RESEARCH ARTICLE

A practical guide to structured expert elicitation using the IDEA protocol

Victoria Hemming[1] [iD] | Mark A. Burgman[2] | Anca M. Hanea[1] | Marissa F. McBride[3] | Bonnie C. Wintle[1,4]

[1]Centre of Excellence for Biosecurity Risk Analysis, University of Melbourne, Melbourne, Vic., Australia

[2]Centre for Environmental Policy, Imperial College, London, UK

[3]Harvard Forest, Harvard University, Petersham, MA, USA

[4]Centre for the Study of Existential Risk, University of Cambridge, Cambridge, UK

Correspondence
Victoria Hemming
Email: hemmingv@student.unimelb.edu.au

Abstract

1. Expert judgement informs a variety of important applications in conservation and natural resource management, including threatened species management, environmental impact assessment and structured decision-making. However, expert judgements can be prone to contextual biases. Structured elicitation protocols mitigate these biases, and improve the accuracy and transparency of the resulting judgements. Despite this, the elicitation of expert judgement within conservation and natural resource management remains largely informal. We suggest this may be attributed to financial and practical constraints, which are not addressed by many existing structured elicitation protocols.

2. In this paper, we advocate that structured elicitation protocols must be adopted when expert judgements are used to inform science. In order to motivate a wider adoption of structured elicitation protocols, we outline the IDEA protocol. The protocol improves the accuracy of expert judgements and includes several key steps which may be familiar to many conservation researchers, such as the four-step elicitation, and a modified Delphi procedure ("Investigate," "Discuss," "Estimate" and "Aggregate"). It can also incorporate remote elicitation, making structured expert judgement accessible on a modest budget.

3. The IDEA protocol has recently been outlined in the scientific literature; however, a detailed description has been missing. This paper fills that important gap by clearly outlining each of the steps required to prepare for and undertake an elicitation.

4. While this paper focuses on the need for the IDEA protocol within conservation and natural resource management, the protocol (and the advice contained in this paper) is applicable to a broad range of scientific domains, as evidenced by its application to biosecurity, engineering and political forecasting. By clearly outlining the IDEA protocol, we hope that structured protocols will be more widely understood and adopted, resulting in improved judgements and increased transparency when expert judgement is required.

KEYWORDS
Delphi, expert elicitation, forecasting, four-step elicitation, IDEA protocol, quantitative estimates, structured expert judgement

1 | INTRODUCTION

Conservation and natural resource management often involve decisions for which data are absent or insufficient and consequences are potentially severe. In such circumstances, the elicitation of expert judgement has become routine and informs a variety of important applications from forecasting biosecurity risks (Wittmann et al., 2015), threatened species management (Adams-Hosking et al., 2016), priority threat management (Chadés et al., 2015; Firn et al., 2015), predictive models (Krueger, Page, Hubacek, Smith, & Hiscock, 2012), environmental impact assessment (Knol, Slottje, van der Sluijs, & Lebret, 2010) and inputs into structured decision-making (Gregory & Keeney, 2017). Expert judgement also underpins some of the most influential global environmental policies including the IUCN Red List and IPCC Assessments (IUCN, 2012; Mastrandrea et al., 2010).

While expert judgement can be remarkably useful when data are absent or incomplete, experts make mistakes (Burgman, 2004; Kuhnert, Martin, & Griffiths, 2010). This is often due to a range of contextual biases and heuristics such as anchoring, availability, and representativeness (Kahneman & Tversky, 1973), groupthink (Janis, 1971), overconfidence (Soll & Klayman, 2004) and difficulties associated with communicating knowledge in numbers and probabilities (Gigerenzer & Edwards, 2003). Inappropriate and ill-informed methods for elicitation can amplify these biases by relying on subjective and unreliable methods for selecting experts (Shanteau, Weiss, Thomas, & Pounds, 2002), asking poorly specified questions (Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986), ignoring protocols to counteract negative group interactions (Janis, 1971) and applying subjective or biasing aggregation methods (Aspinall & Cooke, 2013; Lorenz, Rauhut, Schweitzer, & Helbing, 2011).

Structured elicitation protocols can improve the quality of expert judgements, and are especially important for informing critical decisions (Cooke, 1991; Keeney & von Winterfeldt, 1991; Mellers et al., 2014; Morgan & Henrion, 1990; O'Hagan et al., 2006). These protocols treat each step of the elicitation as a process of formal data acquisition, and incorporate research from mathematics, psychology and decision theory to help reduce the influence of biases and to enhance the transparency, accuracy, and defensibility of the resulting judgements.

Structured protocols have been increasingly adopted in conservation and natural resource management, for example, Cooke's Classical Model (Cooke, 1991) has been applied to case studies on the Great Lakes fisheries in North America (e.g. Rothlisberger, Finnoff, Cooke, & Lodge, 2012; Wittmann et al., 2015) as well as sea-level rise and ice-sheet melt (Bamber & Aspinall, 2013). However, reviews by Burgman (2004), Regan et al. (2005), Kuhnert et al. (2010) and Krueger et al. (2012) highlight that informal methods for expert elicitation continue to prevail. Furthermore, few elicitations provide sufficient detail to enable review, critical appraisal and replication (French, 2012; Krueger et al., 2012; Low Choy, O'Leary, & Mengersen, 2009).

These reviews have highlighted challenges which may present barriers to the implementation of existing structured protocols within conservation and natural resource management. These include difficulties experts face in expressing judgements in quantitative terms (Martin et al., 2012a), the cost and logistics associated with face-to-face elicitations of more than one or two experts (Knol et al., 2010; Kuhnert et al., 2010), as well as challenges experienced by experts in translating their knowledge into quantiles or probability distributions (Garthwaite, Kadane, & O'Hagan, 2005; Kuhnert et al., 2010; Low Choy et al., 2009).

While no structured protocol has yet been proposed to overcome such challenges, a range of individual and practical steps have been developed and tested. For example, Speirs-Bridge et al. (2010) provided a means of eliciting a best estimate and uncertainty in a language that respects the bounded rationality of experts in communicating their knowledge as probabilities and quantities. The approach also reduced judgement overconfidence. Burgman et al. (2011) found that reliable experts in conservation cannot be predicted a priori. Instead, improved judgements result from a diverse group of individuals engaged in a structured modified Delphi process. McBride et al. (2012) demonstrated that it is feasible to elicit judgements from conservation experts across the globe on a modest budget using remote elicitation.

Encouragingly, the steps listed above are being readily adopted by conservation scientists to solve a range of problems. For example, the four-step elicitation has been incorporated by Metcalf and Wallace (2013), Ban, Pressey, and Graham (2014), Chadés et al. (2015), Firn et al. (2015) and Adams-Hosking et al. (2016), while Delphi protocols have been utilised by Runge, Converse, and Lyons (2011), Adams-Hosking et al. (2016), and Chadés et al. (2015). The incorporation of such steps highlights that there is a willingness to adopt more rigorous approaches to expert elicitation; however, that an alternative approach to existing protocols may be required to help overcome the constraints faced by practitioners in conservation and natural resource management.

While an alternative approach is required, such an approach should not be a compromise. Rather, it should meet the requirements of more rigorous definitions of structured elicitation protocols. That is, it should treat the elicitation of expert judgements in the same regard as empirical data, by using repeatable, transparent methods and addressing scientific questions (not value judgements) in the form of probabilities and quantities (Aspinall, 2010; Aspinall & Cooke, 2013; French, 2011; Morgan, 2014). Importantly, it should account for each step of the elicitation including the recruitment of experts, the framing of questions, the elicitation and aggregation of their judgements, using procedures that have been *tested* and clearly demonstrated to improve judgements (e.g. Cooke, 1991; Mellers et al., 2014). Finally, it should enable judgements to be subject to review and critical appraisal (French, 2012).

In this paper, we suggest the IDEA structured protocol provides a much needed alternative approach to structured expert elicitation that meets these requirements, and overcomes many of the constraints faced by conservation and natural resource management practitioners when eliciting expert judgements. The protocol is relatively simple to apply, and has been tested and shown to yield relatively reliable judgements. Importantly, the protocol incorporates a range of key steps that, as noted above, have already been adopted to some extent by the conservation community, such as the four-step elicitation and a modified
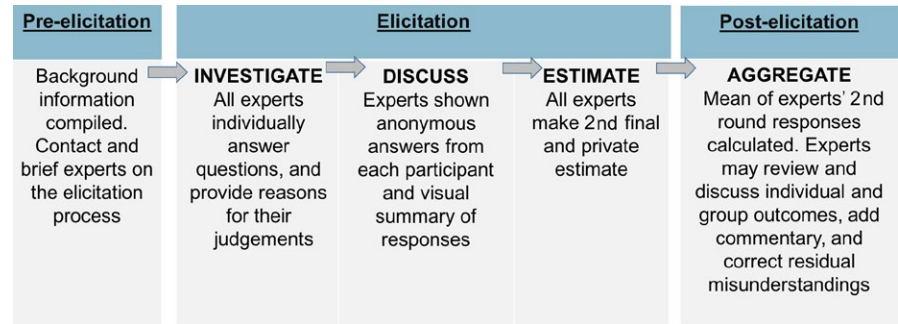
FIGURE 1 The IDEA protocol adapted from Burgman (2015)

Delphi procedure. Its applicability to remote elicitation makes it more cost-effective than methods that rely on face-to-face meetings.

While this paper emphasises the suitability of the approach in conservation and natural resource management, it should be noted that the IDEA protocol is equally suited to a wide variety of scientific and technical domains. This is evidenced by its effective application in geopolitical forecasting (Hanea, McBride, Burgman, & Wintle, 2016; Hanea et al., 2016; Wintle et al., 2012) and engineering (van Gelder, Vodicka, & Armstrong, 2016).

Although the protocol has been introduced elsewhere (Burgman, 2015; Hanea et al., 2016) and closely parallels the approach used by Burgman et al. (2011) and Adams-Hosking et al. (2016), to date, there has been no detailed description of the steps required to carry out the method. This work fills that important practical gap.

2 | THE IDEA PROTOCOL

The acronym IDEA stands for key steps of the protocol: "Investigate," "Discuss," "Estimate" and "Aggregate" (Figure 1). A summary of the basic steps is as follows. A diverse group of experts is recruited to answer questions with probabilistic or quantitative responses. The experts are asked to first *Investigate* the questions and to clarify their meanings, and then to provide their private, individual best guess point estimates and associated credible intervals (Round 1) (Speirs-Bridge et al., 2010; Wintle et al., 2012). The experts receive feedback on their estimates in relation to other experts. With assistance of a facilitator, the experts are encouraged to *Discuss* the results, resolve different interpretations of the questions, cross-examine reasoning and evidence, and then provide a second and final private *Estimate* (Round 2). Notably, the purpose of discussion in the IDEA protocol is not to reach consensus but to resolve linguistic ambiguity, promote critical thinking, and to share evidence. This is based on evidence that incorporating a single discussion stage within a standard Delphi process generates improvements in response accuracy (Hanea, McBride, et al., 2016). The individual estimates are then combined using mathematical *Aggregation*.

The IDEA protocol initially arose in response to Australian Government requests to support improved biosecurity decision-making. Over the past 10 years, individual steps of the protocol have been tested in public health, ecology and conservation (Burgman et al., 2011; McBride et al. 2012; Speirs-Bridge et al., 2010; Wintle, Fidler, Vesk, & Moore, 2013). More recently, the protocol was refined and tested in its entirety as part of a forecasting tournament that commenced in 2011 as an initiative of US Intelligence Advanced Research Projects Activity (IARPA) (Hanea, McBride, et al., 2016; Hanea et al., 2016; Wintle et al., 2012). The results demonstrated the value of many steps of the IDEA protocol including using diverse experts in deliberative groups, of giving experts the opportunity to examine one-another's estimates and to reconcile the meanings of questions through discussion. It verified that prior performance on questions of a similar kind can be used to identify the most valuable experts (Hanea, McBride, et al., 2016).

3 | PREPARING FOR THE IDEA PROTOCOL

Undertaking a structured elicitation requires substantial planning to ensure timelines are met and experts are appropriately engaged (EPA, 2009). The IDEA protocol is no different in this regard; planning is the key to a successful elicitation. Key considerations and timelines are outlined below, and summarised in Figure 2.

3.1 | Develop a timeline

The first step is to develop a timeline of tasks and a schedule of key dates for each step of the elicitation (Figure 2). Enable sufficient time for delays caused by human subject research approval (if necessary), late replies by experts, and delays in the analysis.

In our experience, preparation can take anywhere from 2 weeks to 4 months, depending on how well defined the questions and the purpose of the elicitation are. The subsequent elicitation of quantities from a single group of experts ranges between 2 and 6 weeks, depending on whether face-to-face elicitation or remote elicitation is used, and assuming a maximum of 20–30 questions (Figure 2 and Supplementary Material A).

3.2 | Form a project team

An expert elicitation team typically consists of a coordinator, a facilitator, an analyst and the problem owner (Table 1). If no conflict of interest exists and time permits, then these roles may be undertaken by one person or shared between many (Martin et al., 2012a).

The specific roles of each member are outlined in Table 1; however, it is important that all members have an understanding of the
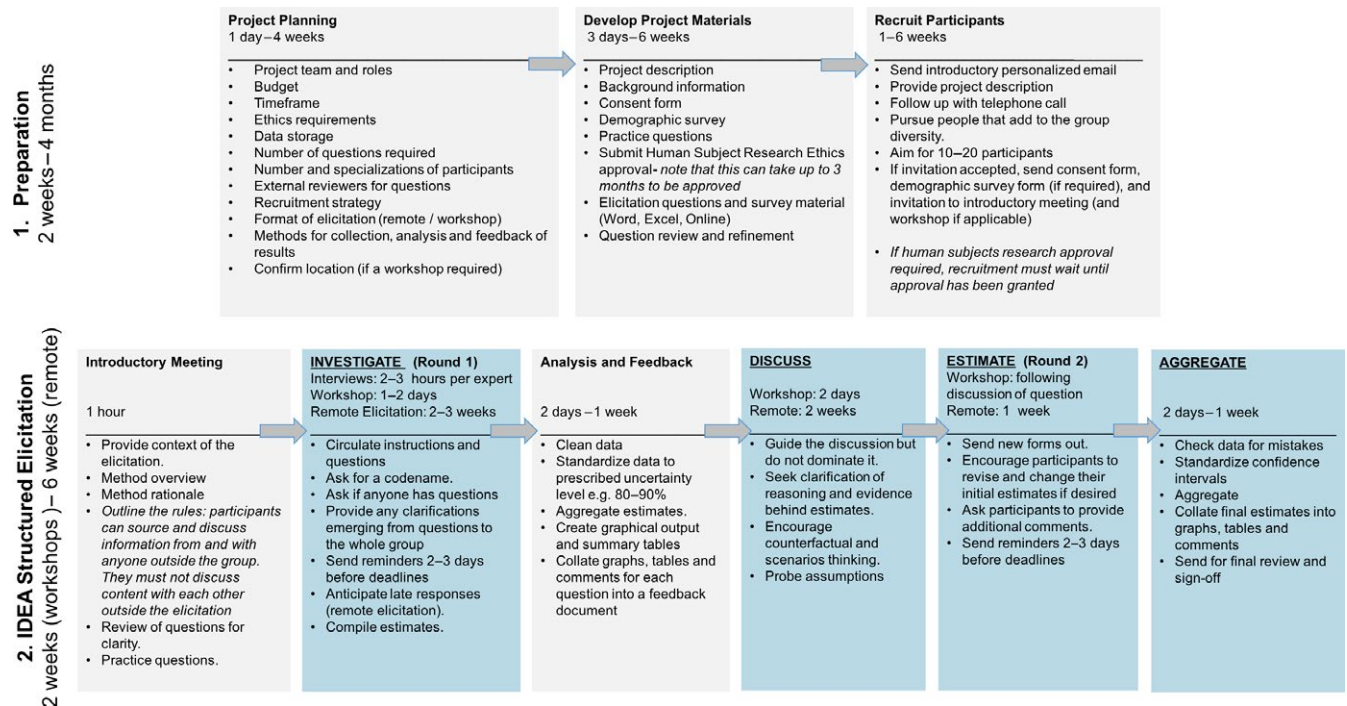
FIGURE 2 The steps and time taken to prepare for and implement the IDEA protocol. Time can be reduced if the questions and the objectives of the elicitation are already well defined, and through the use of workshops (although more expensive). The time shown above assumes experts are volunteers and asked up to 30 technical quantities

TABLE 1 The project team and their roles and responsibilities

Position	Role
Problem owner	Responsible for requesting the elicitation and usually the source of funding for the study. They may inform the selection of suitable questions, help identify experts, anticipate sources of bias and specify time and budget constraints
Coordinator	Manages the elicitation, timelines and the collection of responses. Anticipates and solves emerging problems
Facilitator	Responsible for managing interactions between participants. They should be neutral with regard to the problem and capable of diplomatically handling a wide range of personalities. They should encourage critical thinking and consideration of counterfactuals. They should also have a working technical understanding of the problem, an awareness of the various potential biases, and understand how the IDEA protocol aims to mitigate them. The facilitator also needs to understand how judgements will be transformed and aggregated using the IDEA protocol (Section 4.3)
Analyst	Responsible for processing responses and undertaking the analysis. They will need to standardise and combine estimates, and generate feedback of group estimates

many ways in which biases and heuristics can affect the accuracy and calibration of expert judgements. Common problems include overconfidence (Soll & Klayman, 2004; Speirs-Bridge et al., 2010), anchoring (Furnham & Boo, 2011; Tversky & Kahneman, 1975), failure to adequately consider counterfactual information (Nickerson, 1998), linguistic ambiguity (Kent, 1964; Wallsten et al., 1986) and groupthink (Janis, 1971). Useful introductions to these biases and heuristics can be found in Cooke (1991), O'Hagan et al. (2006), Hastie and Dawes (2010), McBride et al. (2012) and Burgman (2015).

If the team is approaching expert elicitation for the first time, then it is recommended that someone with experience in structured expert elicitation is engaged to review the questions for unintended bias.

3.3 | Decide elicitation format

A key advantage of the IDEA protocol is that it is flexible enough to be conducted face-to-face, in workshops or by remote elicitation. The most appropriate format will ultimately depend on time and budget, and the location and availability of experts.

The inception meeting should be undertaken either via a group tele-conference or workshop. In *Round 1* and *Round 2*, experts must be free to answer the questions posed by the facilitator independently from others in the group. This can be achieved remotely or in a face-to-face environment. The *discussion* phase can take place by teleconference, email, webpage or a combination of platforms. Alternatively, experts can be invited to a workshop with the purpose of discussing Round 1 results.

If using *remote elicitation* for Round 1 and/or Round 2 the questions can be sent in a simple document that can be accessed both off and online. We have used Excel spreadsheets, Word documents and PDF forms (Supplementary Materials A and B). If eliciting judgements *face-to-face*, the facilitator may schedule interviews with individual experts to elicit their judgements, or elicit individual judgements in a group session by asking them to enter estimates on a device or on paper.

Discussion facilitated over email or web forum can be inexpensive and enable all experts to take part regardless of their location and work commitments. However, the drawbacks may include substantial time investment by experts, lower engagement levels and the possibility that some conversations may not be resolved, especially if experts are distracted by local commitments or are late to join the discussion process (McBride et al., 2012). Workshops usually result in better buy-in and acceptance of the outcomes than do exercises that are exclusively remote (Krueger et al., 2012; McBride et al., 2012). However, they can be expensive or logistically infeasible (Knol et al., 2010).

3.4 | Develop clear questions

Like many other structured approaches (Cooke, 1991; Mellers et al., 2014; O'Hagan et al., 2006), IDEA requires experts to estimate numerical quantities or probabilities. The objective is to obtain approximations of facts that can be cross-examined and used to inform decisions and models (Morgan, 2014).

Achieving this requires the formulation of questions that are relatively free from linguistic ambiguity and framing that may generate unwanted bias. This also means providing details such as the time, place, methods of investigation, the units of measurement, the level of precision required and other caveats (Table 2 and Supplementary Material A). Questions should also aim to elicit information in a format that most closely aligns with the domain knowledge and experiences of the experts (Aspinall & Cooke, 2013).

The IDEA protocol incorporates two alternative question formats depending on whether quantities or probabilities are being elicited (Table 2). The four-step format (Speirs-Bridge et al., 2010) is mostly used to elicit quantities (Table 2); however, it can also be used to elicit other types of data such as percentages and ratios. Wherever possible, questions should be framed in their frequency format because it is less prone to linguistic ambiguity than asking experts for these estimates directly (refer to Gigerenzer & Edwards, 2003; Supplementary Material A). The four-step format involves asking for upper and lower plausible bounds, a best guess and a 'degree of belief' (how sure are you). Taken together, an expert's responses to the four-step format are designed to be interpreted as a credible interval (i.e. the degree of belief that an event will occur, given all knowledge currently available). If an expert provides a certainty level of, say, 70% for 10 similar events, then the truth should lie within their credible intervals 7 out of 10 times.

The three-step format (Burgman, 2015; Wintle et al., 2012) is used in place of the four-step format when eliciting single event probabilities (and thus differs from other three-step formats mentioned in the literature, e.g. Speirs-Bridge et al., 2010; Soll & Klayman, 2004). It was developed to avoid difficulties associated with asking experts to specify second order probabilities (i.e. their confidence in their degree of belief). It involves asking experts for their degree of belief that an event will occur by asking for their lowest probability, their highest probability and their best guess of the probability that the described event will occur (Table 2).

Usually, we suggest that experts using the four-step elicitation method constrain their confidence levels to between 50% and 100% (Speirs-Bridge et al., 2010; Table 2). If an expert states they are less than 50% confident that their intervals contain the truth, it implies that they are more confident that the truth lies outside their intervals

TABLE 2 The three-step and four-step elicitation formats used by the IDEA protocol. More examples are provided in Supplementary Material A

Three-step elicitation (probability of an event)	Four-step elicitation (quantities and frequencies)
Question: 'Will Crown of Thorns Starfish be recorded at outbreak densities (1 or above) by the Australian Institute of Marine Science during their two-minute manta tow surveys of Rib Reef, on the Great Barrier Reef, in March 2016?'	Question: 'What will be the average density of Crown of Thorns Starfish recorded by the Australian Institute of Marine Science during their two-minute manta tow surveys of Rib Reef, on the Great Barrier Reef, in March 2016?'
3-step questions: 1. Realistically, what do you think is the lowest plausible probability [event X] will occur?__ 2. Realistically, what do you think is the highest plausible probability that [event X] will occur?___ 3. Realistically, what is your best estimate for the probability that [event X] will occur?___	4-step questions: 1. Realistically, what do you think the lowest plausible value for [event X] will be ?___ 2. Realistically, what do you think the highest plausible value for [event X] will be ?___ 3. Realistically, what is your best guess for [event X]?___ 4. How confident are you that your interval, from lowest to highest, could capture the true value of [event X]? Please enter a number between 50% and 100%___

than within them, which experience suggests is rarely what the expert actually believes.

Both the wording and question order in the three and four-step question formats are important (Table 2). The words 'plausible' and 'realistic' are intentionally used to discourage people from specifying uninformative limits (such as 0 and 1 for bounds on probability estimates). Asking for lowest and highest estimates *first*, encourages consideration of counterfactuals and the evidence for relatively extreme values, and avoids anchoring on best estimates (Morgan & Henrion, 1990).

It is important to highlight that the IDEA protocol centres around the three-step and four-step question formats, as they have been shown to assist experts in constructing and converting their knowledge into quantitative form, however, it does not automatically dictate what this information is, e.g. the best guess could be interpreted as a mean, median, or mode if desired but is not defined. Thus, the basic/standard protocol described is this paper was not designed, on its own, to elicit a probability distribution, but rather a best estimate and uncertainty bounds.

That said, the responses elicited using the IDEA question format can be used to assist in construction of a probability distribution if desired, for example by taking the best guess and interval bounds as moments of a distribution and fitting via least squares (Chadés et al., 2015; McBride, Fidler, & Burgman, 2012). However, the testing and refinement required to develop a standardised protocol for fitting distributions has not yet been done.

If taking this approach, careful consideration must be given to the underlying assumptions involved in converting the information to a distribution, for example, what the best guess represents and how to extrapolate uncertainty bounds, as well as to the choice of distribution(s) (e.g. O'Hagan et al., 2006 and references therein). We suggest that all assumptions are documented and are made clear to experts before and throughout the elicitation.

Alternatively, experts may be asked to provide their estimates as fixed quantiles (e.g. as in Cooke & Goossens, 2000); however, this may be challenging for many experts, particularly if remote elicitation is utilised. For this reason, additional iteration between the project team and experts may be necessary to ensure their beliefs are being adequately represented. Additional research and testing of methods to elicit such information remotely in a language more familiar to experts is needed.

The number of questions will depend on their difficulty, the time available and the motivations of the experts. To avoid expert fatigue we recommend asking no more than 15–20 questions in a single day of face-to-face elicitation (Speirs-Bridge et al., 2010). It is possible to ask more questions when elicitations are run remotely (over the web or by email) although at the risk of reducing participation levels if experts are not sufficiently motivated. If there are many questions, we recommend dividing questions and subsequently eliciting judgements over a number of weeks or months. If time is limited, then we suggest recruiting more experts and dividing questions between multiple groups.

The inclusion of previous data, current trends and useful links can help to clarify questions and inform responses. However, experts may anchor on background information, even on irrelevant quantities that the background information provides (see McBride, Fidler, & Burgman, 2012). This may result in a biased group estimate. We recommend providing background information for the discussion following Round 1, so that experts initially use their own reasoning and source data independently.

Practice questions can be developed and sent to help familiarise experts with the question style and the overall process. Practice questions should have a resolution but need not be taken from the domain of the elicitation and could focus on events for which there will be a short-term outcome, such as weather forecasts, traffic incidents or stock prices. Alternatively, the project team may have access to data that is not available to the experts, which can be used to develop practice questions.

We recommend that at least two subject matter experts are engaged to review draft questions. Ideally, these experts should be sourced from outside of the expert group, although this may not be possible for highly specialised topics. The experts should consider whether the questions are within the domain of expertise of the participants, are free from linguistic ambiguity or biases and can be completed in the time-frame available. The problem owner should also review the questions to ensure they provide the required data.

3.5 | Ethics clearance and useful project documents

Depending on the nature of the elicitation, human subjects research approval (ethics clearance) may be mandated by your institution or funding source. Many journals insist on ethics clearance. Ethics clearance can take some time to organise and should be sought as soon as the project details are specified. Approval should be obtained before experts are recruited. Some important elements outlined below should be considered regardless of whether ethics clearance is required (refer to Supplementary Material A for examples).

The coordinator should consider how they will protect data and the anonymity of experts. We recommend de-identifying expert judgements and personal information using codenames for experts (they can nominate their own), encrypting folders and establishing maximum periods for which data will be stored. If information provided in the questions includes data that are not publicly available, the coordinator may need to obtain permission from the owners to use them.

A project statement written in plain language should be developed. It should be brief, but list key information including that it is voluntary, whether payment will be involved, how much time will be asked of the experts, over what period, how the data will be used, how the anonymity of judgements will be protected, how the data will be stored, who will have access, how they can enquire or complain about the process and that they are free to withdraw at any time.

A consent form provides an opportunity for experts to sign that they acknowledge having read and understood the purpose of the study, and that they are willing to take part. In some applications, it may be important to seek permission from experts to publish their names and credentials as supporting information. If the project team plan to retain the judgements elicited from experts as their intellectual property, they should make this clear in the consent form.

An instruction manual should be developed to help guide the experts through the elicitation process and to reiterate key dates for the elicitation (Supplementary Material A).

3.6 | Selecting and engaging experts

As with other structured protocols (Cooke, 1991; Keeney & von Winterfeldt, 1991; Mellers et al., 2014; O'Hagan et al., 2006), IDEA promotes the use of multiple experts based on empirical evidence that indicates that while criteria such as age, experience, publications, memberships and peer-recommendation can be useful for sourcing potential experts, they are actually very poor guides to determining a priori someone's ability to provide good judgements in elicitation settings (Burgman et al., 2011; Shanteau et al., 2002; Tetlock & Gardner, 2015) and may result in the unnecessary exclusion of knowledgeable individuals (French, 2011; Shanteau et al., 2002). The best guide to expert performance is a person's previous performance on closely related tasks, which is rarely available a priori. The inability to identify the best expert means that groups of multiple experts almost always perform as well as, or better than, the best regarded expert(s) (Burgman et al., 2011; Hora, 2004; Mellers et al., 2014; Surowiecki, 2004).

Because it is usually not possible to predict who has the requisite knowledge to answer a set of questions accurately, the main criterion when selecting experts is whether the person can understand the questions being asked. We recommend that you establish relevant knowledge criteria and create a list of potential participants including their specialisation or skills, and contact details.

Be especially vigorous in pursuing people that add to the group's diversity. Diversity should be reflected by variation in age, gender, cultural background, life experience, education and specialisation. These are proxies for cognitive diversity (Page, 2008). In high profile or contentious cases, a transparent and balanced approach to selecting experts will also be important for circumventing claims of bias (Keeney & von Winterfeldt, 1991). We recommend aiming for around 10–20 participants, based on practicality and experience, and on empirical evidence suggesting only minor improvements in the group's performance are gained by having more than 6–12 participants (Armstrong, 2001; Hogarth, 1978; Hora, 2004).

When recruiting experts, send a short introductory email inviting them to participate in an introductory teleconference or workshop. We do this at least 3 weeks before the proposed teleconference. Try to have the email originate from someone known to the experts. If this is not possible, then include details of how they came to be recommended, and why they should be involved. Personalised communication will also help ensure that experts see and respond to email invitations. Mail-merge software is helpful for personalising emails by linking text fields such as names to email addresses.

Important information, such as the consent form, project statement and timeline, should be included as attachments to the introductory email. Provide contact details and offer experts the opportunity to discuss the project prior to the teleconference. Keep a contact log to track responses and follow up on late replies. Follow-up introductory emails with a telephone call if experts don't respond within 3 to 4 days. Send reminders ahead of due dates.

4 | UNDERTAKING AN ELICITATION

The following outline assumes that judgements will be elicited from experts remotely using the IDEA protocol. However, the same basic approach can be adopted for face-to-face workshops. Although the method specifies that the questions will be sent to experts and answered remotely, we recommend an inception meeting in the form of a teleconference (Section 4.1). If practical, we also recommend a workshop between Round 1 and Round 2 (Section 4.3.2), although this is less important than the inception meeting.

4.1 | Inception meeting

An introductory meeting is vital for establishing a rapport with the experts, explaining the motivations for, and expectations of, the elicitation. It provides an opportunity for the coordinator to explain context, and for the facilitator and analyst to explain how the various steps and constraints contribute to relatively high quality judgements.

Start the meeting by thanking experts for their time and introducing the motivation behind the project, and objectives for the elicitation. Many participants will be sceptical of expert elicitation or of the involvement of others. Acknowledge this scepticism and explain the unavoidable nature of expert judgement: the data we need are not available. Experts are often uncomfortable with stating their judgements in quantitative form. It is important to acknowledge this, while emphasising the primary motivation for using the IDEA protocol is that it applies the same level of scrutiny and neutrality to expert judgement as is afforded to the collection of empirical data. In addition, the elicitation of numbers helps to overcome linguistic ambiguity.

Explain that participants must *not* speak to other participants about the elicitation prior to the discussion phase between Rounds 1 and 2. However, they can and should speak to anyone else they choose and use all relevant information sources.

If time permits, run through the list of intended questions to further ensure that the wording and requirements are clear. Reiterate the timelines and procedures for the Round 1, and allow sufficient time for experts to ask questions. An example transcript is provided in Supplementary Material A.

4.2 | Round 1 estimates

Round 1 commences with an email to the experts containing the questions to be answered by the experts, or a link to the questions (if using a web-based platform), together with instructions on how to complete them (Supplementary Material A).

When undertaking remote elicitation, allow about 2 weeks for experts to complete the exercise. Allow sufficient time for late responses (another one to 2 weeks). Send a reminder to experts before the close of the elicitation.

4.3 | Analysis

Prior to the discussion phase, feedback on the results of Round 1 for experts will need to be prepared. This step involves cleaning and standardising data, aggregating judgements (Table 3), and providing clear and unambiguous graphs of the Round 1 estimates (Figure 3).

4.3.1 | Cleaning the data

Elicited responses will almost always require some level of cleaning. Common mistakes include entering numbers in the wrong boxes (e.g. the lowest estimate entered into the best guess, Table 3), blank estimates, wrong units, e.g. estimates in tonnes instead of kilograms, or as proportions (out of 1) rather than percentages (out of 100), illogical and out-of-range numbers. Clarify with experts whether apparent errors represent mistakes.

4.3.2 | Standardise intervals (four-step only)

In the four-step question format, experts specify credible intervals. The analyst should standardise these intervals, typically to 90% or 80% credible intervals, so that experts view the uncertainties of all experts across questions on a consistent scale (Table 3). We use linear extrapolation (Adams-Hosking et al., 2016; Bedford & Cooke, 2001), in which:

$$\text{Lower standardised interval: } B - ((B - L) \times (S/C))$$

$$\text{Upper standardised interval: } B + ((U - B) \times (S/C))$$

where B = best guess, L = lowest estimate, U = upper estimate, S = level of credible intervals to be standardised to and C = level of confidence given by participant. In cases where the adjusted intervals fall outside of reasonable bounds (such as [0, 1] for probabilities), we truncate distributions at their extremes.

Participants often ask why they should specify confidence levels for their intervals when their credible intervals are subsequently standardised (e.g. to 80%). While counter-intuitive, Speirs-Bridge et al.

(2010) found that overconfidence was reduced if experts were obliged to specify their own level of confidence and the credible intervals were subsequently standardised.

As the main purpose of the adjusted intervals at this stage is to allow for comparison during the discussion phase, linear extrapolation provides an easy-to-implement and explainable approach that minimises the need to make additional distributional assumptions. Our experience is that alternative approaches (e.g. using the elicited responses to fit a distribution such as the beta, betaPERT or log-normal) make little difference to the visual representations that result, or to the discussions that follow. Thus, we use linear extrapolations for simplicity. Experts are encouraged to change their estimates in Round 2 if the extrapolation does not represent their true belief.
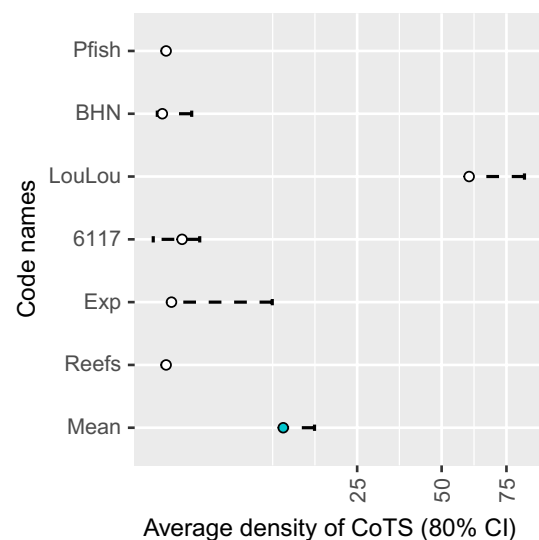

FIGURE 3 Graphical feedback provided in Round 1 (from the data provided in Table 3). The circles represent the best guess of experts, the dashed lines show their standardised 80% uncertainty bounds. All experts except LouLou believe the density will be equal to or lower than 0.50. The arithmetic mean is 10.16

TABLE 3 An example of data cleaning, standardisation and aggregation for a single four-step elicitation question. The raw data provided by experts is shown to the left and the standardised data to the right. The group was asked to estimate the average density of Crown of Thorns starfish on Rib Reef, on the Great Barrier Reef (Table 2). Note that participant 'Exp' has entered data in an illogical order (the lower estimate is higher than the upper), this is corrected in the standardised data. Credible intervals have been standardised to 80%, which changes the upper and lower estimates but not the best guess. Quantile aggregation is then used to derive a group aggregation for lower, upper and best guesses

Raw data					Standardised data				
Name	Lower	Upper	Best	Conf (%)	Name	Lower	Upper	Best	Conf (%)
Pfish	0.06	0.12	0.1	70	Pfish	0.05	0.12	0.10	80
BHN	0	1	0.05	90	BHN	0.01	0.89	0.05	80
LouLou	60	80	60	70	LouLou	60.00	82.86	60.00	80
6117	0	1	0.5	50	6117	0.00	1.30	0.50	80
Exp	8	0.18	0.2	75	Exp	0.18	8.52	0.20	80
Reefs	0.05	0.15	0.1	75	Reefs	0.05	0.15	0.10	80
					Mean	10.05	15.64	10.16	80

4.3.3 | Calculate a group aggregate estimate

Combined estimates are calculated following standardisation of the experts' intervals. Most applications of the IDEA protocol make use of quantile aggregation, in which the arithmetic mean of experts' estimates is calculated for the lower, best and upper estimates for each question (Table 3). Quantile aggregation using the arithmetic mean avoids the need to fit a distribution, and has been found to perform as well as more complex methods by Lichtendahl, Grushka-Cockayne, and Winkler (2013), although there is ongoing debate in the literature over how well this result holds more widely. In particular, recent studies by Eggstaff, Mazzuchi, and Sarkani (2014) and Colson and Cooke (2017), that involved large cross-validation studies, carried out on a massive dataset of expert elicited estimates (a total of 73 independent expert judgement studies from the TU Delft database (Colson & Cooke, 2017)) found that the quantile aggregation method performs poorly when compared to aggregating fitted distributions. These new results suggest further investigation of aggregation methods for the IDEA protocol is warranted. However, until such time, we advocate quantile aggregation, as a fast, straight forward approach that is well-understood by participants and requires no distributional assumptions.

Equally weighted group aggregations can be sensitive to extreme outliers in small groups. Rather than excluding outliers, they should be integrated into discussion, to determine whether there are good reasons for providing them. We believe outliers should only be trimmed if they are clearly and uncontroversially incorrect (for example outside of possible bounds).

4.3.4 | Create and share graphical output

Create graphs for each question to display the estimates of each participant (labelled with their codename) and the group aggregate (Figure 3). If displaying judgements from a four-step elicitation, remind experts that their displayed uncertainty bounds may vary from their original estimates due to standardisation, and that if the adjusted interval doesn't accurately reflect their beliefs, they can and should adjust their uncertainty bounds in Round 2.

Compile graphs, tables, comments for each question and additional information submitted by experts and return to experts.

4.4 | Discussion phase

The discussion phase commences once experts have received feedback on the results of Round 1. The facilitator guides and stimulates discussion, but does not dominate it. The facilitator should choose contrasting results and ask questions which explore sources of variation for example, 'What could occur that would lead the estimates to be high (or low)?'. Where necessary, the facilitator should clarify meaning or better define terms. A set of useful questions for facilitating discussion is provided in Supplementary Material A.

4.5 | Round 2 estimates

Following the discussion, experts make a second, anonymous and independent estimate for each question. Experts who dropped-out following Round 1 should be excluded from the final aggregation. Analyse the results using the methods described in Sections 4.3 and 4.3.1 above.

4.6 | Post-elicitation: Documentation and reporting

Following completion of the elicitation, the experts' final assessments and the group aggregate judgement should be circulated to the group for final review and 'sign off'. All steps taken and results collected during the elicitation should be documented to provide a transparent record of the process and results. In presenting outputs, aggregated point estimates and uncertainty intervals should be communicated along with the individual expert Round 2 estimates (Figure 4) to convey the full level of inter-expert uncertainty (Morgan, 2014). This concludes the elicitation process.

5 | DISCUSSION

This paper provides advice on the IDEA protocol, a structured elicitation method that is well-suited to the time and resource constraints of many conservation problems. The three-step and four-step question formats (Speirs-Bridge et al., 2010) derive numerical estimates in a language accessible to most experts, and summarises
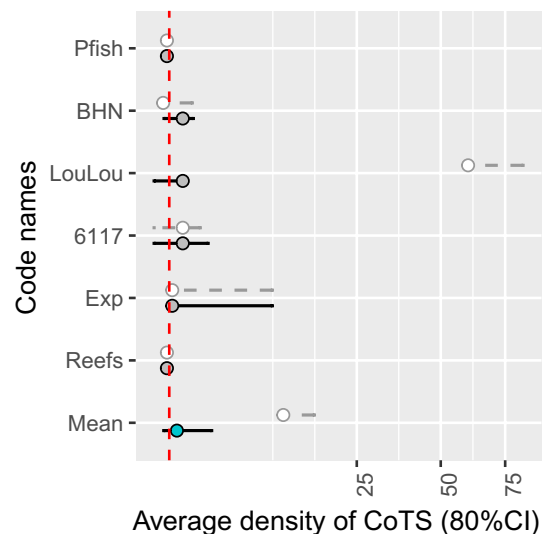
FIGURE 4 Graphical feedback provided in Round 2. The dashed horizontal lines show the Round 1 estimates of each of the experts. The bold horizontal lines show their corresponding Round 2 estimates. Only three experts revised their estimates (BHN, LouLou and 6117). The mean of Round 2 estimates changed substantially from Round 1 (Figure 3), to 0.32 Crown of Thorns Starfish (CoTS). The realised truth (0.14 CoTS) is shown by the dashed vertical line

inter- and intra-expert uncertainty, while the option for remote elicitation accommodates modest budgets.

The protocol is simple to understand and apply and if needed could be undertaken entirely by hand (though most often in Excel), making it an attractive option for those who may have limited resources or time/willingness to familiarise with new techniques and/or software. Importantly, the protocol has been shown to yield relatively reliable judgements in domains as diverse as conservation (Burgman et al., 2011) and geopolitical forecasting (Hanea, McBride, et al., 2016; Hanea et al., 2016; Wintle et al., 2012).

While we advocate that structured protocols can improve judgements, their use does not guarantee accurate estimates. Unfortunately, in most cases, the need for expert judgement arises where empirical data cannot be obtained, and there is no way of evaluating judgements for their accuracy (or calibration), and decisions need to be made urgently (Martin, Camaclang, Possingham, Maguire, & Chadès, 2017; Martin et al., 2012b; McBride, Fidler & Burgman 2012). In such cases, the best means of assessing whether the experts have relevant/useful knowledge, can adapt their knowledge and communicate their knowledge accurately is to test them on carefully crafted test questions (Cooke, 1991). The development of such questions and the incorporation of improved methods for aggregation into the IDEA protocol is the subject of current research by the authors. An additional avenue of research is required to understand the best way to elicit probability distributions in a way that respects the bounded rationality of experts, especially under remote elicitation conditions. We advocate that while all such approaches could improve the IDEA protocol, they would benefit from additional testing prior to formal incorporation into the IDEA protocol.

6 | CONCLUSION

The reliability of expert judgement will always be sensitive to which experts participate and how questions are asked. However, structured protocols such as the IDEA protocol improve the quality of these judgements by taking advantage of the wisdom of the crowd and mitigating a range of the most pervasive and potent sources of bias. This guide explains the rationale behind the IDEA protocol and decomposes the process into manageable steps. It also highlights areas of future research. Regardless of whether the IDEA protocol is adopted, we strongly advocate that structured protocols for expert elicitation must be adopted within conservation, natural resource management and other scientific domains.

AUTHORS' CONTRIBUTIONS

V.H. led the development and writing of the manuscript based on her experience implementing the IDEA protocol. M.B., A.H., M.M. and B.W. provided additional review and advice based on their own experiences with the IDEA protocol and structured expert elicitation. All authors contributed critically to the drafts and gave their final approval for publication.

DATA ACCESSIBILITY

This manuscript does not use data.

REFERENCES

Adams-Hosking, C., McBride, M. F., Baxter, G., Burgman, M., de Villiers, D., Kavanagh, R., … McAlpine, C. A. (2016). Use of expert knowledge to elicit population trends for the koala (*Phascolarctos cinereus*). *Diversity and Distributions*, *22*, 249–262.

Armstrong, J. S. (2001). Combining forecasts. In: J. S. Armstrong (Ed.), *Principles of forecasting. International series in operations research & management science* (pp. 417–439, Vol. *30*). Boston, MA: Springer.

Aspinall, W. P. (2010). A route to more tractable expert advice. *Nature*, *463*, 294–295.

Aspinall, W. P., & Cooke, R. M. (2013). Quantifying scientific uncertainty from expert judgement elicitation. In J. Rougier, S. Sparks, & L. Hill (Eds.), *Risk and uncertainty assessment for natural hazards* (pp. 64–99). Cambridge, UK: Cambridge University Press.

Bamber, J. L., & Aspinall, W. (2013). An expert judgement assessment of future sea level rise from the ice sheets. *Nature Climate Change*, *3*, 424–427.

Ban, S. S., Pressey, R. L., & Graham, N. A. (2014). Assessing interactions of multiple stressors when data are limited: A Bayesian belief network applied to coral reefs. *Global Environmental Change*, *27*, 64–72.

Bedford, T., & Cooke, R. M. (2001). *Mathematical tools for probabilistic risk analysis*. Cambridge, UK: Cambridge University Press.

Burgman, M. A. (2004). Expert frailties in conservation risk assessment and listing decisions. In P. Hutchings, D. Lunney, & C. Dickman (Eds.), *Threatened species legislation: Is it just an Act?* (pp. 20–29). Mosman, NSW, Australia: Royal Zoological Society.

Burgman, M. A. (2015). *Trusting Judgements: How to get the best out of experts*. Cambridge, UK: Cambridge University Press.

Burgman, M. A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., … Twardy, C. (2011). Expert status and performance. *PLoS ONE*, *6*, 1–7.

Chadés, I., Nicol, S., van Leeuwen, S., Walters, B., Firn, J., Reeson, A., … Carwardine, J. (2015). Benefits of integrating complementarity into priority threat management. *Conservation Biology*, *29*, 525–536.

Colson, A. R., & Cooke, R. M. (2017). Cross validation for the classical model of structured expert judgment. *Reliability Engineering & System Safety*, *163*, 109–120.

Cooke, R. M. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. New York, NY: Oxford University Press.

Cooke, R., & Goossens, L. (2000). Procedures guide for structural expert judgement in accident consequence modelling. *Radiation Protection Dosimetry*, *90*, 303–309.

Eggstaff, J. W., Mazzuchi, T. A., & Sarkani, S. (2014). The effect of the number of seed variables on the performance of Cooke's classical model. *Reliability Engineering & System Safety*, *121*, 72–82.

EPA. (2009). *Expert elicitation white paper review* (ed. U.S.E.P.A. Science Council). Washington, DC: EPA.

Firn, J., Martin, T. G., Chadès, I., Walters, B., Hayes, J., Nicol, S., & Carwardine, J. (2015). Priority threat management of non-native plants to maintain ecosystem integrity across heterogeneous landscapes. *Journal of Applied Ecology*, *52*, 1135–1144.

French, S. (2011). Aggregating expert judgement. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, *105*, 181–206.

French, S. (2012). Expert judgment, meta-analysis, and participatory risk analysis. *Decision Analysis*, *9*, 119–127.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*, 35–42.

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

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *BMJ: British Medical Journal*, *327*, 741–744.

Gregory, R., & Keeney, R. L. (2017). A practical approach to address uncertainty in stakeholder deliberations. *Risk Analysis*, *37*, 487–501.

Hanea, A. M., McBride, M. F., Burgman, M. A., & Wintle, B. C. (2016). Classical meets modern in the IDEA protocol for structured expert judgement. *Journal of Risk Research*, 1–17. https://doi.org/10.1080/1 3669877.2016.1215346

Hanea, A., McBride, M., Burgman, M., Wintle, B., Fidler, F., Flander, L., … Mascaro, S. (2016). I$_{nvestigate}$D$_{iscuss}$E$_{stimate}$A$_{ggregate}$ for structured expert judgement. *International Journal of Forecasting*, *33*, 267–269.

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Thousand Oaks, CA: Sage.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, *21*, 40–46.

Hora, S. C. (2004). Probability judgments for continuous quantities: Linear combinations and calibration. *Management Science*, *50*, 597–604.

IUCN. (2012) IUCN red list catergories and criteria: Version 3.1. pp. iv + 32 pp. IUCN, Gland, Switzerland and Cambridge, UK.

Janis, I. L. (1971). Groupthink. *Psychology Today*, *5*, 43–46.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237.

Keeney, R. L., & von Winterfeldt, D. (1991). Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management*, *38*, 191–201.

Kent, S. (1964). Words of estimative probability. *Studies in Intelligence*, 49–65.

Knol, A. B., Slottje, P., van der Sluijs, J. P., & Lebret, E. (2010). The use of expert elicitation in environmental health impact assessment: A seven step procedure. *Environmental Health*, *9*, 1.

Krueger, T., Page, T., Hubacek, K., Smith, L., & Hiscock, K. (2012). The role of expert opinion in environmental modelling. *Environmental Modelling & Software*, *36*, 4–18.

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.

Lichtendahl, K. C. Jr, Grushka-Cockayne, Y., & Winkler, R. L. (2013). Is it better to average probabilities or quantiles? *Management Science*, *59*, 1594–1611.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, *108*, 9020–9025.

Low Choy, S., 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.

Martin, T. G., Burgman, M. A., Fidler, F., Kuhnert, P. M., Low-Choy, S., McBride, M., & Mengersen, K. (2012a). Eliciting expert knowledge in conservation science. *Conservation Biology*, *26*, 29–38.

Martin, T. G., Camaclang, A. E., Possingham, H. P., Maguire, L. A., & Chadès, I. (2017). Timing of protection of critical habitat matters. *Conservation Letters*, *10*, 308–316.

Martin, T. G., Nally, S., Burbidge, A. A., Arnall, S., Garnett, S. T., Hayward, M. W., … Possingham, H. P. (2012b). Acting fast helps avoid extinction. *Conservation Letters*, *5*, 274–280.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., … Zweirs, F. W. (2010). Guidance note for lead authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Jasper Ridge, CA.

McBride, M. F., Fidler, F., & Burgman, M. A. (2012). Evaluating the accuracy and calibration of expert predictions under uncertainty: Predicting the outcomes of ecological research. *Diversity and Distributions*, *18*, 782–794.

McBride, M. F., Garnett, S. T., Szabo, J. K., Burbidge, A. H., Butchart, S. H., Christidis, L., … Watson, D. M. (2012). Structured elicitation of expert judgments for threatened species assessment: A case study on a continental scale using email. *Methods in Ecology and Evolution*, *3*, 906–920.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., … Swift, S. A. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, *25*, 1106–1115.

Metcalf, S. J., & Wallace, K. J. (2013). Ranking biodiversity risk factors using expert groups – Treating linguistic uncertainty and documenting epistemic uncertainty. *Biological Conservation*, *162*, 1–8.

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, *111*, 7176–7184.

Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. New York, NY: Cambridge University Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., … Rakow, T. (2006). *Uncertain judgements: Eliciting experts' probabilities*. West Sussex, UK: John Wiley & Sons.

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. New Jersey: Princeton University Press.

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.

Rothlisberger, J. D., Finnoff, D. C., Cooke, R. M., & Lodge, D. M. (2012). Ship-borne nonindigenous species diminish Great Lakes ecosystem services. *Ecosystems*, *15*, 1–15.

Runge, M. C., Converse, S. J., & Lyons, J. E. (2011). Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. *Biological Conservation*, *144*, 1214–1223.

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*, 253–263.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 299.

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.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. London, UK: Little, Brown.

Tetlock, P., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. New York, NY: Random House.

Tversky, A., & Kahneman, D. (1975) Judgment under uncertainty: Heuristics and biases. Utility, probability, and human decision making: Selected proceedings of an interdisciplinary research conference, Rome, 3–6 September, 1973 (eds D. Wendt & C. Vlek), pp. 141–162. Dordrecht, The Netherlands: Springer.

van Gelder, T., Vodicka, R., & Armstrong, N. (2016). Augmenting expert elicitation with structured visual deliberation. *Asia & The Pacific Policy Studies*, *3*, 378–388.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, *115*, 348.

Wintle, B. C., Fidler, F., Vesk, P. A., & Moore, J. L. (2013). Improving visual estimation through active feedback. *Methods in Ecology and Evolution*, *4*, 53–62.

Wintle, B., Mascaro, S., Fidler, F., McBride, M., Burgman, M., Flander, L., … Manning, B. (2012). The intelligence game: Assessing Delphi groups and structured question formats. In: *The 5th Australian security and intelligence conference (3rd-5th December)*. Perth, WA: SRI Security Research Institute, Edith Cowan University.

Wittmann, M. E., Cooke, R. M., Rothlisberger, J. D., Rutherford, E. S., Zhang, H., Mason, D. M., & Lodge, D. M. (2015). Use of structured expert judgment to forecast invasions by bighead and silver carp in Lake Erie. *Conservation Biology*, *29*, 187–197.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.