Expert frailties in conservation risk assessment and listing decisions

Mark Burgman

School of Botany, University of Melbourne, Parkville, 3010, Australia

Email: markab@unimelb.edu.au

ABSTRACT

Listing decisions are a form of risk assessment supported largely by expert judgement. Expert judgements of rare events in novel circumstances are error prone. Experts are susceptible to social influences and their views are shaped by context, framing, and personal values. Expert judgment is a legitimate source of knowledge, but only if it is fallible. A range of behavioural and numerical methods may be employed to elicit, combine and communicate opinions in deliberation processes by expert groups. Expert groups provide an efficient and convenient way of synthesising a broad range of knowledge but their advice may be improved by training, feedback of performance, consistent communication formats, and awareness of motivational and other sources of bias. A suite of behavioural and numerical methods to aggregate opinions and communicate uncertainty provides new opportunities.

Key words: Conservation, listing, expert judgment, uncertainty, aggregation, expert groups

Introduction

Risk is the chance, within a time frame, of an adverse event with specific consequences. In our personal lives, we perform risk assessments every day and carry the burden of the consequences. Conservation biologists evaluate risks to species, communities, and ecosystem processes. Epidemiologists, toxicologists, engineers, ecologists, geologists, chemists, sociologists, economists, foresters and others also conduct environmental risk assessments routinely.

Under most legislation, listing decisions are recommendations, usually by panels of scientific experts, regarding proposals for species or ecological communities to be afforded protection under threatened species legislation. For example, Table 1 gives the numbers of species of plants and animals thought to be at risk in several countries. Such lists are published and used routinely by governments for a number of purposes (Possingham *et al.* 2002a). They are created largely on the basis of expert opinion.

Table 1. The number of species listed nationally as endangered in 2001 in Australia, USA and China (each of these countries has relatively large numbers of endemic species) compared to the number of species listed as endangered in the UK and the number thought to exist globally (data from Groombridge 1994, Anonymous 1995, May *et al*. 1995, Burgman 2002).

Taxon	Estimated total number of species in the world[1]	Number listed as 'Threatened'[2]			
		Australia	*USA*	*China*	*UK*[2]
Fungi	1,000,000	0	0	0	46
Nonvascular plants	100,000	0	0	0	70
Vascular plants	250,000	1597	1845	343	61
Invertebrates	5,000,000	372	860	13	171
Fish	40,000	54	174	16	7
Amphibians	4,000	20	16	0	3
Reptiles	6,000	42	23	8	6
Birds	9,500	51	46	86	25
Mammals	4,500	43	22	42	18

1. Estimates of the global numbers of species in each taxon were estimated crudely from numbers of currently described species and expert judgments of the proportion remaining to be described (see May *et al*. 1995, Burgman 2002).

2. The numbers provide a good approximation of the totals listed nationally as threatened in each country in 2001.

3. Numbers from the UK represent all those species in the UK's Biodiversity Action Plan for which conservation action plans were written (Anonymous 1995).

Pp 20 - 29 in *Threatened species legislation: is it just an Act?* 2004, edited by Pat Hutchings, Daniel Lunney and Chris Dickman. Royal Zoological Society of New South Wales, Mosman, NSW, Australia.

Decisions to include species on a threatened species list are sometimes mistaken. The uncertainty is reflected in turnover of species on the lists through time. For instance, Briggs and Leigh (1996) created a list of 'threatened' Australian plants, estimating 4955 species of vascular plants considered to be '*at risk (i.e., endangered, vulnerable, rare or poorly known and thought to be threatened)*'. Briggs and Leigh's (1996) list was modified as new information came to hand, as people and agencies responsible for the list changed, and as taxonomists revised species descriptions. In 2001, only 65% of the species earlier listed as extinct, endangered and vulnerable remained on the official national list (Burgman 2002).

The lists are also biased. May *et al.* (1995) called them popularity polls. For instance, the list from the UK (Table 1) has relatively large numbers of fungi, nonvascular plants and invertebrates. In contrast, the proportions of threatened species in these groups in other countries are much smaller. The total number of plants and animals in the UK is relatively modest and survey and taxonomic effort per species has been high. Scientists and others may be guided by funding opportunities and personal interests. Numbers of threatened insects, non-vascular plants and fungi are low in other countries because of scientific and popular lack of interest.

Some of the reasons for changes are illustrated in Table 2. Changes in attitude to extinction and taxonomic uncertainty drove many changes in lists of plants thought to be extinct. In some instances, the listing authority became more convinced of a taxon's extinct status after reconsidering existing literature or specimens, and broader expert consultation. Changes in attitude towards taxonomic status were usually due to advice from taxonomic specialists. Expert judgment weighed heavily on these outcomes.

Table 2. Reasons for changes to federal lists of extinct Australian vascular plants since 1981 (after Keith and Burgman 2004).

Reason for change in listing status	Additions	Deletions
Presumed extinction during 1981-2001	8	
Not previously evaluated	48	
Change in extinction uncertainty[1]	46	1
Taxonomic revision[2]	6	25
Change in taxonomic uncertainty	7	15
Rediscovered		89
Correction		34
Change in opinion on native status		3
Total changes in status	115	167
Originally listed in 1981	113	
Remaining listed in 2001		61[3]
Total listings	**228**	**228**

1. Triggered by new information (unsuccessful searches) or reconsideration of literature and expert opinion.
2. Taxonomic revision refers to newly described taxa based on old collections. Change in taxonomic uncertainty means that the listing authority became more convinced of a species status as a distinct taxon due to expert advice. Species were moved from poorly known or data deficient to presumed extinct on the basis of this advice.
3. Includes 12 species remaining listed despite rediscovery.

The considerable uncertainty in the status of listed species is not communicated. The consequence of failing to acknowledge and communicate uncertainty is that people are misled into believing the lists are reliable and unbiased.

Often, different kinds of information are treated as though they are equivalent. In most listing decisions, direct measurements are almost always missing. Expert opinion is usually available. To work with experts, one must,

1. define necessary knowledge and skills for the problem at hand (decide who might be considered an expert),
2. select the experts,
3. elicit information,
4. evaluate the reliability of the information,
5. combine information from different experts, and
6. use the information to reach decisions.

Scientific experts and groups are expected to make reliable, transparent and consistent judgments on behalf of the broader community. Listing determinations are published but lists are not accompanied by measures of the reliability of the decisions. So, how reliable is expert judgment? Has it been assessed? Is it used in the same way in other disciplines? This paper explores the frailties of expert groups, particularly those involved in conservation risk assessments, and how their errors may propagate in decision-making. It outlines some of the options for using expert judgment that may make it more reliable and transparent.

Kinds of uncertainty

Many taxonomies of uncertainty distinguish between natural variability and 'incertitude'. Incertitude reflects lack of knowledge about nature and can be reduced with additional samples. Natural variation can be better estimated by additional samples, but cannot be reduced (Burgman 2004). Regan *et al.* (2002) created a taxonomy of uncertainty that, at the highest level, distinguished between epistemic and linguistic uncertainty. Epistemic uncertainty exists because of the limitations of measurement devices, insufficient data, extrapolations and interpolations, and variability over time or space. There is a fact, but we do not know it exactly. For instance, a population's size may in fact be 281, but be estimated to be 300, with 95% confidence intervals of 250 to 350. This is the domain of statistics and conventional scientific training.

Linguistic uncertainty, on the other hand, arises because natural language, including scientific vocabulary, is not exact. Linguistic uncertainty can be classified into several distinct types.

Vagueness arises because language permits borderline cases. For example, the term endangered in general usage refers to a continuum of threat but is defined sharply by the IUCN (1994, 2001), for example, as a species that is estimated to number fewer than 250 mature individuals (Rule D). Context dependence arises from a failure to specify the context in which a proposition is to be understood. For example, if a range reduction is said to be 'small', without specifying the context, it is not possible to say whether the reduction is small on the scale of a local region or a continent. Ambiguity arises when a word has more than one meaning and it is not clear which meaning is intended. Knowledge is under-specific when there is unwanted generality in a statement or where data could have been obtained but are no longer available. For example, fauna and flora collections often are associated with very imprecise locations such as 'north of Esperance'.

All risk assessments involve at least some elements of linguistic uncertainty. Imprecise language clouds the perceptions of experts and limits their ability to communicate. Expert assessments may be made substantially more reliable if care is taken to eliminate as many sources of linguistic uncertainty as possible, prior to the commencement of an elicitation process. The preamble developed by the IUCN (2001) is a good example of a careful attempt to minimise linguistic uncertainty.

Who's an expert? … philosophical and legal definitions

Hart (1986; see also Meyer and Booker 1990) identified three attributes of an expert:

1. effectiveness; they use knowledge to solve problems with an acceptable rate of success,
2. efficiency; they solve problems quickly, and
3. awareness of limitations; they are willing to say when they cannot solve a problem.

This definition is superficially appealing but it is difficult to put into operation. It can be very difficult to distinguish between a competent expert and an incompetent one, and to have a clear picture of what an expert knows. Experts usually are assumed to have both substantive expertise resulting from technical experience and training, and normative expertise, which is the ability to communicate and knowledge of the statistical principles and jargon of a field.

Lawson (1900) wrote the foundation rules for expert and opinion evidence for the US legal system. In these rules, opinion is not admissible in evidence except '…*on questions of science …persons instructed therein by study or experience may give their opinions. Such persons are called experts.*' (p. 1-2). Expert opinion is seen as necessary to inform a court completely as possible about facts that might be otherwise unattainable because they are future probabilities, contingencies or facts '*not within positive knowledge*' (Lawson 1900, p. 236). In Australian Federal courts, opinion evidence is admissible if it assists '…*the trier of fact in understanding the testimony, or determining a fact in issue*' (ALRC 1985, p. 739-740). Science is presumed to act as a check on bias or prejudice.

Courts have the luxury of being able to question the reliability of experts and to reject them. The qualifications of an expert may be tested by the opinions of other experts. In adversarial systems, the accuracy of testimony may be tested under cross-examination. The substantive knowledge of an expert may be tested by hypothetical questions. Opinions may be 'impeached' by proof that on a former occasion, an expert expressed a different opinion. The ability to question experts is an essential part of the provision and use of expert judgment.

In the adversarial legal systems, potential expert witnesses are selected overwhelmingly for their credentials and for the strength of their support for the lawyer's viewpoint (Shuman *et al.* 1993, in Freckelton 1995). Lawyers search for appropriate attributes in an expert and develop strategies to maximize their chance of winning a case. This view suggests that scientists are advocates of a scientific position (Walton 1997), whereas scientists usually see themselves as objective interpreters of fact, protected from prejudice by scientific method.

Elicitation

The reliability of expert judgments depends, to some extent, on how the experts are interrogated. Methods for elicitation may include correspondence, personal interviews, traditional meetings, structured group meetings aimed at achieving consensus, and meetings that combine consensus with numerical aggregation. Meyer and Booker (1990), Morgan and Henrion (1990), Cooke (1991), Vose (1996), Ayyub (2001) and Walley and DeCooman (2001) provide practical advice on how to conduct scientific elicitation and analyse the information they produce.

Context may affect the quality of expert judgment. Experts sometimes make judgments just to be helpful, or to retain the semblance of expert respectability (because the experts have been brow-beaten into providing an answer, or won't admit they can't). Experts tire during prolonged elicitation exercises and fatigue may make their knowledge unavailable.

Conservation listing decisions often involve novel circumstances in which data and experience are limited, conditions under which expert judgments are particularly error-prone. Errors are exacerbated when the topic involves very low probability events such as catastrophic failure of a system, extreme weather, coincidences of independent events, and so on. Sometimes, it may be possible to compare the situation with other events that are better defined, or it may be possible to disaggregate the rare event into a sequence of more likely events that are easier to estimate, the combination of which generates the outcome in question (Bier *et al.* 1999).

Expert frailties

In the 1970s, Kahneman and Tversky began experiments on the ways in which people perceive and react to risks. One of their discoveries was that experts are susceptible to context and how questions are framed. A framing effect occurs when a change in the presentation of a choice influences choice behaviour, even when the objective characteristics (outcomes or probabilities) are not changed (Kahneman and Tversky 1979, 1984). Scientific training provides only limited protection against these weaknesses (Fischhoff *et al.* 1982, Fischhoff 1995, Morgan *et al.* 1996, Freudenburg 1996, Slovic *et al.* 2000, Gigerenzer 2002, Trumbo and McComas 2003).

Conservation risk assessments are subject to distorting influences, perhaps more so than other types of scientific analysis, because of the politically charged and value-laden nature of many of the problems. Typically there is considerable pressure to produce results to diffuse social tension. Unfortunately, experts are themselves susceptible to the same set of pressures, even in circumstances in which they believe themselves to be considering only the scientific dimension of a decision. They cannot occupy the independent, objective ground that politicians and policy makers wish them to.

Biases may arise because of the relative ease with which different kinds of information is remembered by experts. The reliability of an estimate is determined by the expert's experience, their ability to recollect accurately and the similarity of past experiences to the problem at hand (Meyer and Booker 1990, Vose 1996).

Optimism is a pervasive feature of expert assessments (Plous 1993, Tufte 1997). Typically there is little relationship between confidence and accuracy, including eyewitness testimony in law, clinical diagnoses in medicine, and answers to general knowledge questions by people without special training. Most experts have a region of overconfidence, a domain between the subset of facts they have learned (Figure 1) and the subset they think they know. It varies between experts and depends on the elicitation process.

Baran (2000) described a 40 ha patch of sclerophyll forest and asked a group of ecologists at a professional meeting to estimate the number of 0.1 hectare quadrats that would be necessary to sample 95% of the perennial, vascular plant

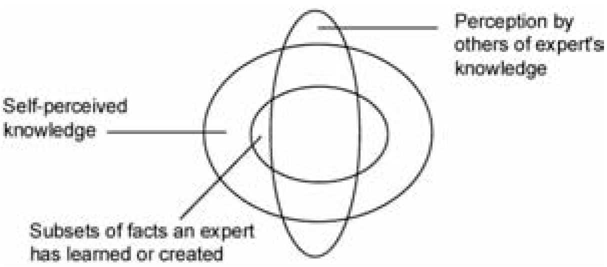

Figure 1. Expert knowledge and the region of overconfidence, which lies between the subset of facts known by the expert and the subset they believe they know (modified from Ayyab 2001; Burgman 2004).

species. She also asked them to provide 90% credible bounds on their estimates. The context was one in which the experts may have been expected to provide reliable answers. Baran (2000) sampled the area intensively to validate their answers. Unfortunately, only 2 of 22 intervals captured the correct answer. The median response substantially underestimated the number of samples required. Experts are often insensitive to sample size and there is no simple feedback between estimation and outcome in plant surveys. As a result, experts may be poor judges of the effort required to achieve a given level of reliability.

The surprise index is the frequency with which a true value falls outside the limits of a judgment. If experts are asked to provide 1^{st} and 99^{th} quantiles, for instance, surprises should occur roughly 2% of the time. Cooke (1991) and Vose (1996, p. 156-158) found in a range of settings that experts provide intervals that contain the truth about 40% of the time, when asked for intervals that will contain the truth about 90% or more of the time. Morgan and Henrion (1990) concluded that all methods for elicitation show a strong tendency towards overconfidence, generating far more surprises than expert judgments lead us to expect. Freudenburg (1999, p. 108) called it 'insufficient humility about what we do not know'.

Motivational biases are tied to ethical positions and experts often are unaware of the value-laden nature of their opinions. Such biases may arise during listing because the scientists involved may have a view of what is to be gained or lost by a decision that influences their interpretation of uncertain information. In a different context, but one that illustrates the issue, Campbell (2002) interviewed experts interested in the conservation and management of marine turtles. They were asked about the sustainable, consumptive use of turtles and their eggs. Campbell's results revealed four 'positions' on use, all of which were defensible on scientific grounds (Table 3).

A total of 19 experts thought that uncertainty dictates caution, so that managers should err on the side of non-use. Three were implacably opposed to commercial use, irrespective of uncertainty. Most experts saw opposing views as influenced by 'emotions', at the same time claiming dispassionate scientific objectivity for their own views, irrespective of their positions. Participants attributed emotional involvement to conservationists without biological training, as though scientific training protects people against emotion. Few propositions are so plainly self-deluded. Almost all experts have personal, value-laden opinions about the outcomes of environmental decisions. It is naïve to think that scientists are anything but advocates of scientific positions and personal value systems.

Table 3. Positions on use and uncertainty, and institutional affiliation of 36 marine turtle experts (after Campbell 2002).

Positions	NGO	University	Government	Total
Consumption is supported, learn through use	3	5	2	10
Support limited consumptive use	2	2	0	4
Uncertainty dictates caution, consumptive use is not supported	5	7	7	19
Consumptive use is unacceptable	0	2	1	3

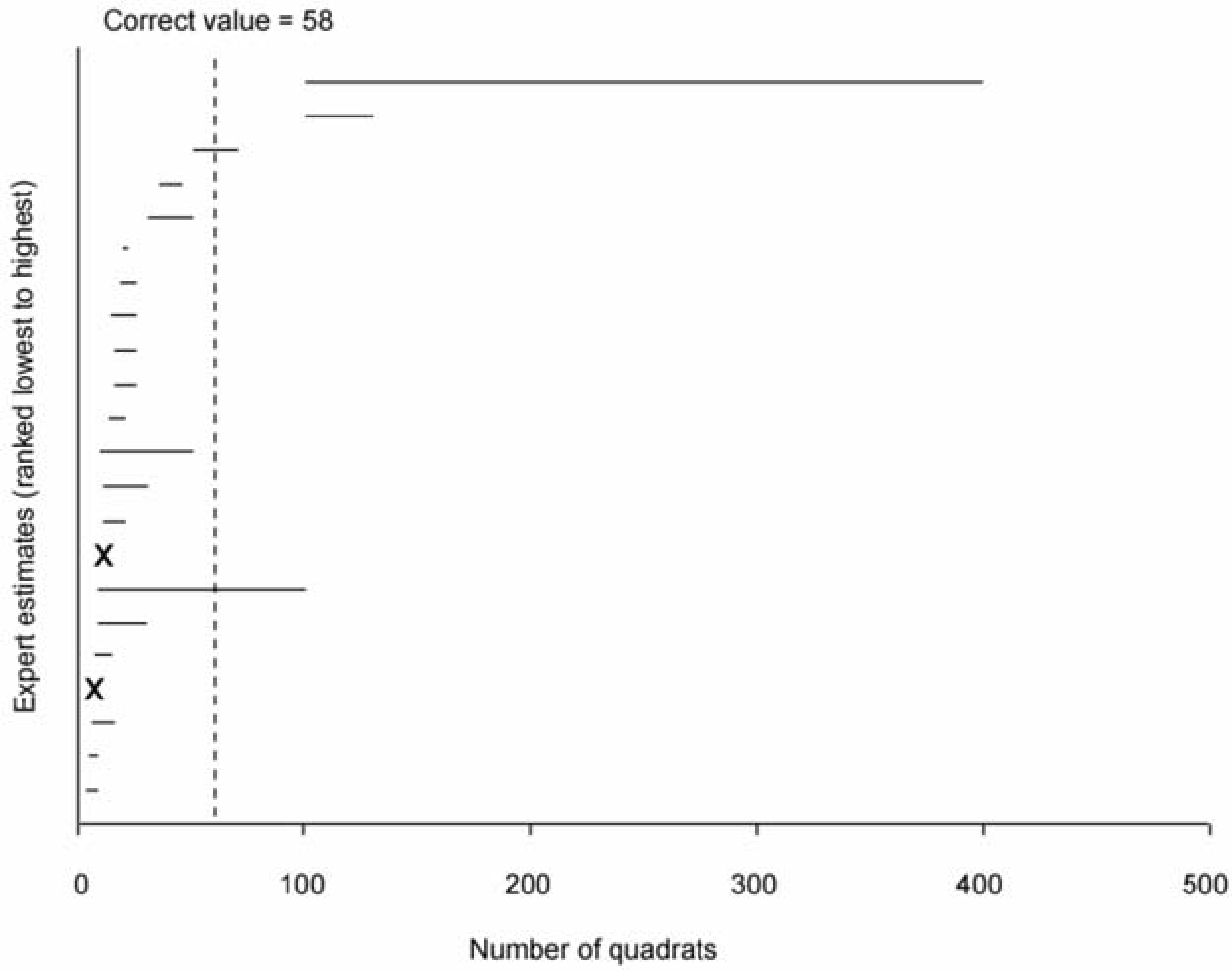

Figure 2. Expert estimates of the number of 0.1 ha quadrats necessary to sample 95% of the vascular plant species in a 40 ha patch of dry sclerophyll forest containing about 120 species, and their 90% credible bounds (from Baran 2000). The analyses assume plants were randomly distributed in the 40 ha patch, with respect to the sample quadrats. Two people failed to provide bounds, marked by a 'x'.

Are expert judgments any good?

Expert judgments are used because estimates are needed in circumstances in which it is difficult or impossible to acquire direct data. There are a few direct comparisons of expert estimates with actual probabilistic outcomes. Most are from engineering (see Cooke 1991). For example, Hynes and Vanmarcke (1975, in Krinitzsky 1993) found experts' risk assessments to be optimistic (their confidence intervals rarely captured the true value) and sometimes substantially biased (Figure 3).

Training can alleviate some of these problems. Routinely reliable estimates have only ever been demonstrated in people who make frequent, repeated, easily verified, unambiguous predictions so that they learn from feedback, such as weather forecasters (Figure 4). Expert engineers underestimated the lifetimes of long-lived components and over-estimated the lifetime of short-lived components. Subjective probabilities made by army doctors that patients have pneumonia were poor predictors of the outcomes of more reliable diagnoses based on radiography. Performances of experts in business and military intelligence are likewise mixed and unimpressive (Morgan and Henrion 1990).

Most technical disciplines have conventional (shared) wisdom. If the technical base of expertise is narrow, or if people are selected by inter-personal relationships, they may represent a single school of thought or be influenced by the same data (Bier 2002).

Experts do somewhat better than untrained people, within their domain of expertise. Wright *et al.* (2002) concluded that if objective models and reliable feedback are available, it may be possible for experts to improve their performance above the performance of an untrained person. However, judgments by some professionals may be unlearnable, especially if they involve novel or infrequent events. Demands for unlearnable (and untested) judgments, together with unquenchable expert optimism, create a spectre of widespread, unreliable expert opinion that wears a mantle of scientific reliability. The documented unreliability of expert opinion in other disciplines signals the possibility that unreliable expert opinion may be prevalent in environmental domains where novel events are common and objective models, monitoring and feedback are rare.

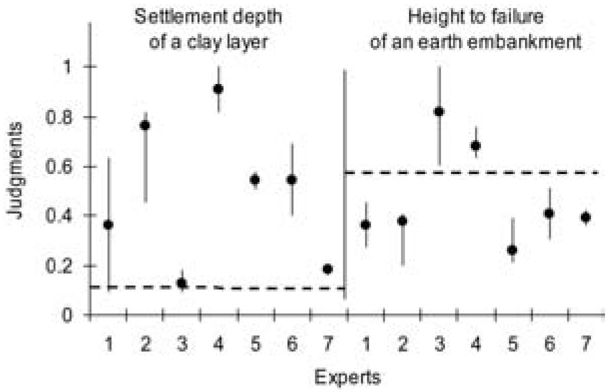

Figure 3. Opinions of geotechnical experts on two standard problems. The correct (measured) values for settlement depth was 1.5 cm and for added height to failure was 4.9 m. The x-axis for both was rescaled so the maximum value was 1. Correct values are shown as dashed horizontal lines. The intervals show 'minimum' and 'maximum' values reported by the experts (after Hynes and Vanmarcke 1975 in Krinitzsky 1993).

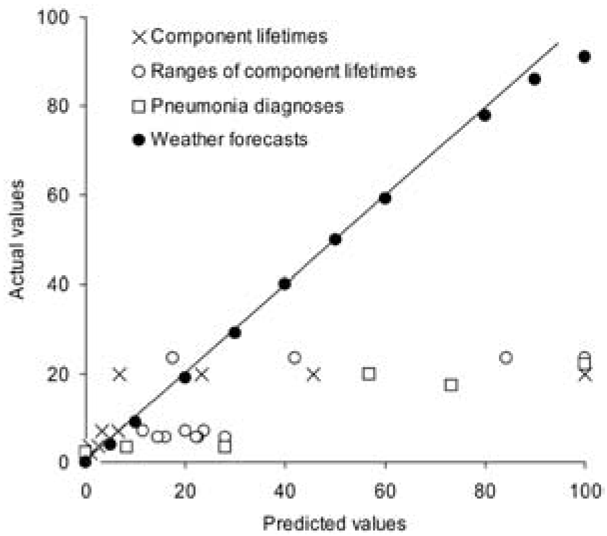

Figure 4. Calibration curves: expert predictions plotted against actual outcomes. Crosses are estimates by engineers of the mean lifetime of components in nuclear power systems, versus measured lifetimes. Open circles are estimates of the ranges for the mean lifetimes of the same components, versus measured ranges. Ranges are expressed as the maximum divided by the minimum. The components included pumps, valves and heat exchange units (after Mosleh 1987, in Cooke 1991). The squares are army doctors' subjective probabilities that patients have pneumonia, versus more reliable diagnoses based on radiography (after Christensen-Szalanski and Bushyhead 1981). Solid circles are meteorologists predictions for the probability of rain on the following day, against the observed relative frequencies (after Murphy and Winkler 1984, in Plous 1993). The diagonal line provides the line of correct estimation for all sets of observations. Values are scaled so that the maximum value in each set is 100.

Salvation in expert groups?

Many disciplines use expert groups to provide independent, technical advice. In conservation biology, disagreements among experts involved in listing decisions are rarely communicated. Regan *et al.* (2004) documented uncertainty among experts in listing assessments. They asked 18 conservation biologists to assess the status of 13 species from a broad range of taxonomic groups. The assessors were provided with standard information sets and interpreted them using the IUCN (2001) logical rules (Figure 5).

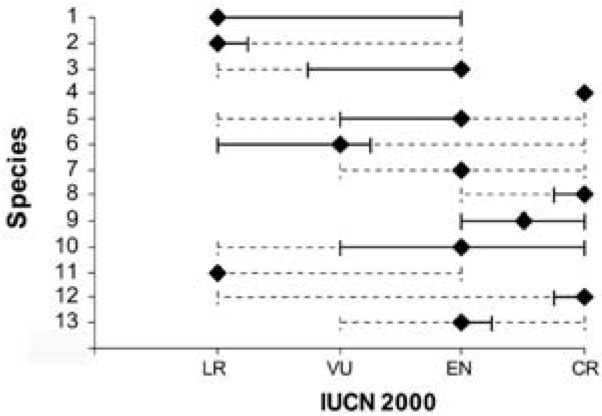

Figure 5. Assessments of the conservation status of 13 species made by 18 assessors (Regan et al. 2004). The diamond represents the median assessment. 50% of assessments are contained within the solid bars. The dashed lines show the full range of assessments. LR is low risk, VU is vulnerable, EN is endangered, and CR is critically endangered.

Some participants were unwilling to make inferences, due to different attitudes and interpretations of language in the guidelines and the information sets. Some parameter estimates spanned thresholds in the rule set. Variation in qualitative judgments, inconsistent logic and mistakes entering data all played a role. The lesson from Regan *et al.*'s (2004) study was that expert judgments can differ, even when the only source of uncertainty is the subjective interpretation of language and data, a largely unavoidable element in almost all conservation risk assessments.

Careless use of expert groups can lead to unacknowledged uncertainty and bias. Uncertainty will be underestimated if the experts involved share common values, experiences, professional norms, context, and cultural background, so that they stand to gain or lose in similar ways from the outcomes of decisions and hold the same motivational biases (Bier 2002). Anonymous opinions may lead to bias because experts cannot be held personally accountable (Krinitzsky 1993). Strong personalities influence outcomes, participants promote decisions before examining the problem, views are anchored, change is resisted, people hold covert opinions, and there is substantial pressure for conformity (e.g., Ruckelshaus *et al.* 2002). Group judgment may lead to overconfidence and polarization of opinions in which groups adopt a position that is more extreme than any individual member (Clemen and Winkler 1999). Despite these frailties, some experts persist in the belief that expert-group elicitation of risks is largely detached from subjective, social and emotional influences (e.g., Kerr 1996).

If these issues are managed explicitly, if participants act rationally (from the perspective of classical probabilities) and are provided with information that reflects utilities and expectations in an unbiased way, then in general, group (consensus) judgments can be expected to be better than individual ones. Multiple experts with different backgrounds increase the extent of knowledge and experience contributing to an assessment (Clemen and Winkler 1999). Groups may improve the interpretation of evidence: experts revise opinions when they are alerted to new and relevant issues or errors in their logic or interpretation of language. Stratifying experts to include a range of demographic and social attributes increases the chances of disagreement, but reduces the chances of bias.

Aggregating opinions

There are two basic forms of aggregation: numerical and behavioural aggregation. Behavioural aggregation techniques include the Delphi approach and variations (Cooke 1991, Ayyub 2001), negotiation, consensus, closure, resolution, and voting or preference voting (Engelhardt and Caplan 1986, Valverde 2001). In general, behavioural aggregation does not guarantee complete consensus and it is not always sensible to seek it. Disagreement cannot necessarily be resolved and consensus may mask important, legitimate differences of opinion. One notable example of a procedure to retain and communicate differences of expert opinion was made by Lunney *et al.* (1996) who displayed the spread of opinions relative to decision thresholds and noted areas of disagreement among experts for each species they evaluated.

It is always useful to eliminate arbitrary elements of disagreement (such as ambiguity and underspecificity). But in the spirit of honest and transparent assessments, differences of opinion about scientific detail, as well as about substantial ethical issues, can be made transparent, without compromising the ability of a group to reach a decision.

Numerical aggregation uses quantitative strategies to arrive at a combined estimate. If the information is probabilistic, then the tools of formal statistics including Bayesian analysis are appropriate. Many approaches including evidence theory and fuzzy sets (see Walley 1991, Klir and Weirman 1998, Sentz and Ferson 2002, Ferson *et al.* 2003) have been developed to cope with non-probabilistic uncertainty.

All aggregation techniques have to deal with the question of weighting the opinions of different experts. Some may be more prone than others to bias and misinterpretation of data, some people have more to lose than others, and so on. For example, some of the opinions in Figure 5 were contributed by people who routinely made more errors than others. Experts may be weighted by consistency, frequency of errors and concordance with known values (Cooke 1991). There have been no explicit applications of weighting strategies in listing decisions for conservation biology.

Akçakaya *et al.* (2000)used non-probabilistic uncertainty methods to combine parameters for the IUCN (1994) criteria. The result of the analysis was a best estimate and bounds that encapsulate uncertainty about parameters and dependencies between the elements of the rule sets (Figure 6).

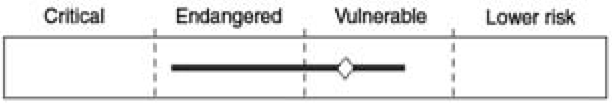

Figure 6. Example output from the method developed by *Akçakaya et al.* (2000).The species is most likely to be Vulnerable, but it is possible that it is Endangered.

Probabilities representing the beliefs of experts may be combined as weighted linear combinations of opinions (Cooke 1991, Vose 1996, Valverde 2001). Numerical averages of individual judgments seem to perform slightly better than group judgments reached by behavioural consensus when the focus is on unambiguous, value free, and sharply defined parameter estimates (Gigone and Hastie 1997, Clemen and Winkler 1999). A few methods have been developed that combine elements of numerical aggregation and behavioural consensus (see Lehrer and Wagner 1981, Cooke 1991, Lehrer 1997). Aggregation methods have not been explored to any great extent in conservation biology and might provide a rich source of ideas for improving the way expert opinions are combined and used in decisions.

Making better use of expert opinion in listing decisions

Overconfidence can be substantially improved by asking people to consider the reasons why they may be wrong (Morgan and Henrion 1990). Performance is enhanced when experts possess appropriate models and are trained to translate subjective assessments into numerical estimates. Experts perform better when they use a consistent format to communicate, when they are trained to provide unbiased estimates and when their judgement is reinforced by immediate, unambiguous feedback (Wright *et al.* 2002). Guidelines for elicitation may be tailored for specific circumstances.

Other strategies might improve the quality of expert judgements. They include a systematic search for motivational biases (Morgan and Henrion 1990), testing expert understanding against known standard problems, making experts aware of potential sources of bias, providing them with training, and monitoring sources of bias during the elicitation process (Meyer and Booker 1990). Cooke (1991) provided some numerical methods for achieving these goals.

Science encourages obedience to scientific authority through its claims of objectivity and expert status. Expert groups, in particular, can be difficult to challenge. Walton (1997) pointed out that scientific argument may be seen as rigid in the sense that it is beyond refutation by non-experts, generating what Walton called a culture of technical control. In conservation biology, the appeal to expert opinion should be a fallible but legitimate strategy (Walton 1997). Anyone with a stake in the outcome should be able to criticize an expert's opinion and be taken seriously. If it is broadly accepted that experts often are advocates of a theory and that they have a stake in the outcome, methods are required that balance their opinions. In addition, honest differences of opinion among experts, those that remain after linguistic uncertainties are eliminated, should be retained and communicated.

Often, experts are the only available source of knowledge to support decisions that would otherwise be made in a vacuum. Group opinions are likely to be better than individual opinions and they provide an efficient and convenient way of compiling and aggregating a broad range of knowledge. But there may be a great deal to be learned from other technical disciplines and from the psychology and sociology of risk perception and decision theory. Some advances have already been made in local assessment protocols (Lunney *et al.* 1996) that would have broad utility. Table 4 summarises some of the potential avenues for improvement. At least some could be implemented almost immediately, with very little cost. Others will take time. Alternative approaches need to be evaluated. However, implementation could substantially improve the reliability, transparency, utility and credibility of listing decisions.

Table 4. Recommendations for expert opinion in listing decisions

Recommendations for providers of expert opinion
• consider why your opinions may be wrong
• use appropriate models to organize ideas and evaluate the adequacy of available data
• seek training to translate subjective assessments into numerical estimates
• use a consistent format to communicate
• seek immediate, unambiguous feedback on your judgements
• seek situations in which your opinions can be tested independently
• ensure you are aware of potential sources of bias
• be aware of overconfidence and practice to reduce it
• be aware of sources of linguistic uncertainty and seek to eliminate them
• declare your interests (your stake in the outcome)
• provide opinions in a form that can be verified as easily as possible
Recommendations for facilitators
• select experts to encompass a broad range of social and technical strata
• decide or negotiate weighting and aggregation strategies before the elicitation process
• consider alternative elicitation strategies (one on one, small groups, workshops)
• create systems that define terms and attributes as clearly and unambiguously as possible
• ensure the elicitation system actively eliminates linguistic uncertainty
• conduct a systematic search for motivational biases
• ensure experts are aware of potential sources of bias
• measure and correct for overconfidence
• use structured rules to assist decisions
• provide experts with immediate, unambiguous feedback on judgements
• anticipate and manage strong personalities, entrenched positions, anchoring, resistance to change, covert opinions, pressure for conformity, and power differences between people
Recommendations for users of expert opinion
• ensure experts encompass a broad range of social and technical strata
• understand terms, parameters and attributes of the system used by experts
• attribute opinions to individuals
• ensure all those who share the burden of the risk of decisions have the opportunity to cross-examine the information and expert opinions
• request that expert performances against standards be made public
• ensure differences of opinion among experts are communicated in lists
• anticipate overconfidence
• evaluate aggregation and weighting schemes for their effect on the consensus position

Acknowledgements

I am indebted to Natalie Baran and Tracey Regan for their permission to use some of the results of their theses at the School of Botany, University of Melbourne. I am grateful to Kuniko Yamada for her assistance in preparing this manuscript, and to David Keith, Dan Lunney, Pat Hutchings and one anonymous reviewer for their comments. Background information in this paper is drawn from a chapter of a book by the author entitled 'Risks and Decisions for Conservation and Environmental Management', to appear through Cambridge University Press. Most of this work was conducted while I was a sabbatical fellow at the National Center for Ecological Analysis and Synthesis at Santa Barbara, California.

References

Akçakaya, H. R., Ferson, S., Burgman, M. A., Keith, D. A., Mace, G. M. and Todd, C. R. 2000. Making consistent IUCN classifications under uncertainty. *Conservation Biology* **14:** 1001–1013.

ALRC. 1985. Evidence. Australian Law Reform Commission Report 26.Volume 1. Australian Government Publishing Service, Canberra. http://www.austlii.edu.au/au/other/alrc/publications/reports/26/

Anonymous. 1995. Biodiversity: the UK Steering Group report. Volume 2. Action Plans, HMSO, London.

Ayyub, B. M. 2001. *Elicitation of expert opinions for uncertainty and risks.* CRC Press, Boca Raton.

Baran, N, 2000. Effective survey methods for detecting plants. MSc Thesis, School of Botany, University of Melbourne.

Bier, V. 2002. Implications of the research on overconfidence for challenge problem solution strategies. Paper presented to the SANDIA Conference on Uncertainty. Albuquerque: New Mexico, August 2002.

Bier, V. M., Haimes, Y. Y., Lambert, J. H., Matalas, N. C. and Zimmerman, R. 1999. A survey of approaches for assessing and managing the risks of extremes. *Risk Analysis* **19:** 83-94.

Briggs, J. D. and Leigh, J. H. 1996. Rare or threatened Australian plants. CSIRO, Melbourne.

Burgman, M. A. 2002. Are listed threatened plant species actually at risk? *Australian Journal of Botany* **50:** 1-13.

Burgman, M. A. 2004. *Risks and decisions for conservation and environmental management.* Cambridge University Press, Cambridge (in press).

Campbell, L. M. 2002. Science and sustainable use: views of marine turtle conservation experts. *Ecological Applications* **12:** 1229-1246.

Christensen-Szalanski, J. and Bushyhead, J. 1981. Physicians use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance* **7:** 928-935.

Clemen, R. T. and Winkler, R. L. 1999. Combining probability distributions from experts in risk analysis. *Risk Analysis* **19:** 187-203.

Cooke, R. M. 1991. *Experts in uncertainty: opinion and subjective probability in science.* Oxford University Press, Oxford.

Engelhardt, H. and Caplan, H. 1986. Patterns of controversy and closure: the interplay of knowledge, values, and political forces. Pp. 1-23 in *Scientific controversies: case studies in the resolution and closure of disputes in science and technology*, edited by H. Engelhardt and H. Caplan. Cambridge University Press, New York.

Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D. and Sentz, K. 2003. Constructing probability boxes and Dempster-Shafer structures. SAND2002-4015. Sandia National Laboratories, Albuquerque: New Mexico.

Fischhoff, B. 1995. Risk perception and communication unplugged: twenty years of progress. *Risk Analysis* **15:** 137-145.

Fischhoff, B., Slovic, P. and Lichtenstein, S. 1982. Lay foibles and expert fables in judgements about risk. *American Statistician* **36,** 240-255.

Freckelton, I. 1995. The challenge of junk psychiatry, psychology and science: the evolving role of the forensic expert. Pp. 58-59 in *Tomorrow's law,* edited by H. Selby. Federation Press, Sydney.

Freudenburg, W. R. 1996. Risky thinking: irrational fears about risk and society. *Annals AAPSS* **545:** 44-53.

Freudenburg, W. R 1999. Tools for understanding the socioeconomic and political settings for environmental decision making. Pp. 94-125. in *Tools to aid environmental decision making*, editied by V. H. Dale and M. R. English. Springer, New York.

Gigone, D. and Hastie, R. 1997. Proper analysis of the accuracy of group judgments. *Psychological Bulletin* **121:** 149-167.

Gigerenzer, G. 2002. *Calculated risks: how to know when numbers deceive you.* Simon and Schuster, New York.

Groombridge, B. (Ed) 1994. Biodiversity data sourcebook. World Conservation Monitoring Centre, Biodiversity Series No 1. World Conservation Press, Cambridge.

Hart, A. 1986. *Knowledge acquisition for expert systems*. McGraw-Hill, New York.

IUCN. 1994. IUCN red list of threatened animals. International Union for the Conservation of Nature, Gland: Switzerland.

IUCN. 2001. IUCN Red List Categories. IUCN Species Survival Commission, The World Conservation Union, Gland: Switzerland.

Kahneman, D. and Tversky, A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* **47:** 263-291.

Kahneman, D. and Tversky, A. 1984. Choices, values, and frames. *American Psychologist* **39:** 342-347.

Keith, D. A. and Burgman, M. A. 2004. The Lazarus effect: can the dynamics of extinct species lists tell us anything about the status of biodiversity? *Biological Conservation* **117**: 41-48.

Kerr, R. 1996. A new way to ask the experts: rating radioactive waste risks. *Science* **274:** 913-914.

Klir, G. and Wierman, M. J. 1998. *Uncertainty-based information: elements of generalized information theory.* Physica-Verlag, Heidelberg.

Krinitzsky, E. L. 1993. Earthquake probability in engineering – Part 1: The use and misuse of expert opinion. *Engineering Geology* **33:** 257-288.

Lawson, J. D. 1900. *The law of expert and opinion evidence* (2nd edition). T. H. Flood, Chicago.

Lehrer, K. 1997. Consensus, negotiation and mediation. Pp. 3-14. in *Consensus under fuzziness,* edited by J. Kacprzyk, H. Nurmi and M. Fedrizzi. Kluwer, Boston.

Lehrer, K. and Wagner, C. 1981. *Rational consensus in science and society*. Reidel, Dordecht.

Lunney, D., Curtin, A., Ayers, D., Cogger, H. G. and Dickman, C. R. 1996. An ecological approach to identifying the endangered fauna of New South Wales. *Pacific Conservation Biology* **2**: 212-231.

May, R. M., Lawton, J. H. and Stork, N. E. 1995. Assessing extinction rates. Pp. 1-24. in *Extinction rates,* edited by J. H. Lawton and R. M. May. Oxford University Press, Oxford.

Meyer, M. A. and Booker, J. M. 1990. Eliciting and analyzing expert judgment: a practical guide. Office of Nuclear Regulatory Research, Division of Systems Research, US Nuclear Regulatory Commission, Washington, DC.

Morgan, M. G. and M. Henrion. 1990. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press, Cambridge.

Morgan, M. G., Fischhoff, B., Lave, L. and Fischbeck, P. 1996. A proposal for ranking risk within Federal agencies. Pp. 111-147. in *Comparing environmental risks*, edited by Davies, J. C. Resources for the Future, Washington, DC.

Plous, S. 1993. *The psychology of judgment and decision making.* McGraw-Hill, New York.

Possingham, H. P. Andelman, S. J., Burgman, M. A., Medellín, R. A., Master, L. L. and Keith, D. A. 2002a. Limits to the use of threatened species lists. *Trends in Ecology and Evolution* **17:** 503-507.

Regan, H. M., Colyvan, M. and Burgman, M. A. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* **12:** 618-628.

Regan, T. J., Burgman, M. A., McCarthy, M. A., Mace, G. M., Master, L. L., Keith, D. A. and Andelman, S. J. 2004. The consistency of extinction risk classifications. *Conservation Biology* (in review).

Ruckelshaus, M. H., Levin, P., Johnson, J. B., and Kareiva, P. M. 2002. The Pacific Salmon wars: what science brings to the challenge of recovering species. *Annual Review of Ecology and Systematics* **33:** 665-706.

Sentz, K. and Ferson, S. 2002. Combination of evidence in Dempster-Shafer theory. SAND Report, SAND2002-0835. Sandia National Laboratories, Albuquerque, New Mexico.

Slovic, P. Fischhoff, B. and Lichtenstein, S. 1984. Perception and acceptability of risk from energy systems. Pp. 115-135. in *Public reactions to nuclear power: are there critical masses?* edited by W. R. Freundenburg and E. A. Rosa. AAAS/Westview, Boulder, Colorado.

Slovic, P., Monahan, J. and MacGregor, D. G. 2000. Violence risk assessment and risk communication: the effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior* **24:** 271-296.

Trumbo, C. W. and McComas, K. A. 2003. The function of credibility in information processing for risk perception. *Risk Analysis* **23:** 343-353.

Tufte, E. 1997. *Visual explanations: images and quantities, evidence and narrative*. Graphics Press, Cheshire, Connecticut.

Valverde, L. J. 2001. Expert judgment resolution in technically-intensive policy disputes. Pp. 221-238. in *Assessment and management of environmental risks*, edited by I. Linkov and J. Palma-Oliveira. Kluwer Academic, Dordrecht, The Netherlands.

Vose, D. 1996. *Quantitative risk analysis: a guide to Monte Carlo simulation modelling*. Wiley, Chichester.

Walley, P. 1991. *Statistical reasoning with imprecise probabilities.* Chapman and Hall, London.

Walley, P. and DeCooman, G. 2001. A behavioral model for linguistic uncertainty. *Information Sciences* **134:** 1-37.

Walton, D. 1997. *Appeal to expert opinion: arguments from authority*. Pennsylvania State University Press, Pennsylvania.

Wright, G., Bolger, F. and Rowe, G. 2002. An empirical test of the relative validity of expert and lay judgments of risk. *Risk Analysis* **22:** 1107-1122.