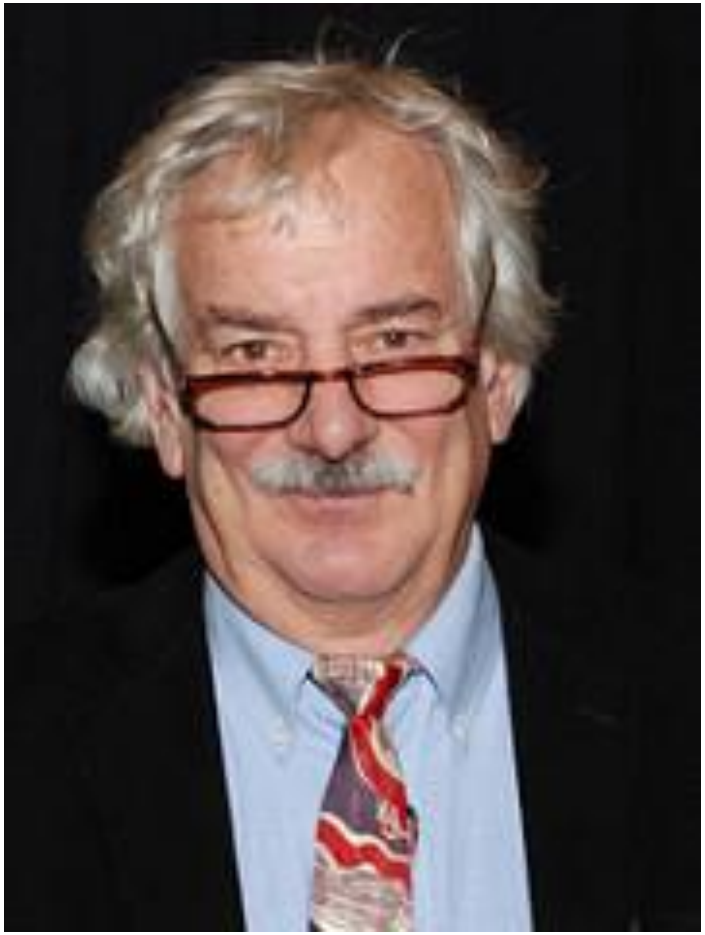
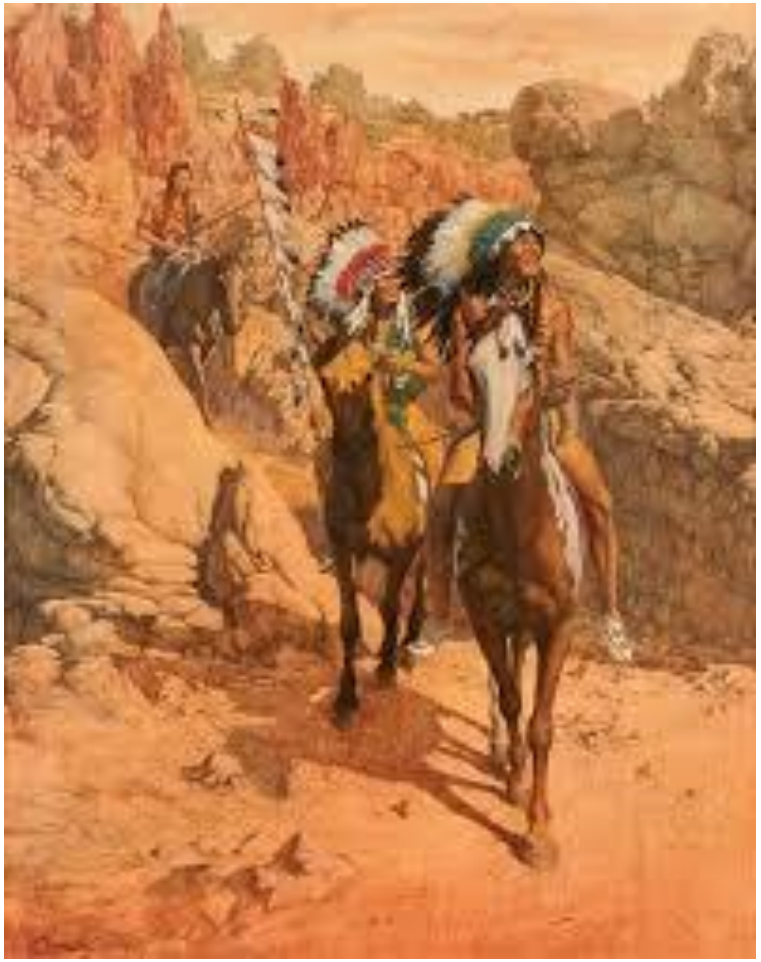
University of Strathclyde Business School

Valuing and evaluating expertise: judgement, policy and business

TIM BEDFORD
Department of Management Science
University of Strathclyde, Glasgow, Scotland



Dedicated to
Roger Cooke



Dedicated to Roger Cooke

and our time working together

Starting the Action…

University of Strathclyde Business School

L'Aquila

- Six top scientists sentenced for "falsely reassuring" the public
- Who would be a seismologist?
- Should seismologists be encouraged to say what they believe?

Italian scientist convicted over L'Aquila earthquake condemns 'medieval' court

Claudio Eva says ruling against him and five others for falsely reassuring statements over 2009 quake was 'eye for an eye'

Tom Kington in Rome
guardian.co.uk, Tuesday 23 October 2012 17.56 BST
Jump to comments (200)

L'Aquila was hit by a devastating earthquake in 2009 that left 308 people dead and thousands homeless. Photograph: Giorgio Cosulich/Getty Images

An Italian physicist handed a six-year jail sentence for giving falsely reassuring statements over an earthquake has condemned as "medieval" the court that convicted him.

The need for EJ… and the problem

- Complexity of society and environment plus need for timely response means we *cannot wait* for long term empirical studies…EJ is indispensable
- BUT
  - Non-structured EJ is not satisfactory
  - Structured EJ is not yet fully developed
  - Social, legal, governmental processes not attuned to the potential or properly calibrated to the limitations

Innovation areas

University of Strathclyde Business School

- Empirical validation of expert data
  - *Huge paradigm shift, leading to peer review, meta-analysis, reporting standards and more*
- Process and problem structuring
  - *Also addressing the way policy makers will adapt to SEJ*
- Dependency assessment
  - *Model outputs highly sensitive to dependence (cf banking)*
- Foundations
  - *Draw together disparate approaches by refocusing on foundations*

- ***Every application area is different***

The Action

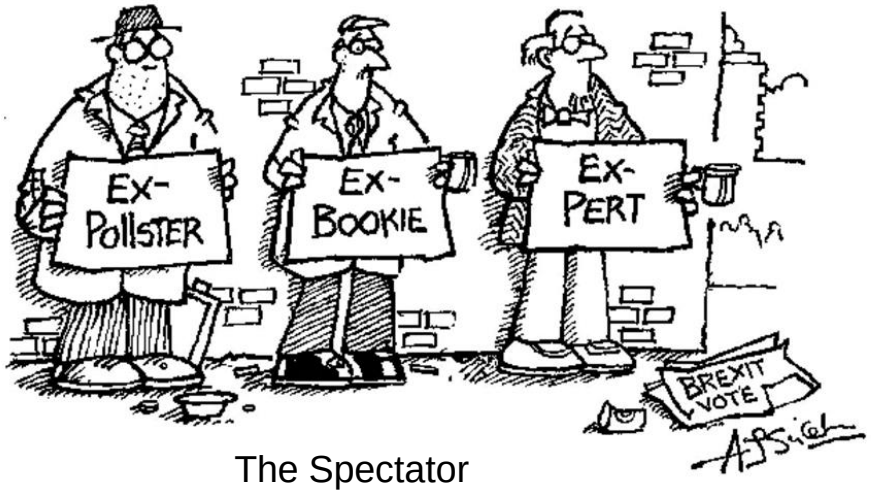
- Major objective of this Action is to be able to encourage senior policy/DMs to use SEJ
- Discussions indicate
  - awareness of EJ, low understanding of SEJ
  - Some awareness of different approaches
- Academic literature
  - Much work on EJ/SEJ from different disciplines
  - Entrenched positions create confusion in users
  - Limited empirical research
  - Limited attempts to incorporate contextual issues into selection of appropriate methods

University of Strathclyde Business School

Current Policy Context

Reputation of experts…

- “People in this country have had enough of experts…” – Michael Gove



The Spectator

- Are we living in a post-factual society?
  - No, but political discourse has become much faster due to social media, and evidence-based consensus has become unfashionable due to political instability

EU Scientific Advice Mechanism

University of Strathclyde Business School

- New Scientific Advice Mechanism introduced (after protests about Junkers abolition of the CSO post) supported by consortium of learned societies and led by high-level group (HLG) of 7 eminent scientists from different fields
- "The HLG provides advice to a European Commissioner ***who has asked for it*** in order to take action in the area that s/he is responsible for. The HLG, in discussion with SAPEA, can also suggest that the College of European Commissioners consults the HLG on a particular topic which is judged of importance. "
- HLG provides advice, but ***"it should not duplicate advice being provided by existing bodies"***
  - Moedas asked for an "explanatory note" explaining the difference between EFSA and WHO opinions on carcinogenic potential of glyphosphate
  - In view of non-duplication, role for SEJ at a level under the SAM HLG

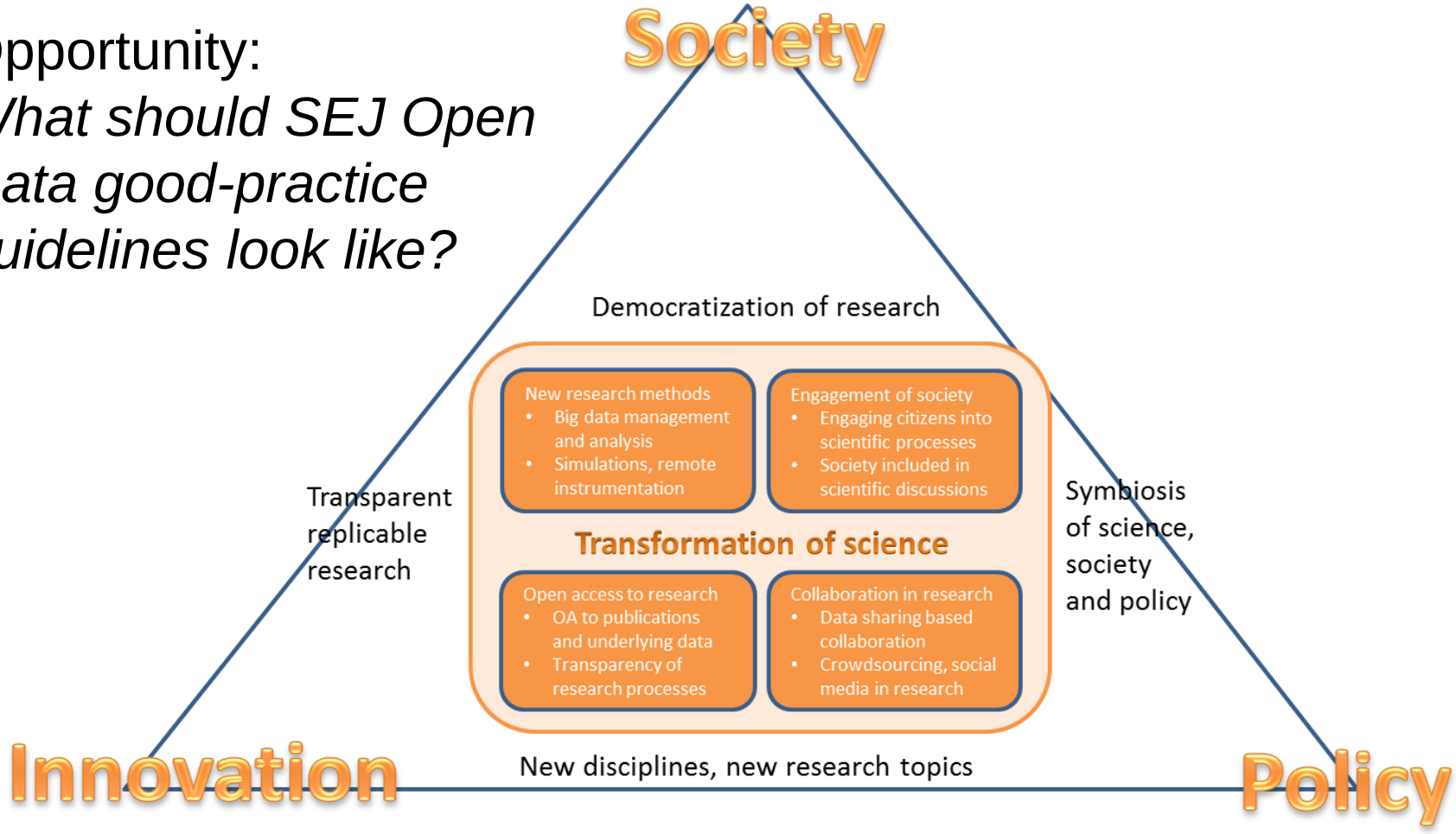
EU Open Data policy

- Key part of EU Open Science policy
- Push to open access journals
- Requirements for FAIR data
  - **Findable, Accessible, Interoperable, and Re-usable**
- Open data requirements including Data Management Plans, for H2020 projects in many areas – justification required for non-participation

- *Prof Karel Luyben, TU Delft Rector, is a member of the EC Open Science Policy Platform*

SEJ and Open Data…

Opportunity:
What should SEJ Open Data good-practice guidelines look like?



University of Strathclyde Business School

Developments in SEJ thinking

Expert Judgement approaches

- Delphi – developed after WW2 by RAND, disavowed, and rehabilitated
- Nominal Group Technique
- Stanford Research Institute Process
- NUREG
- Psychological Scaling
- Classical Model
- SHELF
- Prediction markets
- Superforecasters, IDEA – IARPA ACE competition

Still basically two groups: Behavioural Aggregation and Mathematical Aggregation of quantitative assessments

Key aspects of SEJ

- I will discuss
  - Process
  - Expert Selection/Validation

Process: Expert Rationales: some common ground MA - BA

University of Strathclyde Business School

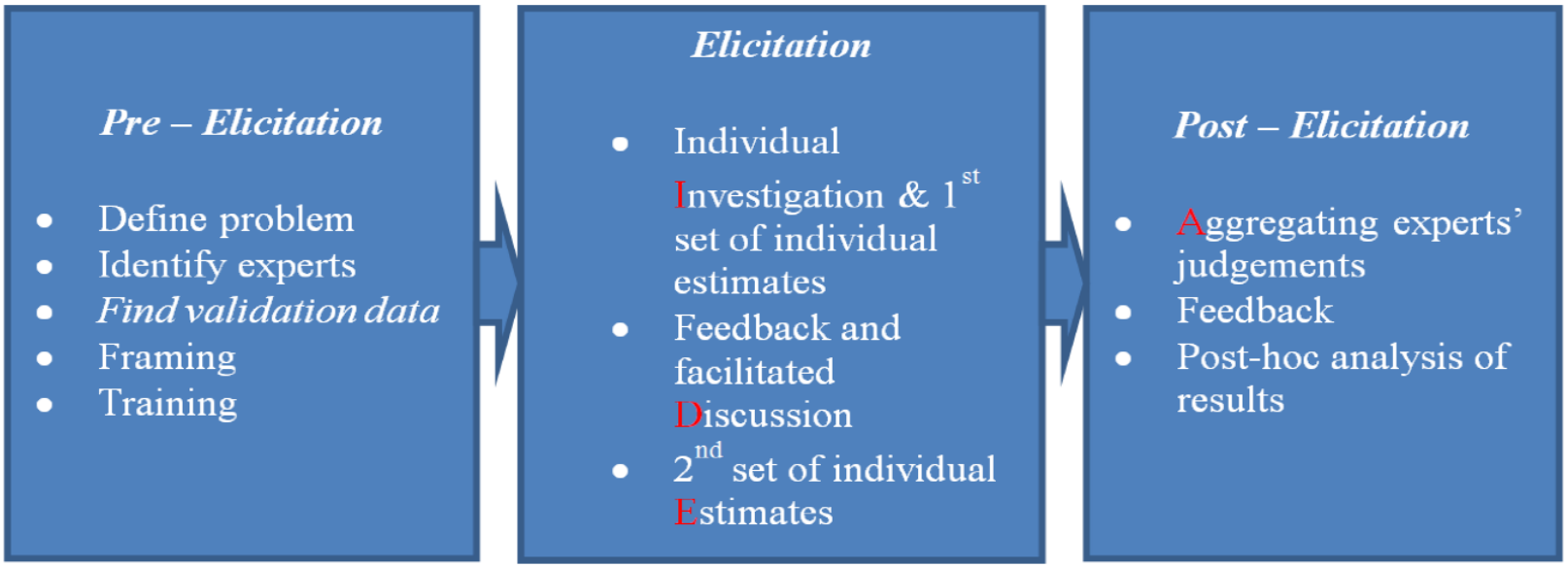
- Provides some reasoning for the assessment
  - Normally qualitative reasoning - if quantitative then may be anchoring on a particular model.
- Provides opportunity to share with other experts
  - Experts may agree on relevant qualitative factors but still disagree on the quantitative effects.
  - Sharing can eliminate potential misunderstandings
- Behavioural aggregation
  - Should work by sharing rationales and then using discussion process to converge to a consensus view on the quantitative aspects
- In all situations need to manage potential bias

Positives and Negatives from Expert interaction

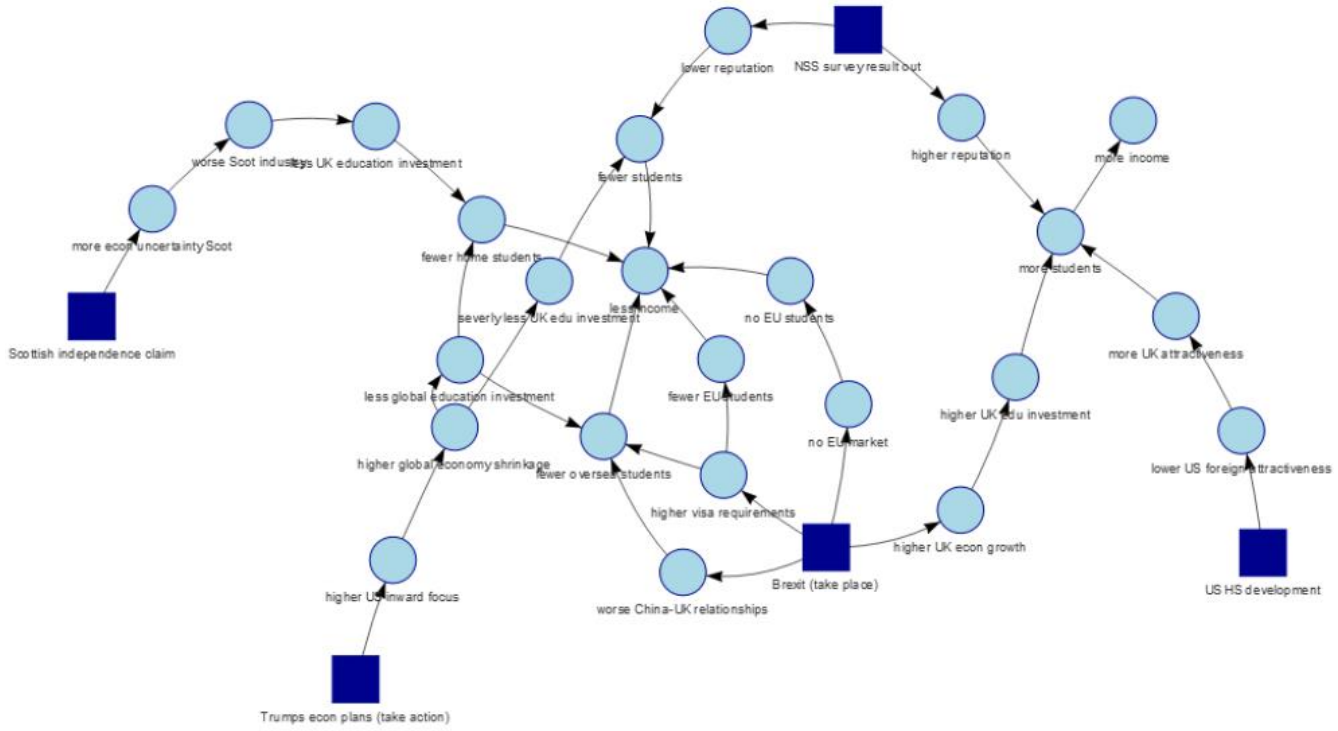
- Positives
  - Ensure all understand the questions and eliminate incorrect (narrowing) assumptions
  - Agree qualitative structure of the problem, hence simplifying the set of questions that need elicitation
  - Discussion about potential mechanisms, base rates, comparative classes etc, highlights aspects that should be considered
- Negatives
  - Development of “groupthink” - Focus on one or two mechanisms, or comparative classes
  - Non-expertise based influences (eg ability to articulate, dominant personality, peer esteem, job level)

Process – Rationales in IDEA

- Due to Burgman et al – a mixed method
- Came second in the IARPA competition



Process – Rationales in Dependency Modelling



- Christoph Werner – representation of rationales in dependence elicitation. Designed to be used remotely.

University of Strathclyde Business School

Key point on rationales:

There are ways for experts to express and share their rationales, giving positives of expert interaction without the negatives of expert interaction

Expert selection, validation and performance weighting

University of Strathclyde Business School

- Relatively little written about expert selection – should represent different scientific schools, different stakeholders..?
  - What are the objectives – are we looking for a consensus or the best possible assessment?
- If doing equal weighting then expert selection drives the outcome of the study
- Expert performance measures can drive either performance based weighting and expert selection

Good Judgement Project

University of Strathclyde Business School

- Came first in the IARPA competition
- Led by Philip Tetlock (University of Pennsylvania)
- Recruited large numbers of potential experts to answer questions
- Different groups 3x4 experimental design
  - Not trained, probability training, scenario training
  - Individual, Crowd-informed individuals, Interactive Group, Prediction Market
- After 1 year, created a Superforecasters group

GJP results

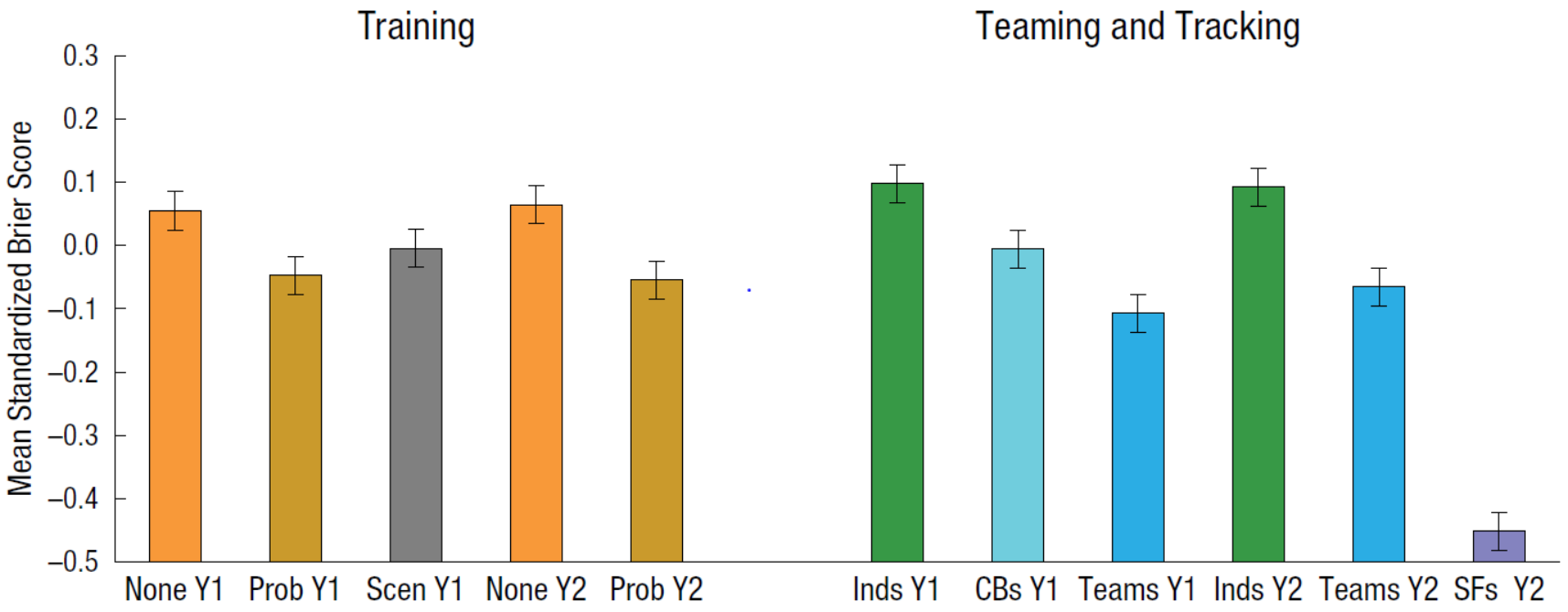


Fig. 1. Effects of training, teaming, and tracking on average Brier scores in Year 1 (Y1) and Year (Y2). The bars at the left show results for the no-training ("None"), probability-training ("Prob"), and scenario-training ("Scen") conditions; the bars at the right show results for independent forecasters ("Inds"), crowd-belief forecasters ("CBs"), team forecasters ("Teams"), and superforecasters ("SFs"). Error bars represent ±2 *SE*s.

B. Mellers, L. Ungar et al, Psychological Strategies for Winning a Geopolitical Forecasting Tournament, Psychological Science 2014, Vol. 25(5) 1106–1115

Classical Method long-term validation studies

- Differences in expert performance; evidence of performance weights improving predictions based on cross-validation – though considered not strong enough evidence by some eg Bolger and Rowe

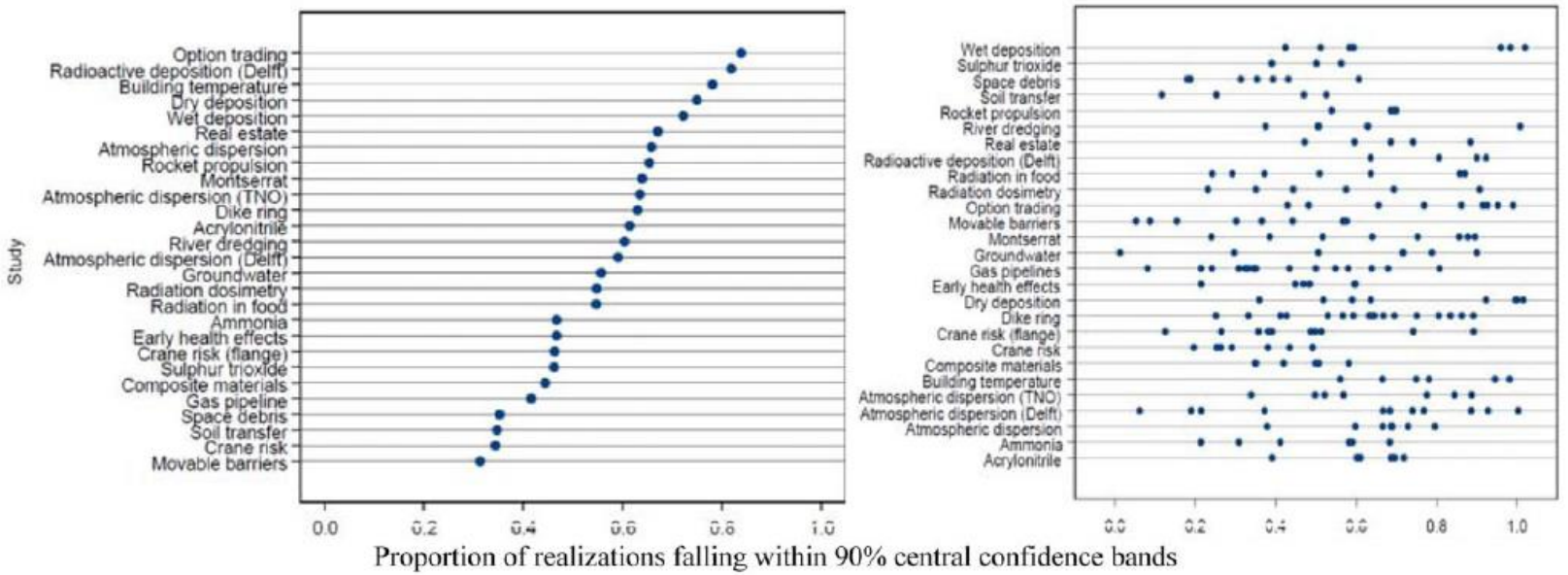


Fig. 2. Average number of realizations falling within experts' 90% confidence bands, per study (left) and per expert per study (right).

University of Strathclyde Business School

Key point of these studies:

Some experts are poor at probability assessments and you are better off without them

Academic controversy

University of Strathclyde Business School

- Critique of Cooke's method in Bolger and Rowe (2015): *The Aggregation of Expert Judgment: Do Good Things Come to Those Who Weight?*
  - Weighting OK, Unequal weighting OK in principle, but Cooke's method ad hoc and atheoretical… (stated in footnote without justification)
  - Might not eliminate gaming
  - Too much emphasis on calibration rather than information
  - Too few calibration questions in typical applications to do anything but identify very poor experts
  - Good scoring a result of normative not substantive expertise
  - Costs of getting seed variables outweigh the benefits
  - Self weighting would be better
  - Experts should discuss the issues together
  - Seed variables are .. "an exam that the statisticians who set it can pass—but which very few domain experts can."
- In the rejoinder they make clear are in favour of behavioural aggregation… and think that tests should be conducted to compare CM to BA….

And Winkler weighed in…

University of Strathclyde Business School

- "**a simple average is readily defensible since it is easy to understand and treats all of the experts equally**, thereby requiring no justification for differential weighting. This can be particularly important in risk analyses of **public decision-making problems, where there are often competing constituencies and resulting challenges to claims and decisions**. This advantage is shared by any combining procedure that treats the experts symmetrically.

More generally, it is important to have different viewpoints represented in the set of experts, preferably with each expert having an understanding of the range of different viewpoints in the larger community of experts and the implications of these viewpoints for the situation of interest. Indeed, the choice of experts is arguably more important than whether or how their forecasts are weighted."

And Granger-Morgan weighed in..

- Agrees with concerns about seed variables, but does not accept their argument “for policy making a single representation of the uncertain quantity, and related probability, is commonly needed.”
- Diversity of expert views needs to be captured if these represent divergent views of the future….Example from global warming.

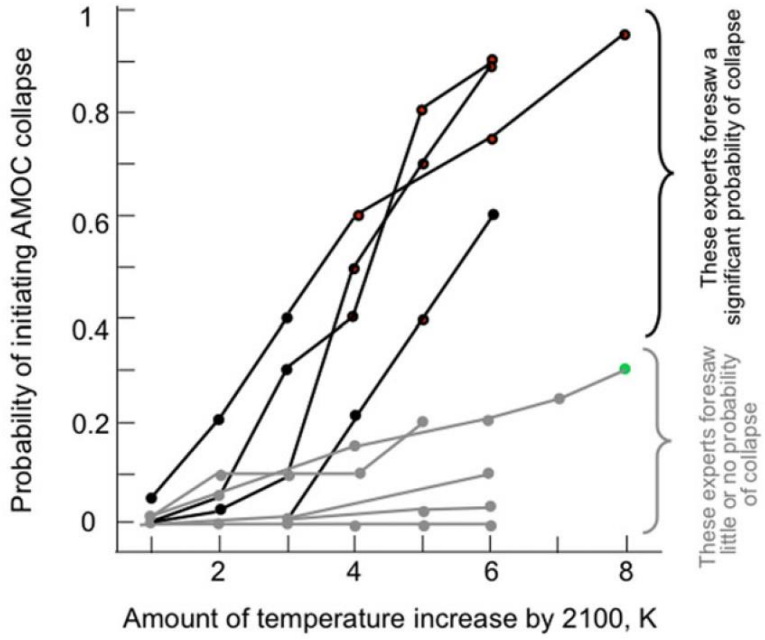


Fig. 1. Example of a situation in which the expert community was divided, with some thinking that the probability of initiating a collapse of the AMOC with plausible amounts for global warming was substantial (upper curves) and some thinking that it was quite low (lower curves). Combining such a group of experts could mask the fundamental disagreement within the community. Figure modified from Zickfeld *et al.*(1)

I claim:

This disagreement arises (partially) because of different contexts in which these authors use SEJ

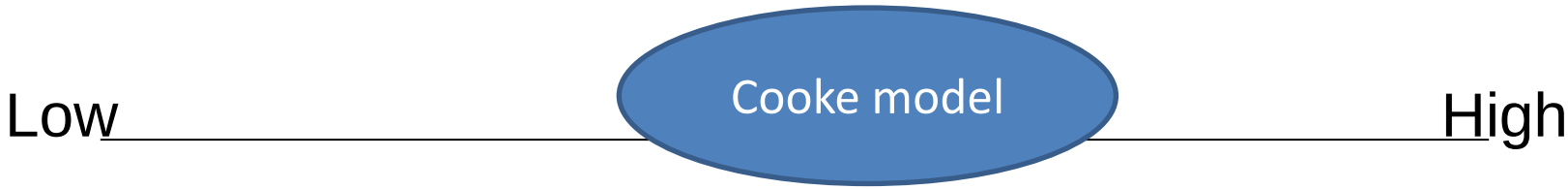
Key SEJ contextual issues

- Extent to which (standard) modelling approach(es) and/or data exists and is relevant
- Speed of application
- Many experts available or highly specialised
- Societal accou
  company/publi
- Game-playing,
  behavioural re
- Consensus- va

Summarise as:

- Degree of Understanding
- Time available for application
- Legitimation burden

Degree of understanding



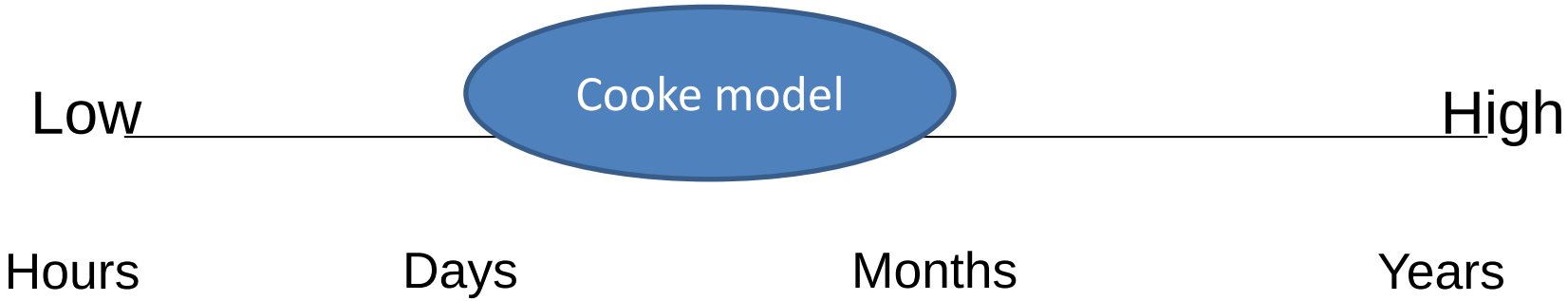
Lack of relevant data or models with explanatory value

Competing models with explanatory value

Models with explanatory value and some relevant empirical data

Excellent explanatory models and relevant empirical data, giving good predictive power in relevant contexts

Time available for application



Legitimation burden

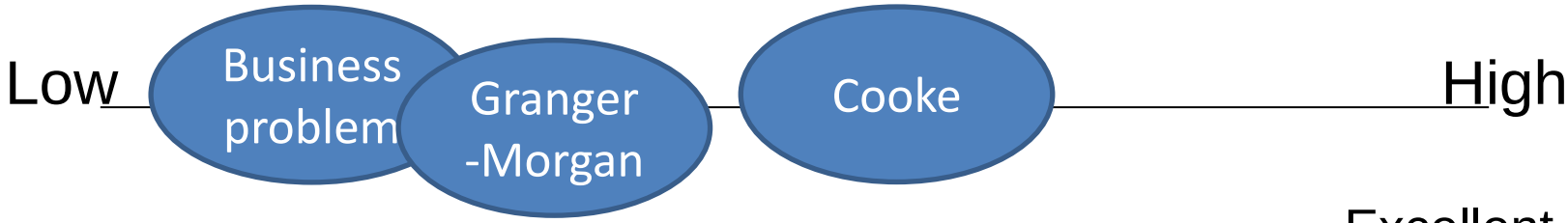
Internal expertise, small numbers of experts with an interest in outcome and no external validation

Consensus driven, but with experts who have no interest in outcome

External validation and quality process but small number of experts

External validation and evidence of quality of the process and validators

Degree of understanding



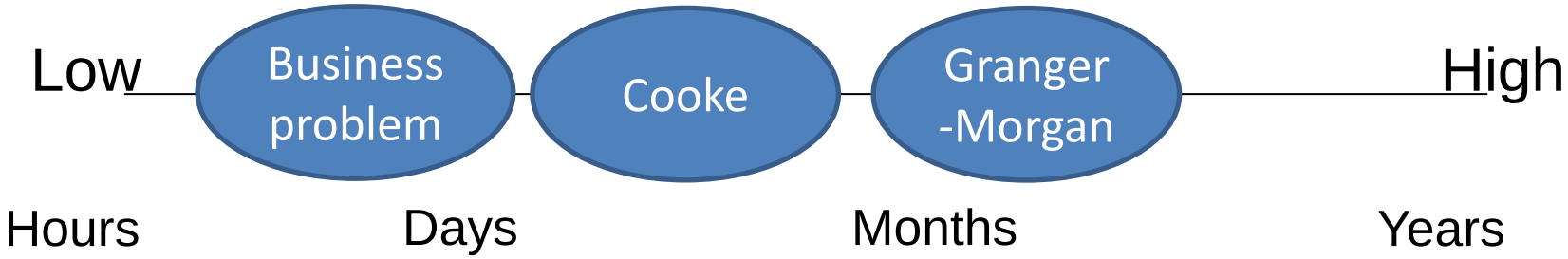
Lack of relevant data or models with explanatory value

Competing models with explanatory value

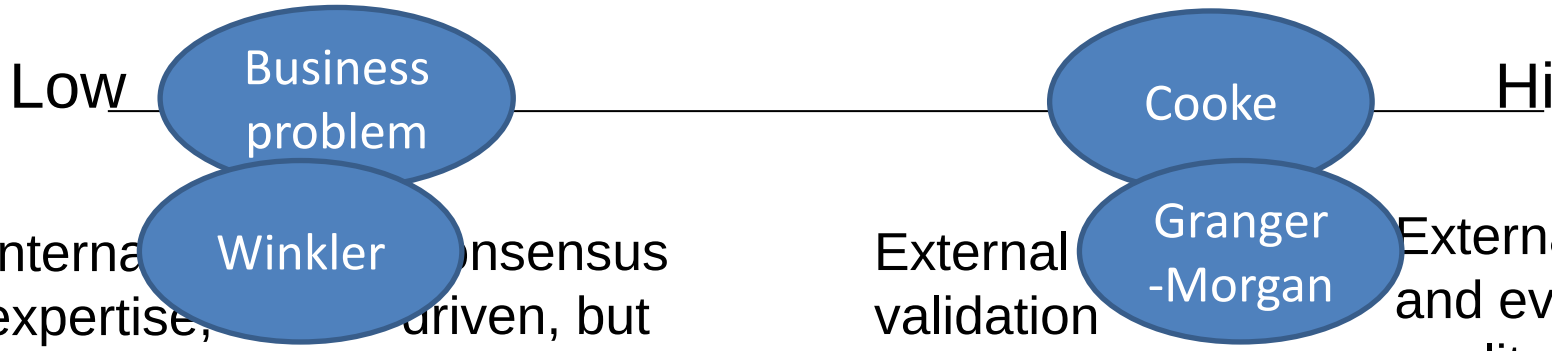
Models with explanatory value and some relevant empirical data

Excellent explanatory models and relevant empirical data, giving good predictive power in relevant contexts

Time available for application



Legitimation burden



Interna expertise, small numbers of experts with an interest in outcome and no external validation

nsensus driven, but with experts who have no interest in outcome

External validation and quality process but small number of experts

External validation and evidence of quality of the process and validators

Finally…

University of Strathclyde Business School

- Real progress made in
  - Confirming that excluding bad experts improves predictions
  - Finding ways for experts to share perspectives without introducing psychological biases
  - Sharing rationales across the internet
  - Understanding how context should shape the requirements for SEJ
  - Influencing better SEJ in different policy areas

University of Strathclyde Business School

Prediction is very difficult, especially about the future

Allegedly due to Niels Bohr