Experts and foresight: review and experience

Denis Loveridge

PREST, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK
E-mail: denis.loveridge@man.ac.uk

Abstract: Experts and their opinions are widely sought, but there are serious questions to be asked about both experts and their opinions. Governments, through their advisory committees, companies and adversarial groups eagerly seek expert opinion with which to influence the formulation of policy and regulations of all kinds. In a different sphere expert opinion is also a powerful feature of many complex court cases where many expert witnesses find themselves in circumstances beyond their normal experience and where their evidence may be manipulated in ways they never anticipated. Here, the nature of expert opinion and its elicitation are first surveyed, necessarily in brief and subsequently the methods and methodological issues of elicitation are discussed, all in relation to foresight activity. Finally, the implications for the practice of foresight, particularly in their manifestation in public foresight programmes, are discussed.

Keywords: experts; foresight; human judgement; practical experience; preference; science in court; uncertainty.


Biographical notes: Denis Loveridge is a Visiting Professor at the University of Manchester, where he is a member of PREST. He has 20 years’ experience in a major international company in foresight activity, the development of new high-technology businesses and long-term corporate development. He was deeply involved in the first UK Foresight programme from its beginnings in 1992 to its completion in 1995. Previously he had 14 years’ experience first in basic research and later in applied process research as a Principal Scientist and Project Manager/Coordinator and was responsible for world firsts in the application of fluidised bed combustion systems and in instrumentation. He has extensive experience of managing and coordinating complex international research programmes in both industry and the public sector.

1 Introduction

Currently there is a paradox where expert opinion is simultaneously avidly sought and lampooned with equal ferocity. Governments seek expert opinion eagerly, through advisory committees, whilst companies and adversarial groups do so with equal ferocity, all with the intention of influencing the formulation of policy and regulations of all kinds. Self-interest is never far away. However, there are serious questions to be asked about experts, their opinions and how they are elicited.
The world now and into the far future faces many situations of the kind described by Weinberg (1972) as trans-science and characterised by questions that are posed to science but that cannot be answered by science. Typically they are questions on disease control, climate change, sustainable development and other complex matters that involve the intermingling of thought streams from the Social, Technological, Economic, Ecological, Political and Values worlds (STEEPV). In giving an opinion an expert manipulates and integrates knowledge subjectively from all six elements of the STEEPV set. In the world in which advisers work and expert opinion is elicited, the rules of the frequentist tradition are transgressed, as large numbers of opinions usually cannot be sought. The time needed and costs involved preclude the observation of frequentist rules. More fundamentally, when dealing with the future there is no system of measurement comparable with physical measurement that frequentist rules usually require. By definition, opinions are subjective and can only be the expert’s best judgement under the terms and rules imposed by the situation. Current notions of risk assessment, presented in a frequentist fashion, are then at odds with the subjective basis of the opinions expressed by experts.

The roots of subjective probability go back to the Ancients. In their modern idiom the rules were laid out from the 1800s onwards and have given rise to an extensive literature and many methods for assessing expert opinion; these will be traced briefly, but not definitively, in the survey that forms the first part of this paper. The later parts concern practical methods; these have proved elusive to formulate as they must be readily appreciated by the interviewee, must not be offensive to him or to her and must not be simply a ‘black box’ applied by and understood only by the interviewer. These matters are of increasing importance as the role of expert advisers and expert witnesses becomes ever more important in government and in the courts where caveats and uncertainties count for little. In court a trial judge and jury are required to reach a judgement to resolve a dispute about a specific set of circumstances in spite of caveats and uncertainties. Similarly, expressions of uncertainty by experts (often scientists) are not welcomed by government whose duty it is to frame regulations, preferably unambiguously. In Experts in Uncertainty, Cooke (1991) has thoroughly reviewed the considerable literature in the field of subjective probability. In this paper, some topics will be added that Cooke does not cover including the selection of people for expert committees, as this is much more hazardous than is generally appreciated. The role of subjective opinion in foresight programmes will be constantly borne in mind, tackling the thorny question of ‘who is an expert?’ in these programmes where the phenomenon of relativism is rife, through questioning whether a non-expert’s opinions are as valid as those of an expert of international standing. Throughout, there is an emphasis on the practical use of subjective opinion as experience has taught me that there is more to using it than theory alone, however elegant that may be. The paper also spreads into other fields that either draw upon or influence subjective opinion. The first of these is court proceedings where expert opinion is often involved. The second field is the nature of preference shifts that may cause volatility in expert opinion and the third draws in matters relating to human judgement. For the narrow specialist the running together of these three fields may be either incomprehensible or anathema, but outside their world the interaction of these three fields is an everyday occurrence met with by decision makers in spheres
ranging from judges to government ministers and their advisers. It may, therefore, be considered an omission not to have dwelt at length on the burgeoning political aspects of expert advice giving that has accompanied the growth of advisory committees and of Quangos. In that field the work of Barker and Peters (1993) has much to commend it, as has the work of others in the related fields of accident and other forms of public enquiries. However, it will be helpful to record, for later interpretation of comments on the selection of members of advisory committees, that Barker and Peters believe there to be six levels of cognitive difficulty that face public policy makers; these they describe in terms of the policy field’s character as follows:

1. elaborate but not difficult detail
2. situations involving complicated (but not complex in Perrow’s terms) matters
3. situations with technical difficulty but that are amenable to non-expert study
4. problems with technical difficulty requiring expert training for their comprehension
5. technical situations bordering on the scientifically unknown and involving competing and conflicting scientific opinions
6. situations where science has nothing to offer; the subject is unknown to science and there are no claims from experts.

The first three have characteristics that make them amenable to study by non-expert policy and decision makers. Thereafter the situation becomes increasingly incomprehensible to the polity in general, whilst the last is clearly in that field where what is later called ‘conjecture’ is the only mode of thought available even to the expert community. In a telling comment, Barker and Peters point to the myth of omnicompetence that policy makers endeavour to maintain throughout all of the above six fields. In this context, they conclude that policy makers indulge in a dangerous self-delusion typified by ‘... a vain hope of maintaining the myth that all problems and issues are safely under their hand, usually with the help of expert advisers’. The dangers of this mindset are enormous when faced with the complexity of trans-science and the characteristics of the last three policy fields listed by Barker and Peters. Added to this must be the ever-shifting nature of society as it evolves from the era of modernity into something different, characterised by an ever decreasing level of trust in policy makers and their advisers.

2 Who is an expert?

The above question is the first that needs to be examined though often it is not asked at all. Here it is nothing to do with identifying people with expertise, but is that lofty inquiry implying doubt about the whole idea of expert opinion and its relevance or place in the modern world. Attitudes to expert opinion have become like Janus’s head being seen simultaneously as highly desirable and greatly mistrusted. The result in foresight studies is paradoxical, as the response to the lofty inquiry may well be that expert opinion ‘of course’ can be sought but it is not to be given any particular value by comparison with everyday experience by the public at large. In the
remainder of this paper, the existence of valuable expert opinion that needs to be
drawn upon by some process is taken for granted, but in the real world the Janus-like
nature of how such opinions are regarded should not be forgotten.

3 Expert opinion – its diversity, evolution and uses

Expert opinion exhibits diversity that has evolved over centuries. The discussion will
proceed as follows:
- subjective opinion
- the courts
- preferences
- human judgement.

For methodological issues the discussion will concern the:
- selection of ‘experts’
- elicitation of expert opinion.

3.1 Subjective probability

Often the distinction between the frequentist (‘objective’) and subjective forms of
probability is ignored to its detriment. Discussions, particularly among the polity, then
become muddled with confusing notions of what any kind of probability is about. As
the history of probability shows the subjective form is by far the older. The frequentist
form grew out of its predecessor when the notion of frequent trials to establish
distributions of likelihood became established. Nevertheless, in the search for advice on
which to act, kings, rulers, company chairmen and boards and other decision takers
cannot do other than work with the subjective opinions of a small group of advisers
for reasons including security, cost or simply the inapplicability of frequentist notions.

Cooke opened his book with the sentence “Broadly speaking, this book is about
the speculations, guesses and estimates of people who are considered experts, in so
far as these serve as ‘cognitive input’ in some decision process.” Cooke considered
subjective probability and opinion under three headings, namely, experts and
opinions; subjective probability and its underlying theoretical concepts; and the
problem of combining expert opinions. Whilst these three broad headings illustrate
the main features of expert opinion they have to be teased apart, as Cooke does, to
reveal the multiplicity of factors involved. Cooke demonstrates how these methods
have been used in some situations, but experts, advisory committees and others who
depend on opinion to make judgements or decisions may not think in the ways he
describes. Communicating the nature of expert opinion remains one of the most
difficult tasks facing anyone who works with advisory bodies, especially when they
are expected to produce advice against tight time scales and in situations resembling
those of the UK’s BSE or similar crises. Cooke’s apt observation is that “An expert
who knows more about a particular subject than anyone else may be able to talk an
opponent under the table. This does not mean that his opinions are more likely to be
true”. Experts are not omniscient, but have their own agendas, predilections and prejudices that become embedded in their opinions.

It would be hard for any facilitator to introduce the many themes underlying expert opinion to any advisory body. Rather the facilitator’s role must be to embody them, ‘black box’ fashion, so that probabilistic thinking, the nature of biases, heuristics and the combination of several sources of opinion are incorporated but not exposed directly. The ‘black box’ problem is the kind that Budescu and Rantilla (2000) approached in a study of how a decision maker aggregates the opinions of multiple advisers and their confidence in the outcome. The problem surrounding the acquisition of knowledge from multiple experts has been explored by Medsker et al. (1995). Their concern was not related primarily to opinion as subjective probability but with the many social aspects such as privacy, geographic dispersion and how computers may be used to automate parts of the process. The latter topic had been explored much earlier by Lipinski et al. (1973) with much the same objectives, to make expert participation independent of geographic separation as well as enabling the elicitation process. Multiple sources of opinion naturally provoke the question of how they can be aggregated. One approach was described by Lipinski and Loveridge (1982) whilst another, the analytical hierarchy process (AHP), has been described by Saaty (1980), Xu (2000), Zio (1996) and Basak (1998) have worked with AHP and the latter claims to have developed a new approach to eliciting and synthesising expert opinion for group processes within the AHP framework.

Hull (1976) reviewed the use of subjective probability for risk assessment in relation to major capital investment under seven methods:

1. fixed interval
2. variable interval
3. relative likelihood
4. psychometric ranking
5. equivalent prior sample
6. hypothetical future sample
7. other.

In the fixed interval method the assessor is asked to state the probability estimate for each interval in a range that covers all the possible values of an uncertain quantity. The order in which the intervals are presented to the assessor is thought to be influential and Huber (1974) has proposed a procedure to overcome this concern. In contrast, the variable interval method requires the assessor to identify a number of intervals of equal probability that cover the entire range of the uncertain quantity. Morrison (1967) has outlined an appropriate procedure, sometimes called the method of successive bisection, but whilst the method can become tedious as greater precision is sought, it is an approach favoured by Peterson et al. (1972) and Murphy and Winkler (1973).

The relative likelihood method expects the assessor to respond to questions to reveal the relative probabilities of different values or ranges of values. Again, procedures have been described by Winkler (1967) and by Barclay and Peterson (1973). The remaining methods do not fall within the context of the present paper.
DeWispelare et al. (1995) describe the use of elicitation to generate cumulative distributions representing the uncertainty in predictions in the USA high-level nuclear waste regulation programme. In a very different field, Dransfield et al. (2000) used weighted expert opinions, in conjunction with a Delphi survey, to study the future of interactive television and retailing.

Lastly, there is the use of subjective probability in scenario development; there are many threads here that have surprising antecedents. Perhaps the most surprising arose in discussions between de Finetti, Keynes and Jeffreys (de Finetti, 1985). Among many claims in the dialogue, Jeffreys’ claims that probability tells us what we ought to believe, might make a connection with other theories of belief enunciated by Shafer in his development of a mathematical theory of evidence. It will also connect with notions from other fields with far longer claims to proposing what we ought to believe. How then do these thoughts relate to the use of subjective probability in scenario building?

Scenarios display sets of beliefs in logical or pseudo logical order; they are to an extent an art form underlain by the theory of subjective opinion. For all that has been written about and in scenarios, very few of them pay overt attention to this lineage and almost none has been prepared using it explicitly. Some authors have been concerned with probability assessment procedures (Bunn, 1979a,b) but not necessarily in ways that relate to scenario building (Bunn and Salo, 1993), whilst Bunn and Salo have been concerned with scenarios and forecasting. However, there are at least two scenario based studies that are based explicitly on the elicitation of subjective opinion; these studies enabled the construction of scenarios for the earth’s climate and the future of the UK. Both used a similar approach in which experts were asked to respond to a defined set of questions through an elicitation process.

For the climate study ten questions dealt with specified climate variables whilst for the UK study nine variables were specified for use in an economic model. The elicitation sought information about the specified variables in three ways: probabilistic forecasts on each variable, the reasons for the quantitative estimates and self and peer expertise rating. The characteristics of the experts in each study were threefold (Lipinski et al., 1973):

1. substantive knowledge in their chosen spheres of interest
2. assessing ability to relate how their sphere may evolve in the future
3. imagination, as this lies behind how the adviser extends his or her substantive knowledge into the future and subsequently assesses it.

Calibration tests exist for the characteristics that ‘measure’ 1 and 2; 3 cannot be assessed directly. However, it is also important that the potential adviser assesses his own level of expertise according to some simple but well defined rules (see Lipinski and Loveridge, 1982) as a further part of the procedure. Lipinski and Loveridge used their self-evaluation of expertise criteria extensively and a similar set, modified to take account of spheres outside science and technology, was used in the 1994-95 UK Technology Foresight Programme (Loveridge et al., 1995).

The climate study used a structured questionnaire. The data for each respondent was converted into a probability density function (PDF) relating to the change foreseen in the particular variable. The PDF was adjusted for the respondent’s expertise weighting and the weighted PDFs were added and normalised. The panel’s
responses to each climate variable were then combined into a set of scenarios that covered the range of uncertainty or range of conditions described by the respondents.

In the UK study a similar but more developed version of the climate study methodology was used; this is partially described later. The refinements were mainly concerned with the calibration of the experts so that their elicited opinions could be converted to approximate a notional ‘perfect assessor’. Generation of the scenarios was highly computerised enabling the interpolation of ‘missing pathways’ through the tree structure involved. Whilst the climate study was placed firmly in the public domain, the outcome of the UK study has remained confidential, only a partial revelation of the methods being permitted by its sponsors.

Making the process of elicitation acceptable to the expert who is to be interrogated has not been discussed here, but will be later as its importance should not be overlooked. Public foresight programmes have so far shown limited understanding of the nature of expert opinion on which they depend. Many are conducted by panels selected by unknown processes and whose members’ claims to expertise often go unexamined. In a few studies attempts have been made to introduce some of the elementary notions of the elicitation of expert opinion, but these have not gone beyond the notions of the self-assessment of expertise. In only one case (Loveridge et al., 1995) did this introduce well-structured criteria, developed from those due to Lipinski and Loveridge (1982), for the respondents to assess themselves against. The outcome was used only to set a minimum level of expertise for inclusion of the respondent’s opinions, a matter referred to obliquely by Dransfield et al. (2000) in their work, but weighting of opinions was not introduced. All foresight studies eventually become political instruments where a practitioner’s wishes for deeper understanding have to be subsumed into a world that sets different criteria and has less patience. The best the practitioner can do is to infect this different world with the ideas of elicitation of expert opinion so that they spread to influence the use of the insights that elicitation and subjective opinion can offer in decision making.

3.2 The courts

The use of expert witnesses has long been a feature of court proceedings. Foster and Huber (1999) have written trenchantly on the role of scientific knowledge and of scientists as expert witnesses, a timely reminder that the pretensions of both must now face the judgements of the world through judges in their courtrooms and the legal process. The authors analysed the *Daubert vs. Merrell Dow Pharmaceuticals Inc.* case in depth to illustrate the various aspects of the use of scientific knowledge in court and the role of scientists as expert witnesses.

*Daubert* illustrates how the nature of scientific evidence may prejudice, confuse or mislead a jury depending on how it is presented. It also exposes how scientific evidence, in relation to the judicial process, may or may not fit with:

- the social dimension
- notions of testability and falsification
- errors in science, reliability and scientific validity
- the role and efficacy of peer review in the scientific community.
The conduct of a trial is antithetical to the purpose of science and to the experience of scientists used to uncertainties even in those areas where core knowledge is believed to be well established. Weinberg’s (1972) notion of trans-science hinges on the phrase ‘... yet cannot be answered by science’ linking time and the associated advance of understanding to the constantly shifting boundaries between questions that can be asked of and answered by science. The formulation of policy is mostly based on the opinions of expert advisers particularly if it proceeds in a quasi-legal fashion. As science and technology intrude ever more deeply into sensitive areas of life and society, judges and juries may frequently define what is expected of scientists and of science. The only certainty is that the frequency of these clashes will increase exposing science, its processes, scientists and the workings of expert advisory committees to public examination. Elsewhere (Foster et al., 1993) Foster and Huber, with others, explored other cases in which their emphasis was on the intersecting themes of the problems of assessing subtle environmental or occupational risks and the havoc this creates in the courtroom.

The foregoing concerns the nature of evidence and in this field Shafer has developed a mathematical theory of evidence (Shafer, 1976). Shafer presented his work as a theory of evidence and of probable reasoning. Bayesian theory is drawn on extensively whilst the difference between chance and ‘partial belief’, which is represented by subjective probability, is emphasised. Shafer avoided the word ‘probability’ as far as possible because of the misunderstandings he claims that it introduces. The notions of chance and degrees of belief are separated, allotting to chance the characteristics of frequentist probability and the subjective variant to degrees of belief. Whilst conditional probability is introduced as a minor effect in the first field, the act of conditioning is of clear importance in influencing degrees of belief, an important point in expert evidence, its presentation and manipulation in court. However, there is a subtle relationship between chance and partial beliefs, since the latter can be influenced if the frequentist probability distribution of some component of belief structure is known. Shafer does not pretend that there is an objective way in which given evidence and a given proposition can be represented numerically, nor does he believe that a degree of belief can similarly be represented. Rather, he proposes that an individual can pronounce, as a matter of judgement, a number that represents a degree of belief about a proposition that is vague, fuzzy and perceived in a confused way. The relationship here is to the inevitable courtroom question to attest a level of probability to a proposition that may subsequently influence the court’s judgement of the case.

The notion of partial belief axiomatically admits ignorance, a point that Shafer deals with explicitly. It is also a point that can be skilfully exploited by an advocate without the expert witness realising the danger of the situation. Also important is the difference between lack of belief and disbelief in relation to ignorance. Throughout, Shafer is careful to distinguish between objective and subjective probability which, when neglected, creates much misunderstanding in courts and advisory committees; the latter has relevance to foresight programmes, their formulation and the implementation of their outcomes.

Williams (1978) reviewed Shafer’s work and concluded that it “... has considerable appeal and naturalness.” and that “... we owe Shafer a debt of gratitude for developing a systematic and comprehensive alternative to Bayes
theory”. The influence of Shafer’s work in the world of practical decision making remains uncertain but Mayo and Hollander’s (1991) confronted the notion of ‘acceptable evidence’ directly.

From case studies, Mayo and Hollander concluded that despite the attention given to the role of values in determining acceptable risks, little attention had been paid to the essential step of examining what type of risk evidence is acceptable. It may be allowed that values influence the acceptability of risk, but that the associated evidence is risk-free, because it is scientific, objective and therefore uncontroversial. In contrast, risk evidence may be seen as permeated by value judgements to the extent that there is no one type of data more objective, valuable or worthy than another, a recourse almost to relativism. Mayo and Hollander were clear in their belief that values influence risk assessment and management, while maintaining the possibility of developing tools for appraising critically the evidence on risk issues. Risk evidence is shot through with value judgements that are socio-cultural and political, but these do not make it meaningless, a practical and theoretical point already encountered in Foster and Huber (1999) and in Shafer (1976). Uncertainty in risk evidence and ways of dealing with it, lie at the heart of Mayo and Hollander’s discourse and of the case studies contributed by different authors. However, only Slovic discusses communicating risk to the polity and goes into any detail of the practicalities involved. Mayo and Hollander’s focus is on the ethical aspect of risk evidence; there are connections to be made with Shafer’s theoretical work, but rather less so with the courts and advisory committees. For the latter there may be more insights to be found in Vickers’ discussion (Vickers, 1963), now some 40 years ago, of what he called appreciative behaviour. Perhaps even more relevant is the series of discussion meetings held over several years by the UK’s Royal Society concerning the perception, analysis and management of risk (The Royal Society, 1981, 1992, 1997) and recent work by Stirling and Meyer in relation to genetically modified crops (Stirling and Meyer, 1999).

3.3 Preferences

We all exhibit preferences; expressing them is the task of all advisory committees; it is their reason for existence. Preferences are closely aligned to Shafer’s degrees of belief and to his notions of how changes in opinion occur to accommodate new evidence. However, the underlying theory of preferences comes from a different stable. In its modern incarnation, the theory of preferences stems from Savage (1954) and is underpinned by symbolic logic. The interpretation put on preference theory here is related to the phenomenon of foresight and to the formalism of public foresight programmes.

Life is made up of events, each of which is itself a set of states; the event that is a single independent state is simply a myth. Similarly, the event that has no states as elements of its set is called vacuous since logically this cannot be accepted. Savage’s example of a vacuous event is the notion of one that is simultaneously good and bad. Clearly, these notions are those of set theory in which Venn diagrams figure strongly expressing, as they do, intersection and unionisation of sets. Venn diagrams are powerful in understanding preferences, their limitations lying in the rapid increase in complication once, say, six intersecting sets or more are involved. Three-dimensional
Venn diagrams have been developed though it is not suggested they can be of use in the foresight studies.

A preference is expressed in the course of making a decision which, on every occasion, implies a choice between two or more acts, each with its own consequence(s). Theoretically, the choice of an act requires that the entire set of possible states of the world be taken into account and also the consequences that are implied by that choice for each of the possible states of the world. These conditions may be met in a small world of a few states, but in the real world in which foresight occurs the theoretical conditions cannot be adhered to if only because of Simon’s (1957) condition of ‘bounded rationality’. A further complication lies in the future uncertainties regarding the consequences of a choice among acts since these, by definition, cannot be specified entirely. Nevertheless, in the real world choices between acts are made and the consequences are lived with. Additionally, the distinction between preference and indifference is important both theoretically and practically. The difference implies a reason for shifts between the two conditions; this is related to the influence of new evidence, where there is also a concern for preference ordering. In the latter any numerical values attached to preferences are not important, it is their ordering that matters, as illustrated by Casti (1992) in his discussion of the 1973 alert crisis that involved the USA, the USSR and the Middle East countries. Savage describes two ways of looking at shifts of this kind. The shift from indifference to preference may be due to the individual recognising that new evidence brings some form of bonus to one or more of the component states. But how does the individual recognise this bonus for what it is? Alternatively, an individual may know introspectively whether an act has been chosen haphazardly or from a definite feeling of preference. The latter is an issue that can arise in priority setting during foresight programmes, due to simple mental exhaustion accompanied by an inconsistent expression of preferences among acts typified by circularity.

Savage examined two proverbs ‘look before you leap’ and ‘you can cross that bridge when you come to it’ as two extremes of policy. The first proverb may be interpreted as requiring anticipation. The act of foresight must envelop the entire set of events and states, in detail, of the future world for that individual and for the selection of one policy for life from the entire set of possible policies for living in the anticipated world. Savage regards this extreme as ‘preposterous’. In contrast, the notion of ‘crossing bridges when you come to them’ denies any form of foresight and is entirely reactive, limiting decisions to an artificially simple and small world where responses are made to events as they are encountered. Savage is “… unable to formulate criteria for selecting these small worlds” believing that “… their selection may be a matter of judgement about which it is impossible to enunciate complete and sharply defined general principles”. Lastly, in this brief excursion into Savage’s work is the idea of admittance which states that if “… several individuals agree among themselves on their preference among consequences, then they must also agree in their preferences among certain acts”. In the current context the relevance of this notion to foresight is clear.

The part played by psychology in preference formation is self-evident. However, preferences and their reversals are dwelt on by several authors in widely different situations. Among these are Lichtenstein and Slovic (1971) who first found evidence of preference reversals in laboratory experiments and later in casino experiments in
Las Vegas (Lichtenstein and Slovic, 1973). Hogarth (1988) similarly discusses preference reversal. Essentially, preference reversals between a pair of choices occur when the probability of the one originally preferred falls below the probability of the alternative. Psychologically, the committee member or expert adviser will make similar assessments either consciously or unconsciously of both the value, which is not necessarily monetary and probability of the outcome of expressing his preference and how that may change under different circumstances. Whilst this conclusion may seem ‘common sense’, because of the frequency with which preference reversals are encountered in everyday life, it becomes more problematic when it is allied with transgressions of the transitivity principle (intransitivity) where illogical preferences are displayed. The transitivity principle requires the absence of circularity among a series of preferences; for example a preference for A over B and B over C should imply a preference for A over C. Intransitivity, where this kind of sequence does not occur, is a common problem, but is important in expert or advisory committees where it requires deliberate recognition and management. Intransitivity was one of Savage’s concerns and should be of concern to anyone involved in formal foresight programmes, particularly when the time comes for prioritisation.

In the outcome of a workshop held at IIASA in the late 1970s Bell et al. (1977) trace the steps to the development of ways to assess and model a decision makers’ preference system. The emphasis is on the construction of decision rules usable in multicriterion problems using heuristic and axiomatic rules; one-step and multi-step rules; and rules leading to full or partial ordering of a set of feasible alternatives. The steps in building complete or partial models of preferences were set out. However, in the closing debate Rivett somewhat testily made the comment that elegant though all the methods described were he rarely met them in practical use, a feature that the expert adviser or committee member cannot avoid since actionable advice is expected of them.

Lastly, preference measurement is often a feature lying behind marketing. In this context Green et al. (1972), Watkins (1984) and Stanek and Mokhtarian (1998) are examples drawn from different eras and different fields. It may seem that much of preference theory boils down to common sense. If that were the case there would be little to be concerned about. However, events in the practical world reveal the presence of intransitivity and preference reversals as commonplace; this evidence alone indicates the need for watchfulness in the areas influenced by expert opinion however it is gathered. For the conduct and implementation of foresight programmes this need is obvious.

### 3.4 Human judgement and bias

The purpose from the outset has been to relate subjective opinion and its probabilistic representation to the current rash of interest in what has become called foresight, so matters that are judged not to relate to that phenomenon are omitted. In foresight programmes, public and private, the time-scale is hardly less than ten years and is often 20 to 30 years, ruling out short-term judgements. In addition, foresight is concerned with possibilities that result from conjunctions and interrelatedness of events and trends, so notions that relate to single events are demoted.
The exercise of judgement is commonplace and is taken for granted, though in some instances its exercise is not immediately welcome or accepted. Vickers (1981) said that “Since the Second World War the most rapidly expanding field of professional study has been the field of human governance”. Foresight has long been a part of that field though often only by implication. Vickers later goes on to quote how Churchill’s decision in 1925 for Britain to revert to the gold standard at its pre-war parity was taken against his judgement because the opinions of all his advisers unanimously favoured the move. Strongly aired opinion, even when based on substantive knowledge, need not turn out to be the most valuable. If subjective probability is the public expression of beliefs, as Jeffreys, Shafer and others might maintain, the factors that lie behind these expressions spread widely into many domains of science and sociology. The connections between the underlying notions of the workings of the human brain, the sociological and ethical notions of values/norms that characterise human behaviour and probability have proved elusive. The point was made cogently by Hammond and Adelman (1976) who saw judgement as the key element in integrating values and science. Now, in a world that 25 years later is even more riven by disagreements over what can be done and what ought to be done, that conclusion is even more important.

There is a long running assumption that human judgement is highly fallible when compared with the prescriptive models of the judgement process. So ‘Is human judgement so fallible that it should be discarded?’ is a question examined by Beach et al. (1987); their conclusions are instructive:

- good judgements are rarely reported (this is the media ‘bad news is news’ phenomenon)\footnote{44}
- experiments on judgement assessment are often contrived and unrepresentative of the real world
- real world assessors often frame tasks in very different ways from experimenters (revealing a difference in task perception)
- the problem framework is often unclear making performance criteria unclear.

Beach et al. point to the “... bewildering array of ‘Biases’, ‘Heuristics’, and ‘Fallacies’ that serve to justify the negative view of judgement and reasoning”. However, they conclude that in recent years there has been “... a change in the tone, if not the methodology, of work in this area ... change emphasises that the models’ inadequacy is the issue of interest rather than the inadequacy of the behaviour”. The change has come late in the day but reflects a major shift that occurred elsewhere in the acceptability of modelling complicated processes. Beach et al.’s final conclusions could hardly be plainer that:

“... the question of the general quality [of human judgement] would become meaningless because it is meaningless. This in turn would allow us to ... get down to work on the ... substantive, problem of how to properly integrate human judgement and reasoning into advanced technologies.”

“... the question of the general quality of judgement and reasoning not only is not settled, it probably never will be and it probably never need be ...”
Both statements are supportable\textsuperscript{13} from experience. The relevance of Beach et al. to foresight and to advisers is clear:

- the use of judgements is inevitable so recognising the fallibility of both judgements and models of decision is important
- that possibly the models are more fallible than human reasoning and that the use of complex computational procedures does not sanctify the outcome.

Human judgement is clearly related to expert opinion; it is what the expert uses when he gives his opinion. Hogarth (1987) catalogues concerns for the fallibility of human judgement, whilst acknowledging that despite these fallibilities human judgement has so far ensured survival of the race. Hogarth examines human fallibility by asking why it is that humans who can and do regularly make considerable physical and conceptual achievements exhibit systematic biases in their judgements. Judgement is concerned with adaptation to survive in the future. Human society has been through several adaptations in which physical survival had to be ensured before mental and conceptual skills could be developed; the similarity with the base of Maslow’s hierarchy is itself evident. The relationship between human judgement and foresight is also self-evident; so what can be expected of human judgement?\textsuperscript{14}

Hogarth says judgements involve biases that result from “... trade-off between different types of error in the design of the human system”; thus the following contributions to bias can be identified:

- Failure to appreciate or identify randomness.
- Inconsistency in judgements made across time; in foresight studies inconsistency over the span of a study of limited duration requires explanation. Volatility between studies is to be expected.
- Learning and the willingness to learn are important aspects of foresight.
- Memory inadequacies and imperfections; predisposition to selective recall and perception, leads to interpreting information in ways to support underlying expectations and hypotheses about the future.
- Computational capability of the human mind is limited.
- Order can be brought out of chaos but this may limit creativity.
- Cognitive myopia, the unwillingness to use imagination or to conduct thought experiments about the future.

Judgement enters into scenario writing and that is the next field for consideration. Reasoning about the future is more art than science; recognition of this point was the reason for Wright’s (Harries, 1999) twofold attack on decision analysis. First, Wright declared that scenario development was creative and based on causal reasoning, whereas decision analysis did not encourage speculation about “... all possibilities”. The last reference is unfortunate since it is a psychological and logical impossibility unless the boundaries to the circumstances\textsuperscript{15} are declared. Wright’s second attack was that probability calculus made decision analysis “... unwieldy and less useful than scenario planning”. Creating visions of the future is the purpose of
scenario planning to which foresight is an essential adjunct. The separation of functions is not pedantry since the introduction of scenario building too early inhibits the search for trends, issues and events that are likely to be of importance in shaping the future and weakens thinking about highly uncertain and contentious matters that should rightly be subject to causal reasoning.

Scenario preparation is not a scientific process, but is an art form with some underlying theoretical principles and association with the disciplines of causal reasoning. The popularity of scenarios lies in their perceived ability to capture the complexity of the real world allowing the nested structure of scenarios within scenarios to be presented. ‘Nesting’ visions or scenarios is common when the future of one entity arises from it being a subset of a wider one.

It is obvious that foresight depends crucially on expert opinion for its substance and vice versa. Some studies seek to understand the substantiveness of the expert opinions they obtain and exhibit in their output. Those studies that ignore this point are likely to be opinionated ‘hand waving’ and the output should deservedly receive limited credibility. Human society being what it is, these strictures often go unheeded so that the selection of experts often proceeds by patronage with consequent effects in the advisory sphere. It seems, to use Holling’s (1977) notions, that unlike engineering systems where the ‘fail-safe’ principle dominates design, human society remains ‘safe when it fails’ through processes of adaptation that lie well outside the domain of subjective probability.

4 Methodological issues – opinion and foresight

There are two practical issues associated with the question ‘who is an expert?’; these are how to find people who are considered to be experts and having done so, how to make the elicitation process acceptable to those involved. The first can be coped with by the co-nomination process (Nedeva et al., 1996) used to find a large number of experts to take part in the Delphi study in the UK Technology Foresight Programme (TFP) of 1994-95. For the second, the practical process used to elicit opinion from a group of experts, in a study of the future of the UK (Lipinski and Loveridge, 1982), will be described. Chronologically these two events were widely spaced; the co-nomination process was used almost 20 years after the study of the future of the UK. If co-nomination had been known earlier it might have made the UK study much easier, but that is mere speculation. Here the two methods will be described in their correct technical sequence.

4.1 Co-nomination – a procedure for finding a group of experts

Finding people with demonstrable expertise is difficult; finding a lot of them is many times more so. There are ways of finding special people in any society by ‘asking around’ after having identified a starting point. The process has a certain inevitably about it tending always to lead towards particular groups of people recognisable anywhere as ‘the great and the good’ of the particular society or community. These recommendations will be vouched for, whilst it is only sensible to do some ‘due diligence’ on each one. Alternatively, the starting point can be with the appropriate
professional institutions and their membership lists, but after a while there is a gravitation toward the ‘asking around’ principle following recommendations from the institution itself. Outwardly, there may seem to be nothing wrong with this process which is used, more often than not, in filling positions in advisory committees and similar bodies. At this point it is as well to look back to the characteristics of an expert set out in the earlier discussion of subjective opinion. With these and the accompanying self-assessment criteria in mind, the ‘asking around’ process does not show up well, particularly if one adds that the search process should, as far as possible, find a slice through the demographic variables of age, gender and occupational position. The co-nomination procedure described here aims to respond to all of these criteria, but it can be thwarted by the difficult step of identifying an initial group.

There were three reasons why the co-nomination method was used in the UK TFP:

1. Political advice that the TFP should seek advice from people beyond those already advising government, especially as this was the first nationwide study of this kind to be held in the UK.

2. Many study panels were to be formed for which members would be needed.

3. The planned Delphi survey would require a correspondingly large number of potential respondents with known characteristics. The group could not be identified in any time-honoured way; this had been demonstrated by experience elsewhere, such as in Germany.

The origins of co-nomination lie in bibliometrics and in the mapping tools used either in classifying clusters of researchers or in identifying academic–industrial networks of researchers (Georghiou et al., 1988). Co-nomination is a survey process that starts with a selected group of people who are thought to be representative of the wider group to be explored. Each of these respondents is asked to identify up to a maximum specified number of people who meet defined criteria, so that if there are N originals each of whom nominate up to a maximum of M others, then the sample grows to a maximum of N*M people. The further group of M is then asked to repeat the nomination process under the same rules with further rounds as necessary (this leads to the notion of ‘snowball’ sampling).

The underlying principle of co-nomination lies in the generation of a network based on recurring pairs of names (revealed from the questionnaire). It is assumed that similarities in the nominated persons work and that of the co-nominees implies a cognitive link. So if:

- A nominates B, C and D
- the presumed links are B-C, B-D and C-D.

If nominations recur in two or more responses, then the likelihood of these links is increased. Nomination itself is an indication of the influence the nominee is having on the respondent (the nominator) which leads naturally to the hypothesis that frequent co-nomination identifies influential actors in a field and through a network illustrates their interrelationship. As the questionnaire makes no reference to the quality of the work of the individuals concerned, the notion of a ‘popularity contest’
is avoided. Co-nomination rather than simple nomination is preferred since it avoids ‘prestige’ effects where respondents link themselves to ‘key’ actors whilst it also enables a scale that allows the strength of links to be examined at different cut-off levels. Those people who are highly co-nominated clearly occupy special positions in the community involved. In this way the identification of potential participants is placed firmly in the hands of the community involved.

The process is outwardly straightforward, but as ever requires much attention to the detail of questionnaire design and survey management. These were increased many times over in the UK TFP by the multiplicity of panels and associated sectors around which it was organised. It is best to start with the design requirements of the questionnaire, ignoring the complications of the UK TFP. When completing their questionnaire the respondent:

1. Gives clear and exact demographic details including title, address, age, gender, affiliation, job role (chosen from a set of defined categories) and other similar information.
2. Declares their field of expertise and level of expertise in it chosen from the self-assessment of expertise schedule (Loveridge et al., 1995).
3. Lists, up to a maximum specified number, people whose work has influenced the way the respondent does their work and estimates the level of expertise of each nominated person; immediate colleagues in their organisation are excluded.
4. Similarly, lists a specified maximum number of colleagues who have influenced the way the respondent works (this is usually a much smaller number than in 3 above).
5. Suggests the name of one person whom they believe to be of outstanding influence, not status.

The questionnaire is intended to be completed by the individual and not under supervision nor in a group setting. The purposes of steps 1 and 2 are self-evident. Steps 3 and 4 are intended to prevent an individual from playing the ‘promotion game’ by quoting colleagues only. Step 5 provides the opportunity to list someone who may lie well outside the field concerned but may still be of outstanding influence.

The survey requires a rigorously constructed and managed database. Two common problems are:

1. The necessity to ensure that each respondent nominated more than once receives only one questionnaire to avoid bias and to avoid giving the impression that the survey is not being well managed and is out of control.
2. Misspellings of names, which could also introduce errors into the process, must be eradicated along with alternative groupings of initials.

The application of co-nomination in the UK TFP is a marker that separates that programme from any other to date. It is hardly surprising that some practical problems were met with and some lessons were learned for the future use of co-nomination to identify a large body of experts; for example:
For an exercise as complicated as the UK TFP, constructing a representative initial list of respondents was difficult and needs to be improved in future.

The time-scale was not ideal. It allowed time for only two iterations of the questionnaire and only a short time for analysis of the outcome.

During later parts of the programme, respondents to the co-nomination survey had a higher propensity to participate than people who were added to the database of participants by other means, perhaps indicating that they felt more in touch with the programme than later additions.

Since the use of the co-nomination process in the UK TFP one of the UK research councils proposed using the method to identify peer-review colleges but time pressure prevented this. More successful was the way co-nomination was used in the South African foresight programme to identify expert participants where political and ethnic constituencies had to be seen to receive due recognition. Recently, in planning a second round, the Swedish programme organisers have been considering using co-nomination to avoid the problem of closed panels experienced in their first programme. The outcome of the co-nomination process also has clear benefits in determining appointments to panels and advisory committees as it has the potential to reveal who is an expert in undeniable terms. However, membership of panels and advisory bodies brings with it the exercise of power, an attraction in itself, as well as, or perhaps even more than expertise, whilst rational methods of assessing who is an expert may also deny patronage to those in whose hands the power to appoint lies. To quote Cooke again ‘An expert who knows more about a particular subject than anyone else may be able to talk an opponent under the table. This does not mean that his opinions are more likely to be true.’

### 4.2 Elicitation of expert opinion – a practical approach

Historically, the methods described here grew out of a partial dissatisfaction with the Delphi and cross-impact methods used in technology forecasting. The first steps were made in the method used in the climate change study, conducted under a DARPA contract, referred to earlier. In that study the first steps in eliciting subjective opinion were introduced via a questionnaire. The methods described here were an improvement on those used in the DARPA contract and were used in a study of the future of the UK made in 1978. It was a particularly apt time for the study as at that time many studies had portrayed the UK as the ‘sick man of Europe’ with a bleak, if not hopeless future. The study was delivered to its sponsoring clients at the end of the infamous ‘winter of discontent’ and within weeks of the 1979 General Election. The study’s 17 clients from the UK and the USA agreed to the partial publication of the methods used (this is what is drawn upon here) but nothing more.

All questionnaire-based elicitation lacks face-to-face contact. The opinions expressed by the respondents are necessarily ‘black boxes’ and remain so unless some subsidiary interviewing takes place. However, the use of questionnaires is not necessarily in tune with the time scale of corporate decision making nor is it secure enough if opinions are sought widely. For government the first condition applies but
the notion of secrecy is being challenged frequently in the industrialised countries. Since governments and companies now face issues that more often than not have the characteristics of trans-science (Weinberg, 1972) how can they proceed to gain benefit from expert opinion?

It would help if they followed the dicta of co-nomination to find people to consult and those of the elicitation of subjective opinion in those consultations, but needless to say they do not. The first requirement of any practical elicitation process has to be its acceptance by people who do not normally think in terms of uncertainty let alone probability. The process outlined in the ensuing paragraphs was intended to meet this problem and was used in some 100 face-to-face elicitations with varying degrees of success (no interview failed its purpose, but some were distinctly ‘difficult’). It is generic and was used in each interview relating to each variable identified by the clients. The principles behind the procedure will be described before setting out the procedural steps.

4.3 Assessing ability

As described earlier, the ‘perfect’ respondent possesses three attributes:17

1 substantive knowledge in a particular field
2 the ability to cope when faced with an uncertain extension of substantive knowledge into the future
3 imagination.

How a respondent thinks about the future depends on the way he or she extends his or her knowledge into a possible but uncertain future. Unprepared, the respondent may perform badly at this task so an assessing ability test was used firstly, to alert each respondent to the mode of thought expected during the later elicitation and secondly, as an attempt to characterise how the respondents coped with their inner uncertainties.

The test tries to measure how respondents allocate probability to uncertain values of a variable. Knowledge in this field was slight in 1978 and probably still is, so no claims for the test are made beyond empirical experience – it seems to work. An important outcome of the test is the subsequent capability to:

* convert all opinions to a common basis of ‘perfect assessment’
* to merge them according to the rules of objective probability.

The test is carried out in two parts as follows:

Step 1 The respondent replies probabilistically to ten questions from the Guinness Book of Records (Morris, 1979). The respondent is asked to quote a high and a low limit so that there is an 80% probability that the answer quoted in the Guinness Book of Records lies between these limits. Whilst the questions are likely to be outside the respondent’s special knowledge, their general knowledge should enable them to set most limits correctly;18 the target for a good assessor is eight out of ten and this is a difficult task.

D. Loveridge
Step 2  The respondent replies in the same manner to a further ten questions, but this time they are drawn from a publication such as the UK's *Social Trends* or a corresponding publication elsewhere. However, in this step the respondent is asked to give two sets of responses corresponding to the 50% and 80% probabilities of embracing the correct result.

‘Perfect assessors’¹⁹ would score eight and five out of ten correct results at the 80% and 50% probability levels respectively. Practical experience shows that experts most frequently capture three out of ten correct results at the 80% level on the first occasion and that with practice scores can be improved consistently; there is a lesson here about the conditions under which all expert advice is sought and given. These points are illustrated in Figures 1 and 2.

**Figure 1** Probability allocation in the assessing ability test. With acknowledgements to Lipinski and Loveridge (1982)

![Probability allocation diagram](image1)

**Figure 2** Distribution of assessing performance. With acknowledgements to Lipinski and Loveridge (1982)

![Distribution curve](image2)
4.4 The elicitation process

During the interview the expert is asked to respond interactively to questions that enumerate a nine point cumulative probability distribution. The distribution represents his or her substantive knowledge, simultaneously tempered by his or her imagination and how he or she copes with subjective uncertainty as imagination leads from the present (where substantive knowledge is valid) into the future (where that knowledge may not apply). It is no accident that this part of the entire interview is placed centrally. Because large numbers of opinions cannot be sought, as would be the case in the frequentist tradition and because when dealing with the future there is no system of measurement comparable to physical measurement, opinions are subjective by definition. The elicited data are the respondent’s best judgement – his or her opinion – about the variable under discussion. Elicitation raises a linguistic problem discussed below and shown in Figure 3.

Figure 3 Extraction of nine point cumulative probability distribution. With acknowledgements to Lipinski and Loveridge (1982)
Respondents often find it difficult to relate to probability levels without a reference framework, a problem recognised by Alpert and Raiffa (1969). The problem was explored by Lipinski et al. (1973) and has latterly caught the attention of other investigators including Beyth-Marom (1982), Delgado et al. (1998) and Chakraborty (2001). However, in the procedure described here Alpert and Raiffa’s terms were the best available and probably remain so. Alpert and Raiffa’s work on word associations is summarised below (Table 1):

<table>
<thead>
<tr>
<th>Probability level</th>
<th>Word association</th>
</tr>
</thead>
<tbody>
<tr>
<td>99% and 1%</td>
<td>Astonished or astounded</td>
</tr>
<tr>
<td>90% and 10%</td>
<td>Surprised</td>
</tr>
<tr>
<td>50%</td>
<td>Even bet</td>
</tr>
</tbody>
</table>

These word equivalents were used often during the interview procedure and there was little doubt that respondents found them helpful in orientating their thinking. As there were no easy word associations with the 75, 66, 33 and 25% intervals, respondents found these more difficult to deal with. There are similarities to the variable interval method discussed earlier, but it is also a demanding step to reason out the difference in opinion between a tercile and a quartile. There is no doubt that these responses were sometimes guesswork born of frustration at the demands of the mental reasoning which was detectable in the subsequent processing of the data.

Whilst the elicited nine-point distribution conveys succinctly the respondent’s opinion about future uncertainties, how respondents reached their conclusions is derived from their ‘thinking out loud’ during this stage of the interview. Useful as it is, the cumulative distribution hides important information that can be derived from its probability density function (PDF). The PDF can be interpreted in the following ways:

- A normal distribution indicates that the respondent’s opinion is based on a single dominant trend. If that peak is sharp and encloses a high percentage of the area under the curve, that implies a great deal of certainty in the respondent’s mind about the future of the particular variable.

However, single peaked functions are rare; this leads to the second and more complicated, but at the same time more interesting and more valuable interpretation as follows:

- Where the PDF exhibits more than one peak, this indicates that the respondent has a mental model of the future for the variable and a way of dealing with the uncertainties that surround it; that leads to the conclusion that more than one dominant trend is possible which is important information for decision makers.

Several independent PDFs are valuable each in their own right, but it is usually more helpful to merge the distributions to create a single PDF that represents the joint opinion of the ‘n’ individuals. In this step, the second part of the assessing ability test is used to account for the differing levels of expertise and differing assessing abilities.
of individual respondents. If wished, the expected value of the joint opinion can be calculated either as a single value for the entire distribution or in, say, three equally likely segments.

4.5 Elicitation process – the practical procedure

The respondent received a description of the interview process in advance which was accompanied by general information about the study so that each respondent could understand his or her role in the study as a whole. On the day, the sequence of events proceeded as follows:

Step 1 Respondent was asked to use the self-assessment of expertise criteria to assess his or her level of expertise.

Step 2 Experience relating to substantive knowledge, assessing ability and imagination was explained. The respondent then undertook the first part of the assessing ability test, based on the Guinness Book of Records, to alert the respondent to their imperfect treatment of uncertainty as they often allocated limits that were too narrow (see Figure 2).

Step 3 The respondent was then taken through the elicitation of the nine point cumulative distribution. Initially the respondent was given information about other impinging variables (if any). The respondent was guided through the elicitation, being encouraged to ‘think out loud’ so that the variables and events being manipulated mentally in framing his or her response could be noted and recorded along with his or her estimate of the value of the variable at each probability level. The interviewers and interviewee were in almost constant dialogue throughout this step for which there was no time limit.

Step 4 The second part of the assessing ability test was administered last and completed the interview. Here the questions were drawn from a publication such as the UK’s Social Trends.

Generally, the interviews lasted one and a half to two hours.

The procedure described was only the initial part of a complicated process of creating scenarios for the future of the UK. However, the subsequent steps are not relevant to the present paper.

5 Implication for practice in foresight programmes

Foresight programmes, public or private, are mostly organised around panels with additional activities such as a Delphi survey; the approach tends to be mechanistic. Much of the original inspiration came from the desire to imitate the Japanese technology forecasting studies, based on Delphi surveys, which at first accentuated the mechanistic approach to what has become known illogically as ‘foresight.’ The term may be a form of shorthand for a wider intent now that recent programmes have tended to migrate from Delphi based surveys, but what that intent is has not been made clear. Consequently, there is a prospect of drifting towards an activity that
gives the impression of being a basis for national planning using scenarios without the
depth of understanding of the underlying processes that is necessary. Outwardly at
least, a deep understanding of the role of expert opinion and its roots in subjective
probability, human judgement and preferences is not overt amongst the current crop
of foresight programmes. Where self-assessment of expertise is included the defining
rules are not rigorous, whilst the interpretation of the outcome of Delphi surveys often
sticks to rigid conventions that are entirely inappropriate, as will be discussed shortly.
Indeed, the interpretation hides more important information than it reveals.

Some recent foresight programmes have abandoned the formality of methods
such as Delphi in favour of panel discussions that amount to brainstorming and
strength, weakness, opportunities and threat analyses. These are conducted by panels
formed by unknown processes that are possibly self-selecting amongst a small group
of people well known to each other with little attention being paid to the desirable
influences of transparency and the strictures implied by the use of the co-nomination
process. Worse still continuity of corporate memory from one study to the next has
not featured highly outside Japan and Germany.

5.1 Some personal experiences

Hull’s review of subjective probability assessment contained some examples that
approximate Loveridge’s practical experience (Loveridge, 1981) in the early 1970s in
forecasting sales of a new product, with no previous history, for several years ahead
on the basis of eliciting subjective opinion. The ‘experts’ were marketing team
members who obtained from potential customers firstly, their likelihood of buying
the product at all and secondly, their likely purchases, not as single figures but as
distributions, in each of the following three years. The team members were ‘trained’
in the elicitation process and were taken through a debriefing after each potential
customer visit. The method was effectively that of fixed intervals but with a number
of complications arising from the number of potential customers, different firm sizes
and different purchasing intentions. These factors complicated merging the
distributions to create a single set of three year distributions and presented a
number of problems as follows:

- Technically the model assumed that there was no competition as the product was
  unique so that the probability of buying was either zero or one. However, this
  situation did not persist so that intermediate probabilities had to be permitted.
- The assessors had difficulty with the elicitation procedure and frequently,
  the distributions did not sum to unity. The choice was then either to repeat their
debriefing or to assume that the form of the distribution was ‘correct’ and to
normalise it without further reference to the assessor. Provided the error in the
probability sum was small the second course was used.
- The greatest difficulty lay in how the elicited data was approved by senior
  managers before the combined forecast was computed. Here the pitfall was the
simple and obvious one with senior managers altering elicited data arbitrarily and
with only minimal consultation. The resulting distributions were chaotic often
with summations far removed from unity. The result was frequent, long and
sometimes fractious meetings causing the approval process to become protracted.
The initial foray into elicitation lasted about two years and led to the formulation of a simpler procedure based on eliciting a three point triangular distribution. The revised procedure recognised the frustration that arose from the tedium, for the assessor, of the elicitation process. The decision to elicit less information was a trade-off to retain the cooperation of the assessors. The revised procedure was facilitated by:

- providing the assessors with pictures of normal and left and right skewed three point distributions
- asking the assessor to obtain just three numbers to define the distribution; the maximum and minimum sales and the ‘best guess’ or mode.

Experience with the second model was limited as the project was soon moved on to another part of the organisation. However, it seemed to fulfil expectations of it as simplification of the elicitation process made it much easier for the assessors to gather intended purchasing data from firms. The question of getting the data approved by senior managers remained unresolved and is likely to remain so. The experience in this project led to a far better understanding of the interpretation of the responses to a Delphi survey, which is the next subject.

Institutional foresight is usually consultative, except in the very rare circumstances of a small expert committee that consults no one but itself. The wider the consultation the stronger the process, but for those charged with reporting the outcome the price is aggregation, negotiation, re-aggregation and so on, until the process is stopped by fiat. When surveys such as Delphi are used, important detail can be hidden for reasons that are far from clear. In Delphi surveys in foresight programmes the time of occurrence of each topic in the questionnaire is the important response. In the 1994–95 UK TFP the sponsors did not allow the respondents to include a probability estimate for the time of occurrence; the probability of the event occurring in the time period declared by the respondent was assumed to be one.

The number of responses to each topic was sufficient to construct a frequency histogram representing the variation of opinion; this is common to most Delphi surveys. Often these histograms exhibit more than one peak, indicating that the respondents’ opinions are divided between two or more beliefs about the likely occurrence of the topic, an eventuality in line with the earlier discussion of Lipinski and Loveridge’s study of the future of the UK. However, instead of presenting this important information, there is a propensity to hide it by adopting the convention of representing the distribution of opinion not as a histogram, but as a calculated three-point distribution (the upper and lower quartiles and the median) which admits of no division of opinion. In interpreting Delphi survey data this is a strange process of denial since, if there are sufficient responses to calculate a distribution of opinion, it is inconceivable that there is not a division of opinion in some instances. The interpretation is more complicated than simply one of divisions of opinion as the report on the Delphi survey (Loveridge et al., 1995) makes plain.

Wherever there was a clear modal value it indicated a strong consensus on those topics. Where responses to adjacent time periods differed by only a small percentage (3%) these were included in the modal value thus tending to spread or flatten the distribution indicating an increase in uncertainty among the population of respondents. Eventually, the outcome can be a flat distribution that falls off sharply at each end, indicating a very uncertain view amongst the population of respondents.
The interpretation of distributions with more than one peak has already been referred to above. With only five point distributions to work with, it may seem risky to introduce these more detailed interpretations. However, provided defined separating criteria are used, as indicated, then the insights gained are valuable and ought not to be obscured since divisions of opinion can have an important influence on priority setting, on policy making and even on the fate of an entrepreneur’s business plan.

As already referred to, the recent public foresight programmes in The Netherlands and the UK have abandoned formal methods in favour of panel discussions where, apart from naming each panel’s area of work, direction from the sponsor has been minimal. The outcomes of The Netherlands programme (Anon., 1996) and that of the UK (Anon., 2001) are now in the public domain. The effect of the shift from structured elicitation of opinions to the vagaries of panel discussions is impossible to assess without any transparent assessment of the expertise of the people involved in the panels. In The Netherlands the Steering Group Report has effectively been ‘laid on the table’ for any interested party to pick up and choose ideas to exploit; there has been no overt policy intervention. What will happen in the UK remains to be seen, but anecdotally the outcome is considered ‘insubstantial’ and, following a review of the programme, has caused the next ‘round’ to be a revival of the maligned notion of ‘picking winners’ (Anon., 2002). Consequently, the new programme will take the form of a rolling programme of themes to be chosen by ‘visionaries’ whose method of selection has remained unclear.

Lastly, in these personal experiences Lipinski and Loveridge were allowed to report from the study of the future of the UK expert opinion on the future price of oil. All the respondents in the elicitation rated themselves in the expert category and their joint opinion for 1985 (remember the elicitations were carried out in 1978) is illustrated in Figure 4.

**Figure 4** Joint opinion on 1985 price of oil. With acknowledgements to Lipinski and Loveridge (1982)
Figure 4 illustrates the level of confidence the experts had in their opinions, which indicated a price for oil in 1985 in real terms of $12.50 per barrel. The same respondents forecast that the price of oil would remain virtually unchanged in real terms for the indefinite future (then up to 1995). Using a simple assumption of a 7% annual inflation rate, the contention made in 1978 has a remarkable correspondence with what has happened over the following 23 years. In 1978 the price of oil was commonly expected to be in the region of $60 per barrel by 1985 and certainly by the 1990s; the actual outcome then provides a strong lesson to avoid ‘knee jerk’ reactions and ‘bandwagon’ effects. There is much that can be learned about the oil market from data of this kind; similar comments could be applied in other fields. It is doubtful if panels in foresight programmes would produce such insights simply from panel discussions.

6 Epilogue

Public foresight studies have created their own heroes and myths in a remarkably short space of time. However, the act of foresight has a long history of achievement in the development of human society despite, as Hogarth commented, the human propensity not to follow ‘the rules’ implied by formal studies of human judgement, subjective probability, preferences and mathematical theories of evidence. As Rivett tersely remarked in the IIASA conference, he did not see many decision theorists making decisions ‘in anger’ in the real world of investment or anywhere else. Whilst this may have been a cruel quip it is a sentiment that seems to be echoed by Beach, Christensen-Szalanski and Barnes in their obvious frustration with the directions being taken by research into human judgement. Maybe this could also be reflected in the contest between human judgement and computer modelling of the earth’s climate as referred to earlier.

Some final reflections on the relevance of subjective opinion to foresight programmes are a fitting way to end. In doing so these will be blended with a paraphrase of Cooke’s conclusions on the notion of ‘experts in uncertainty’. These reflections are as follows where Cooke’s views are given in parentheses with my own additions in italics:

- There is ‘... a need for an overarching methodology for using expert opinion in science’. Whilst this almost goes without saying, how that advice is provided is crucial, given the political interventions that become inevitable.

- Subjective or expert opinion may be a ‘... useful new form of information, but rules for its collection and processing must be developed’ and that these rules need to assist in ‘... building a rational consensus and ... take into account the peculiar feature of subjective data’. For this to happen there needs to be much wider appreciation of the nature of subjective opinion and the impinging fields of evidence in relation to the courts, the propensity for preference shifts and the vagaries and disputes surrounding human judgement. Extensions of knowledge of how to select ‘experts’ and of processes for the elicitation of expert opinion will also be needed.
Savage’s classical theory of rational decision (making) can be used as a basis for representing uncertainty and subjective probability.

The statistical tools available for analysing expert opinion are “... extremely limited”. The available tools are not widely known even in the research community.

When an expert quantifies his uncertainty through a subjective probability distribution, this “... effectively” transforms his opinion into units of probability that permit the use of a wider range of statistical methods, both classical and Bayesian and this is the real reason for “... quantifying expert opinion as subjective probability”. The need to think in terms of distributions of opinion is rarely accepted in advisory committees and is almost entirely absent in foresight programmes.

There is a need for better and user-friendly techniques of eliciting subjective probability distributions. Again the methods available are not widely known, some are highly abstruse without necessarily improving the outcome and there is often contention about the viability of different methods.

Closely related to elicitation is the need for better ways of communicating probabilistic information to policy and decision makers; this is not a simple inversion of the elicitation process.22

In eliciting opinions in foresight studies it is necessary to have ways of ‘seeding’ the expert’s mind since he or she will be expected to provide responses about events that will happen in the future, but where the responses will be dependent on events with known outcomes; the process of seeding is critical, difficult and in need of improvement (author’s addition not due to Cooke (1996)).

When combining expert opinions three operational models exist; these are:

• The classical model that can be applied when seed variables are available; its results are ‘modestly’ robust and the model “... has never yielded strange or intuitively unacceptable results.”

• The Bayesian model is most suited to use with a single expert, but the outcome is very sensitive to “... decisions of the analyst”. The model has the strongest theoretical foundation, but practical implementation remains “... a challenge.”

• Psychological scaling models are very different, are user friendly and consensus building, but theoretical problems remain and little is known (according to Cooke) about their validity in practical applications.

If elicitation has proved to be an art more than a science, then the possibility of training people to be skilled in the art of expressing subjective opinion, the creation of ‘probability assessors’, has remained a matter of speculation. The possibility of ‘calibrating’ probability assessors remains vague whilst the experiments to do so provide conflicting outcomes. Important as the question of expertise is in any foresight programme, it is unlikely to shed much light on the difficult topic that can be broadly headed ‘How good are the expert assessors?’ if only because their opinions are never sought in the form of subjective probability distributions. Suffice it to say that some
experts are much better performers in extending their knowledge into the future than others. Expert opinion is fundamental to foresight but its underlying nature is not well recognised by programme promoters.

References


Beach, L.R., Christensen-Szalanski, J. and Barnes, V. (1987) ‘Assessing human judgement: has it been done, can it be done, should it be done?’ in G. Wright and P. Ayton (Eds.) *Judgmental Forecasting*, John Wiley & Sons Ltd.


Experts and foresight: review and experience 61


**Bibliography**


Collen, R. ‘A study of Sherlock Holmes based on Kostick’s perception and preference inventory’.


Kanizsa, G. ‘Subjective contours’, Unknown.


**Notes**

1 Complex here has the connotation of complexity and not simply complication as described by Charles Perrow (1984) in *Normal Accidents*, Basic Books.

2 It is necessary to treat this term carefully; here it implies information that has been operated on by an individual’s value/norm set and involves the notion of selectivity in relation to the expert’s perception of the scope of the inquiry and (i) the selection of what is perceived to be the relevant information and (ii) the bounds of the individual’s information set.

3 This point will be returned to later in discussing the use of survey methods in foresight programmes.

4 See for example Perrow in Note 1 of this paper and the report into the BSE inquiry, http://www.bse.org.uk/

5 This is described later.

6 Cooke also devotes time to this concern.

7 There is a voluminous literature on probability encoding related to utility, but that is not necessarily related to scenario building.


9 Bendectin was claimed to have caused congenital limb defects to a child due to her mother taking the drug during the early stages of pregnancy.

10 Due to Abraham Wald; see Savage (1954).

11 Note that this had been referred to as inconsistency by Raiffa in 1957 and also by Granberg in 1966.

12 My additions are in parentheses.

In what follows I have made some personal additions but for the sake of ease of reading these are not indicated.

See Popper on ‘holism’ and Simon on ‘bounded rationality’.

I have described a process for scenario building in (1992) The Challenge of the External Environment, Supplementary Reading Book 1, Open Business School Course B885, The Open University.

These stem from empirical experience gained by the Institute for the Future which classified experts under three categories: generalists, persons of thought and persons of present and future action. The first have a spread of interests, perception and a high level of awareness, the second have deep knowledge in a special field and the last are likely decision makers.

It would be very easy to quote absurd limits in order to score ten apparently ‘correct’ results, but that amounts to ill considered ‘hand waving’ and is exactly the kind of response decision makers do not want from expert advisers.

Such people have been found in practical experience.

Whilst this part of the interview had the appearance of a party game its intent was serious enough and the outcome often left a marked impression on the respondent.

In the study of the future of the UK this was important, for a single variable only the historical reference material, sent in advance, was provided.

Author’s note.