Received: 21 July 2021 | Revised: 28 December 2021 | Accepted: 28 January 2022

DOI: 10.1111/faf.12646

GHOTI

FISH and FISHERIES WILEY

Tracing intuitive judgement of experts in fish stock assessment data

Esther Schuch[1,2] | Andries Richter[1,3]

[1]Environmental Economics and Natural Resources group, Wageningen University, Wageningen, The Netherlands

[2]Institute for Advanced Sustainability Studies (IASS), Potsdam, Germany

[3]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway

Correspondence
Esther Schuch, Environmental Economics and Natural Resources group, Wageningen University, Hollandseweg 1, 6706 KN Wageningen, The Netherlands.
Email: esther.schuch@iass-potsdam.de

Funding information
H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 675997

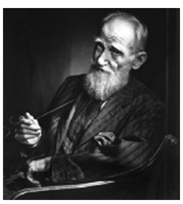
Abstract

Scientific advice is increasingly used to inform policy. Often, experts are asked to give advice when stakes are high, time pressure is severe and uncertainty looms. In such environments, decisions may be guided by instincts and priors, rather than reason. Yet, the extent of these intuitive judgements is unknown. We use a database of fish stock assessments to detect and quantify the systematic tendency to put too much weight on previous information, known as anchoring, in scientific advice. By exploiting exogenous variation in procedures and possibilities to vary model assumptions, we find consistent evidence for intuitive judgement. We find that anchoring is strongest if model choices are flexible and the fish stock is in crisis, potentially increasing pressure and stakes. By providing advice that is biased towards previous results, the stock assessments may be more robust but may also give a false sense of security as more drastic changes may go undetected.

KEYWORDS
anchoring, behavioural bias, expert advice, panel data, policy advice

1 | INTRODUCTION

Many governments strive for evidence-based policy, which typically relies heavily on the involvement of experts. Especially when stakes and uncertainty are high, and data are scarce, the objectivity of expert knowledge is sought after. Yet, while experts have a vast amount of knowledge and experience, their advice is often based on intuitive judgement (Burgman, 2015; Sutherland & Burgman, 2015). After all, experts tend to, consciously or unconsciously, interpret data. How data are perceived or presented depends not only on what is at stake, but also on stress levels, or concerns about reputation (Burgman, 2015; Englich & Soder, 2009; Martin et al., 2012; Sutherland & Burgman, 2015). Intuitive judgements are often shaped by cognitive biases. While experts' advice is ubiquitous in public policy, fisheries management relies particularly strongly on experts, and therefore, cognitive biases are expected to play a large role (Fulton,



Ghoti papers

Ghoti aims to serve as a forum for stimulating and pertinent ideas. Ghoti publishes succinct commentary and opinion that addresses important areas in fish and fisheries science. Ghoti contributions will be innovative and have a perspective that may lead to fresh and productive insight of concepts, issues and research agendas. All Ghoti contributions will be selected by the editors and peer reviewed.

Etymology of Ghoti

George Bernard Shaw (1856–1950), polymath, playwright, Nobel prize winner, and the most prolific letter writer in history, was an advocate of English spelling reform. He was reportedly fond of pointing out its absurdities by proving that 'fish' could be spelt 'ghoti'. That is: 'gh' as in 'rough', 'o' as in 'women' and 'ti' as in palatial.

2021). Out of the vast number of cognitive biases, anchoring, the systematic tendency to put too much weight on prior information, has been widely observed across experts and non-experts (Chrysafi et al., 2019; Englich & Soder, 2009; Furnham & Boo, 2011; Martin et al., 2012; McBride et al., 2012; Montibeller & Winterfeldt, 2015; Oskamp, 1965; Sinkey, 2015; Tversky & Kahneman, 1981).

1.1 | Anchoring in expert advice

Experts may be particularly prone to anchoring if new advice would align poorly with previous advice since such deviations may affect the reputation of being knowledgeable and trustworthy. Adjusting previous results can be perceived as inconsistent, which may damage the reputation of the expert (Ottaviani & Sorensen, 2006; Sinkey, 2015), creating incentives to avoid these 'reputational costs' (Hagafors, 1983; Kirchgassner & Muller, 2006; Nordhaus, 1987). As revising advice would appear inconsistent, new conflicting data may not be taken with the same weight as data that are in line with prior information. As a result, experts often face a trade-off when adjusting previous assessments between running the risk of being perceived as 'flip-flopping' and scientific accuracy. The reluctance to adjust estimates seems to be even stronger if these results, as well as their deviations from the previous ones, attract high publicity (Kirchgassner & Muller, 2006) or the experts are in a stressful situation (Kassam et al., 2009; Starcke & Brand, 2012). For example, during the global Covid-19 pandemic, experts advising national governments have entered celebrity status and kitchen table conversations (Kupferschmidt, 2020). While anchoring (the systematic tendency to put too much weight on prior information) may play a large role in affecting decisions provided by experts, it has to our knowledge never been documented with observational field data. Whilst the role of cognitive biases in fisheries management has been suggested to play a role (Fulton, 2021), we are using European fish stock assessments to test whether anchoring plays a role in expert judgement.

1.2 | Fish stock assessments

Natural resource management is riddled with uncertainties, which gives experts, and their in-depth knowledge, a special role (Martin et al., 2012). Especially in fisheries management, there are huge uncertainties in regard to quantity, distribution, as well as growth rate of fish stocks, which can only be handled via expert knowledge. The International Council for the Exploration of the Sea (ICES) is an international organisation responsible for providing scientific advice to governments, multinational organisations and NGOs to manage marine living resources (ICES, 2016). An important advisory product is recurrent stock assessments that advise on status of the stock, biomass levels, and evaluate exploitation relative to management objective. While ecosystem-based management is increasingly considered, in practice, advice on the single stock level remains the main input for the policy makers for setting a total allowable catch (TAC) for a fish stock in a given year. This advice is based on output from computational models, which are analysed by fisheries scientists, who are often employed by national institutes and universities. The assessment is carried out by working groups during annual meetings following specific procedures that are often slightly mysterious to outsiders, as illustrated by Hilborn (1992): 'Each year in holy sites around the world, huddled in dark caverns in cold, inaccessible and undesirable places with names such as Lowestoft, St Johns and Reykjavik, members of a special priesthood gather for their annual rites that affect the lives of millions [...]. To the uninitiated, the chants of the priests are incomprehensible, but emerging from these holy gatherings are prophesies that are cast down as law [...]: this year's TAC'. The complexity of the process pose challenges in understanding how decisions are made for those who are not privy to the meetings. Each year, the TAC is calculated together with an estimated historical time series of spawning stock biomass (SSB). This allows for the comparison between time series of SSB that were produced in different assessment years (Figure 1).

Determining the TACs is a cornerstone of fishing policy in the European Union. Hence, these assessments have been scrutinised to avoid mistakes, resulting in an extensive literature on biases in fish stock assessments. The main biases analysed are all in the realm of natural science and related to uncertainty in recruitment (Deroba & Miller, 2016; Francis, 2016; Kehler et al., 2002; Lee et al., 2012; Marshall et al., 2006; Methot et al., 2011; Myers & Barrowman, 1995; Walters, 2004), the estimation of natural fishing mortality and fishing pressure (Dickey-Collas et al., 2007; Johnson et al., 2014), the

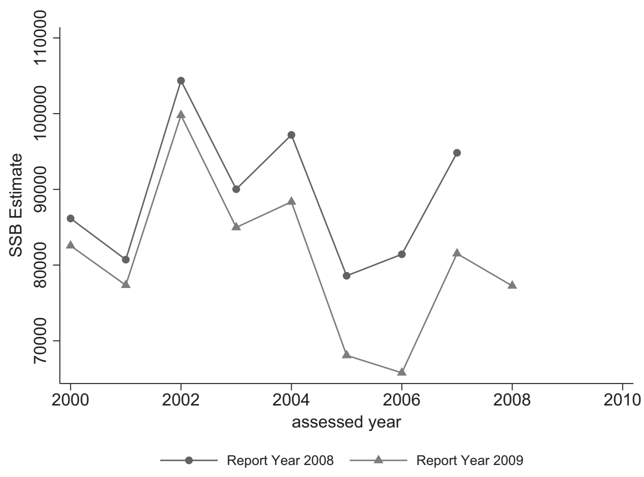


FIGURE 1 The estimates for spawning stock biomass (SSB) of Atlantic herring (*Clupea harengus*, Clupeidae) in the Gulf of Riga in the report years 2008 and 2009. Report years are the years when the assessment was done. Assessed years are all the years for which a SSB estimate is provided. The assessed year 2007 provides an example for the calculation of the relative change between the report years 2008 and 2009. V_{ijt} measures the change in biomass estimate between report years per assessed year, where $V_{ijt} = |\frac{(SSB_{ijt} - SSB_{ij,t-1})}{SSB_{ij,t-1}}|$.

use and quality of data (Ichinokawa et al., 2014; Kraak et al., 2009; Payne et al., 2009), as well as biases related to spatial distribution (Fallon et al., 2015; Soria et al., 1996). Yet, while the human factor of stock assessments is increasingly put on the agenda (Chrysafi et al., 2019; Dankel et al., 2015; Fulton et al., 2011; Wilson, 2010), the role of experts in fisheries stock assessments has only recently been touched upon (Fulton, 2021).

1.3 | Experts in fish stock assessments

Under high time pressure—the meetings have fixed lengths between five and ten days—an advice report is produced. This leads to tension, illustrated by a statement by a stock assessment scientist: 'It's not really science anymore. We're number engineers. We fiddle with numbers to, you know, try to add some scientific credibility to an opinion' (Wilson, 2010). Time pressure is high and so are the stakes. The experts are fully aware that their advice is used for policy making, and that the results generated have real consequences on fishing quota (Wilson, 2010). The current setting, in which scientists are expected to determine the exact amount of fish that can be fished safely, leads to discomfort amongst the scientists. While they are required to produce a TAC for policy makers, they are acutely aware of the underlying uncertainties (Kraak et al., 2010).

In 2004, ICES introduced a new procedure called 'benchmarking' to 'improve assessment quality and enhance credibility' (ICES, 2013). From this point on, assessments were categorised as either 'update' or 'benchmark'. Benchmarks are done in regular intervals (3–5 years) in which the assessment methodology (including the model and its specifications) are pre-agreed upon. In the years between those benchmarks, the assessments are called updates. For updates, the protocol is less strict since the model is not supposed to be revised, and the data only have to be fed into the model to produce the advice. The first benchmark was done in 2004 and was applied to more stocks over the years.

Fisheries management is highly complex and path-dependent (Fulton, 2021). As a result, not every benchmark implies a change in model, and occasionally a model is changed even though the assessment is no benchmark. The benchmark is conducted by stock assessment experts as well as external experts to ensure that the assessment is based on the 'best available method' (ICES, 2013). Very much like a study protocol or pre-analysis plan (Coffman & Niederle, 2015), benchmarks also have the function to specify the methods ex ante to limit the freedom of researchers to 'fiddle' with model input to produce results that align with prior beliefs of the scientists. However, in practice, there are sometimes diversions from the protocol. It may be necessary that experts make judgement calls, e.g. on how to integrate new data into the model. One example for experts' discretion is the weighting of data sets, often referred to as 'adequate weighting' or 'down-weighting', which has been flagged as highly dependent on expert judgement (Francis, 2011). This is of particular importance if two data sources (e.g. scientific surveys and catches) do not point in the same direction, since the experts weigh the data sources according to what they deem to be more plausible.

Since 'truth' is an elusive construct in fisheries science, due to the inherent uncertainty (Palmer & Demarest, 2018), experts always run the risk of getting it wrong after all. Mistakenly adjusting the biomass levels downwards implies that the stock is in reality in better shape than suggested by the advice, and there is a risk of catching less than what would be sustainable, leading to economic losses. However, mistakenly making an upward adjustment is most likely even costlier since it can endanger the sustainability of the stock, putting pressure on the experts.

1.4 | Hypotheses

To assess whether expert judgements are impacted by anchoring, we focus on three hypotheses. In general, we assume that anchoring is stronger than the more freedom experts have to apply judgement calls. To test this, we focus on the difference between updates and benchmarks. Furthermore, we analyse whether we can observe an anchoring effect when data weighting is applied. Here, we focus on whether we see differences between assessments based on the availability of one or two data sources. Lastly, we hypothesise that high stakes lead to stronger anchoring effects. In fisheries management, the stakes are particularly high if a fish stock is considered to be overfished already.

To sum up, the hypotheses to be tested are

Hypotheses 1 'Room to Fiddle'—Updates show stronger anchoring effects (estimates close to the previous ones) than benchmarks.

Hypotheses 2 'Adequate Weighting'—Using two data sources lead to stronger anchoring effects than using only one source.

Hypotheses 3 'High Stakes'—Assessments of endangered stocks show stronger anchoring effects than stocks that are not in a critical state.

2 | MATERIALS AND METHODS

2.1 | Data

The analysis was conducted with a database containing information on the stock assessments done by ICES. The data were compiled by Pastoors (2020) and include the published stock assessments by ICES. These assessments not only provide a spawning stock biomass (SSB) estimate for the most recent year but also a time series for previous years based on the current stock assessment model. The lengths of the time series per fish stock differ, ranging from one to 57 years. These assessments also provide information on the model, reference points and data used. Pastoors (2020) combined all assessments published by ICES between the years 1988 and 2015. We refer to the year the assessment is conducted, as 'report year' and 'assessed year' refer to the years for which the estimates were provided. The original database covers 164 fish stocks. Keeping only the assessments for which the assessment model is known, we are left with 102 stocks. Further data cleaning was necessary as not every assessment provides absolute SSB estimates and fishing pressure F. We drop all cases that do not have both (absolute SSB and F) available. Lastly, we drop all the cases for which we have less than three assessments per fish stock. This leaves us with an unbalanced panel (missing observations for some years) of 71 fish stocks and 25591 SSB estimates (see Table 1 for overview).

We also have data on stock-specific reference points such as MSYbtrigger, B_{lim}, F_{MSY} and F_{lim}. MSYbtrigger refers to the SSB level that, when reached, triggers the advice to adjust fishing pressure to be in line with maximum sustainable yield (MSY) in the long run. Fishing at F_{MSY} is one of the management objectives in the European Union and thus covers the majority of stocks (51 out of 71 stocks). Fishing at F_{lim} will lead to the biomass level of B_{lim}, which is the minimum size of a fish stock to ensure reproductive capacity. Thus, fishing above F_{lim} endangers the fish stock. For 49 out of 71 stock, B_{lim} is reported. ICES introduced a buffer to avoid reaching B_{lim}, namely the biomass precautionary reference point B_{PA}. If the biomass drops below B_{PA}, management action should be taken to reduce fishing pressure to avoid reaching B_{lim} (Lassen et al., 2013). Reference points are fairly static values, but they are sometimes updated (see Table A1 for summary statistics).

TABLE 1 Overview of data used for the analysis

Variable		
Number of stocks	71	
Report year	1988–2015	
Assessed year	1946–2013	
Number of assessments	777	
Number of updates	210	
Number of benchmarks	66	
Number of unclassified assessments	501	
Assessments based only on catches	93	
Assessments based only on survey	25	
Assessments based on catches and survey	114	
SSB estimates	25591	
SSB estimates based on catches and/or survey data	7062	
Per fish stock	**Min**	**Max**
Number of report years/assessments	3	23
Number of benchmarks	0	2
Number of updates	0	10
Number of unclassified assessments	1	17
Backdated time series lengths per assessment	4	56
Number of assessments based on catches and/or survey data	0	14
Number of SSB estimates	68	853

For each report year, we have one biomass estimate for each assessed year, which allows us to see how much the stock assessment has changed from one report year to another (see Figure 1). We are interested in the interannual change of estimates depending on the report year. First, we measure the magnitude of the interannual change by converting the differences between estimates into relative changes. We use the absolute value of relative changes to measure the deviation between report years independent of the direction of the changes. The relative change V_{ijt} is defined as

$$V_{ijt} = \left| \frac{(SSB_{ijt} - SSB_{ij,t-1})}{SSB_{ij,t-1}} \right| \tag{1}$$

with t being the report year, j the assessed year and i the stock assessed (Figure 1). Second, apart from the size of the interannual change, we are also controlling for the direction of change since we hypothesise that an upward shift is stronger influenced by anchoring than a downward shift. Figure 2 illustrates that not every fish stock was assessed every year. Naturally, the number of assessed years increases with more recent report years.

2.2 | Variables

To analyse the effects of different assessments, we use indicator variables for update (one for update and zero otherwise) and one for benchmarking (one for benchmark and zero otherwise). We also have an indicator variable, which takes the value one in all years after the stock has been benchmarked for the first time. To account for model changes, we use an indicator variable that takes the value one if a model change occurred. We use two indicator variables to control for how data were collected. We focus on two data sources, namely catches and scientific surveys. Due to the inconsistent reporting of the data sources, we combine all the available data on discards, catches and landings in one variable named 'catches', as it is not transparent when catches actually include discards and landings. The indicator variable for catches takes the value one if catches were part of the data source and the one for surveys takes the value one if surveys were part of the data source. We use interaction effects to determine the impact on interannual change when both data sources are used.

Regarding the health of a fishery, we focus on fishing pressure and biomass in relation with reference points (Methot Jr., 2015; Ricard et al., 2011). While we have different fishing pressure reference points in the database, we use maximum sustainable yield (F_{MSY}). Since we do not have F_{MSY} for all years of each stock and F_{MSY} is more or less static, we use the average F_{MSY} per fish stock. Relative mortality is defined as fishing pressure divided by average F_{MSY}. We use a continuous variable which indicates that the higher is a ratio, the higher is the pressure on the fish stock. A value above 1 indicates overfishing (Hilborn & Stokes, 2010). With regard to biomass status, we use B_{lim} as reference point since it is the most severe reference point. As in the case of F_{MSY}, we use the average of B_{lim} for each of

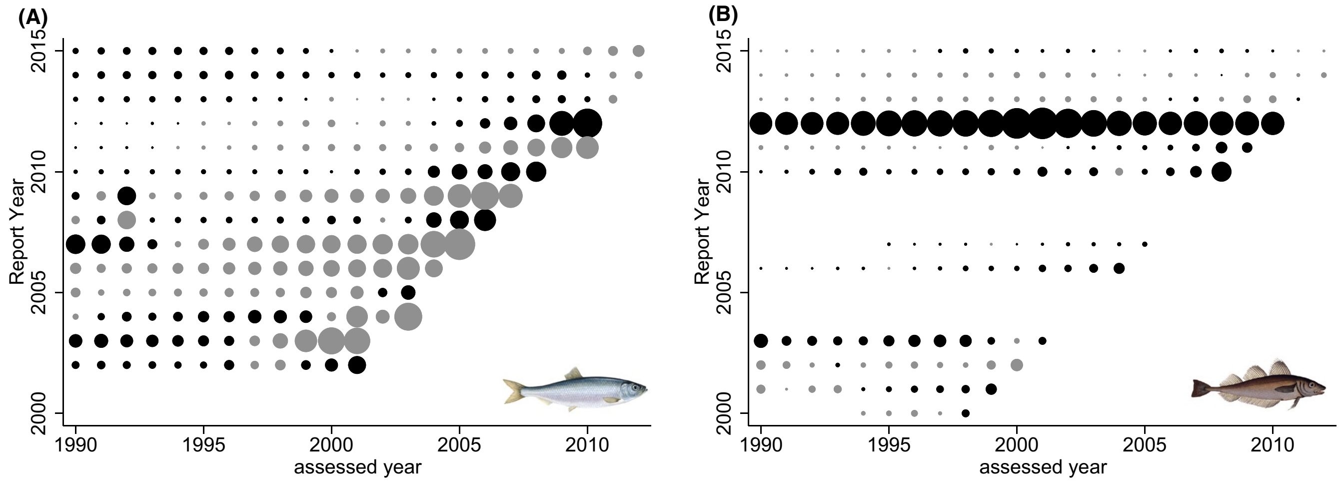


FIGURE 2 Relative changes in fish stock assessments between report years. (a) Assessments of herring in the Gulf of Riga. (b) Assessments for whiting (*Merlangius merlangus*, Gadidae) in the North Sea. In each report year, a time series of estimates for the previous years is generated. Each column indicates an assessed year. The size of the circle demonstrates the size of the relative change of spawning stock biomass (SSB) estimates between report years. If the estimate of the current report year yields the same SSB for a year as the previous report year, the relative change is zero. Black circles indicate that the new estimate is higher than the previous one (upward change) and grey marks the cases when the new estimate is lower than the previous one (downward change)

the stocks. In line with Lindegren et al. (2009), we define the case when SSB is below B_{lim} as critical.

To control for information and learning, we include additional data in our analysis. The longer a stock has been assessed, the better is the understanding of the stock. Hence, we construct a variable that measures the time difference between the first and the current report year, which we refer to as 'history of stock assessments'. In a similar manner, we control for the range of SSB values a stock has experienced. A stock for which very low and very high SSB values have been observed may be inherently more volatile thus posing challenges—but also learning opportunities—to scientific advice (Magnusson & Hilborn, 2007). We construct variables for highest and lowest SSB, which measure the highest/lowest SSB recorded up to year j in report year t. We use the difference between highest and lowest SSB as an indicator for 'SSB range'. In the same manner, we measure the range of recruitment. The last control variable we use is species interaction. We combine SSB of all assessed species per year per ecosystem and divide it by the number of species assessed in the ecosystem (we use the ecoregion reported in the assessments such as North Sea as proxy). This proxy allows us to control for ecosystem complexity and species interaction effects, which might influence changes in SSB.

2.3 | Statistical model

Since the assessed years of a fish stock are not independent, we use a linear fixed-effects model in which we group assessed year and fish stocks. This allows us to account for structural differences between stocks. We use report years (the year the assessment was conducted) as time variable t, i denotes the fish stock and j refers to the assessed year. In each report year t, a time series is produced with estimates for previously assessed years j for a given fish stock i. The introduction of benchmarks marks a change in how the fish stock assessments are conducted. The year when benchmarks were introduced differs between fish stocks. To test whether this change of doing stock assessments is also reflected in the estimated SSB values, we rely on a statistical technique called 'Difference in differences' (DiD) (Angrist & Pischke, 2008, 2015; Cunningham, 2021). This approach is commonly used in economics and looks at observational data through the lens of a 'natural experiment'. We study the effect of a so-called 'treatment' on a 'treatment group' (the stocks that are benchmarked) versus a 'control group' (stocks that are not benchmarked yet). Since not all stocks are benchmarked at the same time, we use a staggered difference in difference model (DiD) with fixed effects of assessed year per stock. This model allows us to compare the stocks that were already subjected to the new policy of benchmarks with the stocks that are assessed with a business-as-usual approach. Formally, benchmarks were introduced in different reports years for different fish stocks (the stocks were 'treated', $\beta_1 \text{TREAT}_{ij}$); see Eq. 2. The time period from when a stock is treated is referred to as 'post' treatment introduction ($\beta_2 \text{POST}_t$). The effect of benchmarking is estimated with an interaction effect between treatment in a given year and the fish stocks that were treated ($\beta_3(\text{TREAT}_{ij} \times \text{POST}_t)$)

$$y_{ijt} = \alpha_{ij} + \beta_1 \text{TREAT}_{ij} + \beta_2 \text{POST}_t + \beta_3(\text{TREAT}_{ij} \times \text{POST}_t) + \varepsilon_{ijt} \quad (2)$$

Since we use a fixed-effects model that controls for stock and assessed year, time-invariant variables can be omitted and we rewrite the equation above to estimate the following model

$$y_{ijt} = \alpha_{ij} + \beta D_{jit} + \lambda_t + \varepsilon_{ijt} \quad (3)$$

with y_{ijt} denoting the relative interannual change and λ_t being an indicator variable for report years. D_{jit} indicates whether a stock has been benchmarked or not. The variable takes the value 1 if the stock has been assessed by a benchmark in the current or an earlier report year *t*. To account for the fact that there are unobserved factors that affect the dependent variable and change over time, which leads to the time-variant idiosyncratic error ε_{ijt}, we use a robust estimator. We cluster the error term on assessed year of a fish stock to account for intragroup correlations (assessed years for a fish stock always depend on the report year they were estimated; thus, it is implausible to assume a correlation of zero). All estimations were done with STATA (*xtreg*, *fe*).

While the introduction of benchmarking marked a major change in how stock assessment modelling is done, which lends itself to being estimated by a staggered DiD model, we are also interested in factors that have a more subtle impact on the estimates. To test our other hypotheses, we estimate the following fixed-effects model

$$y_{ijt} = \alpha_{ij} + \beta \mathbf{X_{ijt}} + \lambda_t + \varepsilon_{ijt} \tag{4}$$

with y_{ijt} being the relative change of estimated biomass in absolute values, $\mathbf{X_{ijt}}$ the independent variables (kind of assessment, data source, stock status and control variables), λ_t is report year dummies and the time-variant idiosyncratic error ε_{ijt}.

3 | RESULTS

The benchmarking procedure has been introduced at different points of time across fish stock assessments. We conduct a staggered DiD analysis (Eq. 3) to analyse how variation changes after benchmarking has been introduced. To make sure that our results indeed measure causal effects of benchmarks, we investigate by looking at 'parallel trends' whether the stocks that are benchmarked are not already performing differently prior to the benchmark. The assumption of parallel trends (without the introduction of benchmarking, there would be no differences) holds, since the analysis using time trends shows no significant difference between time trends (using *tvdiff* in STATA) of fish stocks without benchmark ($F(1, 2676)$, $p = 0.0791$), but there is a clear difference to be seen when benchmarks were introduced (see Figure A1). We run the analysis looking at the relative change between report years in absolute values (see definition of V_{ijt} in Eq. 1), as well as separating between downward change (the new estimate is lower than the previous one) and upward change (the new estimate is higher than the previous one). Clearly, the introduction of benchmarking leads to significantly different estimates in the assessment independent of whether it is higher or lower than the previous estimate. The effect of benchmarking is strongest in the case of an upward change with a relative deviation of 0.0411 compared to a relative deviation of 0.0328 in the downward changes (Table 2). On average, the introduction of benchmarking leads to a deviation of 3.78 percentage points of relative change compared to having no benchmark.

Next, we investigate whether we can observe anchoring in a given year. Comparing years where updates have taken place with years in which benchmarks were done, we find that an update reduces the variability while a benchmark increases it (Table 3, for robustness, see Table A3). Positive coefficients indicate in how far the new SSB estimate diverges in comparison with the relative change in the report year before. Negative coefficients imply that the relative change in the current report year is smaller than that in the previous one. Hence, a negative coefficient signals anchoring. The flexible nature of an update leads to results that are close to the previous one (the coefficient −0.0144 indicates that the relative change is 1.4 percentage points smaller than the previous estimate), while the rigid structure of a benchmark leads to an increase in variability (the relative change is 8.8 percentage points bigger than the previous estimation). Larger changes seen in benchmark years could also just reflect more rigorous revisions, such as a change in stock assessment model. To control for this, we include a variable that indicates whether the underlying model was changed. Since models are not only changed during benchmarks but also during updates, we include the interaction effect with benchmarks. We find that a model change during a benchmark reinforces the effect on variability while model changes without the rigorous structure of benchmarks reduce variability. Thus, it appears that model changes intensify the anchoring in updates. These findings are in line with our hypothesis

TABLE 2 Staggered difference in difference (DiD) model

	(1) Relative change, full	(2) Relative change, downward	(3) Relative change, upward
Stock benchmarked	0.0378*** (0.00298)	0.0328*** (0.00263)	0.0411*** (0.00507)
Constant	0.0871** (0.0289)	0.0697*** (0.000718)	0.0914** (0.0349)
Fixed report year effect	√	√	√
Fixed assessed year per stock	√	√	√
Observations	20,994	9644	11,350
Adjusted R^2	0.035	0.109	0.039

Note: Robust standard errors in parentheses, clustered on assessed year per fish stock.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 3 Expert judgement in fish stock assessments

	(1) Relative change, full	(2) Relative change, downward	(3) Relative change, upward
Type of assessment			
Update	−0.0144***	0.00218	−0.0196***
	(0.00221)	(0.00230)	(0.00361)
Benchmark	0.0654***	0.0363***	0.0882***
	(0.00389)	(0.00346)	(0.00575)
Model change	−0.0153***	−0.0108***	−0.00888*
	(0.00260)	(0.00306)	(0.00387)
Benchmark × Model change	0.0758***	0.0296***	0.107***
	(0.00989)	(0.00718)	(0.0182)
Health of fishery			
Critical stock status	−0.0482***	0.0166	−0.0762***
	(0.0130)	(0.0119)	(0.0204)
Relative fishing pressure	−0.0222***	0.00957	−0.0446***
	(0.00556)	(0.00550)	(0.00921)
Data source			
Catches	0.0277***	0.0179***	0.0183***
	(0.00436)	(0.00476)	(0.00549)
Survey	0.0314***	0.0485***	0.0358***
	(0.00512)	(0.00696)	(0.00734)
Catches × survey	−0.0346***	−0.0591***	−0.0150
	(0.00738)	(0.00895)	(0.0108)
Control variables			
History of stock assessment	−0.00159	−0.00216***	−0.00204
	(0.00135)	(0.000219)	(0.00156)
SSB range	2.83e−09	6.52e−09	−2.32e−08***
	(5.32e−09)	(4.90e−09)	(6.88e−09)
Recruitment range	−0.00000493***	−0.00000117**	−0.00000399***
	(0.000000608)	(0.000000388)	(0.000000872)
Species interaction	−3.73e−10**	2.63e−10	−7.67e−10***
	(1.27e−10)	(1.47e−10)	(1.61e−10)
Fixed report year effect	√	√	√
Fixed assessed year per stock	√	√	√
Observations	16735	7706	9029
Adjusted R^2	0.202	0.139	0.289

Note: Linear fixed-effects regression with absolute values of relative change as dependent variable. Full sample as well as upward and downward estimate. (1) Full sample, (2) estimates done on the subsample when the new estimate was lower than the previous one, (3) estimates done on the subsample when the new estimate was higher than the previous one.

Robust standard errors in parentheses, clustered on assessed year per fish stock.

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

that updates leave room to 'fiddle', giving rise to anchoring while the structured approach used in benchmarking reduces the anchoring effect.

To estimate whether adequate weighting of various data sources has an effect on variation, we compare variation of having only one source of data (survey data or catches) with having both data sources available. We see a clear difference between the input of only one data source in comparison with the use of both data sources. If only one data source is available, a larger diversion from previous results can be observed (Table 3). If two data sources are available, the results are closer to previous assessments, thereby supporting our hypothesis that adequate weighting provides room for stronger anchoring of results.

As hypothesised, a critical stock status leads to lower variation and estimates to be anchored closer to the previous ones (Table 3). Note that this finding only applies for upward adjustments, i.e. announcing that a critical stock is recovering. The same effect can be observed when using high fishing pressure as a measure for overfishing, and using MSYBtrigger as an indicator for critical stock status (Table A2). Generally, anchoring seems larger if the new estimate indicates a higher fish abundance, calling for an upward adjustment (Table 3). Potentially, the effects of assessment procedure (benchmark or update) and adequate weighting of various data sources could interact with the critical stock status (see Table A4). We find that anchoring plays an even bigger role in stocks that have a critical status. With regard to adequate weighting, we do not see any interaction effect.

We have shown that outputs from stock assessment models are affected by intuitive judgement by experts. In general, we observe anchoring, which implies that new estimates tend to be very close to previous estimates, if stock scientists can choose the model, the model settings or model input without too many constraining rules. Obviously, any recurrent assessment will not be made from scratch and will take previous knowledge into account allowing for learning. Indeed, the stock of knowledge changes over time within working groups and the fisheries science community as a whole and may differ between different stocks. We have taken several steps to probe into the robustness of our results and account for learning. First, all of our results are estimated with fixed effects, so that our findings reflect differences within a certain stock, rather than differences between stocks since the latter may just reflect that some stocks are better understood. Second, to control for learning, we included various control variables, e.g. observed historical range of biomass levels or the history of the stock being assessed, but these were insignificant. Third, we conduct placebo tests in which we randomise the type of assessment while keeping everything else constant. In the first test (Table A5 columns 1–3), we randomly assign updates and benchmarks across report years and within fish stock. In the second test (Table A5 columns 4–6), we randomise across stocks and within report years. As expected, our placebo tests does not show any evidence of anchoring (Table A6).

4 | DISCUSSION

We find that anchoring is larger whenever there is 'room to fiddle' and/or the stock is in a critical state and particularly so within upward adjustments. A potential explanation of this finding lies in the political nature of fisheries stock assessments. Since the experts know that higher biomass estimates translate into higher total allowable catch, they are under high pressure to be certain that the spawning stock biomass really is higher. It seems that, in this case, experts follow (consciously or subconsciously) a precautionary road.

One of the key challenges in fisheries management is that one never learns about the true state of the world, calling for careful experimentation and trial and error (Jensen et al., 2012). While the lack of immediate feedback makes learning much harder, it makes it also much more difficult to assess the quality of the assessments. While we documented how intuitive judgement affects experts advice, we do not know whether this brings us closer to the truth or not. There is ample evidence suggesting that relying on intuition, gut feeling and heuristics is not only efficient, but often also produces better decisions than relying on careful reasoning and objective procedures (Gigerenzer & Gaissmaier, 2011; Gigerenzer & Todd, 1999; Gilovich et al., 2002; Kahneman, 2011). The use of heuristics is a valid strategy if they are applied in a 'high validity' environment (valid cues and the opportunity to learn those cues) (Kahneman & Klein, 2009). Arguably, fisheries experts do not always have the benefit of acting within a high-validity environment, but whenever they do, their use of heuristics can be a big advantage. So, removing all possibilities for experts to make judgement calls will most likely not lead to better stock assessments. After all, experts do have the ability to judge if something does not seem to add up and making those judgement calls requires experience.

Our findings provide some lessons that can contribute to an ongoing discourse how fisheries management can be improved. First, treating diverging data as outlier and relying on what we have seen before may be appropriate in many cases, but it may also lead to the rejection of very unlikely and implausible data, which would require immediate attention (Taleb, 2007). The collapse of Northern cod in Canada is a particular painful example where experts had trouble making sense of data that suggested the stock was in realms they had never even deemed possible before (Finlayson & McCay, 1998; Steele et al., 1992). In that light, it is actually comforting that anchoring plays a larger role when it comes to making upward shifts, which is more precautionary. Second, while work of experts—fishery scientists included—is often seen as a task carried out with surgical precision, there is a decisive human factor involved. Yet, we need to understand much better how this human factor works especially since experts, as well was policy makers, are often unaware of its existence. There is also more need to evaluate not only which procedures, but also composition of working groups, affect the ability to reflect and ultimately the quality of the advice. Third, there is also an important role for science communication. In complex environments, such as fisheries, scientific evidence changes over time and what was solid advice yesterday, may need to be revised today. As long as policy makers and the broader public perceive changes of advice in light of new information—but also learning about inherent uncertainty in advice—as 'experts have no clue', advice may not reflect the best available science.

ACKNOWLEDGEMENTS

We are grateful to Martin Pastoors as well as two anonymous reviewers for valuable comments on a previous version of this manuscript. This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 675997.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data used are freely accessible from the ICES reports.

ORCID

Esther Schuch https://orcid.org/0000-0001-9049-4064

REFERENCES

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Angrist, J. D., & Pischke, J.-S. (2015). *The path from cause to effect: mastering 'metrics*. Princeton University Press.

Burgman, M. A. (2015). *Trusting Judgements: How to Get the Best out of Experts*. Cambridge University Press.

Chrysafi, A., Cope, J. M., & Kuparinen, A. (2019). Eliciting expert knowledge to inform stock status for data-limited stock assessments. *Marine Policy*, *101*, 167–176. https://doi.org/10.1016/j.marpol.2017.11.012

Coffman, L. C., & Niederle, M. (2015). Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives*, *29*, 81–98. https://doi.org/10.1257/jep.29.3.81

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

Dankel, D. J., Stange, K., & Nielsen, K. N. (2015). What hat are you wearing? On the multiple roles of fishery scientists in the ICES community. *ICES Journal of Marine Science*, *73*, 209–216. https://doi.org/10.1093/icesjms/fsv199

Deroba, J. J., & Miller, T. J. (2016). Correct in theory but wrong in practice: Bias caused by using a lognormal distribution to penalize annual recruitments in fish stock assessment models. *Fisheries Research*, *176*, 86–93. https://doi.org/10.1016/j.fishres.2015.12.002

Dickey-Collas, M., Pastoors, M. A., & Van Keeken, O. A. (2007). Precisely wrong or vaguely right: simulations of noisy discard data and trends in fishing effort being included in the stock assessment of North Sea plaice. *ICES Journal of Marine Science*, *64*, 1641–1649. https://doi.org/10.1093/icesjms/fsm155

Englich, B., & Soder, K. (2009). Moody experts–How mood and expertise influence judgmental anchoring. *Judgment and Decision Making*, *4*, 41.

Fallon, N. G., Collins, M. A., Marshall, C. T., & Fernandes, P. G. (2015). Assessing consistency of fish survey data: uncertainties in the estimation of mackerel ice fish (*Champsocephalus gunnari*) abundance at South Georgia. *Polar Biology*, *39*, 593–603. https://doi.org/10.1007/s00300-015-1810-0

Finlayson, A. C., & McCay, B. J. (1998). *Crossing the threshold of ecosystem resilience: the commercial extinction of northern cod', Linking social and ecological systems: management practices and social mechanisms for building resilience* (pp. 311–337). Cambridge University Press.

Francis, C. (2011). Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*, 1124–1138. https://doi.org/10.1139/f2011-025

Francis, C. (2016). Explaining bias caused by the wrong recruitment penalty in stock assessment models. *Fisheries Research*, *181*, 114–115. https://doi.org/10.1016/j.fishres.2016.04.001

Fulton, E. A. (2021). Opportunities to improve ecosystem- based fisheries management by recognizing and overcoming path dependency and cognitive bias. *Fish and Fisheries*, *22*, 1–21.

Fulton, E. A., Smith, A. D. M., Smith, D. C., & van Putten, I. E. (2011). Human behaviour: the key source of uncertainty in fisheries management. *Fish and Fisheries*, *12*, 2–17. https://doi.org/10.1111/j.1467-2979.2010.00371.x

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*, 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Todd, P. M. (1999). *Simple Heuristics that Make Us Smart*. Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.

Hagafors, R. (1983). When experts are influential: Comparisons with respect to success and consistency in judgment tasks. *Scandinavian Journal of Psychology*, *24*, 293–301. https://doi.org/10.1111/j.1467-9450.1983.tb00503.x

Hilborn, R. (1992). Current and future trends in fisheries stock assessment and management. *South African Journal of Marine Science*, *12*, 975–988. https://doi.org/10.2989/02577619209504756

Hilborn, R., & Stokes, K. (2010). Defining overfished stocks: Have we lost the plot. *Fisheries*, *35*, 113–120. https://doi.org/10.1577/1548-8446-35.3.113

ICES (2013). *Introduction to Benchmarks at ICES*, Technical report, International Council for the Exploration of the Sea (ICES). URL: https://www.ices.dk/community/Documents/Advice/Introduction%20to%20Benchmarks%20at%20ICES.pdf

ICES. (2016). *ICES Advice Basis*. URL: http://www.ices.dk/sites/pub/PublicationReports/Advice/2016/2016/Introductiontoadvice2016.pdf

Ichinokawa, M., Okamura, H., & Takeuchi, Y. (2014). Data conict caused by model mis-specification of selectivity in an integrated stock assessment model and its potential effects on stock status estimation. *Fisheries Research*, *158*, 147–157. https://doi.org/10.1016/j.fishres.2014.02.003

Jensen, O. P., Branch, T. A., & Hilborn, R. (2012). Marine fisheries as ecological experiments. *Theoretical Ecology*, *5*, 3–22. https://doi.org/10.1007/s12080-011-0146-9

Johnson, K. F., Monnahan, C. C., McGilliard, C. R., Vert-pre, K. A., Anderson, S. C., Cunningham, C. J., Hurtado-Ferro, F., Licandeo, R. R., Muradian, M. L., Ono, K., Szuwalski, C. S., Valero, J. L., Whitten, A. R., & Punt, A. E. (2014). Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. *ICES Journal of Marine Science*, *72*, 137–150. https://doi.org/10.1093/icesjms/fsu055

Kahneman, D. (2011). *Thinking, Fast and Slow, Farrar, Straus and Giroux: New York, NY, USA*.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kassam, K. S., Koslov, K., & Mendes, W. B. (2009). Decisions under distress: Stress profiles inuence anchoring and adjustment. *Psychological Science*, *20*, 1394–1399. https://doi.org/10.1111/j.1467-9280.2009.02455.x

Kehler, D. G., Myers, R. A., & Field, C. A. (2002). Measurement error and bias in the maximum reproductive rate for the Ricker model. *Canadian Journal of Fisheries and Aquatic Sciences*, *59*, 854–864. https://doi.org/10.1139/f02-060

Kirchgassner, G., & Muller, U. K. (2006). Are forecasters reluctant to revise their predictions? Some German evidence. *Journal of Forecasting*, *25*, 401–413. https://doi.org/10.1002/for.995

Kraak, S. B. M., Daan, N., & Pastoors, M. A. (2009). Biased stock assessment when using multiple, hardly overlapping, tuning series if fishing trends vary spatially. *ICES Journal of Marine Science*, *66*, 2272–2277. https://doi.org/10.1093/icesjms/fsp179

Kraak, S. B. M., Kelly, C. J., Codling, E. A., & Rogan, E. (2010). On scientists' discomfort in fisheries advisory science: the example of simulation-based fisheries management-strategy evaluations. *Fish and Fisheries*, *11*, 119–132. https://doi.org/10.1111/j.1467-2979.2009.00352.x

Kupferschmidt, K. (2020). The Coronavirus Czar: The COVID-19 pandemic has made German virologist Christian Drosten an unlikely cult figure. *Science*. https://doi.org/10.1126/science.abc5095

Lassen, H., Kelly, C., & Sissenwine, M. (2013). ICES advisory framework 1977–2012: from Fmax to precautionary approach and beyond. *ICES Journal of Marine Science*, *71*, 166–172. https://doi.org/10.1093/icesjms/fst146

Lee, H.-H., Maunder, M. N., Piner, K. R., & Methot, R. D. (2012). Cansteepness of the stock-recruitment relationship be estimated in fishery stock assessment models? *Fisheries Research*, *125–126*, 254–261. https://doi.org/10.1016/j.fishres.2012.03.001

Lindegren, M., Mollmann, C., Nielsen, A., & Stenseth, N. C. (2009). Preventing the collapse of the Baltic cod stock through an ecosystem-based management approach. *Proceedings of the National Academy of Sciences*, *106*, 14722–14727. https://doi.org/10.1073/pnas.0906620106

Magnusson, A., & Hilborn, R. (2007). What makes fisheries data informative? *Fish and Fisheries*, *8*, 337–358. https://doi.org/10.1111/j.1467-2979.2007.00258.x

Marshall, C. T., Needle, C. L., Thorsen, A., Kjesbu, O. S., & Yaragina, N. A. (2006). Systematic bias in estimates of reproductive potential of an Atlantic cod (Gadus morhua) stock: implications for stock-recruit theory and management. *Canadian Journal of Fisheries and Aquatic Sciences*, *63*, 980–994. https://doi.org/10.1139/f05-270

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. https://doi.org/10.1111/j.1523-1739.2011.01806.x

McBride, M. F., Garnett, S. T., Szabo, J. K., Burbidge, A. H., Butchart, S. H., Christidis, L., Dutson, G., Ford, H. A., Loyn, R. H., Watson, D. M., & Burgman, M. A. (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. https://doi.org/10.1111/j.2041-210X.2012.00221.x

Methot, R. D. Jr (2015). *Prioritizing fish stock assessments*, *Technical Report NMFS-F/SPO-152*, U.S. Dep. Commer., NOAA.

Methot, R. D., Taylor, I. G., & Chen, Y. (2011). Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*, 1744–1760. https://doi.org/10.1139/f2011-092

Montibeller, G., & Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, *35*, 1230–1251. https://doi.org/10.1111/risa.12360

Myers, R. A., & Barrowman, N. J. (1995). Time series bias in the estimation of density-dependent mortality in stock-recruitment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *52*, 223–232. https://doi.org/10.1139/f95-022

Nordhaus, W. D. (1987). Forecasting efficiency: concepts and applications. *The Review of Economics and Statistics*, *69*, 667–674. https://doi.org/10.2307/1935962

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, *29*, 261–265. https://doi.org/10.1037/h0022125

Ottaviani, M., & Sorensen, P. N. (2006). Professional advice. *Journal of Economic Theory*, *126*, 120–142. https://doi.org/10.1016/j.jet.2004.08.005

Palmer, M. C., & Demarest, C. (2018). Getting to good enough: Performance of a suite of methods for spatially allocating fishing effort to management areas. *Fisheries Research*, *204*, 26–32. https://doi.org/10.1016/j.fishres.2018.02.003

Pastoors, M. A. (2020). *Evaluation of the performance of stock assessments in the Northeast Atlantic using a new metric for historical retrospective analyses*. Working Paper.

Payne, M. R., Clausen, L. W., & Mosegaard, H. (2009). Finding the signal in the noise: objective data-selection criteria improve the assessment of western Baltic spring-spawning herring. *ICES Journal of Marine Science*, *66*, 1673–1680. https://doi.org/10.1093/icesjms/fsp185

Ricard, D., Minto, C., Jensen, O. P., & Baum, J. K. (2011). Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and Fisheries*, *13*, 380–398. https://doi.org/10.1111/j.1467-2979.2011.00435.x

Sinkey, M. (2015). How do experts update beliefs? Lessons from a nonmarket environment. *Journal of Behavioral and Experimental Economics*, *57*, 55–63. https://doi.org/10.1016/j.socec.2015.04.003

Soria, M., Freon, P., & Gerlotto, F. (1996). Analysis of vessel influence on spatial behaviour of fish schools using a multi-beam sonar and consequences for biomass estimates by echo-sounder. *ICES Journal of Marine Science*, *53*, 453–458. https://doi.org/10.1006/jmsc.1996.0064

Starcke, K., & Brand, M. (2012). Decision making under stress: a selective review. *Neuroscience & Biobehavioral Reviews*, *36*, 1228–1248. https://doi.org/10.1016/j.neubiorev.2012.02.003

Steele, D. H., Andersen, R., & Green, J. M. (1992). The managed commercial annihilation of northern cod. *Newfoundland and Labrador Studies*, *8*.

Sutherland, W. J., & Burgman, M. (2015). Policy advice: use experts wisely. *Nature*, *526*, 317–318. https://doi.org/10.1038/526317a

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*, Vol. *2*. Random House.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458. https://doi.org/10.1126/science.7455683

Walters, C. (2004). Simple representation of the dynamics of biomass error propagation for stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *61*, 1061–1065. https://doi.org/10.1139/f04-120

Wilson, D. C. (2010). *The paradoxes of transparency: Science and the ecosystem approach to fisheries management in Europe*. Amsterdam University Press.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Schuch, E., & Richter, A. (2022). Tracing intuitive judgement of experts in fish stock assessment data. *Fish and Fisheries*, 23, 758–767. https://doi.org/10.1111/faf.12646