

Structured elicitation of expert judgments for threatened species assessment: a case study on a continental scale using email

Marissa F. McBride^{1*}, Stephen T. Garnett², Judit K. Szabo², Allan H. Burbidge^{3,4}, Stuart H. M. Butchart⁵, Les Christidis^{6,7}, Guy Dutton², Hugh A. Ford⁸, Richard H. Loyn⁹, David M. Watson¹⁰ and Mark A. Burgman¹

¹Australian Centre of Excellence for Risk Analysis, School of Botany, University of Melbourne, Parkville, Victoria 3010, Australia; ²Research Institute for the Environment and Livelihoods, Charles Darwin University, Northern Territory 0909, Australia; ³Department of Environment and Conservation, PO Box 51, Wanneroo, Western Australia 6946, Australia; ⁴School of Natural Sciences, Edith Cowan University, Joondalup Drive, Joondalup, Western Australia 6027, Australia; ⁵BirdLife International, Wellbrook Court, Cambridge, CB3 0NA, UK; ⁶National Marine Science Centre, Southern Cross University, Coffs Harbour, New South Wales 2054, Australia; ⁷Department of Genetics, University of Melbourne, Parkville, Victoria 3052, Australia; ⁸University of New England, Armidale, New South Wales 2351, Australia; ⁹Department of Sustainability & Environment, Arthur Rylah Institute, PO Box 137, Heidelberg, Victoria 3084, Australia; and ¹⁰Institute for Land, Water and Society, Charles Sturt University, PO Box 789, Albury, NSW 2640, Australia

Summary

1. Expert knowledge is used routinely to inform listing decisions under the IUCN Red List criteria. Differences in opinion arise between experts in the presence of epistemic uncertainty, as a result of different interpretations of incomplete information and differences in individual beliefs, values and experiences. Structured expert elicitation aims to anticipate and account for such differences to increase the accuracy of final estimates.
2. A diverse panel of 16 experts independently evaluated up to 125 parameters per taxon to assess the IUCN Red List category of extinction risk for nine Australian bird taxa. Each panellist was provided with the same baseline data. Additional judgments and advice were sought from taxon specialists outside the panel. One question set elicited lowest and highest plausible estimates, best estimates and probabilities that the true values were contained within the upper and lower bounds. A second question set elicited yes/no answers and a degree of credibility in the answer provided.
3. Once initial estimates were obtained, all panellists were shown each others' values. They discussed differences and reassessed their original values. Most communication was carried out by email.
4. The process took nearly 6 months overall to complete, and required an average of 1 h and up to 13 h per taxon for a panellist to complete the initial assessment.
5. Panellists were mostly in agreement with one another about IUCN categorisations for each taxon. Where they differed, there was some evidence of convergence in the second round of assessments, although there was persistent non-overlap for about 2% of estimates. The method exposed evidence of common subjective biases including overconfidence, anchoring to available data, definitional ambiguity and the conceptual difficulty of estimating percentages rather than natural numbers.
6. This study demonstrates the value of structured elicitation techniques to identify and to reduce potential sources of bias and error among experts. The formal nature of the process meant that the consensus position reached carried greater weight in subsequent deliberations on status. The structured process is worthwhile for high profile or contentious taxa, but may be too time intensive for less divisive cases.

Key-words: cognitive bias, consensus, Delphi process, expert elicitation, expert panel, IUCN Red List, subjective knowledge, virtual teams, 4-point estimation method

*Correspondence author. E-mail: mfc.mcbride@gmail.com

Introduction

Conservation managers and practitioners frequently operate with short timelines and limited resources. Particularly in contexts where empirical information is sparse or unobtainable, they may rely on experts as a useful, alternative source of knowledge for decision-making (Sutherland 2006; Martin *et al.* 2012). Experts have acquired learning and experience that allows them to provide valuable insight into the behaviour of environmental systems (e.g. Fazey *et al.* 2006), and they may estimate 'facts' such as population sizes, rates of change or life-history parameters, consolidate and synthesise existing knowledge, determine problem framing and solution methods, and offer predictions about the future (Kuhnert, Martin & Griffiths 2010; Perera, Johnson & Drew 2011; Martin *et al.* 2012).

However, experts may be subject to cognitive and motivational biases that impair their abilities to accurately report their true beliefs. Expert judgments of facts may be influenced by values and conflicts of interest (Krititzsky 1993; Shrader-Frechette 1996; O'Brien 2000) and are sensitive to a host of psychological idiosyncrasies and subjective biases (Table 1), including framing, overconfidence, anchoring, halo effects, availability bias and dominance (Fischhoff, Slovic & Lichtenstein 1982; Kahneman & Tversky 1982; Slovic 1999; Gilovich, Griffin & Kahneman 2002). Structured protocols for elicitation have been developed that attempt to counter these biases. These protocols employ formal, documented and systematic procedures for elicitation, and encourage experts to cross-examine evidence, resolve unclear or ambiguous language, think about where their own estimates may be at fault or superior to those of others and generate more carefully constructed uncertainty bounds. A substantial body of evidence supports the assertion that structured elicitation methods produce more reliable and better-calibrated estimates of facts than do unstructured or naïve questions (e.g. Spetzler & Stael von Holstein 1975; Keeney & Von Winterfeldt 1991; Stewart 2001; O'Hagan 2006).

Within ecology, the uptake of structured methods has been gaining traction (see Choy, O'Leary & Mengersen 2009; Kuhnert, Martin & Griffiths 2010; Burgman *et al.* 2011a; Martin *et al.* 2012 for recent reviews). It is generally agreed that face-to-face interviews and workshop-based methods are the most likely to elicit high-quality responses (e.g. Morgan & Henrion 1990; Clemen & Reilly 2001; O'Hagan 2006; Choy, O'Leary & Mengersen 2009; O'Leary *et al.* 2009; Kuhnert 2011). However, it is not always desirable or feasible to assemble experts together, and a role also exists within ecological applications for methods that facilitate elicitation and interaction among members that are spatially and temporarily distributed (e.g. Donlan *et al.* 2010; Teck *et al.* 2010; Eycott, Marzano & Watts 2011).

In ecology, the elicitation of opinions via remote means is typically conducted with email or postal mail via a tra-

ditional, single iteration questionnaire (e.g. White *et al.* 2005) or an iterative Delphi-style process (e.g. Kuhnert, Martin & Griffiths 2010). In the classical Delphi process (Dalkey & Helmer 1963; Linstone & Turoff 1975; Rowe & Wright 2001), experts make an initial estimate, are provided with anonymous feedback about the estimates of the other group members and then make a second, revised estimate, with the estimate and feedback rounds continuing for some set number of rounds or until a pre-specified level of agreement is reached. The Delphi process is well-established in ecology (e.g. Crance 1987; MacMillan & Marshall 2006; O'Neill *et al.* 2008; Eycott, Marzano & Watts 2011), and it has the advantage when compared with single iteration e-questionnaires and unstructured groups, of allowing judges to revise their judgments in the light of others in the group while alleviating some of the most pervasive social pressures that emerge in unstructured discussion settings (e.g. Kerr & Tindale 2004, 2011; Table 1) through its use of structured interaction and maintenance of participant anonymity.

However, recent reviews and research on the Delphi process suggest that to achieve improvements in accuracy from round to round, experts must be provided with rationales to accompany the feedback they receive about the responses from other group members, and that in the absence of these rationales, their responses will tend to converge only towards a majority position (Rowe & Wright 1999; Rowe, Wright & McColl 2005; Bolger *et al.* 2011; Dalal *et al.* 2011). Incorporation of discussion into the feedback stage of the elicitation is one natural and effective means for providing rationales. Burgman *et al.* (2011b) provide one such example where incorporating a Delphi-based 'talk-estimate-talk' approach into a face-to-face expert workshop resulted in revisions that did indeed contribute to improvements in overall response accuracy. Such structured discussion-based methods are typically incorporated into elicitation as part of workshops (e.g. Delbecq, Van de Ven & Gustafson 1975), but could feasibly be adapted for use in remote elicitation to improve on the standard Delphi methodology (e.g. Turoff 1972; Linstone & Turoff 2011).

The purpose of this paper is to adapt a modified Delphi approach that incorporates facilitator-assisted discussion for use via electronic mail. We apply this method to an assessment of threatened Australian birds. We aimed to test the feasibility of applying such an approach via email and demonstrate the value of structured elicitation techniques for identifying and reducing potential sources of bias and error among experts. Our procedure facilitates the interaction and aggregation of opinions from multiple, distributed experts, and is, we believe, accessible to practitioners and suitable for elicitation in a wide variety of applied ecological settings. The outcomes provide both a motivation for the use of structured procedures and a roadmap to guide future elicitors in the process of conducting structured elicitation successfully.

Table 1. Subjective biases commonly encountered in expert elicitation. Adapted from Supplementary Information Table S3 in Martin *et al.* (2012)

Bias	Description	Illustration	Suggested reading
<i>Individual biases</i>			
Anchoring	Final estimates are influenced by an initial salient estimate, either generated by the individual or supplied by the environment	People give a higher estimate of the length of the Mississippi River if asked whether it is longer or shorter than 5000 miles, than if asked whether it is longer or shorter than 200 miles	Jacowitz & Kahneman (1995); Mussweiler & Strack (2000)
Anchoring and adjustment	Insufficient adjustment of judgments from an initial anchor, known to be incorrect but closely related to the true value	People's estimates of the boiling point of vodka are biased towards the self-generated anchor of the boiling point of water	Tversky & Kahneman (1974); Epley & Gilovich (2005, 2006)
Availability bias	People's judgments are influenced more heavily by the experiences or evidence that most easily come to mind	Tornadoes are judged as more frequent killers than asthma, even though the latter is 20 times more likely	Tversky & Kahneman (1973); Lichtenstein <i>et al.</i> (1978); Schwarz & Vaughn (2002)
Confirmation bias	People search for or interpret information (consciously or unconsciously) in a way that accords with their prior beliefs	Scientists may judge research reports that agree with their prior beliefs to be of higher quality than those that disagree	Lord, Ross & Lepper (1979); Koehler (1993);
Framing	Individuals draw different conclusions from the same information, depending on how that information is presented	Presenting probabilities as natural frequencies (e.g. 6 subpopulations out of 10) helps people reason with probabilities and reduce biases such as overconfidence	Gigerenzer & Hoffrage (1995); Levin, Schneider & Gaeth (1998)
Overconfidence	The tendency for people to have greater confidence in their judgments than is warranted by their level of knowledge	People frequently provide 90% confidence intervals that contain the truth on average only 50% of the time	Lichtenstein, Fischhoff & Phillips (1982); Soll & Klayman (2004); Moore & Healy (2008)
<i>Group biases</i>			
Dominance	Social pressures induce group members to conform to the beliefs of a senior or forceful member of the group	Groups spend more of their time addressing the ideas of high-status members than they do exploring ideas put forward by lower-status members	Maier & Hoffman (1960)
Egocentrism	Individuals tend to give more weight to their own opinions than to the opinions of others than is warranted	Individuals attribute weights of on average 20–30% to advisor opinions in revising their judgments, when higher weights would have been optimal	Yaniv & Kleinberger (2000); Yaniv (2004)
Groupthink	When groups become more concerned with achieving concurrence among their members than in arriving at carefully considered decisions	The invasion of North Korea and the Bay of Pigs invasion have been attributed to decision makers becoming more concerned with retaining group approval than making good decisions	Janis (1972)
Halo effects	When the perception of an attribute for an individual or object is influenced by the perception of another attribute or attributes	Attractive people are ascribed more intelligence than those who are less attractive	Nisbett & Wilson (1977); Cooper (1981); Murphy, Jako & Anhalt (1993)
Polarisation	The group position following discussion is more extreme than the initial stance of any individual group members	Punitive damages awarded by juries tend to be higher than the median award decided on by members prior to deliberation	Myers & Lamm (1976); Isenberg (1986); Sunstein (2000)

Materials and methods

CASE STUDY

This study was undertaken as part of the assessment of the IUCN Red List status of all species and subspecies of Australian birds. This is the third time this exercise has been undertaken in the last two decades. In 1990 (Garnett 1992) and 2000 (Garnett & Crowley 2000), individual experts were contacted and the information they provided was assessed against the prevailing IUCN criteria with uncertainties adjudicated by the authors. However, as the significance of the IUCN Red List has grown (e.g. Butchart *et al.* 2010; United Nations 2011), and as more money has been invested in threatened species conservation, there has been a requirement to develop a more formal process to evaluate the IUCN Red List status of taxa about which there is uncertainty.

The IUCN Red Listing process

The assessment of the conservation status of species worldwide is most frequently carried out using the IUCN Red List protocols (IUCN 2001; IUCN Standards and Petitions Subcommittee 2010). The IUCN system consists of a set of criteria with quantitative thresholds for each category of extinction risk (ranging from Least Concern through to Critically Endangered and Extinct). Species are classified at the highest category for which they meet the thresholds under any one of five rule sets. Classification requires quantitative estimates for numerous parameters in relation to these thresholds (Table 2). Expert judgments form an essential part of the listing process, because direct data on the parameters for listing are often outdated, incomplete, approximate, uncertain or unavailable (e.g. Newton & Oldfield 2008; Luky, Crawford & Gillis 2010). Previous studies have examined the effects of expert assessors on the listing process and observed high levels of operator error and variation in listing decisions among experts (e.g. Keith *et al.* 2004; Regan *et al.* 2005). However, while expert judgment is a valid method for assessing IUCN Red List status, the IUCN guidelines provide minimal guidance on how best to elicit expert opinion.

CHOICE OF EXPERTS

Two groups of experts took part in the elicitation: a group of 16 panel members who completed the full elicitation process and provided assessments for multiple species, and a second group of 12 taxon specialists who provided assessments for their specialty taxon. The expert panel (panellists) comprised 16 ornithologists identified by their track record, experience, knowledge of the birds of particular regions or specialist skills (taxonomy, IUCN Red Listing, particular bird taxa). Most of the panellists had previously worked together on the assessment of IUCN status for Australian birds. All panellists had published extensively on Australian birds and were selected from what we believe to be a relatively small (<100) pool of people with similarly high levels of experience.

Nine of the authors (AB, SB, LC, GD, HF, SG, RL, JS, DW) were members of this expert panel. All panellists involved have had a long-standing commitment to the conservation of Australian birds, although they also demonstrated a desire that the determination of their status be well grounded in science. Not all panellists assessed all taxa. Time restrictions prevented some panellists from assessing all taxa, and some panellists

omitted taxa for which they felt they possessed inadequate knowledge. Ten of the panellists assessed all nine taxa, two assessed eight, one assessed four and three assessed three.

A second group of 12 taxon specialists was invited to inform the expert panel of their views on the parameters, one of whom was also a panel member. Taxon specialists were identified by their association with interest groups and scientific societies, or by their relevant publications. Invitations were made to all researchers known to have undertaken research and published on each taxon in the last decade, although not all invitations were accepted and not all taxa had been the subject of research.

BACKGROUND INFORMATION

All experts (panellists and taxon specialists) were provided with background information derived from an account of the status of the taxon published a decade earlier (Garnett & Crowley 2000), subsequent literature published on the taxon, government assessments of the taxon's status and submissions by taxon specialists. Background information varied greatly in quantity and quality between taxa.

STRUCTURED ELICITATION PROTOCOL

We used a structured procedure (Fig. 1) for questioning experts, adapted from the workshop-based procedure used in Burgman *et al.* (2011b) for implementation via email. The key novel elements integrated into this procedure are:

(i) A four-point question format (Speirs-Bridge *et al.* 2010) for eliciting quantities to mitigate the overconfidence effects typically observed in expert estimates of uncertainty (e.g. Lichtenstein, Fischhoff & Phillips 1982; Russo & Schoemaker 1992; Soll & Klayman 2004). This approach has been applied so far only in Burgman *et al.* (2011b). It draws on research from psychology on the effects of question formats, and while structurally similar, differs from existing methods that have been applied for eliciting quantitative estimates of uncertainty in ecology (e.g. O'Neill *et al.* 2008; Murray *et al.* 2009; O'Leary *et al.* 2009; Rothlisberger *et al.* 2010) which typically involve the use of greater numbers of questions per parameter and more statically complex concepts.

(ii) The structured interaction of experts via email discussion, based on increasing evidence that suggests that providing reasoning behind the estimates from other group members is required in order to enable the revisions made during Delphi process iterations to improve accuracy (Bolger *et al.* 2011). While the Delphi process and modified variants including expert discussion at face-to-face workshops are commonplace in ecology (e.g. MacMillan & Marshall 2006), the inclusion of facilitated email discussion between experts at the feedback stage has not previously been explored. While anonymity is usually maintained in Delphi processes, in this study participants could elect to waive anonymity (though it would be possible to also conduct email discussions anonymously if necessary).

PROCEDURE

The procedure for the elicitation was as follows:

Pre-elicitation

Both panellists and taxon specialists were contacted via email by the organisers (one of us, SG), who described the process and expected

Table 2. Summary of the parameters elicited from experts for each taxon for assessment against the IUCN Red List criteria. Experts provided assessments for up to 16 quantitative parameters and up to 33 categorical parameters.

Quantitative parameters	Categorical parameters
Generation time	Fluctuations and fragmentation
Age at first breeding	Are there extreme fluctuations (more than 10-fold) in any of EOO, AOO, number of locations or number of subpopulations
Oldest bird in the wild	Are there extreme fluctuations (more than 10-fold) in number of mature individuals
Percentage of adults surviving from one year to the next	Is the population extremely fragmented (> 50% of AOO in isolated patches too small for long term persistence)
Population size	In the last 10 years/3 generations
Total population size (number of mature individuals)	(i) Has there been a decline in EOO?
Number of subpopulations	If there has been a decline, is it (ii) continuing, (iii) reversible, (iv) understood, or (v) ceased?
Number of mature individuals in largest subpopulation	(i) Has there been a decline in AOO?
Trends in mature individuals	If there has been a decline, is it (ii) continuing, (iii) reversible, (iv) understood, or (v) ceased?
Population size	(i) Has there been a decline in habitat area/extent/quality?
(i) 1 generation/3 years ago	If there has been a decline, is it (ii) continuing, (iii) reversible, (iv) understood, or (v) ceased?
(ii) 2 generations/5 years ago	(i) Has there been a decline in number of locations/subpopulations?
(iii) 3 generations/10 years ago	If there has been a decline, is it (ii) continuing, (iii) reversible, (iv) understood, or (v) ceased?
Population size in	(i) Has there been a decline in the number of mature individuals?
(i) 1 generation/3 years	If there has been a decline, is it (ii) continuing, (iii) reversible, (iv) understood, or (v) ceased?
(ii) 2 generations/5 years	
(iii) 3 generations 10 years	
Biggest change in any period of 3 generations/10 years (% change)	
Geographic parameters	
Extent of occurrence (EOO) (km ²)	Has any change in the number of mature individuals been
Area of occupancy (AOO) (number of km ² grid cells occupied)	(i) Observed, (ii) Estimated, (iii) Projected, (iv) Suspected, or (v) Inferred*
Number of locations (distinct areas that could be engulfed rapidly by a single threat)	

*Each of these terms is specifically defined under the IUCN criteria.

outcomes, including the objective to reach consensus assessments for nine taxa over which there was disagreement concerning conservation status.

A telephone meeting between panellists and another of us (MB) was used to outline the structure and details of the elicitation process and to answer technical questions. Further communication with the group, including all discussion between experts, was thereafter by email, although some individual panellists were telephoned to elicit missing information or to resolve inconsistent responses.

Elicitation

Stage 1: estimation – round one. Experts were emailed a spreadsheet containing sets of up to 125 questions required for assessment of the IUCN categorisation parameters for each of the taxa (up to nine) to be assessed (Table 2).

The questions requested estimates of quantities and percentages using a four-point procedure (Speirs-Bridge *et al.* 2010; Burgman *et al.* 2011b) in the following format:

1. What is the lowest the value could be? (α)
2. What is the highest the value could be? (β)
3. What is your best estimate? (the most likely value)? (γ)
4. How confident are you that the interval you provided contains the truth (provide an answer in the range of 50–100%)? (ρ)

Participants were asked to interpret this as a frequency of repeated events, such that their judgments over many questions should be

accurate with the prescribed frequency. For example, if they provided 10 intervals in response to 10 questions and specified 50% certainty for each interval, then the truth should fall within their intervals, five times out of 10.

The questions for binary (yes/no) questions used a two-step procedure:

1. Is the statement true or false? (I)
2. How sure are you that your answer is correct (provide an answer in the range of 50–100%)? (p)

Participants were asked to interpret this as a bet. That is, state the probability which reflected the odds they would accept in a gamble on the outcome of the judgment (true or false).

Experts completed each question in the spreadsheet, resulting in an initial, private estimate of each parameter, for each taxon. Experts answered up to 125 questions per taxon, including 16 quantitative questions (each of which had four parts) and 33 binary questions (28 with two parts). These were required to determine IUCN Red List categories (Table 2; the number of questions varied because some answers precluded the need to answer contingent questions). They also recorded the time taken to complete each taxon assessment. Experts were allowed 2 weeks to complete the task. Some experts chose not to answer some questions on some taxa and left their responses blank.

Stage 2: feedback. The full set of individual estimates was compiled by one of us (JS). Estimates for quantities and percentages were standardised to fit 80% credible bounds around each individual's best esti-

mate using linear extrapolation (Bedford & Cooke 2001; see 'characterisation of uncertainty' section for details). The results were displayed in graphs in a spreadsheet and then distributed back to the panellists so that they could compare their estimates with others. A facilitator (SG) drew attention to major differences between panellists, particularly where these had an impact on the IUCN Red List category, some of which were then discussed by group email. Any new information from taxon specialists or other sources was distributed. Panellists were given the opportunity to resolve ambiguities over the meanings of terms, specify context and introduce and cross-examine relevant information.

Stage 3: estimation – round two. At the end of 3 weeks, all panellists completed a second set of final confidential assessments for each of the questions assessed in round 1, in which they were asked to reconsider their previous assessments in the light of the discussion. These revised assessments were then used in the final determination of status.

Post-elicitation

Panellist responses were used to generate individual expert IUCN listings for each taxon. The final (second round) responses and listings were circulated to participants for comment and final approval. A set of aggregate estimates was calculated using the mean of panellist assessments for each parameter. These were used to determine a group 'consensus' IUCN listing for each taxa. These, along with the individual IUCN listings generated from each individual experts' responses, were circulated to panellists for comment and approval. An additional post-elicitation discussion, separate from the formal structured elicitation process, took place among participants about the validity of the listing outcomes for three of the nine taxa.

CHARACTERISATION OF UNCERTAINTY

A framework has been developed for incorporating uncertainty into the IUCN listing process using fuzzy numbers (Akçakaya *et al.* 2000; Mace *et al.* 2008), which we adopt here. However, our elicitation methodology is also suitable for the elicitation of probabilities and probability distributions. Fuzzy numbers (Kaufmann & Gupta 1985; Zimmermann 2001) are a non-probabilistic approach to representing uncertainty that avoids the need to represent uncertainty via a statistical distribution (Akçakaya *et al.* 2000). Using fuzzy numbers, the expert-assigned estimates of uncertainty (ρ for quantitative questions and p for categorical questions) are interpreted as possibility measures, a measure of

the degree of plausibility of a statement or reliability of the evidence associated with it. Quantitative and categorical parameter estimates from experts were used to construct fuzzy triangular measures, with parameters $[a,b,c]$ defining the minimum (a), most likely (b) and maximum (c) values (Fig. 2). Constructed fuzzy numbers were inputted into the RAMAS Red List software (Akçakaya & Ferson 2007), which propagates uncertainty through the listing process to determine the range of possible final IUCN categorisations.

Quantitative estimates elicited using the four-point estimation method were normalised using linear extrapolation (Bedford & Cooke 2001) to absolute lower (α_{abs}) and upper (β_{abs}) bounds within which 100% of all estimates might be expected to fall, such that

$$\alpha_{\text{abs}} = \gamma - (\gamma - \alpha)(c/\rho)$$

$$\beta_{\text{abs}} = \gamma + (\beta - \gamma)(c/\rho)$$

where c is the required possibility level (100%) and ρ is the experts' stated confidence. These 100% interval bounds were used as the minimum (a) and maximum (c) values for the triangular fuzzy numbers, and the best estimate (γ) was taken as the most likely value (b) (Fig. 2a). The same approach, taking c as 80%, generated the standardised 80% intervals for experts to view and compare each other's responses at the feedback stage of the elicitation. Categorical estimates were represented as triangular distributions with parameters $[p,p,1]$ for 'yes' responses, for (p) the expert-assigned level of confidence and as $[0,p,p]$ for 'no' responses (Fig. 2b,c).

AGGREGATION OF OPINIONS

Even with multiple deliberation and reassessment stages, it is rarely possible to arrive at a complete agreement between experts on parameter values. However, combining disparate opinions raises several methodological difficulties, principally because there is no objective basis for combining multiple expert opinions (Keith 1996; Clemen & Winkler 1999). For the IUCN Red List assessments, experts' estimates for each parameter are not independent, but conditional on their mental models about the taxon's ecology and status. Experts' estimates of the range of uncertainty about current and future population sizes, for example, may be contingent on beliefs about the severity of impact of a given threat. To account for this, we focused first on evaluating listings individually for each panellist's set of responses (e.g. Titus & Narayanan 1996). This provided a set of possible listing classifications under the panellists' different men-

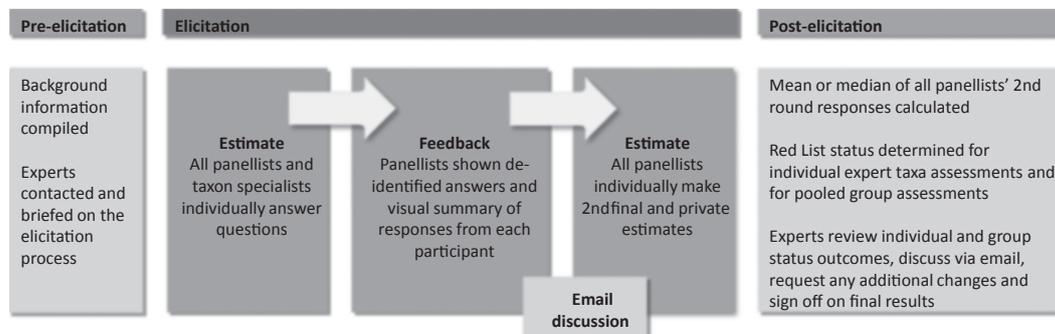


Fig. 1. Summary flow-chart of the structured elicitation procedure. All stages were conducted remotely via email or telephone. In the 'elicitation' stage, dark grey boxes show steps conducted by panellists individually, and light shaded boxes show stages conducted as a (virtual) group. Pre- and post-elicitation stages involved both individual and group interaction components.

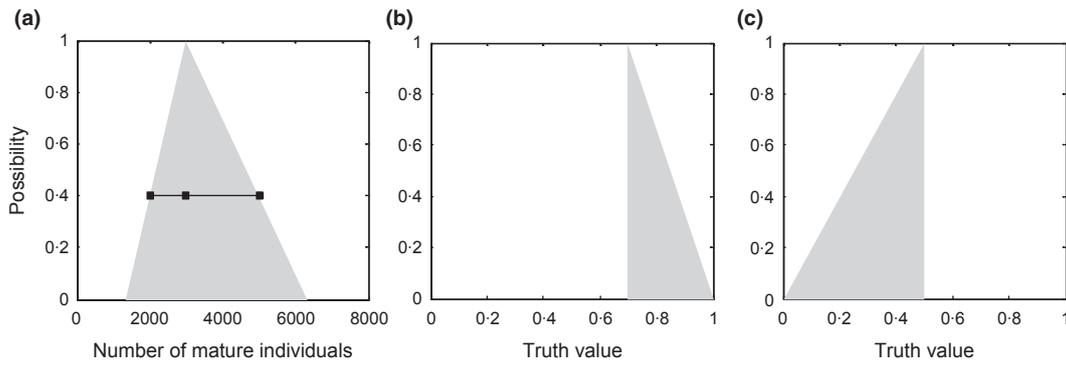


Fig. 2 Illustration of triangular fuzzy numbers constructed for the panellists' responses for input into the RAMAS Red List software for (a) quantitative and (b, c) categorical questions. The y-axis measures the possibility level, which corresponds inversely to the confidence with which we believe the true value lies within the bounds of the fuzzy number at that level of possibility. In (a), black square markers show the expert's assessment of the lower (α), upper (β) and best estimate (γ) ([2000, 3000, 5000]), for number of mature individuals in the population plotted on the possibility scale at 0.4, the inverse of the expert's stated confidence level (ρ) of 60%. Linear extrapolation is used to determine the minimum and maximum values at which the bounds of the fuzzy number bounds cross the x-axis (1333 and 6333 respectively). In (b), the panellist's 'no' response with assigned confidence (ρ) of 50% is represented by the fuzzy number [0, 0.5, 0.5]. In (c), the panellist's 'yes' response with assigned confidence (ρ) of 70% is represented as the fuzzy number [0.7, 0.7, 1].

tal models (Table 3) and allowed us to focus on understanding and resolving the differences in parameter estimates that lead to conflicting classifications. However, as a final listing was still required, we also determined a group IUCN Red List assessment for each taxon using the mean of the normalised responses (that is, the mean of the group's best estimates and the means of the normalised upper and lower bounds), the standard approach in Delphi-style elicitation exercises.

RESPONSE ANALYSIS AND FOLLOW-UP

Following completion of the elicitation, the expert responses were reviewed for evidence of bias (e.g. Table 1). Instances of possible bias were documented and, where possible, followed up by further discussion with panellists about the reasoning behind their responses. Additional analyses were undertaken to characterise some instances of bias for inclusion in this manuscript.

To characterise the effects of the discussion stage on responses, we examined changes in:

- (i) Levels of confidence in responses, tabulated for each taxon across all questions and experts;
- (ii) Patterns of agreement and disagreement between expert responses, measured in terms of the 'proportion of non-overlap', the proportion of all possible expert-to-expert pairings for each question for which their intervals (once normalised to 100% intervals) were non-overlapping, totalled across quantitative questions, and the average coefficient of variation (CV) between panellist responses for each question, which is defined as:

$$CV_{question\ i} = \sigma_i / \mu_i$$

where μ_i and σ_i are the mean and standard deviation of panellist responses for question i . For quantitative questions, CVs were calculated separately for each of the lower, upper and best estimate responses. For categorical questions, the centre of mass of the triangular fuzzy numbers was used to translate the fuzzy numbers into single, crisp values, a standard approach for 'defuzzification' (e.g. Yager & Filev 1993; Van Leekwijck & Kerre 1999).

Experts provided comments on the process and the results in a follow-up email questionnaire. They were also offered the chance to revise for a second time their responses for the three most contentious taxa.

Table 3. Disagreement in final taxa classifications among panellists. Table values indicate the percentage of panellists arriving at each categorisation level for each taxon based on final round assessments

Taxon	<i>n</i>	C	E	V	NT	LC	DD
A	13		15	77	8		
B	13		38	23	31	8	
C	11			27		73	
D	12		17	33	25	8	17
E	16		13	19	19	44	6
F	12			17	8	75	
G	15		19	50	25	6	
H	12			42	8	50	
I	15		50	21		29	

n, number of expert assessments for the taxon; C, critically endangered; E, endangered; V, vulnerable; NT, near threatened; DD, data deficient.

Results

PROCESS

The process, from the initial invitation to the final assembly of results following the submission of second-round estimates took 10 weeks, and the subsequent additional post-elicitation discussion took a further month. Discussion after the first round centred on a few major differences between estimates by the panellists and taxon specialists, but there was no attempt to reach consensus. A more active discussion took place following the elicitation when the final responses were condensed into a recommendation for an IUCN Red List category for each taxon. At this stage, panellists discussed the apparent appropriateness of the category, rather than the underlying parameter estimates. Rapid consensus of opinion occurred among panellists on six taxa during the discussion stage of the elicitation, during which minor differences were clarified and an agreed position reached. Of the remaining three, one was dis-

cussed at length in the post-elicitation discussion. This could be largely attributed to strong advocacy from individuals outside the group with some responsibility for the conservation and management of the species. These discussions were eventually terminated by the facilitator to meet pre-determined time-tables for evaluation of taxon status.

EMAIL FORMAT

A notable feature of this deliberative elicitation process was that it was conducted by email. Email groups are highly flexible and allow individuals to participate at their convenience with ready access to outside resources and without the requirement that members be assembled simultaneously (Martins, Gilson & Maynard 2004). Feedback about the process from panellists was largely favourable. Panellists cited the written format as providing time to digest the information and opinions provided by others before making a reply, and allowing 'more time to spend pondering the issues and considering comments and responses than would have occurred during a workshop' [feedback comment from panellist]. The automatic documentation of all communication was also seen as an advantage, as it allowed panellists to revisit the evidence and differing opinions at any stage and encouraged them to provide what appeared to be 'more considered and reasoned responses in what they knew would be a permanent record of their information or views'.

The time involved in compiling responses and the protracted period of consultation and discussion were seen as the key drawbacks. Electronic groups tend to require more time than a face-to-face group to complete tasks (e.g. Hiltz, Johnson & Turoff 1986; Hollingshead 1996; Baltes *et al.* 2002), and most panellists commented on the large amount of work involved (e.g. the amount of reading, becoming familiar with the approach, multiple rounds of assessments etc.). Panellists took 3.8–13.5 h to complete assessments in the round 1 assessment process alone (not including reading time), with an average of 60 min (± 14.8 min SD) taken by panellists and taxon specialists to assess each taxon (Fig. 3). The same process would usually be achieved in a 2- to 3-day workshop in a group setting.

Panellists also regretted the lack of the opportunity that a group setting provides to discuss issues in person, particularly for the more contentious issues for which the ability to give and receive behavioural cues might assist in reducing misunderstandings and conveying greater nuance (e.g. Hinds & Bailey 2003; Maruping & Agarwal 2004). Despite this, the general consensus appeared to be that there was an adequate level of discussion and debate, that the email format 'reduce[s] dominance and gives less confident and articulate people a better opportunity to contribute meaningfully' and that 'no one voice or view dominated, and frank and balanced discussion were achieved'.

BIAS AMELIORATION

The elicitation procedure was designed to offset a range of predictable biases to which experts are prone when considering

uncertainty. The expected biases emerged and were remediated with varying levels of success.

Anchoring

Anchoring occurs when an expert uses a published estimate or the estimate of a colleague as a reference point for their own judgment. This is apparent when, for example, a number of experts use published values to construct their intervals of uncertainty. For the IUCN Red List assessment, anchoring could take one of two forms: adherence to published values, or estimates that reflect quantities associated with thresholds in the IUCN Red List criteria. Both types of anchoring were observed in this study.

Example 1. The only published estimate of the population size for taxon H is 6500 birds, although this estimate was considered to be of low reliability and accuracy. Eight of 12 panellists initially stated that their best estimate was between 6000 and 7000 mature individuals (Fig. 4), even for those who also believed that substantial declines have been occurring since 2001 when the estimate of 6500 was published. The published figures were accompanied by a description of the ways in which the data were collected,

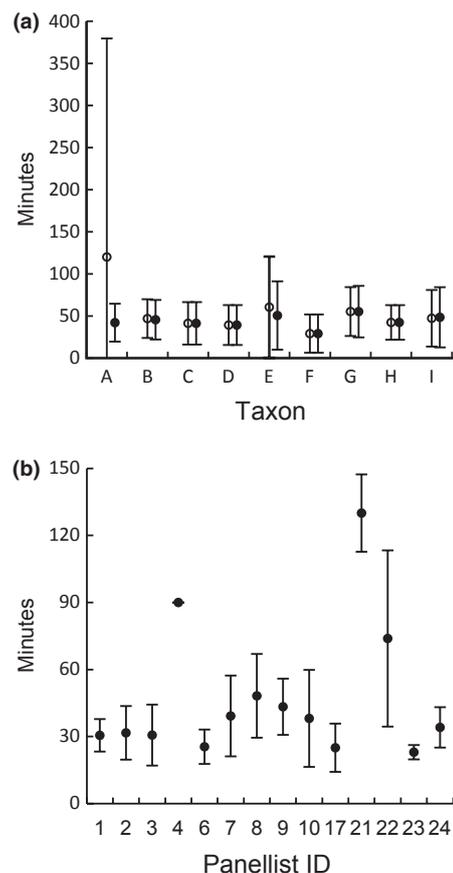


Fig. 3. (a) Average time taken per taxon by panellists to complete IUCN Red List assessments. (a) Average assessment time per taxon for taxon specialists (open markers) and panellists (filled markers), and (b) Average assessment time per taxon for each panellist.

which were essentially guesses. Information was available that would have led to other conclusions but which was not associated with precise figures. Thus, the figures on which people appear to have anchored were far less certain than would appear from the estimates made. A published analysis of additional data was circulated subsequently and led to agreement by the group that the population is likely to be well in excess of 10 000 mature individuals.

Example 2. One panellist consistently estimated values of population sizes such that they were one individual below threshold values for the taxon to be considered at a lower level of extinction risk. For example, the upper bound of their estimate of mature population size for a poorly known bird was estimated at 2499, and the threshold for classification as Vulnerable is <2500. While time and the number of questions precluded this, a more in depth questioning procedure forcing the expert to reconsider and justify their reasoning may have encouraged the expert to revise their estimates and remove the influence of the listing threshold on their responses.

Dominance

Dominance effects arise when a senior or forceful individual in a group setting makes pronouncements about facts, leading others to gravitate to their position. The use of email to conduct all interactions reduced many of the potential effects of dominance (e.g. Turoff & Hiltz 1982; Hollingshead 1996; Martins, Gilson & Maynard 2004). There was no face-to-face group setting and limited opportunity for any individual to use non-written communication to exert dominance. When a forcefully expressed external submission came to the committee, there was discussion of the ideas it included, but the tone of the submission was assessed by the group as inappropriate.

One panellist suggested that dominance could potentially be exerted by use of an aggressive tone in emails, and that there can be a reluctance to be the first person to disagree with the dominant proponent as it could lead to confrontation. In a group context, an aggressive approach can be mitigated by group dynamics or by the facilitator, while unintended aggression or misinterpretation of statements can more readily be corrected if there are visual cues (Martins, Gilson & Maynard 2004; Maruping & Agarwal 2004). Such actions are more difficult via email. In this study, no aggression was exhibited by any of the panellists.

Overconfidence

Overconfidence in interval estimation is the tendency for experts to assign unrealistic reliability to their intervals. Ideally, expert intervals will be well-calibrated, and stated levels of confidence will correspond to the frequency with which intervals contain the observed or 'true' value (Lichtenstein, Fischhoff & Phillips 1982). For example, a set of 90% intervals from a well-calibrated expert should contain the correct value on average 90% of the time. In practice, both expert and non-expert esti-

mates frequently exhibit overconfidence (Lin & Bier 2008), although the effects tend to be mitigated by question format (Wilson 1994; Soll & Klayman 2004; Jenkinson 2005). The methods employed in this elicitation process were designed to reduce expert overconfidence, although they would not entirely remove it (Speirs-Bridge *et al.* 2010).

In this study, it was not possible to test directly for overconfidence as the answers to the IUCN classification questions are unknown. A comparison of 100% confidence bounds from panellists for each taxon revealed that many confidence intervals did not overlap or did so only partially (Table 4). This suggests that at least some of the panellists must have been overconfident in their estimates of uncertainty. High levels of inter- vs. intra-expert variability are common (e.g. O'Neill *et al.* 2008; Czembor & Vesik 2009), and eliciting judgments from just one or a few experts might have masked the true level of uncertainty. The extent of overconfidence in the estimates of population size for taxon H was suggested when alternative data relating to population size were discovered, resulting in revised estimates for the panellists that were considerably larger than the original estimates.

Framing effects

Trend estimates are more likely to be accurate when made using natural numbers from which percentage changes can be calculated, rather than from direct estimates of percentage change (Gigerenzer & Hoffrage 1995). In the assessment process, experts estimated past and future percentage population declines over three generations using natural frequencies and made estimates directly of the greatest percentage change expected in any three-generation period encompassing the present. In 13% of cases, the percentage decline estimated from natural numbers for the entire six-generation period was greater than the greatest decline for any three-generation per-

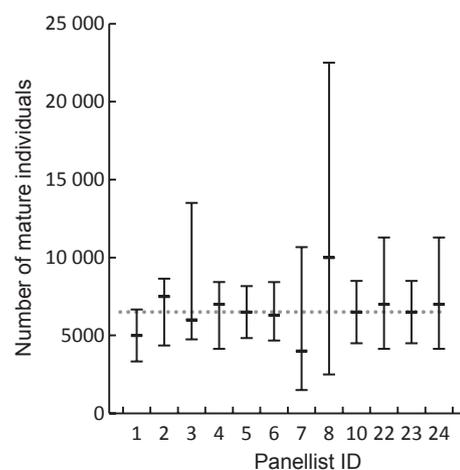


Fig. 4. Initial estimates of population size for taxon H with 100% credible intervals and best estimate (black bar). The dotted line corresponds to the published estimate of 6500 individuals [on which panellists may have anchored].

iod, implying that current declines are less than in the past, the future or both. Similarly, 27% of the panellists' estimates for past population reduction and 23% of the panellists' estimates for future population reduction in the next three generations calculated from natural numbers were greater than the respective panellist's estimate of the maximum three-generation period decline when estimated directly as a percentage. While it is possible that the panellists believed that rates during the middle period were faster or slower than at the beginning or end, we judge that this is less likely than the possibility that the panellists failed to appreciate the internal inconsistency in their estimates. These inconsistencies were spread among panellists: each panellist provided at least one example of such inconsistency (we omitted one panellist who saw the potential inconsistency and simply calculated the percentages using the maximum for any three-generation period, estimated from natural numbers).

Availability bias

Availability bias refers to the phenomenon where experts' judgments are conditioned on the basis of recent or high-profile events, weighting them more heavily than is warranted by the data. While the procedures outlined previously provide the opportunity for panellists to cross-examine and reflect on the relative importance of data, availability bias may emerge nevertheless. For example, one submission from taxon specialists advocated that a taxon be classified as threatened (i.e. in one of the categories of Vulnerable, Endangered or Critically Endangered) when the committee had concluded it was of Least Concern. The submission emphasised the political implications of de-listing (failing to appreciate that these are excluded from the IUCN Red List system), but the only new information presented was a personal communication that the birds had declined in a particular forest: 'evidence suggests that the sub-population in [name omitted] National Park has declined in

recent years'. Count data subsequently became available which did not corroborate this opinion and indicated population stability at this site over the last 18 years. However, following the submission but before the count data became available, three panellists revised their assessments of rates of decline and four urged a more precautionary approach to the data, although it is unclear to what extent they were influenced by the submission, by their re-evaluation of the original data (we note this might also be an example of dominance effects), or by the political implications of a wrong decision. This can also be interpreted as an example of initial overconfidence, which was then shaken when challenged, even without substantial new data.

Language-based misunderstanding

Linguistic uncertainties are pervasive in language-based deliberations and qualitative risk assessments (Regan, Colyvan & Burgman 2002). Vagueness is when categories associated with words have borderline cases, and an entity may belong to more than one category. Ambiguity occurs when words have more than one meaning, and it is not clear which is intended. Context dependence refers to situations in which the meanings of words depend on the contexts in which they are meant to be understood, and this context is not clear or consistent. Under-specificity occurs when information critical for understanding what is meant by words has not been provided. In most situations, linguistic uncertainties may be resolved with careful, facilitated discussion.

Language-based misunderstandings arise when experts misinterpret technical terms used in the IUCN Red List system, despite definitions being provided (IUCN 2001), along with detailed guidelines for their interpretation (IUCN Standards and Petitions Subcommittee 2010). For example, according to the IUCN guidelines, the term 'location' defines a geographically or ecologically distinct area within which a single threatening event can rapidly affect all individuals of the taxon (IUCN 2001). The term is often confused with 'subpopulation', which is defined as distinct groups in the population between which there is little demographic or genetic exchange (IUCN 2001).

In this study, the frequency distributions of the estimates for number of locations for widespread taxa were dichotomous, with some panellists stating simply that there were more than 10 locations (a threshold for listing as Vulnerable) and other panellists estimating that there were very few (one and five locations are the thresholds for Critically Endangered and Endangered respectively). For four of the panellists, it was possible to detect a positive correlation between the number of subpopulations and the number of locations estimated across the nine taxa (Pearson's $r = 0.36-1.0$), regardless of whether distributions were continuous or fragmented, where it might be expected. When panellists were asked to name the locations explicitly as well as the threats used to define them, there was considerable variation in how panellists had chosen to interpret 'location' in determining their estimates, and some panellists had failed to interpret

Table 4. Degree of overlap of intervals before and after discussion (pairwise) for each of the nine assessed taxa

Taxon	Proportion of non-overlap		% reduction in non-overlap
	Before discussion (<i>n</i>)	After discussion (<i>n</i>)	
A	6.4	2.5	60
B	5.5	2.8	48
C	22.8	12.9	43
D	9.9	6.7	32
E	8.8	4.5	49
F	25.0	14.5	42
G	34.2	11.1	68
H	8.2	3.7	55

The proportion of non-overlap is calculated as the proportion of all possible expert-to-expert pairings for each question for which their intervals (once normalised to 100% intervals) were non-overlapping, totalled across the 16 quantitative questions answered by experts for each taxa.

the term correctly. One panellist withdrew all their estimates of location because they felt that they did not understand the term sufficiently well.

Variability among experts

Experts' responses may include evidence of systematic bias, with experts sometimes producing systematically optimistic or pessimistic responses, or systematically larger or smaller estimates of uncertainty compared with other members of the group (e.g. Cooke 1991; Meyer & Booker 1991). For example, a multiple comparisons test across panellists for differences in estimated future declines over three generations for each taxon found that one panellist gave consistently higher estimates of population declines, and two panellists gave consistently lower estimates of population declines (Fig. 5). When discussing their responses, some panellists stated that they had taken an optimistic or conservative stance to uncertainty when specifying their responses. For example, one panellist noted that 'several of my "upper limits" were higher than others, and that's partly because I'm a natural optimist, but also because I think we sometimes kid ourselves about [i.e. exaggerate] our ability to find most of the birds in a population'. Another stated that in specifying their uncertainty bounds, '[given the high levels of uncertainty] the exact estimates then seem to come down to how precautionary we wish to be'.

Convergence

A key feature of the protocol is that people typically respond to the results of other panellists and move towards a consensus range of values that is more likely to include the correct response than any expert is likely to reach alone (the so-called 'theory of errors', see Dalkey 1975). In this exercise, we had two exemplars of this. The coefficient of variation (CV) for participants' upper bounds was more likely to be reduced after discussion, creating tighter bounds for estimates from which to judge taxon status (Fig. 6a). The levels of overlap between panellist estimates were also higher in the second round (Table 4); outlying estimates, in particular, tended to move towards the group average.

However, in many cases, considerable variation between expert parameter estimates persisted even after the second round of assessments. There was no appreciable convergence among questions requiring a yes/no response (Fig. 6b) and estimates of confidence around numbers changed little (Fig. 7). The IUCN categories determined based on individual second-round assessments also revealed significant levels of inter-expert variation (Table 3).

Discussion

This study demonstrates the successful implementation of a structured elicitation process conducted via email for assisting in reducing predictable judgmental biases and pooling knowledge across multiple, dispersed experts. The efficacy of the elicitation process can be considered separately from the medium through which it was undertaken.

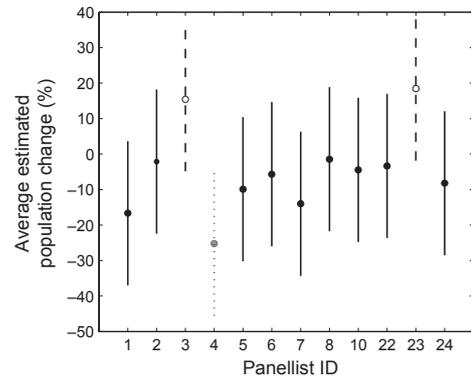


Fig. 5. Results of a multiple comparisons test for panellist average estimates of 'percentage change in population size over the next three generations' across the nine taxa assessed. Bars show 95% confidence intervals after Bonferroni correction for multiple comparisons. One panellist was found to be consistently more pessimistic about future population changes (dotted line), and two panellists to be consistently more optimistic about future changes (dashed lines).

tation process can be considered separately from the medium through which it was undertaken.

ELICITATION PROCESS

The elicitation process clarified the parameters and evidence critical to listing and encouraged the panellists to make considered assessments of every parameter required for Red List assessment, instead of making quick assessments of only those parameters considered critical for each taxon. At least four of the nine taxa are likely to have been assessed differently had a less formal process been followed or had the elicitation process been reduced to a single step. Despite the evidence of bias, non-overlap of some values and high levels of uncertainty, the panel and most of the taxon specialists considered the final recommendations on the IUCN Red List categories to be closer to the truth than had the assessment been carried out by a single individual. An evaluation of the efficacy of the IUCN listing process by Keith *et al.* (2004) found that the median of assessed IUCN ranks across experts resulted in a higher proportion of correct assessments than any individual expert assessor, suggesting that such a belief even in the face of the high levels of disagreement observed, may be well founded.

It is possible that the group converged in line with shared perceptions that were untrue (Kahan 2010). We speculate that where panellists did revise their initial responses, they did so because they (i) had no knowledge of the parameter themselves, (ii) aligned responses to those of a notable taxon specialist, or (iii) were swayed by a particular line of reasoning during the discussion. Assessing and enhancing the degree to which estimate revisions lead to improved accuracy represents an important goal in future applications (e.g. Rowe, Wright & McColl 2005). In the majority of elicitation contexts it will be impossible to assess the accuracy of responses. However, it is possible to test levels of accuracy and calibration using test ('seed') questions, domain relevant questions for which there is

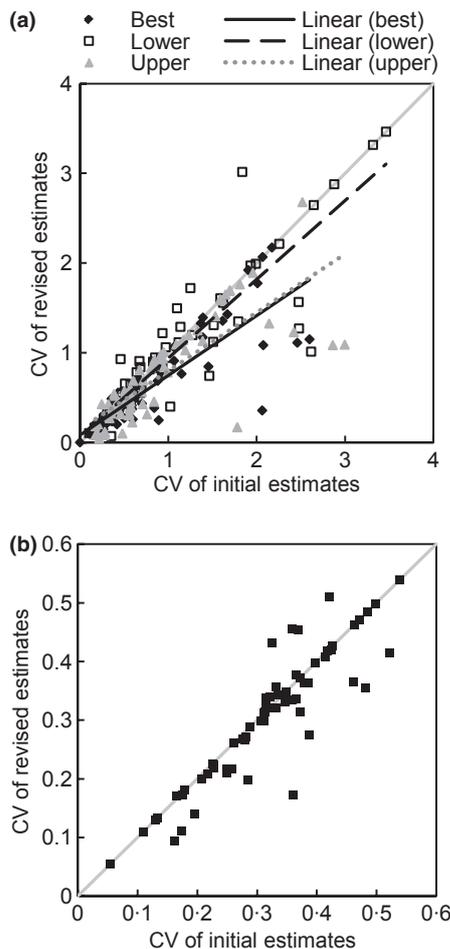


Fig. 6. Average coefficient of variation (CV) among panellist responses for initial and revised estimates for (a) lower, best and upper values for the 16 quantitative parameters and (b) yes/no questions. Each dot measures the variation across all panellist responses to an individual question for a single taxon. Values above the equal value line (solid grey line) suggest answers diverged after discussion, and values below suggest they converged.

a determinable ‘truth’, which the facilitators know but the experts do not (Cooke 1991; Aspinall 2010). However, finding relevant context-specific seed questions on which to base assessments can be challenging. In this study, we omitted seed questions partly because the number of IUCN categorisation questions already requiring assessment was prohibitively large, and partly because of the difficulty in finding appropriately relevant questions for which answers were not already known to at least one of the panellists.

DEBIASING TECHNIQUES

Our procedure demonstrated that the same sources of bias apparent among other professional groups are present in judgments from environmental scientists. The transparency of our approach allowed biases to be identified, and it is likely that they are equally present in listings made using non-structured approaches, if not equally observable. A good protocol will aim to reduce the effects of biases, and our method

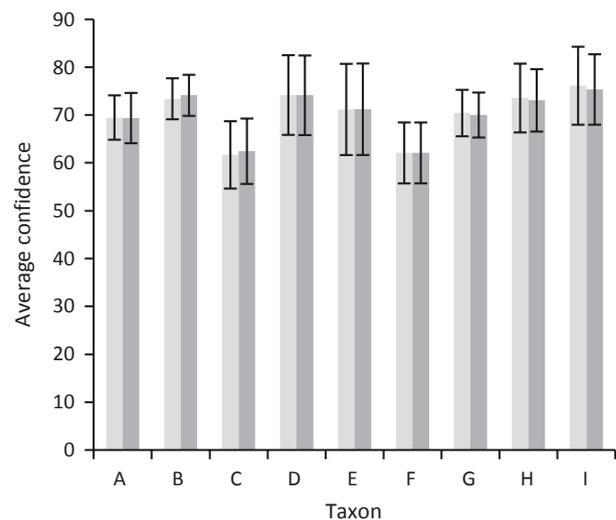


Fig. 7. Average response confidence for each taxon before discussion (pale grey) and after discussion (dark grey). Error bars show one standard deviation.

provided practical examples of how biases can be identified and resolved. However, it is impossible to anticipate all biases and corrective measures do not always fully correct unwanted effects (Fischhoff 1982; O’Hagan 2006).

The successful debiasing of judgments requires adequate vigilance and interrogation (Larrick 2004): in most cases, interrogating the experts about their beliefs can help (e.g. Fischhoff 1982; Arkes *et al.* 1987; Morgan & Henrion 1990). Other commonly employed strategies include discussion of potential bias at the pre-elicitation phase (Morgan & Henrion 1990), as happened here, and rapid analysis of results for evidence of bias to allow for feeding back to experts and informing of their subsequent judgmental revisions. The majority of debiasing techniques are achieved by expending greater effort and time in the elicitation, and this fact highlights the presence of a time-accuracy trade-off in elicitation (e.g. Murray *et al.* 2009; Kuhnert, Martin & Griffiths 2010; Tulloch, Possingham & Wilson 2011). While for the IUCN application the procedure implemented here represents a more comprehensive assessment process than is typically undertaken, the biases observed suggest that still more intensive procedures may further improve the quality of responses.

EFFECT OF THE COMMUNICATION MEDIUM

Email discussions have a number of advantages over face-to-face workshops or telephone conferences. Removing restrictions on both number and location of experts through the use of email is an important factor when assessing the taxa on continental or larger scales. While the email format entailed a substantial commitment in computer time from panellists, it was probably far less than if everyone had travelled to meet together: the minimum combined travel distance for the Australian panellists alone to a consolidated meeting would have been more than 20 000 km, a further three panellists were in

Europe at the time. It certainly cost much less to use email. That said, panellists may have struggled to achieve the same levels of motivation and focus that is possible in a workshop setting (e.g. Rhoads 2010).

The group emails also made discussions transparent. Experts were aware of each other's identities during the discussion, and the high level of transparency was only possible because of trust among panellists. Trust between panellists allowed greater space to express uncertainty, something encouraged by the process, and to be able to make, admit and correct mistakes without recrimination. It is recognised as a critical factor for the operation of successful electronic groups (e.g. Lipnack & Stamps 1997; Jarvenpaa & Leidner 1999). The panellists had a particular advantage having worked together via email for some years before this exercise was initiated, and trust-building exercises may be beneficial where this is not the case (e.g. Alge, Wiethoff & Klein 2003; Aubert & Kelsey 2003; Thompson & Coovert 2003).

Reduced levels of discussion appear to have been the key drawback to use of email. Levels of communication are often lower in electronic groups than in face-to-face settings (e.g. Hiltz, Johnson & Turoff 1986; Hollingshead 1996; Straus 1996), and panellist discussion tended to be stilted, intermittent owing to time delays between answers, and narrowly focused because only a small number of issues can be dealt with in any single email. Facilitation was difficult. Panellists were often left to compare their own answers with those of others, and judge for themselves whether they should adjust their responses. Thus, group workshops may be superior where interactive discussion is required on a large number of issues. We suggest that facilitated group workshops employing similar techniques for structured elicitation should remain the tool of choice for assessments in geographically confined areas where panellists can gather without the excessive costs of long-distance travel.

As technology continues to develop, methods such as live chat and videoconferencing that avoid some of the drawbacks of email represent increasingly viable alternatives (e.g. Kirkman & Mathieu 2005; Ferran & Watts 2008; Rhoads 2010; Mesmer-Magnus *et al.* 2011). Hybrid approaches, which incorporate multiple mediums for communication, are another promising avenue, allowing elicitors to overcome the limitations of the individual formats (e.g. Dennis & Valacich 1999; Martins, Gilson & Maynard 2004; Han *et al.* 2011). Such an approach might incorporate an initial face-to-face or video-conference meeting followed by email correspondence. There are significant benefits to remote elicitation methods that induce minimal process-related loss over face-to-face interaction (e.g. Donlan *et al.* 2010; Teck *et al.* 2010), and it is likely that such 'virtually facilitated' elicitations (e.g. Turoff 1972; Gordon & Pease 2006; Linstone & Turoff 2011) will play an increasing role in expert elicitation in the future.

Resources and software available for conducting a structured elicitation are available at the website of the Australian Centre of Excellence for Risk Analysis: <http://www.acera.unimelb.edu.au/materials/>.

Acknowledgements

The authors would like to thank all the panellists and taxon specialists who devoted their time and expertise to participating in this exercise. We also thank Resit Akçakaya for his advice on framing the elicitation questions and Raquel Ashton for helpful ideas and advice. This research was supported by the Australian Centre of Excellence for Risk Analysis (ACERA), Australian Research Council Linkage Grant LP0990395 and Charles Darwin University.

References

- Akçakaya, H.R. & Ferson, S. (2007) *RAMAS Red List: Spatial and Temporal Data Analysis for Threatened Species Classifications under Uncertainty (Version 3.0 Professional)*. Applied Biomathematics, Setauket, New York.
- Akçakaya, H.R., Ferson, S., Burgman, M.A., Keith, D.A., Mace, G.M. & Todd, C.R. (2000) Making consistent IUCN classifications under uncertainty. *Conservation Biology*, **14**, 1001–1013.
- Alge, B.J., Wiethoff, C. & Klein, H.J. (2003) When does the medium matter? Knowledge-building experiences and opportunities in decision-making teams. *Organizational Behavior and Human Decision Processes*, **91**, 26–37.
- Arkes, H.R., Christensen, C., Lai, C. & Blumer, C. (1987) Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, **39**, 133–144.
- Aspinall, W. (2010) A route to more tractable expert advice. *Nature*, **463**, 294–295.
- Aubert, B.A. & Kelsey, B.L. (2003) Further understanding of trust and performance in virtual teams. *Small Group Research*, **34**, 575–618.
- Baltes, B.B., Dickson, M.W., Sherman, M.P., Bauer, C.C. & LaGanke, J.S. (2002) Computer-mediated communication and group decision making: a meta-analysis. *Organizational Behavior and Human Decision Processes*, **87**, 156–179.
- Bedford, T. & Cooke, R.M. (2001) *Mathematical Tools for Probabilistic Risk Analysis*. Cambridge University Press, Cambridge.
- Bolger, F., Stranieri, A., Wright, G. & Yearwood, J. (2011) Does the Delphi process lead to increased accuracy in group-based judgmental forecasts or does it simply induce consensus amongst judgmental forecasters? *Technological Forecasting and Social Change*, **78**, 1671–1680.
- Burgman, M., Carr, A., Godden, L., Gregory, R., McBride, M., Flander, L. & Maguire, L. (2011a) Redefining expertise and improving ecological judgment. *Conservation Letters*, **4**, 81–87.
- Burgman, M.A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L. & Twardy, C. (2011b) Expert status and performance. *PLoS One*, **6**, e22998.
- Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chener, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leve- rington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.-C. & Watson, R. (2010) Global biodiversity: indicators of recent declines. *Science*, **328**, 1164–1168.
- Choy, S.L., O'Leary, R. & Mengersen, K. (2009) Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology*, **90**, 265–277.
- Clemen, R.T. & Reilly, T. (2001) *Making Hard Decisions with Decision Tools*. Duxbury Press, Pacific Grove.
- Clemen, R.T. & 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, New York, New York.
- Cooper, W.H. (1981) Ubiquitous halo. *Psychological Bulletin*, **90**, 218–244.
- Crance, J.H.B.R. (1987) Guidelines for using the Delphi technique to develop habitat suitability index curves. *Biol. Rep.*, pp. 21. U. S. Fish & Wildlife Service.
- Czembor, C.A. & Vesik, P.A. (2009) Incorporating between-expert uncertainty into state-and-transition simulation models for forest restoration. *Forest Ecology and Management*, **259**, 165–175.
- Dalal, S., Khodyakov, D., Srinivasan, R., Straus, S. & Adams, J. (2011) Expert-Lens: a system for eliciting opinions from a large pool of non-collocated experts with diverse knowledge. *Technological Forecasting and Social Change*, **78**, 1426–1444.

- Dalkey, N.C. (1975) Towards a theory of group estimation. *The Delphi Method: Techniques and Applications* (eds H. Linstone & M. Turoff), pp. 231–256. Addison-Wesley, London.
- Dalkey, N. & Helmer, O. (1963) An experimental application of the Delphi method to the use of experts. *Management Science*, **9**, 458–467.
- Delbecq, A.L., Van de Ven, A.H. & Gustafson, D.H. (1975) *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Scott, Foresman and Company, Glenview, Illinois.
- Dennis, A.R. & Valacich, J.S. (1999) Rethinking media richness: towards a theory of media synchronicity. Proceedings of the 32nd Hawaii international Conference on System Sciences- 1999, 10 IEEE, 10 pp.
- Donlan, C.J., Wingfield, D.K., Crowder, L.B. & Wilcox, C. (2010) Using expert opinion surveys to rank threats to endangered species: a case study with sea turtles. *Conservation Biology*, **24**, 1586–1595.
- Epley, N. & Gilovich, T. (2005) When effortful thinking influences judgmental anchoring. *Journal of Behavioral Decision Making*, **18**, 199–212.
- Epley, N. & Gilovich, T. (2006) The anchoring-and-adjustment heuristic. *Psychological Science*, **17**, 311–318.
- Eycott, A.E., Marzano, M. & Watts, K. (2011) Filling evidence gaps with expert opinion: the use of Delphi analysis in least-cost modelling of functional connectivity. *Landscape and Urban Planning*, **103**, 400–409.
- Fazey, I., Fazey, J.A., Salisbury, J.G., Lindenmayer, D.B. & Dovers, S. (2006) The nature and role of experiential knowledge for environmental conservation. *Environmental Conservation*, **33**, 1–10.
- Ferran, C. & Watts, S. (2008) Videoconferencing in the field: a heuristic processing model. *Management Science*, **54**, 1565–1578.
- Fischhoff, B. (1982) Debiasing. *Judgment Under Uncertainty: Heuristics and Biases* (eds D. Kahneman, P. Slovic & A. Tversky), pp. 422–444. Cambridge University Press, Cambridge.
- Fischhoff, B., Slovic, P. & Lichtenstein, S. (1982) Lay foibles and expert fables in judgments about risk. *American Statistician*, **36**, 240–255.
- Garnett, S.T. (1992) *The Action Plan for Australian Birds*. ANPWS, Canberra, ACT.
- Garnett, S.T. & Crowley, G.M. (2000) *The Action Plan for Australian Birds*. Environment Australia, Canberra, ACT.
- Gigerenzer, G. & Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, **102**, 684–704.
- Gilovich, T., Griffin, D. & Kahneman, D. (2002) *Heuristics and Biases: The Psychology of Intuitive Judgement*. Cambridge University Press, Cambridge.
- Gordon, T.J. & Pease, A. (2006) RT Delphi: an efficient, “round-less”, almost real time Delphi method. *Technological Forecasting and Social Change*, **73**, 321–333.
- Han, H.J., Hiltz, S.R., Fjermestad, J. & Wang, Y. (2011) Does medium matter? A comparison of initial meeting modes for virtual teams. *IEEE Transactions on Professional Communication*, **54**, 376–391.
- Hiltz, S.R., Johnson, K. & Turoff, M. (1986) Experiments in group decision-making: communication process and outcome in face-to-face versus computerized conferences. *Human Communication Research*, **13**, 225–252.
- Hinds, P.J. & Bailey, D.E. (2003) Out of sight, out of sync: understanding conflict in distributed teams. *Organization Science*, **14**, 615–632.
- Hollingshead, A.B. (1996) Information suppression and status persistence in group decision making. *Human Communication Research*, **23**, 193–219.
- Isenberg, D.J. (1986) Group polarization: a critical review and meta-analysis. *Journal of Personality and Social Psychology*, **50**, 1141–1151.
- IUCN (2001) IUCN Red List categories and criteria. Version 3.1. IUCN Species Survival Commission, Gland, Switzerland and Cambridge, United Kingdom. ii + 30 pp.
- IUCN Standards and Petitions Subcommittee (2010) Guidelines for using the IUCN Red List categories and criteria. Version 8.1 Prepared by the Standards and Petitions Subcommittee in March 2010.
- Jacowitz, K.E. & Kahneman, D. (1995) Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, **21**, 1161.
- Janis, I.L. (1972) Groupthink. *Psychology Today*, **5**, 43–???
- Jarvenpaa, S.L. & Leidner, D.E. (1999) Communication and trust in global virtual teams. *Organization Science*, **10**, 791–815.
- Jenkinson, D. (2005) The elicitation of probabilities: a review of the statistical literature. *Technical Report*. University of Sheffield, Sheffield.
- Kahan, D. (2010) Fixing the communications failure. *Nature*, **463**, 296–297.
- Kahneman, D. & Tversky, A. (1982) *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge.
- Kaufmann, A. & Gupta, M.M. (1985) *Introduction to Fuzzy Arithmetic: Theory and Applications*. Van Nostrand Reinhold, New York, New York.
- Keeney, R.L. & Von Winterfeldt, D. (1991) Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management*, **38**, 191–201.
- Keith, D.W. (1996) When is it appropriate to combine expert judgments? *Climatic Change*, **33**, 139–143.
- Keith, D.A., McCarthy, M.A., Regan, H., Regan, T., Bowles, C., Drill, C., Craig, C., Pellow, B., Burgman, M.A., Master, L.L., Ruckelshaus, M., Mackenzie, B., Andelman, S.J. & Wade, P.R. (2004) Protocols for listing threatened species can forecast extinction. *Ecology Letters*, **7**, 1101–1108.
- Kerr, N.L. & Tindale, R.S. (2004) Group performance and decision making. *Annual Review of Psychology*, **55**, 623–655.
- Kerr, N.L. & Tindale, R.S. (2011) Group-based forecasting?: a social psychological analysis. *International Journal of Forecasting*, **27**, 14–40.
- Kirkman, B.L. & Mathieu, J.E. (2005) The dimensions and antecedents of team virtuality. *Journal of Management*, **31**, 700–718.
- Koehler, J.J. (1993) The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, **56**, 28–55.
- Krinitzsky, E.L. (1993) Earthquake probability in engineering – part I: the use and misuse of expert opinion. *Engineering Geology*, **33**, 257–288.
- Kuhnert, P.M. (2011) Four case studies in using expert opinion to inform priors. *Environmetrics*, **22**, 662–674.
- Kuhnert, P.M., Martin, T.G. & Griffiths, S.P. (2010) A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, **13**, 900–914.
- Larrick, R.P. (2004) Debiasing. *Blackwell Handbook of Judgment and Decision Making* (eds D. J. Koehler & N. J. Harvey), pp. 316–338. Wiley-Blackwell, Oxford.
- Levin, I.P., Schneider, S.L. & Gaeth, G.J. (1998) All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, **76**, 149–188.
- Lichtenstein, S., Fischhoff, B. & Phillips, L.D. (1982) Calibration of probabilities: the state of the art to 1980. *Judgment Under Uncertainty: Heuristics and Biases* (eds D. Kahneman & A. Tversky), pp. 306–334. Cambridge University Press, Cambridge.
- Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M. & Combs, B. (1978) Judged frequency of lethal events. *Journal of Experimental Psychology-Human Learning and Memory*, **4**, 551–578.
- Lin, S.W. & Bier, V.M. (2008) A study of expert overconfidence. *Reliability Engineering & System Safety*, **93**, 711–721.
- Linstone, H.A. & Turoff, M. (1975) *The Delphi Method: Techniques and Applications*. Addison-Wesley, Reading.
- Linstone, H.A. & Turoff, M. (2011) Delphi: a brief look backward and forward. *Technological Forecasting and Social Change*, **78**, 1712–1717.
- Lipnack, J. & Stamps, J. (1997) *Virtual Teams: Reaching Across Space, Time and Organizations with Technology*. Wiley, New York, New York.
- Lord, C.G., Ross, L. & Lepper, M.R. (1979) Biased assimilation and attitude polarization: effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, **37**, 2098–2109.
- Lukey, J.R., Crawford, S.S. & Gillis, D. (2010) Effect of information availability on assessment and designation of species at risk. *Conservation Biology*, **24**, 1398–1406.
- Mace, G.M., Collar, N.J., Gaston, K.J., Hilton-Taylor, C., Akcakaya, H.R., Leader-Williams, N., Milner-Gulland, E.J. & Stuart, S.N. (2008) Quantification of extinction risk: IUCN’s system for classifying threatened species. *Conservation Biology*, **22**, 1424–1442.
- MacMillan, D.C. & Marshall, K. (2006) The Delphi process: an expert-based approach to ecological modelling in data-poor environments. *Animal Conservation*, **9**, 11–19.
- Maier, N.R.F. & Hoffman, L.R. (1960) Using trained developmental leaders to improve further the quality of group decisions. *Journal of Applied Psychology*, **44**, 247–251.
- Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M. & Mengersen, K. (2012) Eliciting expert knowledge in conservation science. *Conservation Biology*, **26**, 29–38.
- Martins, L.L., Gilson, L.L. & Maynard, M.T. (2004) Virtual teams: what do we know and where do we go from here? *Journal of Management*, **30**, 805–835.
- Maruping, L.M. & Agarwal, R. (2004) Managing team interpersonal processes through technology: a task-technology fit perspective. *Journal of Applied Psychology*, **89**, 975–990.
- Mesmer-Magnus, J.R., DeChurch, L.A., Jimenez-Rodriguez, M., Wildman, J. & Shuffler, M. (2011) A meta-analytic investigation of virtuality and

- information sharing in teams. *Organizational Behavior and Human Decision Processes*, **115**, 214–225.
- Meyer, M. & Booker, J. (1991) *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Academic Press, New York, New York.
- Moore, D.A. & Healy, P.J. (2008) The trouble with overconfidence. *Psychological Review*, **115**, 502–517.
- Morgan, M.G. & Henrion, M. (1990) *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, New York, New York.
- Murphy, K.R., Jako, R.A. & Anhalt, R.L. (1993) Nature and consequences of halo error: a critical analysis. *Journal of Applied Psychology*, **78**, 218–225.
- Murray, J.V., Goldizen, A.W., O'Leary, R.A., McAlpine, C.A., Possingham, H.P. & Choy, S.L. (2009) How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. *Journal of Applied Ecology*, **46**, 842–851.
- Mussweiler, T. & Strack, F. (2000) Numeric judgments under uncertainty: the role of knowledge in anchoring. *Journal of Experimental Social Psychology*, **36**, 495–518.
- Myers, D.G. & Lamm, H. (1976) The group polarization phenomenon. *Psychological Bulletin*, **83**, 602–627.
- Newton, A.C. & Oldfield, S. (2008) Red Listing the world's tree species: a review of recent progress. *Endangered Species Research*, **6**, 137–147.
- Nisbett, R.E. & Wilson, T.D. (1977) Halo effect: evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, **35**, 250–256.
- O'Brien, M. (2000) *Making Better Environmental Decisions: An Alternative to Risk Assessment*. MIT Press, Cambridge, Massachusetts.
- O'Hagan, A. (2006) *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley, Chichester.
- O'Leary, R.A., Choy, S.L., Murray, J.V., Kynn, M., Denham, R., Martin, T.G. & Mengersen, K. (2009) Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby *Petrogale penicillata*. *Environmetrics*, **20**, 379–398.
- O'Neill, S.J., Osborn, T.J., Hulme, M., Lorenzoni, I. & Watkinson, A.R. (2008) Using expert knowledge to assess uncertainties in future polar bear populations under climate change. *Journal of Applied Ecology*, **45**, 1649–1659.
- Perera, A.J., Johnson, C.J. & Drew, C.A. (2011) *Expert Knowledge and Its Application in Landscape Ecology*. Springer-Verlag, New York, New York.
- Regan, H.M., Colyvan, M. & 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., Master, L.L., Keith, D.A., Mace, G.M. & Andelman, S.J. (2005) The consistency of extinction risk classification protocols. *Conservation Biology*, **19**, 1969–1977.
- Rhoads, M. (2010) Face-to-face and computer-mediated communication: what does theory tell us and what have we learned so far? *Journal of Planning Literature*, **25**, 111–122.
- Rothlisberger, J.D., Lodge, D.M., Cooke, R.M. & Finnoff, D.C. (2010) Future declines of the binational Laurentian Great Lakes fisheries: the importance of environmental and cultural change. *Frontiers in Ecology and the Environment*, **8**, 239–244.
- Rowe, G. & Wright, G. (1999) The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, **15**, 353–375.
- Rowe, G. & Wright, G. (2001) Expert opinions in forecasting: the role of the Delphi technique. *Principles of Forecasting: A Handbook for Researchers and Practitioners* (ed J.S. Armstrong), pp. 125–144. Kluwer Academic Publishers, Norwell, Massachusetts.
- Rowe, G., Wright, G. & McColl, A. (2005) Judgment change during Delphi-like procedures: the role of majority influence, expertise, and confidence. *Technological Forecasting and Social Change*, **72**, 377–399.
- Russo, J.E. & Schoemaker, P.J.H. (1992) Managing overconfidence. *Sloan Management Review*, **33**, 7–17.
- Schwarz, N. & Vaughn, L.A. (2002) The availability heuristic revisited: ease of recall and content of recall as distinct sources of information. *Heuristics and Biases: The Psychology of Intuitive Judgement* (eds T. Gilovich, D. Griffin & D. Kahneman), pp. 103–119. Cambridge University Press, New York.
- Shrader-Frechette, K. (1996) Value judgments in verifying and validating risk assessment models. *Handbook for Environmental Risk Decision Making: Values, Perception and Ethics* (ed C.R. Cothran), pp. 291–309. CRC Lewis Publishers, Boca Raton, Florida.
- Slovic, P. (1999) Trust, emotion, sex, politics and science: surveying the risk-assessment battlefield. *Risk Analysis*, **19**, 689–701.
- Soll, J.B. & Klayman, J. (2004) Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **30**, 299–314.
- Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G. & Burgman, M. (2010) Reducing overconfidence in the interval judgments of experts. *Risk Analysis*, **30**, 512–523.
- Spetzler, C.S. & Stael von Holstein, C.A.S. (1975) Probability encoding in decision analysis. *Management Science*, **22**, 340–358.
- Stewart, R.T. (2001) Improving reliability of judgmental forecasts. *Principles of Forecasting: A Handbook for Researchers and Practitioners* (ed J.S. Armstrong), pp. 753–770. Kluwer Academic Publishers, Norwell, Massachusetts.
- Straus, S.G. (1996) Getting a clue: the effects of communication media and information distribution on participation and performance in computer-mediated and face-to-face groups. *Small Group Research*, **27**, 115–142.
- Sunstein, C.R. (2000) Deliberative trouble? Why groups go to extremes. *The Yale Law Journal*, **110**, 71–119.
- Sutherland, W.J. (2006) Predicting the ecological consequences of environmental change: a review of the methods. *Journal of Applied Ecology*, **43**, 599–616.
- Teck, S.J., Halpern, B.S., Kappel, C.V., Micheli, F., Selkoe, K.A., Crain, C.M., Martone, R., Shearer, C., Arvai, J., Fischhoff, B., Murray, G., Neslo, R. & Cooke, R. (2010) Using expert judgment to estimate marine ecosystem vulnerability in the California Current. *Ecological Applications*, **20**, 1402–1416.
- Thompson, L.F. & Coovert, M.D. (2003) Teamwork online: the effects of computer conferencing on perceived confusion, satisfaction and postdiscussion accuracy. *Group Dynamics: Theory, Research, and Practice*, **7**, 135.
- Titus, J.G. & Narayanan, V. (1996) The risk of sea level rise. *Climatic Change*, **33**, 151–212.
- Tulloch, A., Possingham, H.P. & Wilson, K. (2011) Wise selection of an indicator for monitoring the success of management actions. *Biological Conservation*, **144**, 141–154.
- Turoff, M. (1972) Delphi conferencing: computer-based conferencing with anonymity. *Technological Forecasting and Social Change*, **3**, 159–204.
- Turoff, M. & Hiltz, S. (1982) Computer support for group versus individual decisions. *IEEE Transactions on Professional Communication*, **30**, 82–91.
- Tversky, A. & Kahneman, D. (1973) Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207–232.
- Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: heuristics and biases. *Science*, **185**, 1124–1131.
- United Nations (2011) *Millennium Development Goals 2011 Report*. United Nations, New York, New York.
- Van Leekwijck, W. & Kerre, E.E. (1999) Defuzzification: criteria and classification. *Fuzzy Sets and Systems*, **108**, 159–178.
- White, P.C.L., Jennings, N.V., Renwick, A.R. & Barker, N.H.L. (2005) Questionnaires in ecology: a review of past use and recommendations for best practice. *Journal of Applied Ecology*, **42**, 421–430.
- Wilson, A.G. (1994) Cognitive factors affecting subjective probability assessment. *ISDS Discussion Paper* Institute of Statistics and Decision Sciences, Duke University.
- Yager, R.R. & Filev, D. (1993) On the issue of defuzzification and selection based on a fuzzy set. *Fuzzy Sets and Systems*, **55**, 255–271.
- Yaniv, I. (2004) Receiving other people's advice: influence and benefit. *Organizational Behavior and Human Decision Processes*, **93**, 1–13.
- Yaniv, I. & Kleinberger, E. (2000) Advice taking in decision making: egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, **83**, 260–281.
- Zimmermann, H.J. (2001) *Fuzzy Set Theory – and Its Applications*, 4th edn. Kluwer Academic Publishers, Norwell, Massachusetts.

Received 9 August 2011; accepted 17 April 2012

Handling Editor: Daniel Faith