



CONTRIBUTED PAPER

Evaluating elicited judgments of turtle captures for data-limited fisheries management

William N. S. Arlidge^{1,2} | Joanna Alfaro-Shigueto^{3,4,5} |
Bruno Ibanez-Erquiaga⁶ | Jeffrey C. Mangel^{3,4} | Dale Squires⁷ |
Eleanor J. Milner-Gulland¹

¹Interdisciplinary Centre for Conservation Science, Department of Zoology, University of Oxford, Oxford, UK

²Pembroke College, University of Oxford, Oxford, UK

³ProDelphinus, Lima, Peru

⁴Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Penryn, Cornwall, UK

⁵Facultad de Biología Marina, Universidad Científica del Sur, Lima, Peru

⁶Departamento de Química y Biología, Universidad San Ignacio de Loyola, Lima, Peru

⁷NOAA Fisheries Southwest Fisheries Science Center, La Jolla, California

Correspondence

William N. S. Arlidge, Interdisciplinary Centre for Conservation Science, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.
Email: william.arlidge@zoo.ox.ac.uk

Funding information

Commonwealth Scholarship Commission, Grant/Award Number: NZCR-2015-174; National Fish and Wildlife Foundation, Grant/Award Number: 550011; Pew Charitable Trusts

Abstract

We compare judgments of green turtle (*Chelonia mydas*) captures elicited from local gillnet skippers and not-for-profit conservation organization employees operating in a small-scale fishery in Peru, to capture rates calculated from a voluntary at-sea observer program operating out of the same fishery. To reduce cognitive biases and more accurately quantify uncertainty in our experts' judgments, we followed the IDEA ("Investigate," "Discuss," "Estimate," and "Aggregate") structured elicitation protocol. The elicited mean monthly estimates of green turtle gillnet captures within summer and winter fishing seasons were higher than the equivalent green turtle capture rates calculated from the fisheries observer data; however, no statistically significant differences were identified when comparing the means of the datasets using bootstrap hypothesis tests (winter observed difference-in-means: 83.15, adj mean \pm SD = 42.39 \pm 32.59; summer observed difference-in-means: 68.58, adj mean \pm SD = 54.06 \pm 41.22). We investigated respondent performance in relation to the observer data capture rates. The not-for-profit employees scored high on accuracy and calibration performance metrics. The gillnet skippers' judgments ranked higher on informativeness yet lower on accuracy and calibration, potentially reflective of overconfident judgments. This research presents a new context for using the IDEA protocol, which may prove helpful for rapid, exploratory evaluations of capture and bycatch impact in data-limited small-scale fishery management scenarios.

KEYWORDS

bootstrap, bycatch, Delphi, expert judgment, IDEA, observer data, small-scale fisheries, turtle

1 | INTRODUCTION

The incidental capture and subsequent mortality (*bycatch*) of vulnerable marine species, such as turtles, is a known

conservation issue in many small-scale fisheries around the world (Alfaro-Shigueto et al., 2018; Peckham et al., 2007). The extent to which species of conservation concern are adversely impacted by many coastal fisheries, however,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Conservation Science and Practice published by Wiley Periodicals, Inc. on behalf of Society for Conservation Biology

remains a major knowledge gap—data paucity having been identified as one of the key challenges to address in the management of the small-scale fisheries subsector (FAO, 2018). Small-scale fisheries encompass traditional, low-technology, low-capital fishing methods. A single small-scale fishery can comprise a diverse array of vessels (often of small but varying sizes), participants, locations, resource, and gears (Khalil, Conforti, Ergin, & Gennari, 2017). This heterogeneity can make gathering comprehensive empirical data on capture and bycatch rates a challenge due to the complexity that these social–ecological systems represent (Dietz, Ostrom, & Stern, 2003).

Obtaining reliable data on the incidental capture and mortality of vulnerable species is, nonetheless, necessary to achieve ecologically and socioeconomically sustainable fisheries (Suuronen & Gilman, 2019). At-sea, human observer programs can produce accurate data on the incidental rates of capture and bycatch in fisheries. However, independent validation of the data is essential, and effective implementation of observer programs in small-scale fisheries can be complex and expensive (Bartholomew et al., 2018; Suuronen & Gilman, 2019). Note that here we define capture as everything that is caught and retained in fishing gear, and bycatch as capture that is discarded at sea, dead or injured to an extent where death is the result; following the definitions in Hall (1996). Further complexity arises because observer coverage is rarely available for an entire fleet; capture and bycatch estimates are often inferred from a subset of fishing trips, typically using models to help control for sampling biases (Benoît & Allard, 2009). Electronic monitoring programs are increasingly trialed and implemented in both large- (Ames, Leaman, & Ames, 2007; Needle et al., 2014) and small-scale fisheries (Bartholomew et al., 2018)—trials in small-scale fisheries appear promising, with accuracy similar to at-sea human observers at lower cost. Despite clear potential in the use of electronic monitoring in fisheries, multiple studies evaluating the technology note that improvements are still needed to detect certain species, and recording of catch released below the water level or in areas outside the camera view remains a major limitation (Bartholomew et al., 2018; Gilman et al., 2019; Suuronen & Gilman, 2019). Post-trip interviews of skipper and crews also quantify capture and bycatch in small-scale fisheries (Alfaro-Shigueto et al., 2018; Goetz, Read, Santos, Pita, & Pierce, 2013). These data often take the form of questionnaires and provide the cheapest and most rapid source of information—at times supporting near real-time management measures (Drew, 2005). Quickly obtaining a broad understanding of protected species' capture and bycatch rates can be particularly useful in small-scale fisheries, in which data are often limited (or entirely absent). While significant potential exists

for post-trip questionnaires to support rapid evaluations of incidental captures to inform bycatch impact in data-limited fishery scenarios, questionnaire-based interviews are often considered to be less reliable than data collected using observer programs as they are subject to individual respondent's biases and heuristics (Suuronen & Gilman, 2019).

In conservation, expert knowledge (substantive information on a particular topic that is not widely known by others; Martin et al., 2012) can be used to inform the decision-making process. This is due to the need to make timely management decisions about complex and dynamic environments, particularly in data-limited scenarios, unique circumstances, or when predictions under uncertainty are required (Burgman et al., 2011; Cook, Hockings, & Carter, 2010). When drawing on expert knowledge, however, it is essential to account for the contextual biases and heuristics that individuals bring with them, as these can affect the validity of the information they give and the subsequent management actions that result (Kynn, 2008; O'Hagan et al., 2006). The need to design and implement effective conservation strategies that are rigorous, robust, repeatable, and include an estimate of uncertainty has resulted in structured, evidence-based elicitation protocols such as the widely used Delphi method, which provides feedback from experts over successive questionnaire rounds (Cooke, 1991; Helmer-Hirschberg, Brown, & Gordon, 1966).

While elicited data are not a substitute for empirical data, structured protocol techniques prove informative in various fishery management settings when empirical data are not readily available. For example, this approach has been used in a risk assessment for New Zealand's critically endangered Māui dolphin (*Cephalorhynchus hectori maui*; Currey, Boren, Sharp, & Peterson, 2012), and to evaluate and rank threats to sea turtle populations in several fishing systems (Klein et al., 2017; Riskas, Tobin, Fuentes, & Hamann, 2018; Williams, Pierce, Hamann, & Fuentes, 2017). When fisheries lack the data and resources to implement more comprehensive observer procedures, significant potential exists to apply structured elicitation protocols to expert opinion to reduce personal biases and heuristics, and to quantify the associated uncertainty.

Peru's small-scale fisheries significantly impact marine biodiversity through capture and bycatch (Alfaro-Shigueto et al., 2010). The gillnet, the most commonly utilized gear (Castillo, Fernandez, Medina, & Guevara-Carrasco, 2018), has been identified as a major sink for several species of sea turtle of conservation concern (Alfaro-Shigueto et al., 2011, 2018; Alfaro-Shigueto, Dutton, Van Bresseem, & Mangel, 2007). Peru's current regulatory structure does little to help mitigate fishing-related mortalities of protected species like sea turtles in

the country's small-scale fisheries. With limited government efficacy, not-for-profit organizations play a role in filling data gaps and implementing conservation interventions to minimize protected species capture and bycatch. For example, not-for-profit organizations in Peru implement and maintain volunteer observer programs with small-scale fishers, and undertake post-trip interviews of skippers and crews (Alfaro-Shigueto et al., 2011, 2018). Conservation efforts such as these highlight the need for further management actions to help reduce vulnerable species captures in small-scale fishing systems. To improve rapid, exploratory evaluations of marine megafauna captures in data-limited small-scale fishery management scenarios, we compare at-sea human observer data from a small-scale fishery to the incidental capture estimates of green turtles obtained using a structured elicitation protocol.

In this study, we use the IDEA protocol ("Investigate," "Discuss," "Estimate," and "Aggregate"), which follows a modified Delphi method, incorporating many suggested adaptations to structured elicitation protocols that have been used in previous conservation research. Specifically, the protocol uses a four-step elicitation process to reduce overconfidence (Speirs-Bridge et al., 2010), encourages consultation with a diverse group of experts (Burgman et al., 2011), affords experts the opportunity to examine one another's estimates and to reconcile the meanings of questions through discussion, and uses performance-based mathematical aggregation of judgments (Hanea, McBride, Burgman, & Wintle, 2016). To date, the protocol has produced robust estimates in several studies (e.g., Hanea et al., 2016; Hanea, McBride, Burgman, & Wintle, 2016; Hemming, Walshe, Hanea, Fidler, & Burgman, 2018; van Gelder, Vodicka, & Armstrong, 2016), and shows promise as an effective tool for rapidly assessing capture and bycatch rates in small-scale fisheries.

The aims of this research were to (a) investigate if the IDEA protocol could support a rapid assessment of incidental captures of protected species (green turtles *Chelonia mydas*) occurring in a coastal gillnet fishery where sea turtle mortalities are a known conservation issue, (b) quantify the associated uncertainty, and (c) evaluate participant performance by comparing these estimates to incidental capture rates calculated from at-sea observer data obtained from the same fishery.

2 | METHODS

2.1 | Study system

San Jose, Lambayeque, Peru (6°46' S, 79°58' W) is a coastal fishing community with a high density of gillnet

vessels (Alfaro-Shigueto et al., 2010). The fishing-related mortality of several turtle species is known and problematic, including the East Pacific population of green turtles (*C. mydas*) and the critically endangered East Pacific population of leatherback turtles (*Dermochelys coriacea*; Alfaro-Shigueto et al., 2011; Alfaro-Shigueto et al., 2018). A voluntary at-sea, human observer program has been running with San Jose's gillnet skippers since 2007; however, coverage has not been comprehensive (Alfaro-Shigueto et al., 2007, 2010). Structured questionnaires have also been used to further knowledge of turtle capture and bycatch rates in the area (Alfaro-Shigueto et al., 2011, 2018). Several fishers in the San Jose community have been exposed to conservation interventions, working with a not-for-profit organization on at-sea technology trials to mitigate and record turtle captures (Bartholomew et al., 2018; Ortiz et al., 2016), and partaking in workshops to teach better handling procedures for turtle releases post capture.

The inshore-midwater gillnet fleet comprises vessels with small closed bridges that range in capacity from 5 to 32 gross registered tonnage (GRT), locally known as "lancha" (Guevara-Carrasco & Bertrand, 2017). Vessel numbers fluctuate both seasonally and annually, as fishers migrate from inland areas seeking fishing work during favorable weather conditions. Over the past decade the San Jose inshore-midwater gillnet fleet has been reducing in size as fishers shift their vessels from handling gillnets to jiggling gear to catch giant Humboldt squid *Dosidicus gigas*. Fleet size of the San Jose inshore-midwater gillnet fishery was approximately 60 vessels in 2008, with numbers decreasing to between 28 and 18 in the summer and winter of 2017, respectively (Alfaro-Shigueto et al., 2010; Supporting Information). A winter survey in San Jose in 2017 (July–September estimated that 15 inshore-midwater vessels actively fished, primarily with gillnets, while three additional vessels used gillnets but primarily fished with another gear type. Another small-scale gillnet fleet comprised of small, open-welled vessels known as "chalana," with a capacity range of 1–8 GT, also operates from San Jose in the inshore fishing area (Arlidge et al., 2020). All respondents in the current study were part of a wider elicitation survey investigating the efficacy of turtle capture and bycatch reduction strategies in the San Jose fishing system. Only the inshore-midwater fleet is the focus of this comparative study because observer data were not available for the inshore gillnet fleet.

We separately assessed two seasonal categorizations due to the differences in fishing effort between winter and summer conditions in the Lambayeque coastal fisheries. Summer is usually considered to be December–February (3 months), but information provided by a

government fisheries scientist in San Jose during a key informant interview noted that summer-like conditions span December–May (Arlidge et al., 2020), with this longer seasonal division supported by capture reports from the Lambayeque region (Guevara-Carrasco & Bertrand, 2017). Here we classify the San Jose winter fishing season as June–November and the summer fishing season as December–May.

2.2 | Estimates of turtle encounters

To elicit judgments of incidental captures of green turtles in gillnets set by San Jose inshore-midwater vessels, participants were asked to consider a counterfactual scenario in which a total gear switch occurred, from gillnets to a fishing gear that results in very little chance of turtle captures (such as lobster potting or trolling—a form of handline fishing). Estimates were provided as a monthly reduction in green turtle encounters with gillnets for the entire San Jose inshore-midwater fleet. Capture reduction estimates for leatherback turtles were also elicited, but small numbers make these less reliable than the green turtle estimates (Supporting Information). Participants were asked to assume 100% compliance with the counterfactual scenario. Judgments were given for summer and winter fishing periods. These data were collected as part of a wider study that elicited expert judgments on the efficacy of a range of turtle capture and bycatch reduction strategies that will be used to inform a marine megafauna mitigation model (Milner-Gulland et al., 2018).

2.3 | Expert elicitation procedure

We use the IDEA protocol with a combination of face-to-face group meetings and individual interviews over two elicitation rounds.

2.3.1 | Participant selection

We used simple random sampling by number generator to select gillnet skippers from a census list ($n = 168$) of skippers that were actively fishing during a wider survey period of July 1–September 30, 2017. The expert group ($n = 5$) comprised three local gillnet skippers of inshore-midwater vessels (representing 20% of the actively fishing inshore-midwater gillnet skippers in San Jose), and two not-for-profit conservation organization employees (JAS & JCM). Both of the not-for-profit employees have carried out regular research and conservation action in the study site area and more widely along the western

South American coastline. They have expertise in turtle ecology and the implementation of management strategies to reduce protected species mortalities in small-scale fisheries.

2.3.2 | Elicitation format

Data were elicited through individual face-to-face interviews over two elicitation rounds. Because lancha fishers spend little time on land between fishing trips of 1–13 days (averaging 7 days; Alfaro-Shigueto et al., 2010), this resulted in no time in which the lancha gillnet skippers were all on land, following an initial scoping meeting. Hence the decision was made to interview them separately.

2.3.3 | Stage 1: Introductory meeting

The first stage of the elicitation procedure was undertaken in a face-to-face group meeting. We met with the invited participants and discussed the context of the elicitation procedure with them, including providing an overview of the IDEA protocol, the method, study rationale, and the rules of participation. We ensured that free, prior, informed consent to participate was given, in accordance with our ethics permission (CUREC 1A; Ref No: R52516/RE001 and R52516/RE002).

2.3.4 | Stage 2: Investigate (Round 1)

Question format followed a four-point estimation method that has been shown to reduce overconfidence when eliciting individual judgments (Speirs-Bridge et al., 2010). This involves giving a (a) lower bound, (b) upper bound, (c) best guess, and (d) a level of confidence that the real value lies between these limits. Participants were asked to give estimates of the expected reduction in green turtle captures in gillnets within the winter and summer fishing seasons, for the scenario shifting gillnets to lobster potting or trolling. Estimates were given as monthly gillnet encounters, unless another time period was specified by the participants (e.g., turtle gillnet encounters per season). In cases when turtle captures per season were given, estimates were divided by the total number of months in the season.

2.3.5 | Stage 3: Analysis and feedback

In the four-step question format, participants implicitly specify credible intervals for their estimates. For example,

if in response to the question about how confident they are about their estimate, a participant says that they expect the true value to fall between their stipulated lower and upper limits in 7 of 10 cases; that implies a 70% credible interval. Prior to providing the first round of feedback, we standardized the participants' estimated intervals to 90% credible intervals to allow them to see the uncertainties across their estimates on a consistent scale. Linear extrapolation was used to standardize participants' elicited lower (l) and upper (u) uncertainty bounds to 90% credible bounds (Hemming, Burgman, et al., 2018). The standardized lower (l_{si}) and upper (u_{si}) bounds were calculated as:

$$l_{si} = B - \left((B - L) \times \left(\frac{S}{C} \right) \right) \quad (1)$$

$$u_{si} = B + \left((U - B) \times \left(\frac{S}{C} \right) \right) \quad (2)$$

where l_{si} is the standardized lower estimate, u_{si} is the standardized upper estimate, B is the best guess, L is the lowest estimate, U is the upper estimate, S is the level of credible intervals to be standardized to, and C is the level of confidence given by participant. Any adjusted intervals that fell outside of reasonable bounds (i.e., negative values) were truncated at their extremes (i.e., to zero).

Following standardization, estimates were combined using quantile aggregation, in which the arithmetic mean of participants' estimates is calculated for the lower, best, and upper estimates for each question (Hemming, Burgman, et al., 2018). Graphs for each question were generated to display the estimates of each participant (labeled with codenames that each respondent was individually aware of) and the group aggregate mean. This output was presented to the participants for use in the discussion and re-estimation phase that followed (Supporting Information).

2.3.6 | Stages 4 and 5: Discussion and re-estimation (Round 2)

The discussion and re-estimation phase took place through individual face-to-face interviews, led by the facilitator (BIE) with support from the coordinator and analyst (WNSA; Hemming, Burgman, et al., 2018). We provided hard copies of each question's graphical output to the participants; this included justification comments from the other participants (when given) and any questions from the analyst (Supporting Information). No participants declined to partake in the second elicitation round.

2.3.7 | Stage 6: Final aggregation and review

Following the second elicitation round, the revised data were analyzed and aggregated. We presented first and second round estimates, along with the arithmetic mean for the group's best, lower, and upper estimates to each participant in plot and table form for a final review. Participants were allowed to make fine-scale adjustments to their own estimates if desired; no participants did this.

2.4 | Statistical analysis

2.4.1 | Fisheries observer data

To obtain information on the turtle captures per trip, that is, capture per unit effort rates, for San Jose gillnet vessels against which to compare elicited estimates, we analyzed longitudinal panel data. These data were recorded by onboard observers operating in the inshore-midwater gillnet fleet from San Jose as part of a wider at-sea volunteer observer program run by our local not-for-profit collaborators (JAS, JCM). Captures per trip ($n = 461$) were averaged across seasonal (summer and winter) and annual periods ($n = 10$). Observed trips were across 32 different inshore-midwater gillnet vessels with varying vessel and net sizes. Historical vessel numbers for the inshore-midwater gillnet fleet were obtained from shore-based surveys (Alfaro-Shigueto et al., 2010; Escudero, 1997); for years with no known vessel size, an interpolated approximation was used (Supporting Information). Mean green turtle captures per trip/per season were then converted to mean captures per trip/per month within each season by averaging across each season's months (Supporting Information). Descriptive statistics are presented as mean, standard deviation (SD), and minimum and maximum 90% confidence intervals (CI).

Using the observer dataset ($n = 461$), we extrapolated green turtle capture rates from the proportion of the inshore-midwater fleet covered by observers to the wider gillnet fleet. We categorized vessel GRT into size classes, and then weighted these size classes using binomial logit Generalized Linear Mixed Models (GLMMs) using maximum likelihood estimation and AIC model selection criteria. GLMMs were constructed in R version 3.6.1 (R Core Team, 2019) using the `nlme` package (Pinheiro et al., 2012). Explanatory variables were selected a priori and included GRT, season, year, gillnet soak time (the time the net spends in the water), net length (km), and crew number as fixed effects. Vessel identification was included in the model as a random effect. We tested for fixed versus random effects using the Hausman test in

the **p1m** package in R, failing to reject the null hypothesis of random effects (against fixed effects; Croissant & Millo, 2008; Supporting Information). To avoid collinearity among variables in the model, Spearman's rho (r_s) correlation coefficients were calculated for pairs of variables (Akoglu, 2018). Any highly correlated variables ($r > .8$) would not be used together in the models. None of the variables selected a priori were correlated enough to warrant removal from the model (Supporting Information). After regressing sea turtle capture rates upon the independent variables, we tested for serial correlation and present serial correlation consistent standard errors. We then used the model's coefficients to weight the overall probability of capture of each turtle species by weight class (GRT) across the inshore-midwater gillnet fleet. We also modeled leatherback turtle capture rates; however, the low capture rate recorded ($n = 7$) resulted in little predictive power in the model (Supporting Information).

2.4.2 | Comparing data sources

The small sample size in our elicitation group precludes directly comparing the dataset to the capture rates calculated from the observer dataset using a large-sample test such as an independent two-sample *t*-test. Instead, we used a bootstrap method to simulate the expected distribution of monthly turtle capture rates calculated per season from the elicitation dataset and the observer dataset, and compare the two (Efron & Tibshirani, 1993). The bootstrap methodology (Supporting Information) consists of generating a null data set that has the same number of subjects as in the original data set by randomly selecting subjects from the control group with replacement and using the whole series of repeated measurements from each randomly selected control subject (Nadziejko, Chi Chen, Nádas, & Hwang, 2004). We tested the null hypothesis that, within each fishing season, the mean monthly number of green turtle captures in the San Jose inshore-midwater gillnet fleet calculated from the elicitation exercise is the same as the capture rate calculated from the observer data. All analysis was carried out using core packages in R version 3.6.1 (R Core Team, 2019).

2.4.3 | Performance-based metrics for elicitation estimates

Participants were not asked to define whether their best estimates represent a mean, mode, or median, nor were they asked to specify the quantiles of distribution (i.e., how the residual uncertainty their interval judgments were distributed outside of their bounds; Hemming, Walshe, et al.,

2018). Under more standard elicitation circumstances, mean, median, or mode data may be requested from respondents. In the current study, however, it was not deemed socially appropriate to ask gillnet skippers to specify these measures. We therefore chose metrics that are not based on continuous probability distributions. Instead, participants' performance was evaluated using three performance-based metrics: (a) accuracy of point (best) estimates, (b) calibration of interval judgments, and (c) informativeness of interval judgments (after McBride, Fidler, & Burgman, 2012; Hemming, Walshe, et al., 2018; Figure 1).

Accuracy of point estimates (Accuracy) is classified as the distance of the respondent's best estimate from the turtle capture rates calculated from the observer data (typically referred to as the realized truth; Einhorn, Hogarth, & Klemperer, 1977; Larrick & Soll, 2006). Accuracy was measured by calculating the average log-ratio

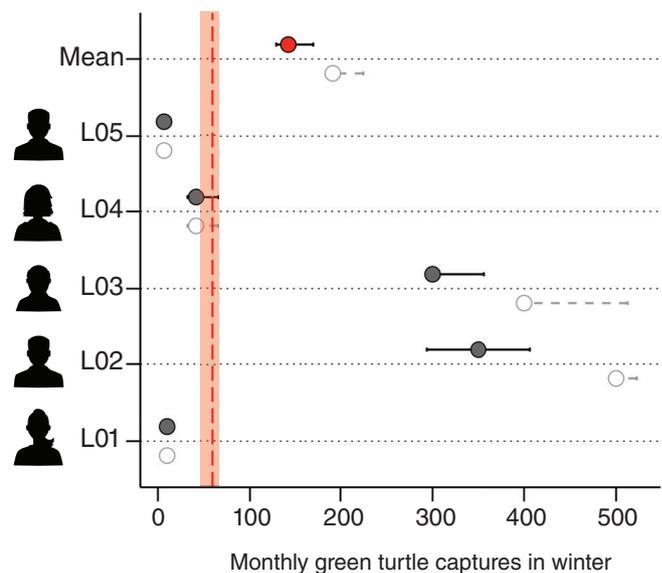


FIGURE 1 Respondents' elicitation estimates for monthly green turtle gillnet captures in winter that are used to explain the accuracy, calibration, and informativeness performance metrics. Participants present estimates (L01–L05) with Round 2 best estimates (grey circles) and associated credible intervals (horizontal lines). The group mean is represented by the red circle. The red dotted line represents the capture rates estimated from the observer data. Participants (L01, L05) are the most informative (smallest credible interval) and their informativeness intervals do not capture the realized truth (which if done over multiple questions would mean they are poorly calibrated). Participant (L02) is the least accurate (best estimate is furthest from the realized truth) and the least informative (largest credible interval). Participant (L04) has the most accurate estimate (closest best estimate to the realized truth), and their credible interval encompasses the realized truth (which if done over multiple questions will result in a good calibration score). Inspired by Hemming, Walshe, et al. (2018)

error (ALRE) for participants' judgments. To calculate ALRE, we first standardized each response by the range of responses for that question, known as range-coding (Hemming, Walshe, et al., 2018; McBride et al., 2012). Range-coding minimizes the effect that one or a few very divergent responses have on the accuracy measure (Burgman et al., 2011). Calibration of interval judgment (Calibration) measures the proportion of questions answered by a respondent for which their intervals capture the realized truth, with a score of 0.9 representing perfect calibration. The perfect calibration threshold is set at 0.9 because participants were asked to provide 90% credible intervals, therefore a participant would be considered perfectly calibrated if they capture the truth for 9 out of 10 questions answered. We used the standardized upper and lower values of participants' intervals and the standardized level of confidence associated with those intervals (Hemming, Hoffman, et al., 2018). Informativeness of interval judgment (Informativeness) measures the width (i.e., maximum minus minimum) of the participant's intervals relative to the total range provided by participants for a question (the highest maximum across all respondents, minus the lowest minimum across all respondents; Supporting Information). The performance-based metric analysis was undertaken in R using quantile aggregation code available on the open-science framework (Hemming, Hoffman, et al., 2018).

3 | RESULTS

Five respondents comprising three gillnet skippers and two not-for-profit employees participated in the elicitation procedure for the inshore-midwater fleet. The group comprised four males and one female. Respondent age was 27–50 years. Fishing experience for skippers was 11–17 years (Supporting Information).

3.1 | Elicited judgments for turtle captures

The group's green turtle confidence bounds were 129–227 individuals per month (Table 1). We used participants' monthly green turtle capture rates with gillnets to infer capture rates for the six-monthly summer (mean = 850, range = 771–1,022) and winter (mean = 1,234, range = 1,105–1,363) seasons. We then summed the seasonal estimates to obtain an annual capture rate (mean = 2,084, range = 1,876–2,385; Table 1). As a supplementary analysis, participants' judgment of leatherback capture was also explored (Supporting Information).

3.2 | Comparison of participant judgments with onboard observer data

We analyzed onboard observer records from the inshore-midwater gillnet fleet in San Jose from August 2007 to March 2019. Over 461 inshore-midwater fishing trips, observers recorded the capture of 379 turtles in gillnets. Species proportions were 86.8% green sea turtles ($n = 329$), 9.2% olive ridley turtles ($n = 35$), 1.8% leatherback sea turtles ($n = 7$), and 2.1% unidentified ($n = 8$). Of the 379 turtles captured, 62% were released alive without visible injury, 28% alive with minor injuries, and 8% were returned dead (Table 2). Observer coverage for the fleet is low, representing approximately 1–4% of net deployments over the 11-year, 7-month monitored period (Supporting Information). As observer deployments occur on a volunteer basis with skippers, sampling selection bias is likely. No vessels were observed in the 2010–2012 fishing years.

The most parsimonious model for green turtle capture included the variables GRT, season (winter and summer), fishing year, soak time, and a random effect for skipper-

TABLE 1 Extrapolated mean estimates of green turtle captures in San Jose inshore-midwater gillnets in summer and winter, between expert elicitation and at-sea observer datasets

Temp. Grp.	Expert elicitation data ($n = 5$)			Observer data ($n = 461$)		
	Mean best (B)	Std. lower 90 CI (l_{si})	Std. upper 90 CI (u_{si})	Mean	Min 90 CI	Max 90 CI
Monthly/winter	141.67	128.54	58.51	58.51	46.49	66.02
Monthly/summer	205.67	184.14	227.09	137.09	108.93	154.69
Total winter	850	771.25	1,022.12	351.07	278.96	396.14
Total summer	1,234	1,104.85	1,362.55	822.54	653.59	928.13
Annual	2,084	1,876.1	2,384.67	1,173.61	932.55	1,324.28

Notes: Values are based on elicited monthly estimates of the efficacy of the management strategy of gear switching from gillnets to potting or trolling, and onboard observer data obtained from the period August 2007–May 2019. Temp. Grp., temporal grouping; winter represents the cold weather months of June to November, summer represents the warm weather months of December to May.

TABLE 2 Turtle bycatches and captures per trip in gillnets of inshore-midwater vessels launching from San Jose in the period August 2007–May 2019, based on an onboard observer program, using trip as the unit of effort

Turtle species	<i>n</i>	Released without injury	Released injured	Dead	State unknown	Captures per trip (<i>n</i> = 461)			
						Mean	SD	Min 90% CI	Max 90% CI
Green	329	199	100	23	7	0.71	1.98	0.53	0.89
Leatherback	7	6	0	1	0	0.02	0.12	0.01	0.03
Olive ridley	35	24	6	5	0	0.08	0.46	0.04	0.12
Unidentified	8	4	2	0	2	0.02	0.21	0.00	0.04
Total	379	233	108	29	9	0.82	2.10	0.63	1.01

Abbreviation: CI, confidence interval.

Green turtle capture				
Reference	Random effects	Intercept	Residual	<i>n</i>
Vessel	SD	0.02837082	0.2556781	32
	Fixed effects	Coefficient	<i>SE</i> ^a	<i>p</i> -value
	Intercept	−0.014	0.026	.59
GRT				
Reference = 0 < 4 GRT	4 < 8 GRT	0.036	0.030	.2362
	8 < 12 GRT	0.050	0.026	.0646
	>12 GRT	0.104	0.083	.2193
Year				
Reference = Year (2007–2010)	Year (2011–2014)	0.153	0.022	.0000*
	Year (2015–2019)	0.012	0.014	.3862
Season				
Reference = season (winter)	Season (summer)	0.036	0.012	.0028*
	Soak time	0.001	0.001	.60

^aSerial correlation-consistent standard errors.

*Significant at the 0.05 confidence level.

vessel (Table 3). The skipper-vessel effect includes the effect of both the vessel and the skipper, the latter which cannot be measured or distinguished from the available data. There may also be a relationship between the skipper and vessel size. Larger vessels were more likely to capture turtles in a given trip than those with small capacities, after controlling for fishing effort. This may be a result of larger vessels having the ability to hold larger nets and stay at sea fishing for longer periods, as well as covering a larger fishing area because they can carry more petrol and oil, larger quantities of ice for their catch, and more supplies for the crew. Fishing across a larger fishing area may result in larger vessels having access to different fishing grounds where there are more turtles. Based on this model, we extrapolated the observer data to produce a mean annual gillnet capture estimate of 1,174 (range 933–1,324) green turtle individuals (Table 1).

TABLE 3 Coefficients of the best fit model for predicting probability of turtle capture

We ran two bootstrap hypothesis tests (each of 10,000 resamples with replacement) for the mean monthly estimates of green turtle gillnet captures within summer and winter fishing seasons. For both winter and summer, we found no statistically significant difference at the 95% confidence level in the mean monthly capture estimates of green turtle between the elicited data and the observed data (winter observed difference-in-means: 83.15, adj mean \pm SD = 42.39 \pm 32.59; *p* = .1177; summer observed difference-in-means: 68.58, adj mean \pm SD = 54.06 \pm 41.22; *p* = .309).

Participant L05 (not-for-profit) judged lower capture rates for green turtles than estimated from the observer data. Participant L04's (not-for-profit) judgment intervals encompassed the observer data for both seasonal estimates (Figure 2). In contrast, participants L02 and L03 (gillnet skippers) estimated significantly higher capture

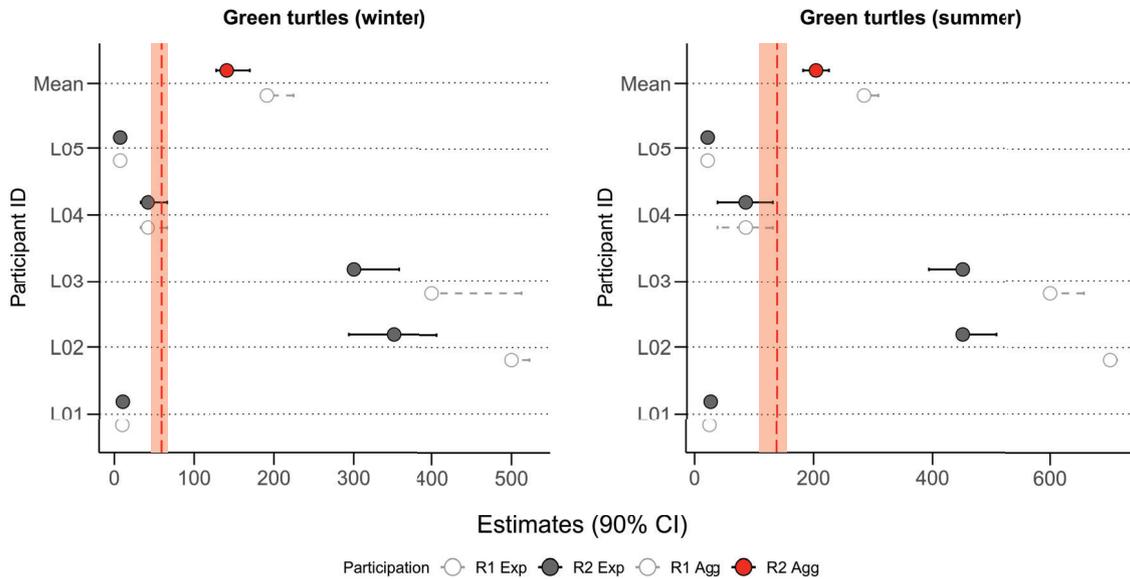


FIGURE 2 Respondents mean monthly estimates of green turtle captures in gillnets compared to extrapolated catch rates calculated from the observer data (red dotted line) with associated uncertainty bounds for the observer data (light red band). Elicited estimates are based on consideration of a possible gear switch from gillnets to trolling or lobster potting for all vessels in the San Jose inshore-midwater gillnet fleet. Monthly estimates were made for summer and winter fishing seasons. Experts assumed 100% compliance with the total gear switch scenario. Uncertainty bars have been adjusted to reflect 90% credible intervals for each expert’s response

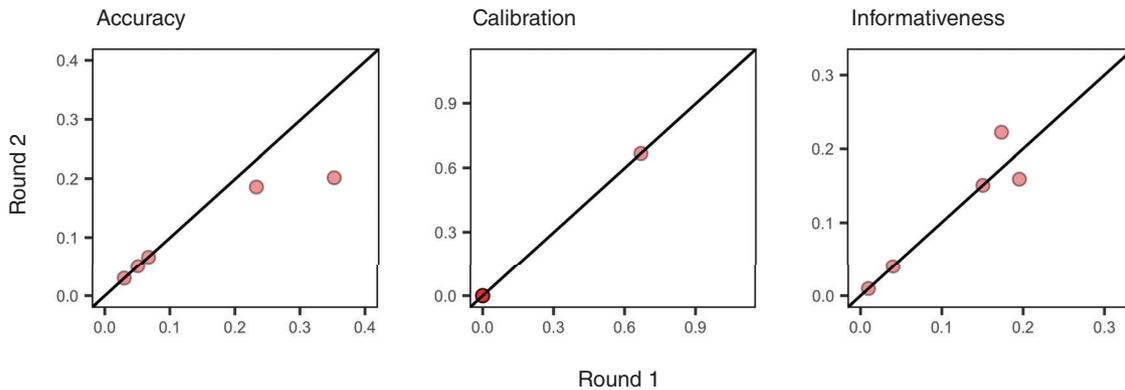


FIGURE 3 Scatterplots show the change in each individual’s estimates ($n = 5$) between Round 1 and Round 2, where they were assessing the total number of turtle captures across the inshore-midwater fleet (from the total gillnet ban scenario) using three performance variables (accuracy, calibration, and informativeness). If dots fall below the line in the “accuracy” or “informativeness” plots, individuals improved their scores on these measures. In the “Calibration” plot, dots above the line indicate individuals who increased the number of realized truths captured between their upper and lower bounds (a score of 0.9 represents perfect calibration)

rates across both winter and summer seasons. Participants L02 and L03 adjusted their estimates downwards between Round 1 and Round 2 in the modified Delphi method, to be closer to the value estimated from the observer data. This indicates that new information from the discussion between elicitation rounds influenced participant L02 and L03’s calibration and accuracy of judgment. Participant L01 (gillnet skipper) estimated closer to the realized truth and to the not-for-profit participants than the other two skippers (Figure 2).

3.3 | Performance metrics

Participant performance was evaluated by occupation groupings (skippers versus not-for-profit), comparing elicited estimates for the total gillnet ban to the capture rates calculated from the observer data (Figure 3). The not-for-profit employees were on average more accurate (lower ALRE score), better calibrated (their credible intervals encompassed the realized truth over more questions elicited), but less informative (they specified larger

credible intervals) than the skippers. The skippers scored higher on informativeness than the not-for-profit employees, but lower on accuracy. Two of the five participants improved the accuracy of their estimates between the two elicitation rounds, and one improved informativeness. Participants did not improve the calibration of their estimates between the elicitation rounds (there was no increase in the number of realized truths captured between their upper and lower bounds). This is potentially reflective of overconfidence or attitudes toward risk from the skippers, leading to them submitting estimates with tight confidence bounds (high informativeness) that underestimate uncertainty (low accuracy).

4 | DISCUSSION

Our estimates of green turtle captures in the San Jose inshore-midwater gillnet fleet, obtained from both the observer data and the group estimates from the expert elicitation, indicate detrimental bycatch rates for turtle populations like the endangered green turtle and the critically endangered East Pacific leatherback turtle population (assessed in Supporting Information) as both species are highly vulnerable to fishing pressure (Lutcavage, 2017; Spotila, Reina, Steyermark, Plotkin, & Paladino, 2000). While our elicited estimates focused on capture rates, the observer data found 7% of captured green turtles died and 38% were returned to sea injured, indicating the potential for a high percentage of estimated captures to result in mortality (Table 2). Green and leatherback turtles are far ranging and traverse multiple nations' waters in their lifetimes. The southeast Pacific waters that these species swim through (Bailey et al., 2012; Eckert, 2012) are fished by multiple small-scale fisheries where observer programs are limited or not currently established (Salas, Chuenpagdee, Seijo, & Charles, 2007; Sara, 2011). For example, questionnaire-based surveys estimated that small-scale fisheries-related turtle mortality across seven Ecuadorian harbors was 13,302 turtles per year (Alfaro-Shigueto et al., 2018). The IDEA protocol offers potential to improve data paucity on incidental capture and bycatch rates in data-limited fisheries such as those in the southeast Pacific by offering a decision-making process to more accurately quantify uncertainty and control for respondents' personal biases and heuristics.

The bootstrap hypothesis testing approach allowed us to compare the means of our two datasets despite small sample sizes. A high level of variation in total fleet size across observed fishing years meant that for observed years where no quantitative total fleet size estimates were available from shore-based surveys (Alfaro-Shigueto

et al., 2010; Escudero, 1997), we were required to use an interpolated approximation of fleet size. This uncertainty must be considered when interpreting the results. Despite the need to approximate fleet size, the methods used in the current study demonstrate that the IDEA protocol can provide broad estimates of protected species captures in small-scale fishery systems that are informative.

Both of the not-for-profit employees' judgments of green turtle captures were consistently closer to the observer data than the group mean. This finding contrasts with a number of Delphi-based elicitation studies that found that pooled group judgments consistently outperform individuals (Burgman, 2015; Burgman et al., 2011). These results may be reflective of two of the three gillnet skippers who were consistently overestimating when compared to the observer data. Due to the elicitation group's small sample size these estimations had a measurable effect on the pooled group means. Related to the small sample size, sample selection bias could also impact the results. Overestimation has been observed when gathering data from both small-scale fishers (O'Donnell, Pajaro, & Vincent, 2010) and scientific experts (Burgman et al., 2011; Oedekoven, Fleishman, Hamilton, Clark, & Schick, 2015). While the four-step elicitation method we employed is more likely to reduce overconfidence than three-point procedures (Speirs-Bridge et al., 2010), it is possible that an overconfident attitude towards risk influenced several of our fishers' judgments (Figure 2). One of the three gillnet skippers (participant L01) estimated closer to the observer data and not-for-profit employees than the other two gillnet skippers. Therefore, the overestimating gillnet skippers' judgments could also be reflective of their actual experience, given the spatially and temporally dynamic nature of turtle captures.

Participant L01 was the only skipper in the study who was a member of a protected species capture and bycatch reduction cooperative currently being trialed in San Jose by the not-for-profit conservation organization with which we were working. Exposure to conservation-oriented fishing practices aimed at reducing impact to sea turtles may have increased this fisher's awareness of fishing-related turtle mortality and contributed to this participant's estimates more accurately reflecting fleet-wide capture rates calculated from the observer data.

In addition to potential biases being present in the respondents' estimations, it is possible that inferences made when extrapolating the observed capture rate to the wider fleet using captures per trip weightings from the GLMM did not accurately approximate turtle captures across the fleet. For example, estimates could be biased or inaccurate due to the rarity of positive turtle capture events, which can be sensitive to extrapolation

from low percentage coverage rates because the data are often zero inflated (Babcock, Pikitch, & Hudson, 2003). The GLMM focused on the potential for a deployment effect (i.e., sample selection bias) as a result of a non-random assignment of observers on vessels within the inshore-midwater fleet. Observer programs in which participation is voluntary, such as our current case study, are often more prone to deployment biases than programs that require vessels to routinely take onboard observers when fishing licenses are issued and in which observers are randomly assigned (Borges, Zuur, Rogan, & Officer, 2004). GRT was selected a priori as a good variable to account for the potential deployment bias, which can arise due to difficulty in placing observers on the smallest vessels, varying range distributions of vessels that results in different spatial and temporal overlap with turtle species, time at sea, and weather. As expected, both green and leatherback turtle capture estimates increased slightly with the captures per trip weighted by GRT class, compared to a straight extrapolation of the captures per trip rate by month (Supporting Information). There is also the possibility of an observer effect that results from fishers changing their behavior as a result of an observer being present onboard (Liggins, Bradley, & Kennelly, 1997). Because this effect occurs at the vessel level it can be hard to detect, especially when modeling a small amount of observer data as in the current study. The presence of an observer onboard a vessel can cause skippers to fish away from their traditional sites, modify their fishing effort, operate their gear differently, retain catch that may have previously been discarded, or release bycatch that may have previously been retained. While observer effects have been found to be more distinct in fisheries with trip quotas (Gillis, Peterman, & Pikitch, 1995), few studies have attempted to disentangle deployment and observer effects on monitoring fishing trips. While our GLMM helps to account for potential non-random sample selection bias (Cotter & Pilling, 2007), any bias from an observer effect ultimately must be addressed during data collection rather than post hoc during data analysis (Benoît & Allard, 2009).

We successfully implemented the IDEA protocol in our case study fishery system. However, protocol adaptations were necessary due to the inshore-midwater gillnet skippers rarely overlapping with one another during the few days they spent on shore during our 3-month survey period. The methodological modification included holding two elicitation rounds facilitated through face-to-face interviews rather than a face-to-face group meeting or over email or web forum. Participants were provided with comprehensive comments and questions from the other participants both between Round 1 and Round 2, and after Round 2, on printed paper in their native

language (Spanish) and they then discussed these verbally with the facilitator (BIE). Continual discussion about specific questions was restricted as a result of the modified format. In addition, the gillnet skippers interviewed were not comfortable writing their responses, preferring to have the questions read aloud, followed by discussion of potential misinterpretation, verbally noting their answer, and then asking the facilitator to record their response. This may be due to some of the gillnet skippers in our case study fishery having difficulty reading and writing. Scenarios preventing group meetings can be numerous in the field and while far less than ideal, the notes and facilitator were able to assist in clarifying uncertainties or misinterpretations held by respondents. In addition, requests were made to record interviews. Respondents were also encouraged by the facilitator to provide comprehensive explanations for their reasoning behind each estimate as detailed explanations support other respondents in understanding the knowledge and rationale behind each respondent's estimate, and therefore help to better weigh each estimate against their own (Hemming, Burgman, et al., 2018). While the IDEA protocol is simple to understand and we were able to undertake it in this individualized way with resource users in our case study system, further investigation into possible local resource user-specific adaptations to modern elicitation protocols would be a beneficial area of future research.

This research has applied the IDEA protocol in a new context of conservation research and natural resource management, to estimate the total number of green turtles captured in a small-scale gillnet fishery and compare these estimates to capture rates calculated from observer data obtained from the same fleet. Our analysis reveals high green turtle capture rates in the San Jose inshore-midwater gillnet fleet. We demonstrate that the IDEA protocol can be implemented to quantify uncertainty and control for personal biases and heuristics when interviewing respondents in small-scale fishing systems, and highlight that both observer data and elicitation estimates are approximations of an unknown truth. While the IDEA protocol was implemented successfully, minor methodological modifications were necessary to obtain participants' judgments. Future research could investigate how best to adapt the protocol to a range of local resource user contexts. Furthermore, comparing elicitation estimates to an observed value obtained from an observer program provided informative data on participant performance when combined with a bootstrap hypothesis testing of means analysis. We encourage researchers and practitioners implementing elicitation studies with local resource users to draw on multiple sources of comparable data.

ACKNOWLEDGMENTS

We would like to thank Stephanie Brittain for her advice. We thank the US National Ocean and Atmospheric Administration, National Marine Fisheries Service, Southwest Fisheries Science Center, and Pew Charitable Trusts through a Pew Marine Fellowship to E.J.M.-G., for supporting this research. W.N.S.A. was supported by a PhD Commonwealth Scholarship from the Commonwealth Scholarship Commission in the U.K. and the University of Oxford (PhD scholarship NZCR-2015-174). J.A.S. and J.C.M. were supported by the National Fish and Wildlife Foundation (550011).

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The survey data in anonymized form and R code underlying this research can be found at <https://osf.io/58sf4/>.

ORCID

William N. S. Arlidge  <https://orcid.org/0000-0002-1807-4150>

REFERENCES

- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18, 91–93.
- Alfaro-Shigueto, J., Dutton, P. H., Van Bressema, M., & Mangel, J. (2007). Interactions between leatherback turtles and Peruvian artisanal fisheries. *Chelonian Conservation and Biology*, 6, 129–134.
- Alfaro-Shigueto, J., Mangel, J. C., Bernedo, F., Dutton, P. H., Seminoff, J. A., & Godley, B. J. (2011). Small-scale fisheries of Peru: A major sink for marine turtles in the Pacific. *Journal of Applied Ecology*, 48, 1432–1440.
- Alfaro-Shigueto, J., Mangel, J. C., Darquea, J., Donoso, M., Baquero, A., Doherty, P. D., & Godley, B. J. (2018). Untangling the impacts of nets in the southeastern Pacific: Rapid assessment of marine turtle bycatch to set conservation priorities in small-scale fisheries. *Fisheries Research*, 206, 185–192.
- Alfaro-Shigueto, J., Mangel, J. C., Pajuelo, M., Dutton, P. H., Seminoff, J. A., & Godley, B. J. (2010). Where small can have a large impact: Structure and characterization of small-scale fisheries in Peru. *Fisheries Research*, 106, 8–17.
- Ames, R. T., Leaman, B. M., & Ames, K. L. (2007). Evaluation of video Technology for Monitoring of multispecies longline catches. *North American Journal of Fisheries Management*, 27, 955–964.
- Arlidge, W. N. S., Squires, D., Alfaro-Shigueto, J., Booth, H., Mangel, J., & Milner-Gulland, E. J. (2020). A mitigation hierarchy approach for managing sea turtle captures in small-scale fisheries. *Frontiers in Marine Science*. <https://doi.org/10.3389/fmars.2020.00049>
- Babcock, E. A., Pikitch, E. K., & Hudson, C. G. (2003). *How much observer coverage is enough to adequately estimate bycatch?* Miami, FL: Pew Institute of Ocean Science.
- Bailey, H., Benson, S. R., Shillinger, G. L., Bograd, S. J., Dutton, P. H., Eckert, S. A., ... Foley, D. G. (2012). Leatherback turtle movement patterns. *The Bulletin of the Ecological Society of America*, 93, 165–169.
- Bartholomew, D. C., Mangel, J. C., Alfaro-Shigueto, J., Pingo, S., Jimenez, A., & Godley, B. J. (2018). Remote electronic monitoring as a potential alternative to on-board observers in small-scale fisheries. *Biological Conservation*, 219, 35–45.
- Benoit, H. P., & Allard, J. (2009). Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? *Canadian Journal of Fisheries and Aquatic Sciences*, 66, 2025–2039.
- Borges, L., Zuur, A. F., Rogan, E., & Officer, R. (2004). Optimum sampling levels in discard sampling programs. *Canadian Journal of Fisheries and Aquatic Sciences*, 61, 1918–1928.
- Burgman, M. A. (2015). *Trusting Judgements: How to get the best out of experts*. Cambridge: Cambridge University Press.
- Burgman, M. A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., ... Twardy, C. J. P. O. (2011). Expert status and performance. *PLoS One*, 6, e22998.
- Castillo, G., Fernandez, J., Medina, A., & Guevara-Carrasco, R. (2018). Third structural survey of the artisanal fishery in the Peruvian littoral. *Informe IMARPE*, 45, 299–388.
- Cook, C. N., Hockings, M., & Carter, R. (2010). Conservation in the dark? The information used to support management decisions. *Frontiers in Ecology and the Environment*, 8, 181–186.
- Cooke, R. M. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. New York, NY: Oxford University Press.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Cotter, A., & Pilling, G. (2007). Landings, logbooks and observer surveys: Improving the protocols for sampling commercial fisheries. *Fish and Fisheries*, 8, 123–152.
- Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27, 1–43.
- Currey, R., Boren, L., Sharp, B., & Peterson, D. (2012). *A risk assessment of threats to Maui's dolphins*. Wellington: New Zealand Ministry for Primary Industries and Department of Conservation.
- Dietz, T., Ostrom, E., & Stern, P. C. (2003). The struggle to govern the commons. *Science*, 302, 1907–1912.
- Drew, J. A. (2005). Use of traditional ecological knowledge in marine conservation. *Conservation Biology*, 19, 1286–1293.
- Eckert, K. L. (2012). *Synopsis of the biological data on the leatherback sea turtle (Dermochelys coriacea)*. Washington, DC: U.S. Department of Interior, Fish and Wildlife Service, Biological Technical Publication BTP-R4015-2012.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.
- Einhorn, H. J., Hogarth, R. M., & Klempner, E. (1977). Quality of group judgment. *Psychological Bulletin*, 84, 158–172.
- Escudero, H. L. (1997). Encuesta estructural de la pesquería artesanal del litoral peruano Progressive Report, IMARPE.
- FAO. (2018). *The state of world fisheries and aquaculture 2018—Meeting the sustainable development goals*. Rome: FAO.
- Gillis, D. M., Peterman, R. M., & Pikitch, E. K. (1995). Implications of trip regulations for high-grading: a model of the behavior of

- fishermen. *Canadian Journal of Fisheries and Aquatic Sciences*, *52*, 402–415.
- Gilman, E., Legorburu, G., Fedoruk, A., Heberer, C., Zimring, M., & Barkai, A. (2019). Increasing the functionalities and accuracy of fisheries electronic monitoring systems. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*, 901–926.
- Goetz, S., Read, F. L., Santos, M. B., Pita, C., & Pierce, G. J. (2013). Cetacean–fishery interactions in Galicia (NW Spain): Results and management implications of a face-to-face interview survey of local fishers. *ICES Journal of Marine Science*, *71*, 604–617.
- Guevara-Carrasco, R., & Bertrand, A. (2017). Atlas de la pesca artesanal del mar del Perú. Edición IMARPE-IRD:183.
- Hall, M. A. (1996). On bycatches. *Reviews in Fish Biology and Fisheries*, *6*, 319–352.
- Hanea, A., McBride, M., Burgman, M., & Wintle, B. (2016). Classical meets modern in the IDEA protocol for structured expert judgement. *Journal of Risk Research*, *21*, 417–433.
- Hanea, A., McBride, M., Burgman, M., Wintle, B., Fidler, F., Flander, L., ... Mascaro, S. (2016). Investigate Discuss Estimate aggregate for structured expert judgement. *International Journal of Forecasting*, *33*, 267–279.
- Helmer-Hirschberg, O., Brown, B. B., & Gordon, T. J. (1966). *Social technology*. New York: Basic Books.
- Hemming, V., Burgman, M. A., Hanea, A. M., McBride, M. F., & Wintle, B. C. (2018). A practical guide to structured expert elicitation using the IDEA protocol. *Methods in Ecology and Evolution*, *9*, 169–180.
- Hemming, V., Hoffman, M., & Lane, S. (2018). RMarkdown Code_Quantile Aggregation. Retrieved from osf.io/t39wv.
- Hemming, V., Walshe, T. V., Hanea, A. M., Fidler, F., & Burgman, M. A. (2018). Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. *PLoS One*, *13*, e0198468.
- Khalil, C. A., Conforti, P., Ergin, I., & Gennari, P. (2017). *Defining small scale food producers to monitor target 2.3 of the 2030 agenda for sustainable development*. Rome: FAO.
- Klein, C. J., Beher, J., Chaloupka, M., Hamann, M., Limpus, C., & Possingham, H. P. (2017). Prioritization of marine turtle management projects: A protocol that accounts for threats to different life history stages. *Conservation Letters*, *10*, 547–554.
- Kynn, M. (2008). The ‘heuristics and biases’ bias in expert elicitation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*, 239–264.
- Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*, 111–127.
- Liggins, G. W., Bradley, M. J., & Kennelly, S. J. (1997). Detection of bias in observer-based estimates of retained and discarded catches from a multi species trawl fishery. *Fisheries Research*, *32*, 133–147.
- Lutcavage, M. E. (2017). *Human impacts on sea turtle survival* (Vol. I, pp. 387–409). Boca Raton, FL: The biology of sea turtles, CRC press.
- 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.
- McBride, M. F., Fidler, F., & Burgman, M. A. J. D. (2012). Evaluating the accuracy and calibration of expert predictions under uncertainty: Predicting the outcomes of ecological research. *Biodiversity Research*, *18*, 782–794.
- Milner-Gulland, E. J., Garcia, S., Arlidge, W., Bull, J., Charles, A., Dagorn, L., ... Squires, D. (2018). Translating the terrestrial mitigation hierarchy to marine megafauna bycatch. *Fish and Fisheries*, *19*, 547–561.
- Nadziejko, C., Chi Chen, L., Nádas, A., & Hwang, J. S. (2004). The ‘fishing license’ method for analysing the time course of effects in repeated measurements. *Statistics in Medicine*, *23*, 1399–1411.
- Needle, C. L., Dinsdale, R., Buch, T. B., Catarino, R. M. D., Drewery, J., & Butler, N. (2014). Scottish science applications of remote electronic monitoring. *ICES Journal of Marine Science*, *72*, 1214–1229.
- O'Donnell, K., Pajaro, M., & Vincent, A. (2010). How does the accuracy of fisher knowledge affect seahorse conservation status? *Animal Conservation*, *13*, 526–533.
- 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*. Chichester: John Wiley & Sons.
- Oedekoven, C., Fleishman, E., Hamilton, P., Clark, J. S., & Schick, R. S. (2015). Expert elicitation of seasonal abundance of North Atlantic right whales *Eubalaena glacialis* in the mid-Atlantic. *Endangered Species Research*, *29*, 51–58.
- Ortiz, N., Mangel, J. C., Wang, J., Alfaro-Shigueto, J., Pingo, S., Jimenez, A., ... Godley, B. (2016). Reducing green turtle bycatch in small-scale fisheries using illuminated gillnets: The Cost of Saving a Sea Turtle. *Marine Ecology Progress Series*, *545*, 251–259.
- Peckham, S. H., Diaz, D. M., Walli, A., Ruiz, G., Crowder, L. B., & Nichols, W. J. (2007). Small-scale fisheries bycatch jeopardizes endangered Pacific loggerhead turtles. *PLoS One*, *2*, e1041.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2012). *nlme: Linear and nonlinear mixed effects models. R package version 3*. Vienna, Austria: R Foundation for Statistical Computing.
- Riskas, K. A., Tobin, R. C., Fuentes, M. M., & Hamann, M. J. B. C. (2018). Evaluating the threat of IUU fishing to sea turtles in the Indian Ocean and Southeast Asia using expert elicitation. *Biological Conservation*, *217*, 232–239.
- Salas, S., Chuenpagdee, R., Seijo, J. C., & Charles, A. (2007). Challenges in the assessment and management of small-scale fisheries in Latin America and the Caribbean. *Fisheries Research*, *87*, 5–16.
- Sara, S. K. (2011). *Bycatch and foraging ecology of sea turtles in the eastern Pacific*. Durham, NC: Duke University.
- Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis: An International Journal*, *30*, 512–523.
- Spotila, J. R., Reina, R. D., Steyermark, A. C., Plotkin, P. T., & Paladino, F. V. (2000). Pacific leatherback turtles face extinction. *Nature*, *405*, 529–530.
- Suuronen, P., & Gilman, E. (2019). Monitoring and managing fisheries discards: New technologies and approaches. *Marine Policy*, 103554. <https://doi.org/10.1016/j.marpol.2019.103554>

- van Gelder, T., Vodicka, R., & Armstrong, N. (2016). Augmenting expert elicitation with structured visual deliberation. *Asia & the Pacific Policy Studies*, 3, 378–388.
- Williams, J. L., Pierce, S. J., Hamann, M., & Fuentes, M. M. (2017). Using expert opinion to identify and determine the relative impact of threats to sea turtles in Mozambique. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29, 1936–1948.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Arlidge WNS, Alfaro-Shigueto J, Ibanez-Erquiaga B, Mangel JC, Squires D, Milner-Gulland EJ. Evaluating elicited judgments of turtle captures for data-limited fisheries management. *Conservation Science and Practice*. 2020;e181. <https://doi.org/10.1111/csp2.181>