing Team, R. K. Pachauri and A. Reisinger (2008). Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment. Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland, Intergovernmental Panel on Climate Change: 104 pages.
- Intergovernmental Panel on Climate Change Core Writing Team, M. L. Walser and S. C. Nodvin (2010). IPCC Fourth Assessment Report, Working Group I: Uncertainty Guidance Note for the Fourth Assessment Report. *Encyclopedia of Earth*. C. H. Cleveland. Washington, D.C., Environmental Information Coalition, National Council for Science and the Environment.
- A. J. Maule (1989). Positive and negative decision frames: a protocol analysis of the Asian disease problem of Kahneman and Tversky. *Process and Structure in Human Decision Making*. O. Svenson and H. Montgomery. Chichester, John Wiley and Sons.
- A. J. Maule (2008). Risk communication and organisations. *The Oxford Handbook of Organizational Decision Making*. W. Starbuck and G. Hodgkinson. Oxford, Oxford University. Press.
- F. Mellor (2010). "Negotiating Uncertainty." *Public Understanding of Science* **29**(1): 16-33.
- G. Montibeller, H. Gummer and D. Tumidei (2006). "Combining scenario planning and multi-criteria decision analysis in practice." *Journal of Multi-Criteria Decision Analysis* **14**: 5-20.

- G. Morgan, B. Fischhoff, A. Bostrom and C. Atman (2002). *Risk Communication: A Mental Models Approach*. Cambridge, Cambridge University Press.
- M. G. Morgan (2003). "Characterising and dealing with uncertainty: insights from integrated assessment of climate change." *Integrated Assessment* 4(1): 46-55.
- R. Moss and S. Schneider (2000). Uncertainties. *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC*. R. Pachauri, T. Taniguchi and K. Tanaka. Geneva, Intergovernmental Panel on Climate Change.
- NEOMAP (2004). Space mission priorities for near-earth object risk assessment and reduction; a report by the Near Earth Object Mission Advisory Panel, European Space Agency.
- 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*. Chichester, John Wiley and Sons.
- J. Pfeffer and P. J. Sutton (2006). "Evidence-based management." *Harvard Business Review* 84(1): 62-76.
- O. Renn (2008). *Risk Governance*. London, Earthscan.
- D. Rios Insua and S. French, Eds. (2010). *e Democracy: a Group Decision and Negotiation Perspective*. Group Decision and Negotiation. Dordrecht, Springer.
- A. Saltelli, K. Chan and E. M. Scott, Eds. (2000). *Sensitivity Analysis*. Chichester, John Wiley and Sons.
- L. J. Savage (1972). *The Foundations of Statistics*. New York, Dover.
- I. Shemilt, M. Mugford, L. Vale, K. Marsh and C. Donaldson, Eds. (2010). *Evidence-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice*. Oxford, Oxford University Press.

- T. J. Stewart, S. French and J. Rios (2010). Scenario-Based Multi-criteria Decision Analysis. *URPDM2010: Uncertainty and Robustness in Planning and Decision Making*. C. H. Antunes. Coimbra, Portugal.
- T. J. Stewart, S. French and J. Rios (2011). "Integration of Multicriteria Decision Analysis and Scenario Planning." *Omega*(submitted for publication).
- G. Wright and P. Goodwin (1999). "Future-focused thinking: combining scenario planning with decision analysis." *Journal of Multi-Criteria Decision Analysis* **8**(6): 311-321.