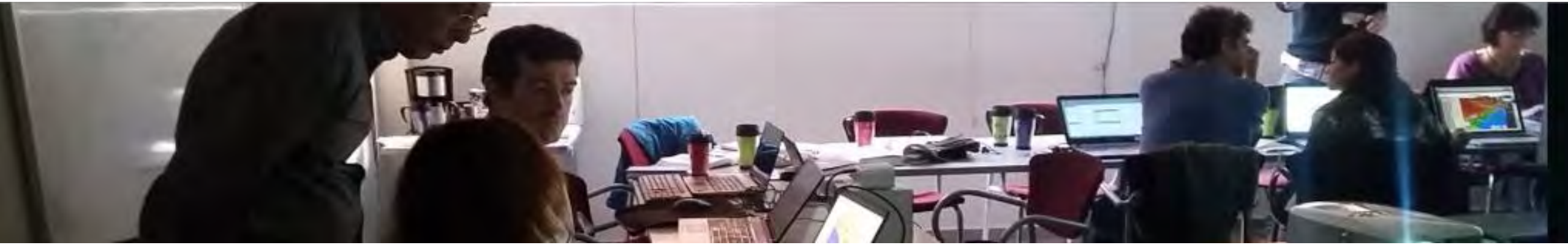
Lessons from structured expert elicitation using the IDEA protocol



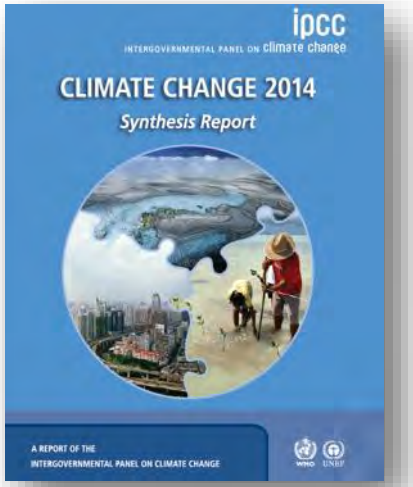
Marissa McBride
m.mcbride@imperial.ac.uk
@riss_mcb

cep
CENTRE FOR ENVIRONMENTAL POLICY

Imperial College London

Why expert judgement


SOON ON CNN
NEW CIA DIRECTOR SPEAKS LIVE
BREAKING NEWS
LIVE
FEARS NORTH KOREA COULD ARM MISSILES WITH SARIN GAS
Gordon Chang | Columnist, The Daily Beast
CNN
2:48 PM ET
FIRST ON CNN
ORCE AIRCRAFT TARGETING ISIS TUNNELS & PERSONNEL IN AF
NEWSROOM


ipcc
INTERGOVERNMENTAL PANEL ON climate change
CLIMATE CHANGE 2014
Synthesis Report
A REPORT OF THE
INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE


Vaccine


THE PROJECT
Deepening required in some places
Melbourne
Pt Cooke
Pt Wilson
Corio
PORT PHILLIP BAY
Mentone
Frankston
Geelong
Swan Is.
Mud Is.
Rosebud
Proposed area for dredged material
Phillip Is.

Experts make mistakes…

Cognitive bias: overconfidence

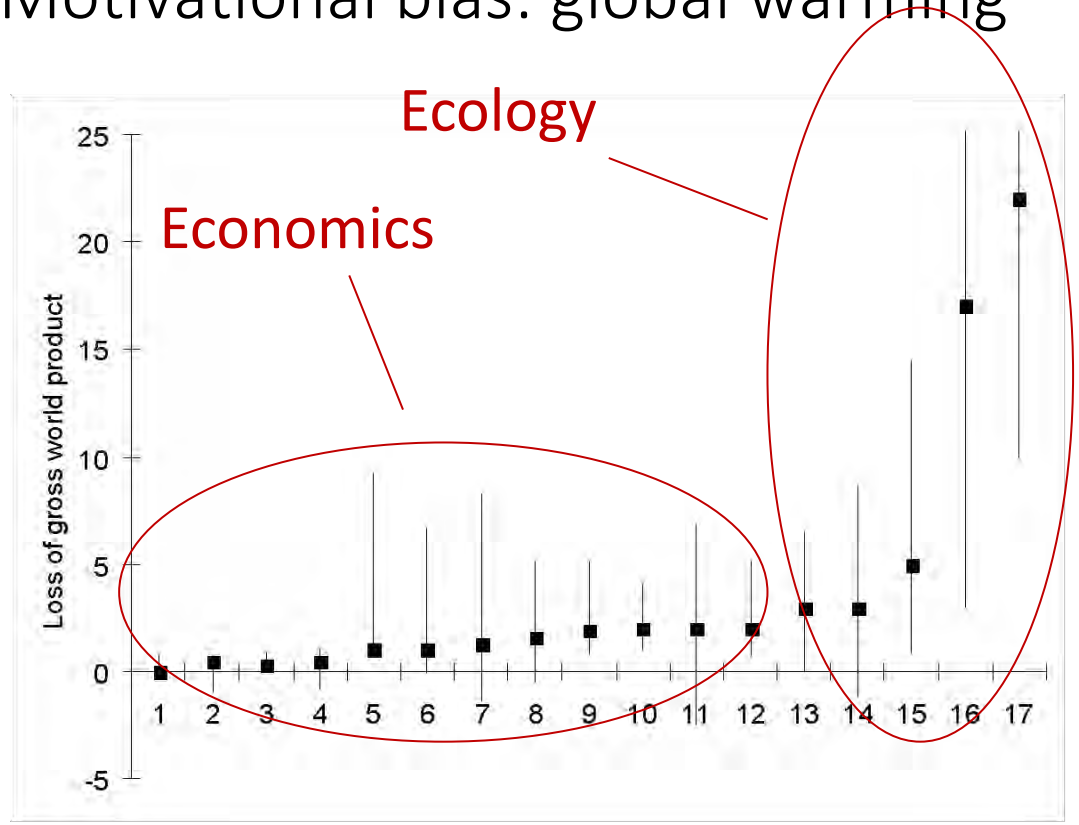
Challenger

NASA 1985: *'the risk of catastrophic failure is 1 in 100,000 launches'*

Chernobyl

Ukranian Minister of Power 1987: *'risk of a meltdown is 1 in 10,000 years*

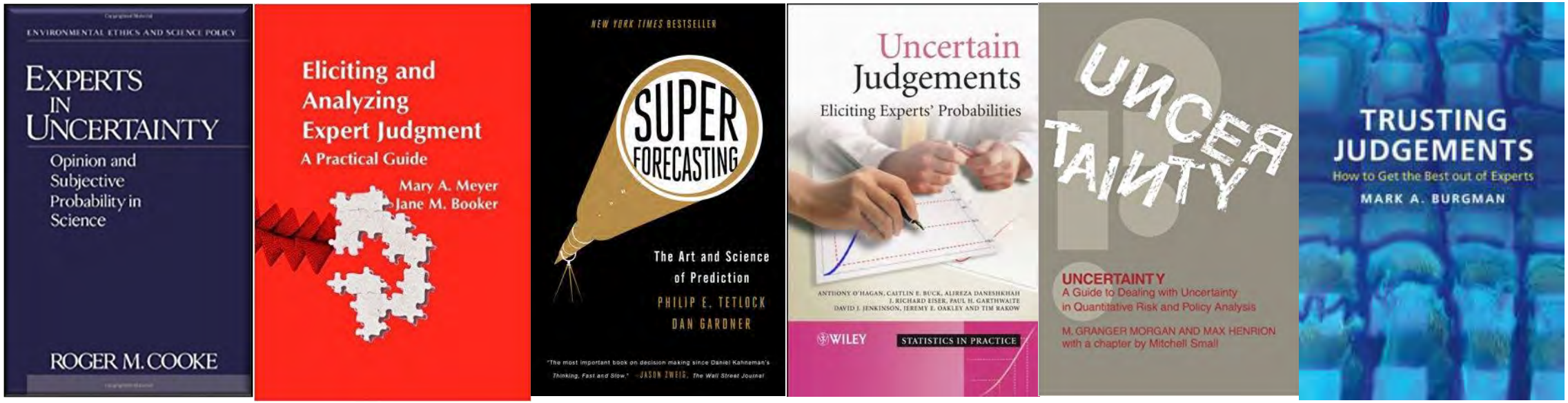
Motivational bias: global warming



Loss of gross world product resulting from a doubling of atmospheric CO_2 by 2050

(Kammen and Hassenzahl 1999)

Structured Elicitation Protocols

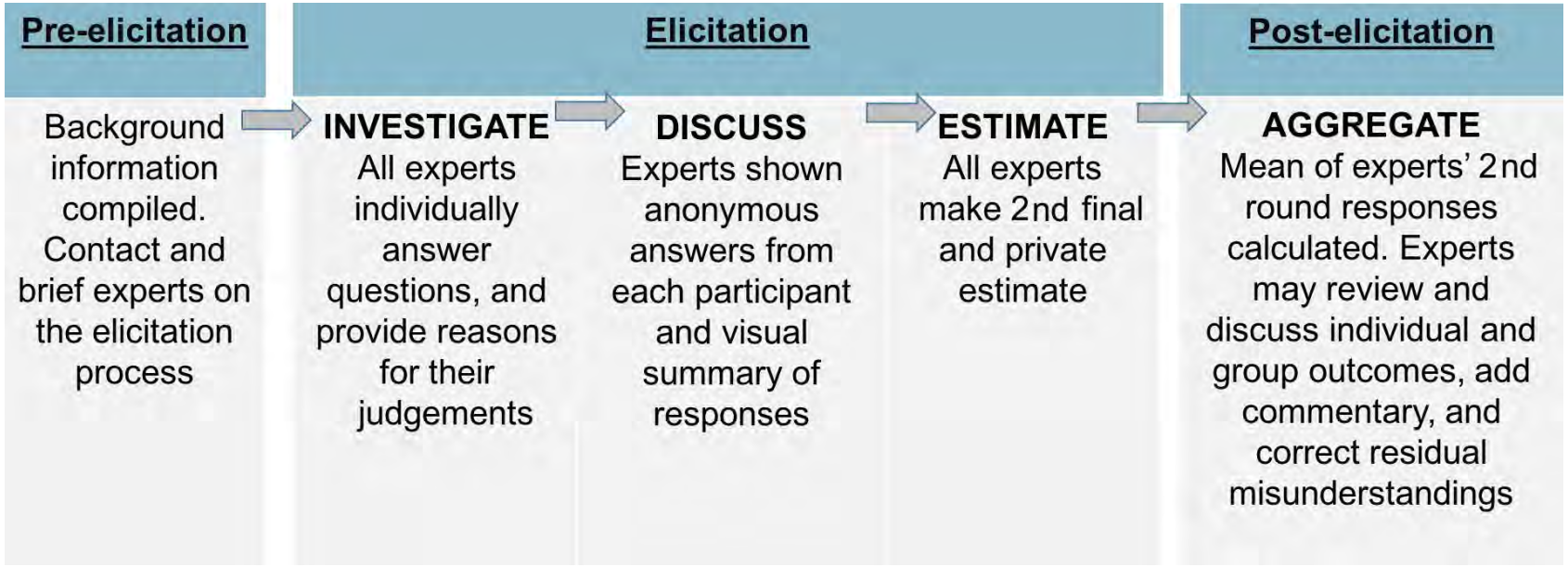


IDEA: A Delphi-like protocol with a twist

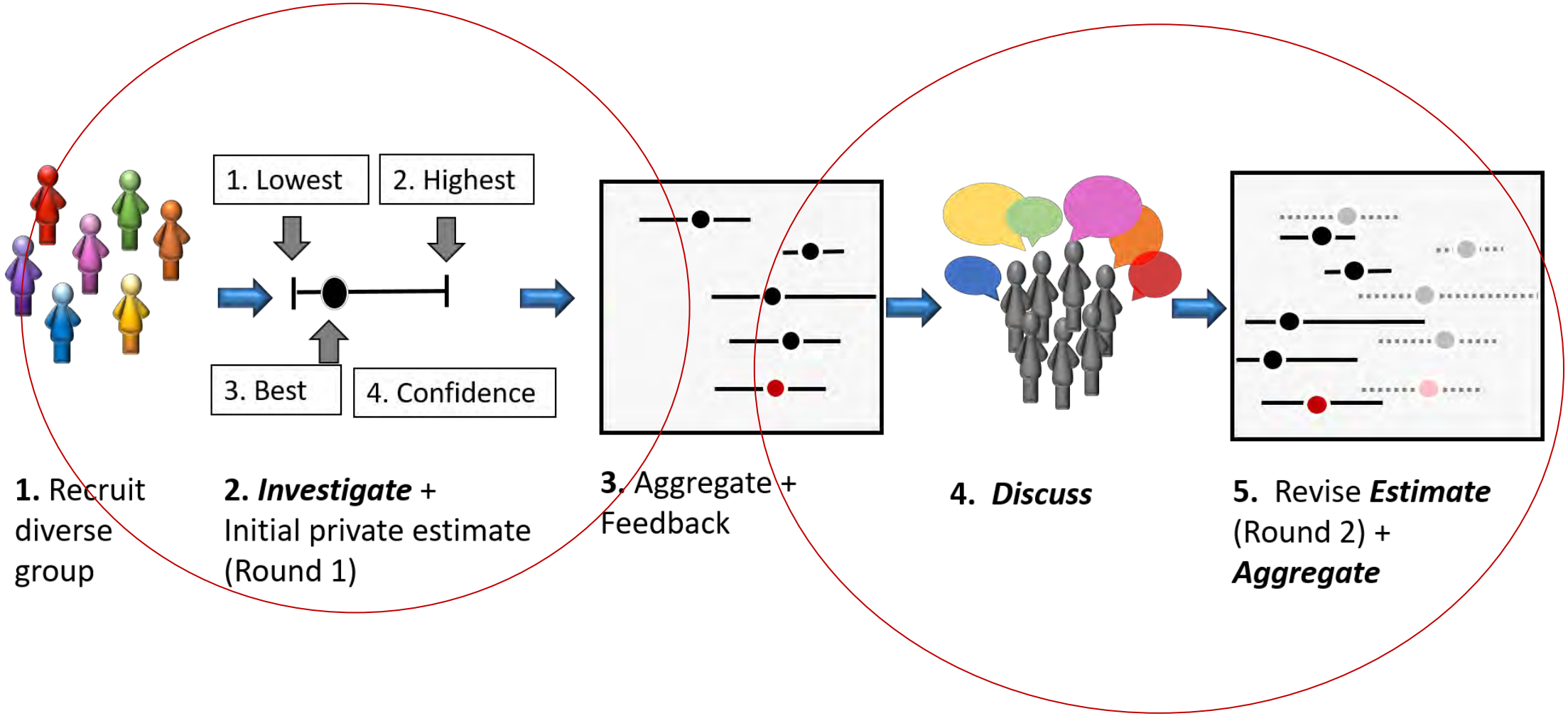
Strategies for better expert judgement

(i) A **four-point question format** for eliciting quantities to mitigate the overconfidence

(ii) The **structured interaction** of experts via facilitated discussion to promote 'better' group judgments



The IDEA protocol (Investigate, Discuss, Estimate, Aggregate)



1. Recruit diverse group

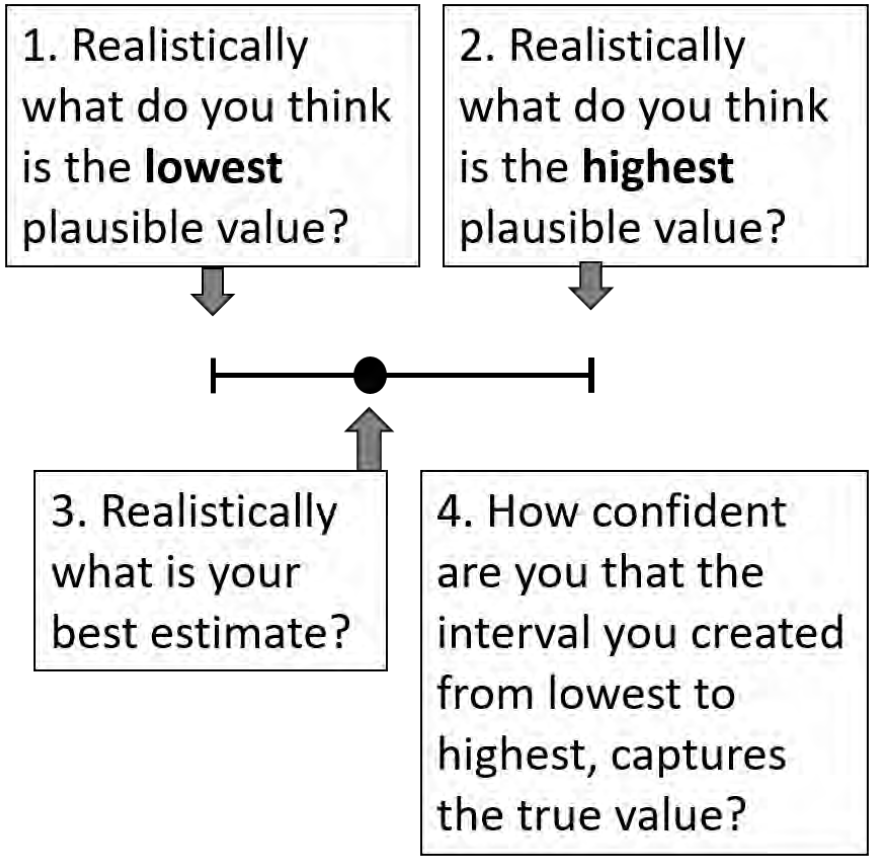
2. ***Investigate*** + Initial private estimate (Round 1)

3. Aggregate + Feedback

4. ***Discuss***

5. Revise ***Estimate*** (Round 2) + ***Aggregate***

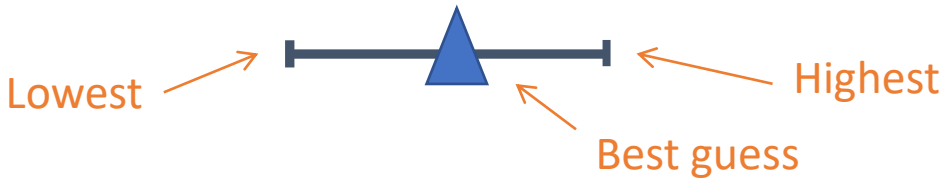
Four-step question format



Expert	5th (lower)	50th (best)	95th (upper)
1	2	12	34
2	4	15	50
3	7	9	40
4	20	22	23
Average	8.25	14.5	36.75

Interval judgements

- First step towards better judgements: express uncertainty through intervals (common in environmental science)



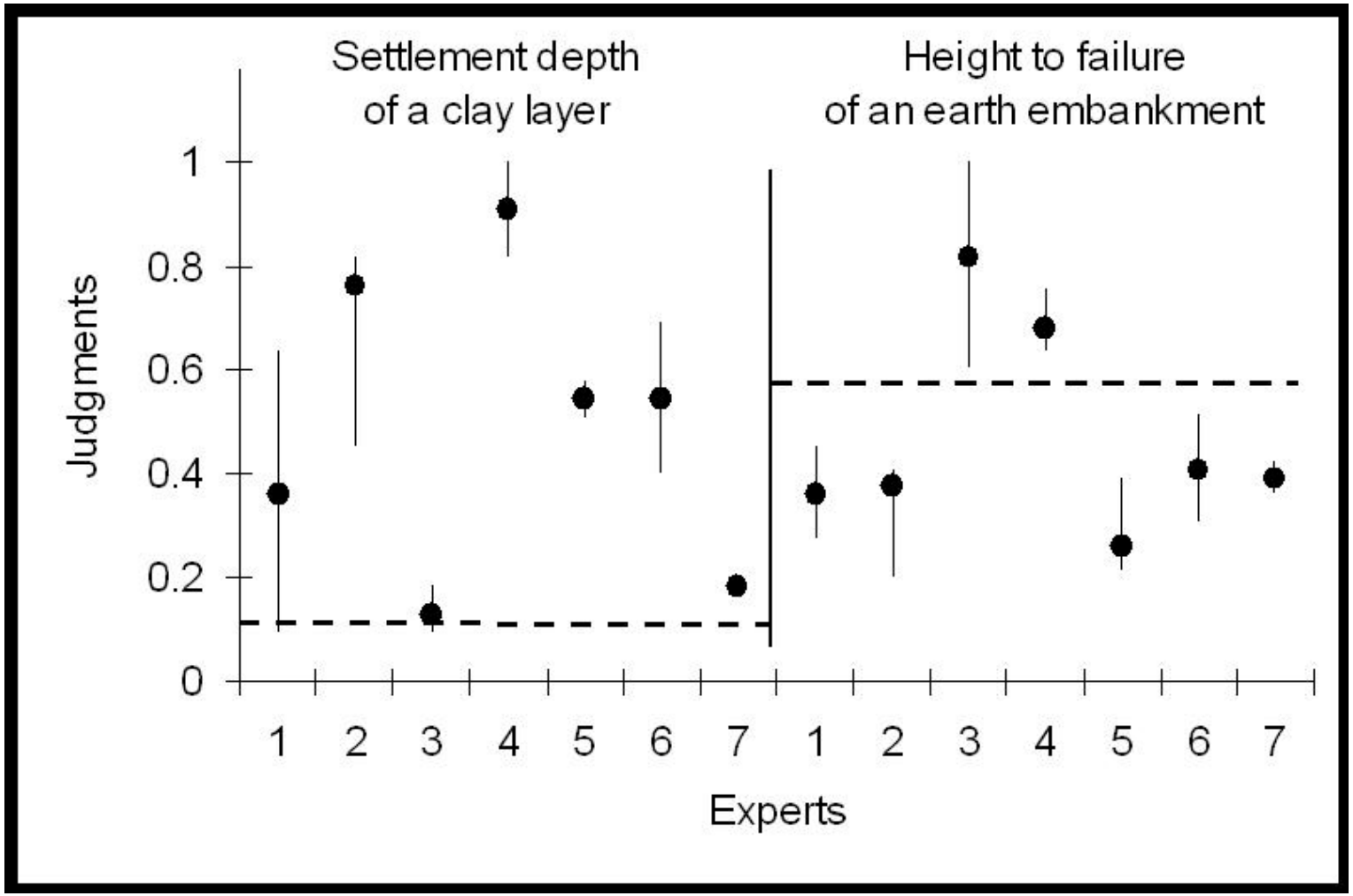
- Generally has a pre-assigned confidence,

e.g. “Provide an interval which you are 90% sure contains the true number of native fish species listed in the Murray River catchment”.

- BUT people are insensitive to pre-assigned confidence levels

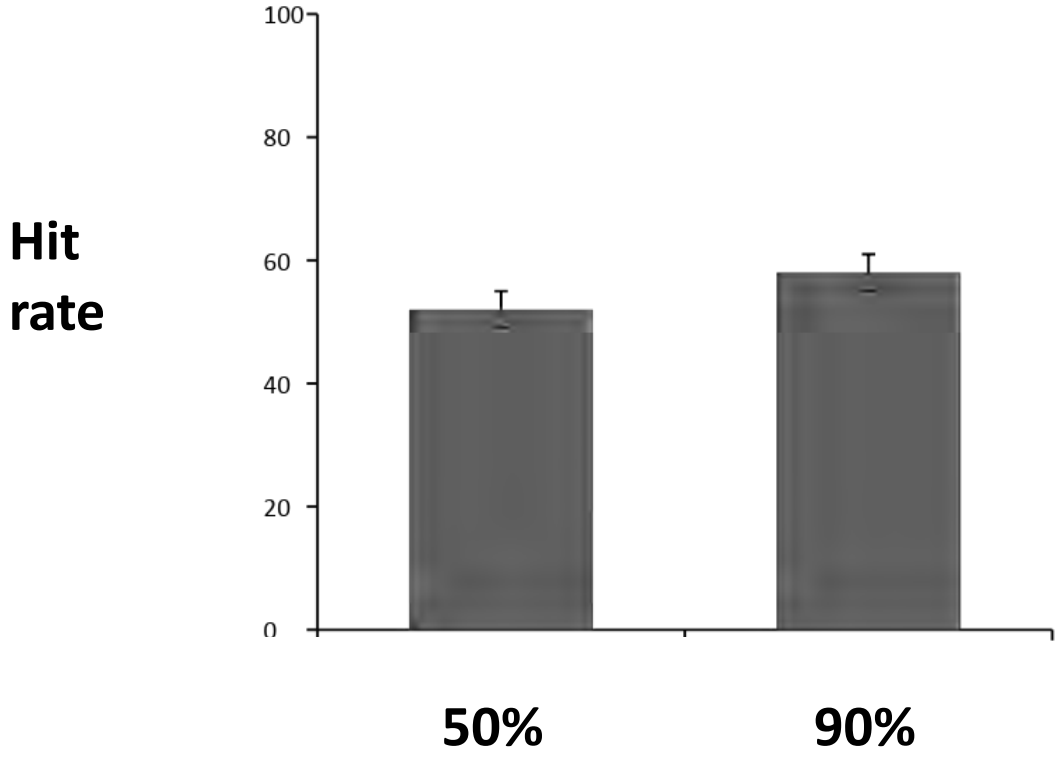
90% intervals tend to contain the answer only 50% of the time
→ **Overconfidence** (undue confidence in the intervals provided)

Overconfident intervals



(Hynes and Vanmarche 1977)

When 90% Confidence Intervals are 50% Certain:



Teigen and Jorgenson 2004

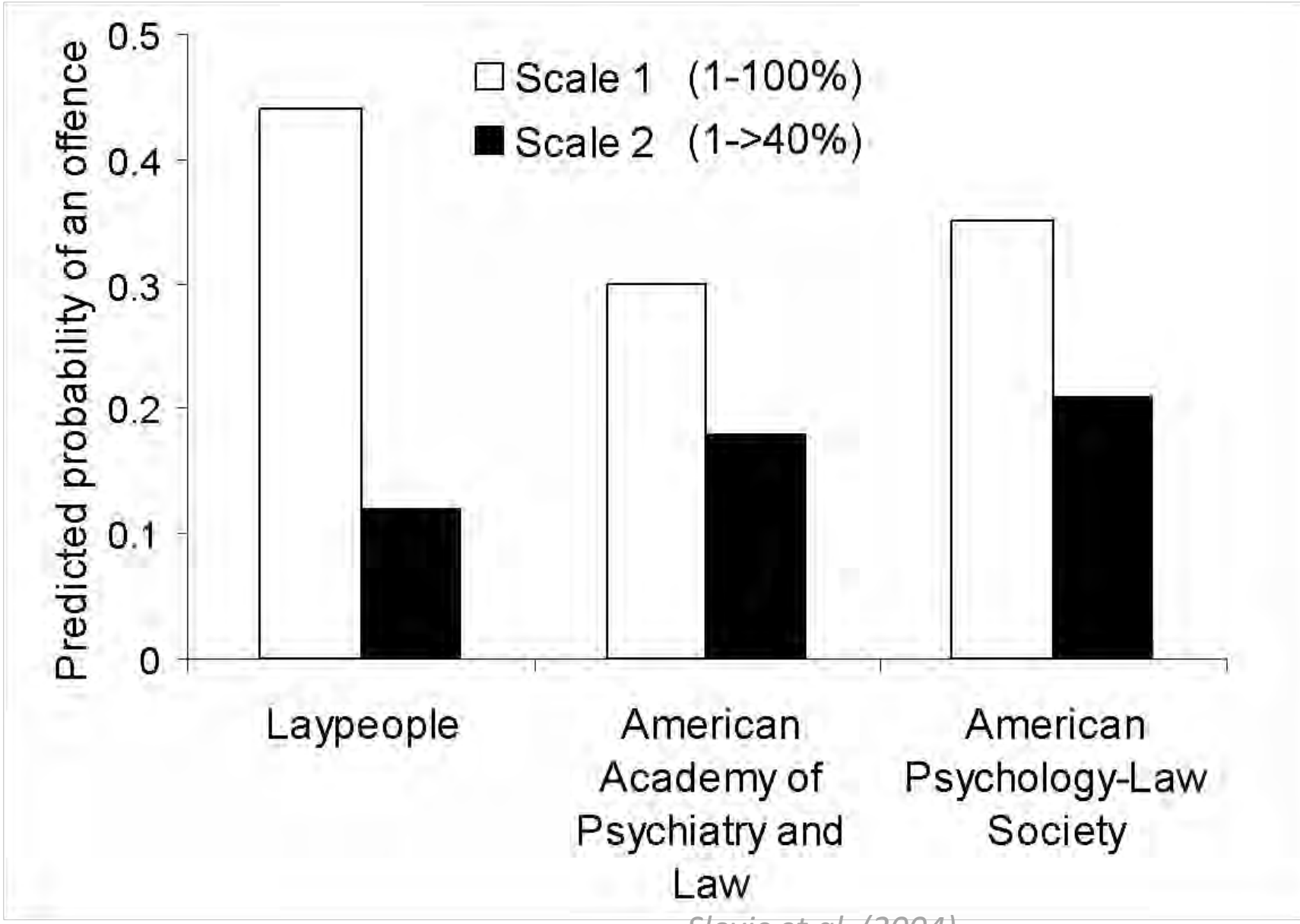
The influence of question format

1 10 20 30 40 50 60 70 80 90 100%

Scale 1: Range 1 - 100%

1 2 5 10 15 20 25 30 35 40 >40%

Scale 2: Range 1 - >40%

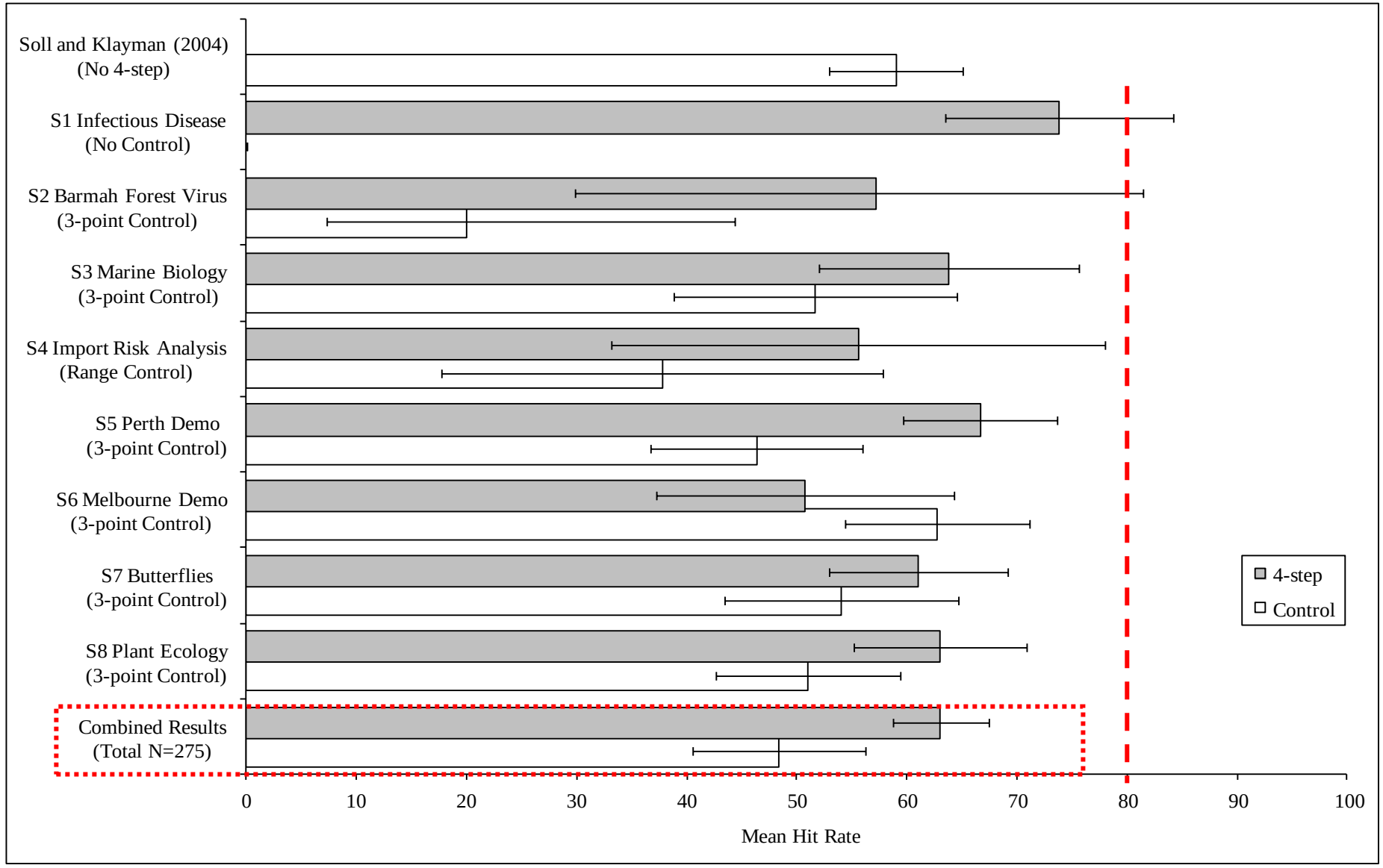


Slovic et al. (2004)

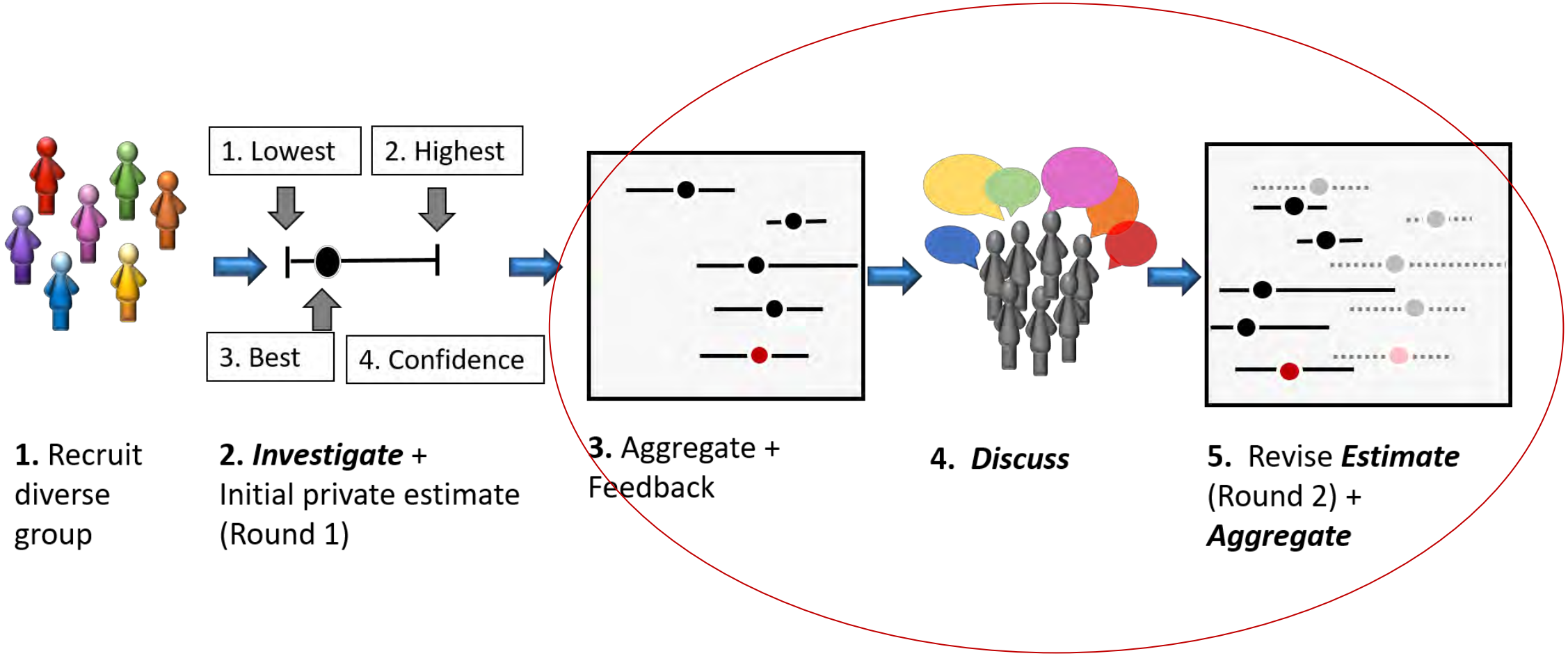
Elicitation Format with Examples	Average overconfidence
Range (one-point) I am 80% sure that this happened between ____and ____	41%
Two-point I am 90% sure that this happened after ____ I am 90% sure that this happened before ____	23%
Three-point I am 90% sure that this happened after ____ I am 90% sure that this happened before ____ I think it's equally likely that this happened after or before ____	14%
Four-step Realistically, what do you think the lowest plausible value is? ____ Realistically, what do you think the highest plausible value is? ____ Realistically, what is your best estimate? ____ How confident are you that the interval you created, from lowest to highest, could capture the true value? ____	12%

Speirs-Bridge et al. 2010

Some results – 3 v 4 step (for quantities)



The IDEA protocol (Investigate, Discuss, Estimate, Aggregate)



1. Recruit diverse group

2. ***Investigate*** + Initial private estimate (Round 1)

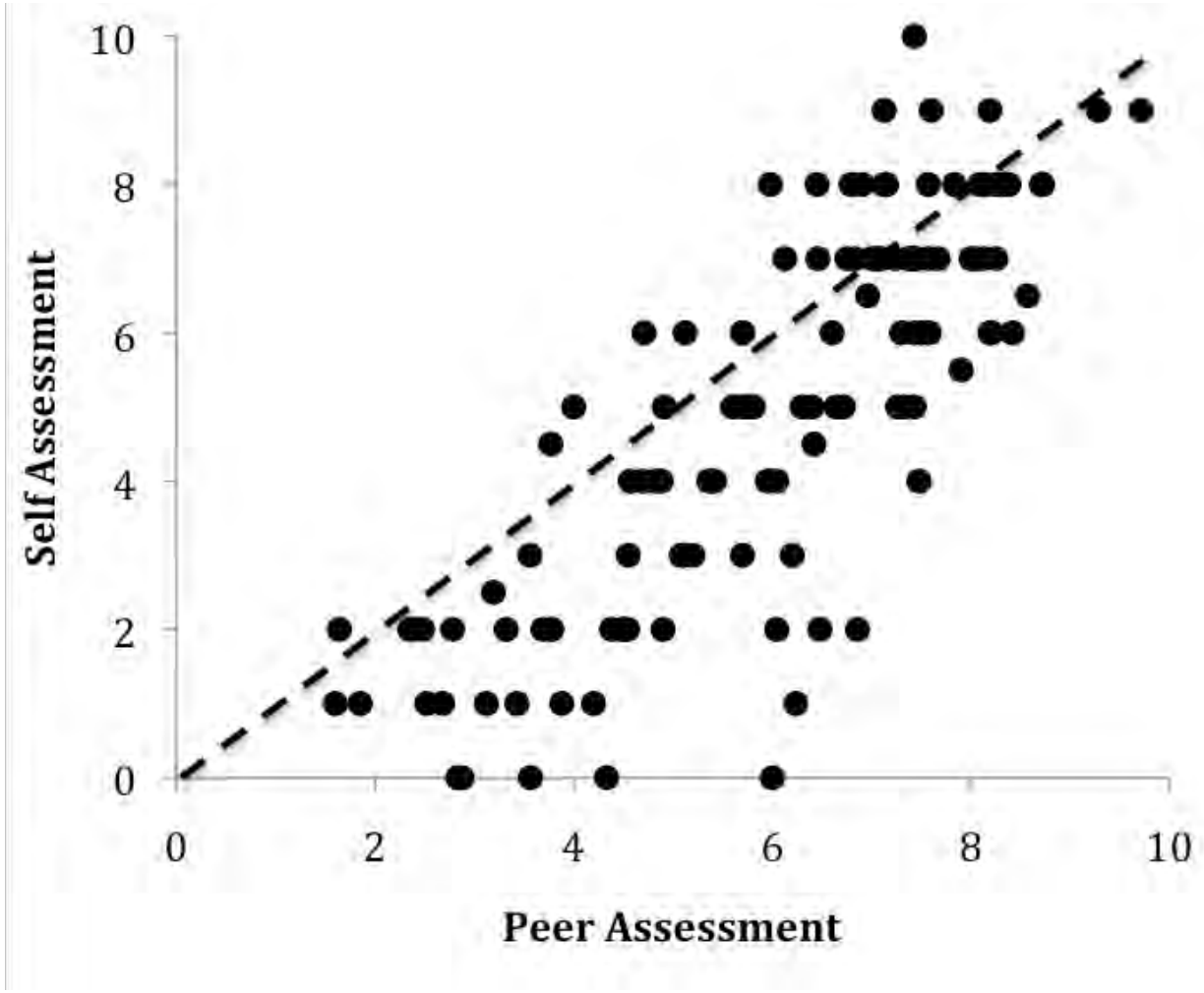
3. Aggregate + Feedback

4. ***Discuss***

5. Revise ***Estimate*** (Round 2) + ***Aggregate***

Worksh op	Discipline	# exper ts	Range of years of relevant experience (median)	Range of qualifications	Range of number of publications
1	Animal and plant Biosecurity and Quarantine	21	0 – 37 (17.5)	BSc, BASc, BVSc, BCom, Grad. Dip., MSc, PhD	0 – 113
2	Animal and plant biosecurity	24	0 – 39 (12)	BSc, MSc, MBA, MCom, PhD	0 – 270
3	Conservation biology, herpetology	13	0 – 42 (15)	BA, BSc, BSc (Hons), M Env Studies, PhD	0 – 45
4	Public health, medicine and epidemiology	25	0 – 45 (12)	BEng, BSc, BEcon, LLB, MBBS, Grad. Dip., MA, MSc, MBA, PhD	0 – 220
5	Risk analysis, biosecurity	20	0 – 40 (6)	BEng, BSc, BVSc, MBBS, Grad. Dip., MA, MBA, PhD	0 – 225
6	Weed risk assessment	14	0 – 50 (6.5)	BSc, MSc, PhD	1 – 220

Peer versus self assessments

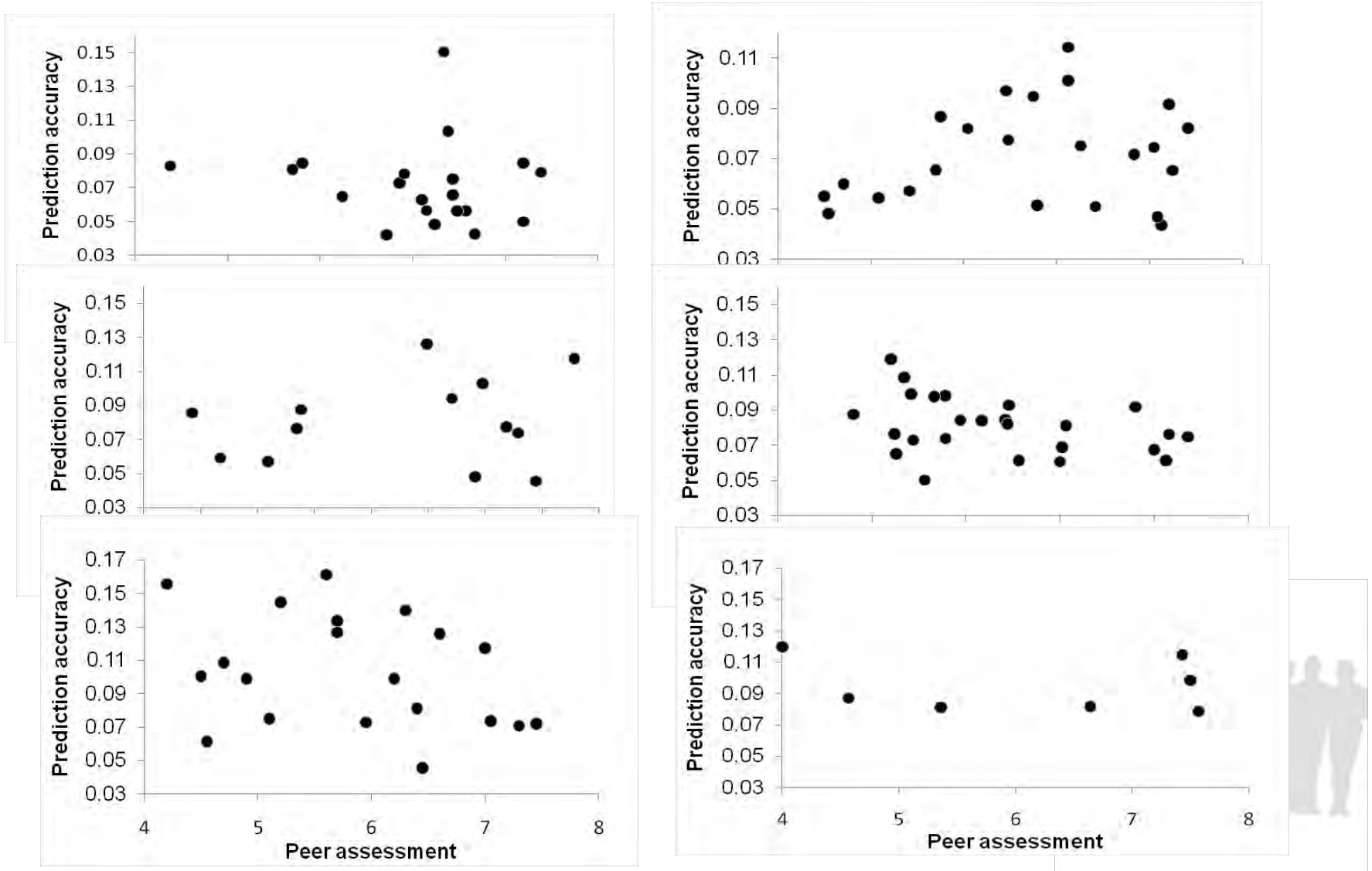


***r* = 0.85**

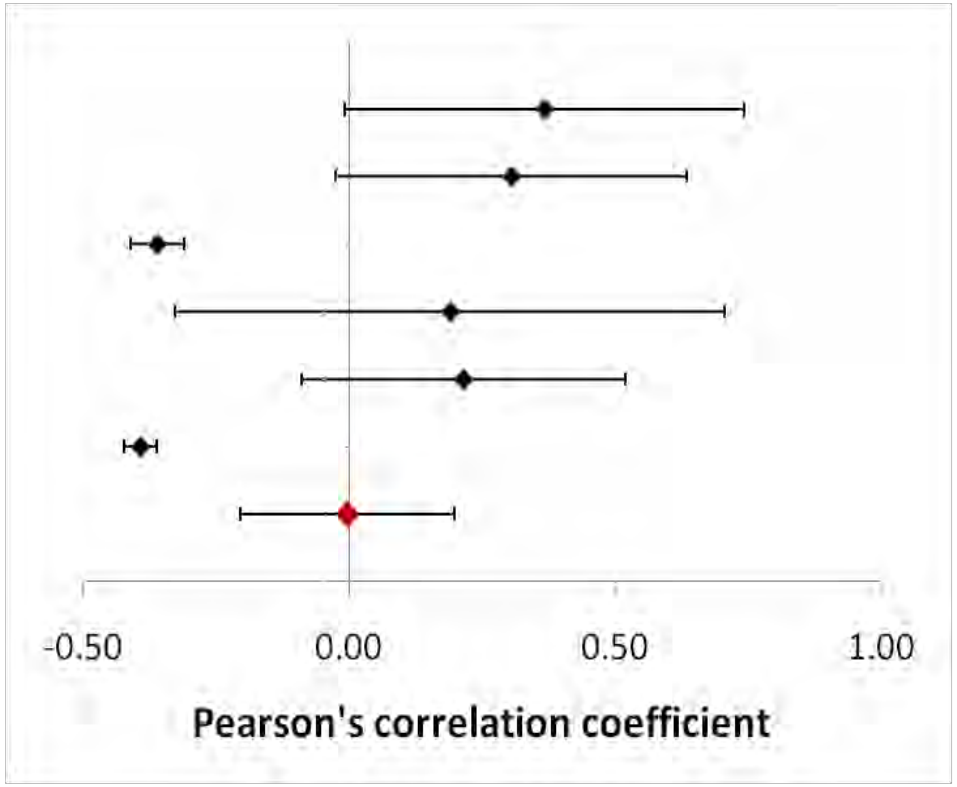
Range across workshops

[0.675 to 0.944]

Do peer assessments correlate with performance?



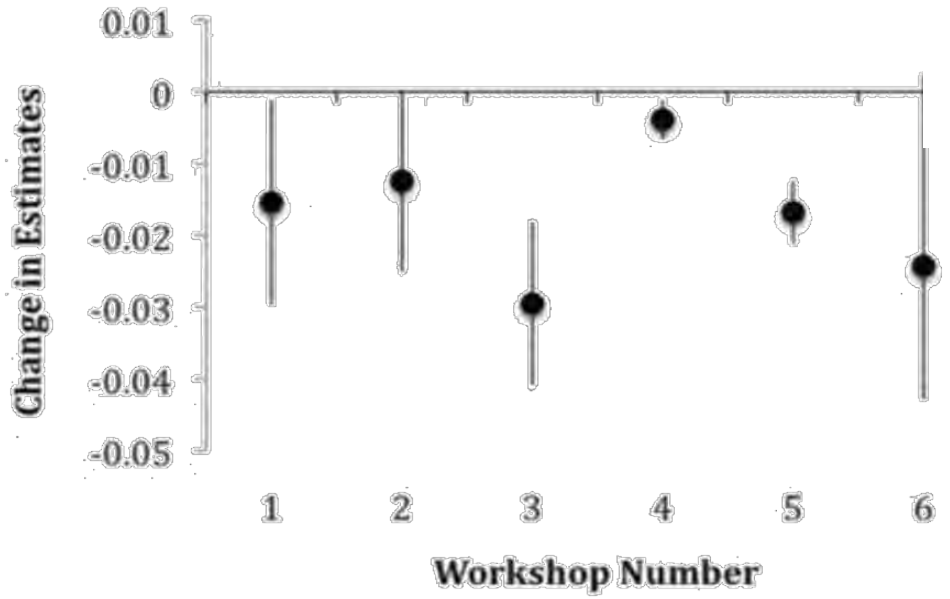
Peer assessment versus performance



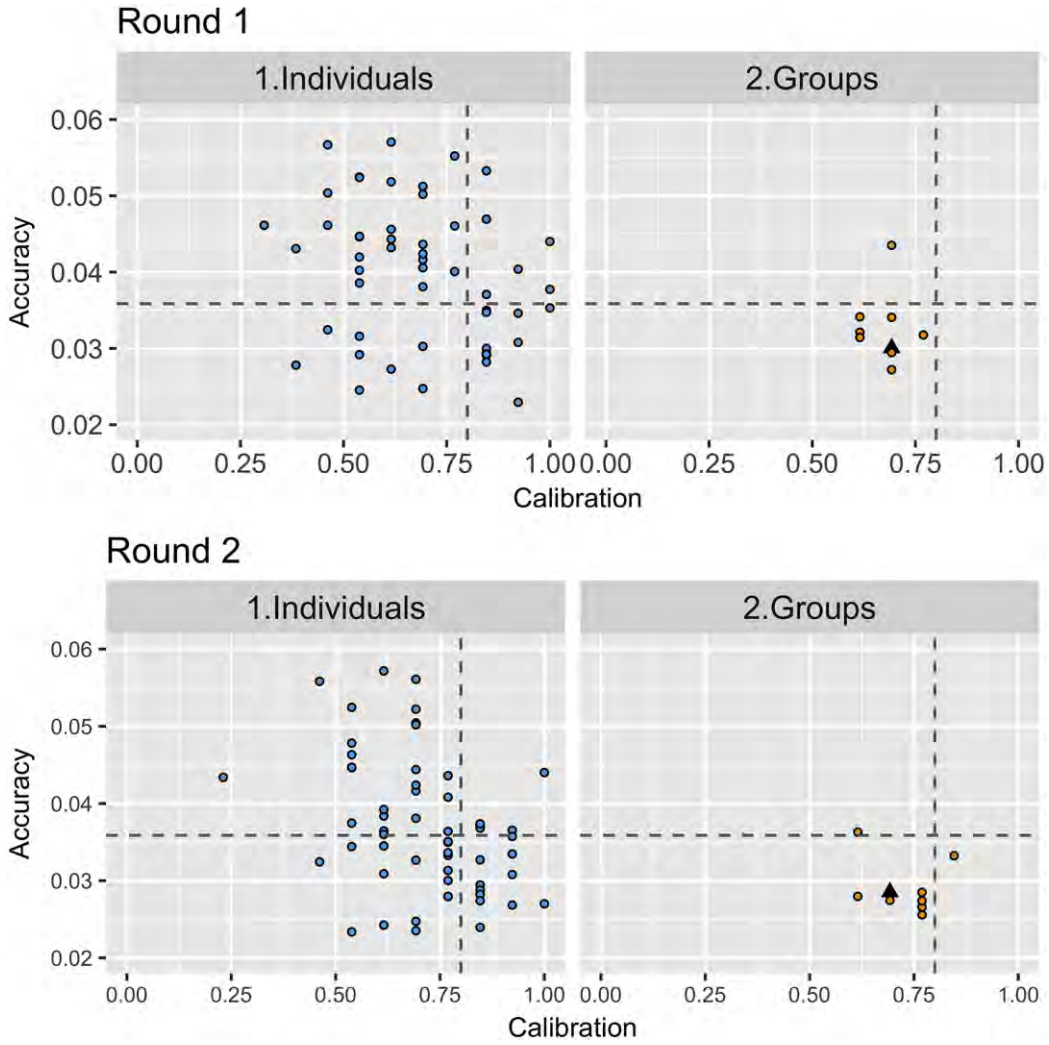
Workshop	Peer assessment versus prediction accuracy
1	-0.391 (n=20)
2	0.215 (n=19)
3	0.190 (n=13)
4	- 0.360 (n=25)
5	0.305 (n=20)
6	0.367 (n=14)

Benefits of (diverse) groups and facilitated discussion:

Round 1 to Round 2 reduction in error



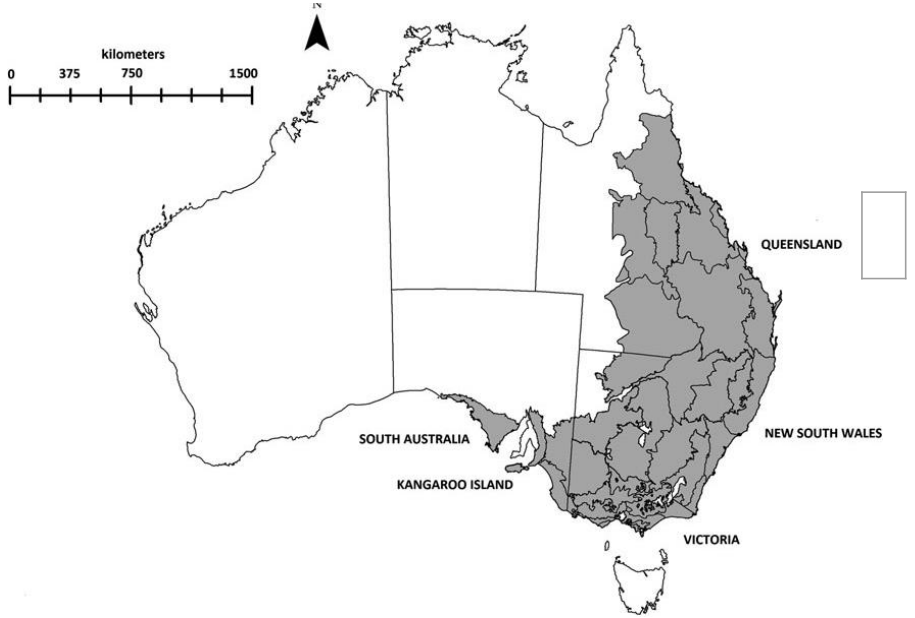
The group average improvement in accuracy (ALRE) following discussion.

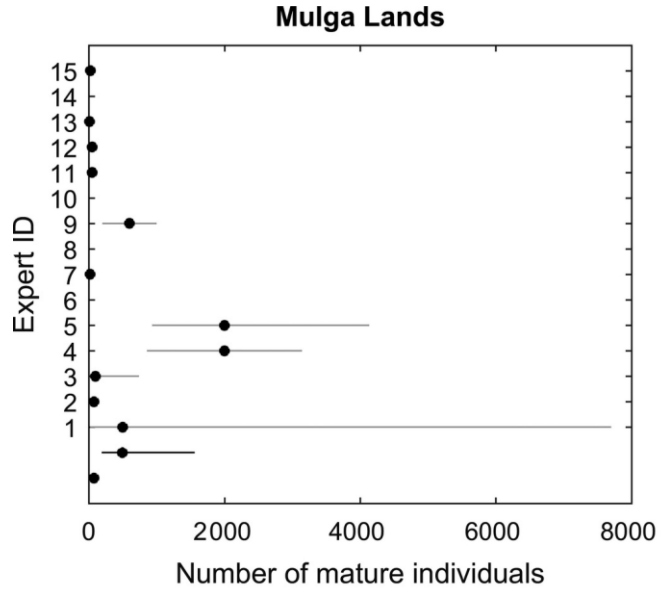


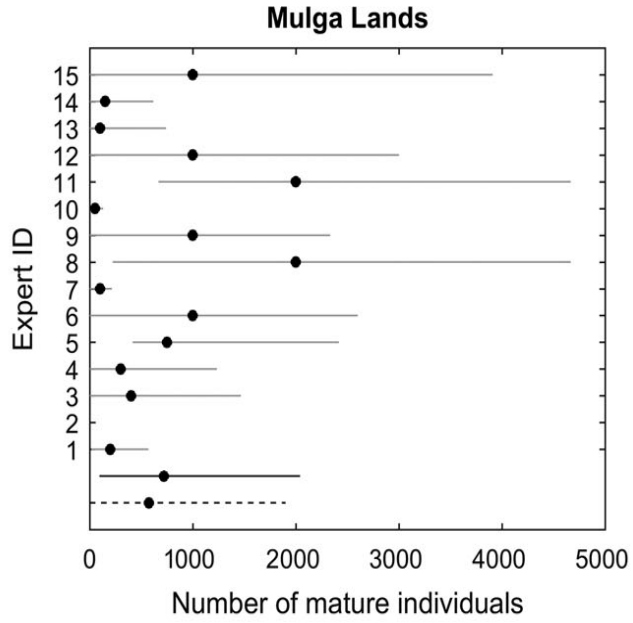
Estimating population sizes

What is your estimate of the current population size (number of mature individuals) for the Mulga Lands Bioregion?

© Australian Koala Foundation

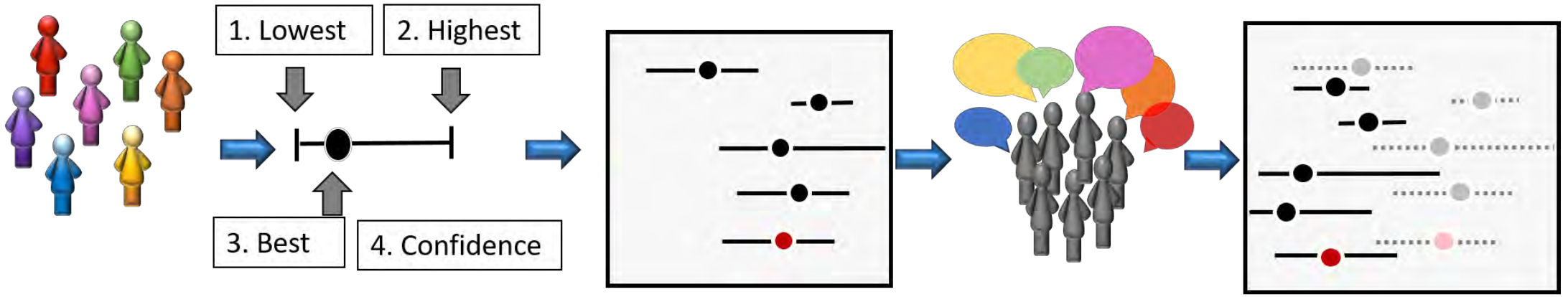






Source: Adams-Hosking et al. (2016

The IDEA protocol (Investigate, Discuss, Estimate, Aggregate)



1. Recruit diverse group

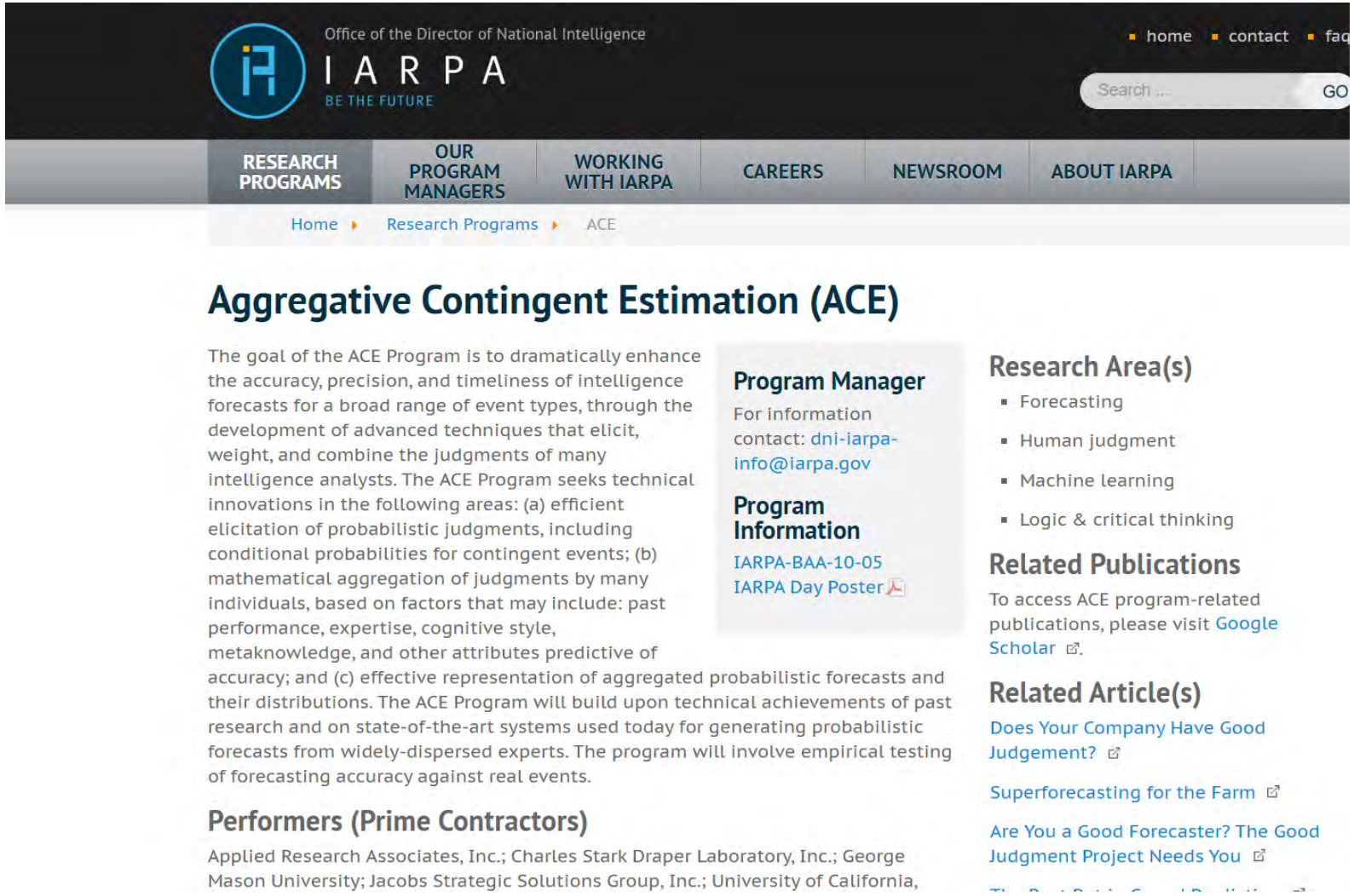
2. ***Investigate*** + Initial private estimate (Round 1)

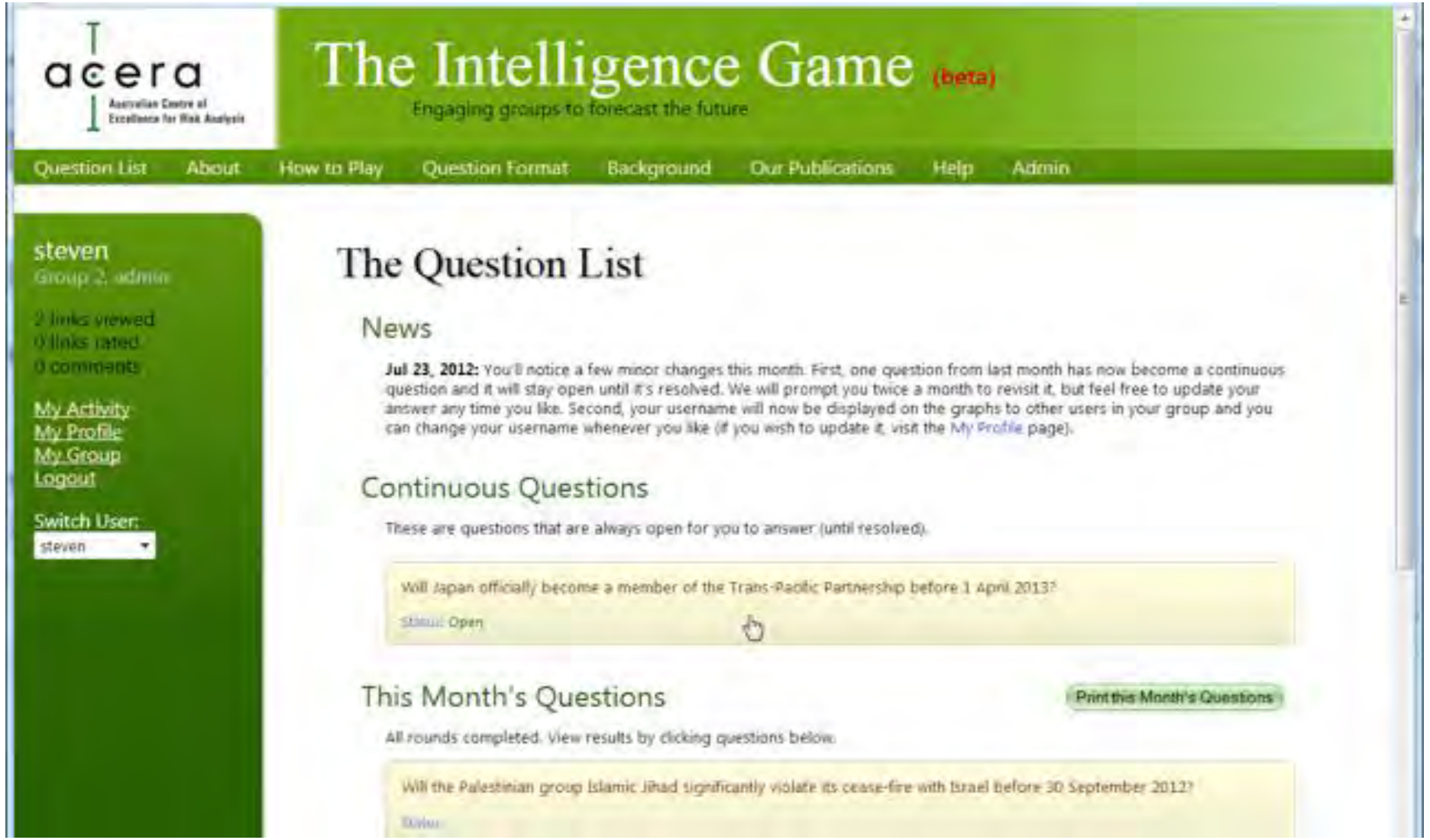
3. Aggregate + Feedback

4. ***Discuss***

5. Revise ***Estimate*** (Round 2) + ***Aggregate***

Geopolitical forecasting tournament




acera
Australian Centre of Excellence for Risk Analysis
The Intelligence Game (beta)
Engaging groups to forecast the future
Question List
About
How to Play
Question Format
Background
Our Publications
Help
Admin
steven
Group 2, admin
2 links viewed
0 links rated
0 comments
My Activity
My Profile
My Group
Logout
Switch User:
steven
The Question List
News
Jul 23, 2012: You'll notice a few minor changes this month. First, one question from last month has now become a continuous question and it will stay open until it's resolved. We will prompt you twice a month to revisit it, but feel free to update your answer any time you like. Second, your username will now be displayed on the graphs to other users in your group and you can change your username whenever you like (if you wish to update it, visit the My Profile page).
Continuous Questions
These are questions that are always open for you to answer (until resolved).
Will Japan officially become a member of the Trans-Pacific Partnership before 1 April 2013?
Status: Open
This Month's Questions
Print this Month's Questions
All rounds completed. View results by clicking questions below.
Will the Palestinian group Islamic Jihad significantly violate its cease-fire with Israel before 30 September 2012?
Status:

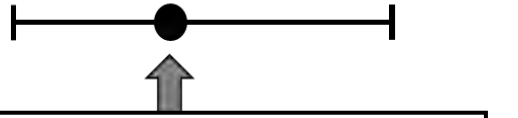
Three-step Question Format (for probabilities)

e.g. Will Liu Yandong be selected as a member of the next Politburo Standing Committee of the Communist Party of China?

1. When you think of reasons that make this unlikely to happen, how sure do you feel that X will happen?

2. When you think of reasons that make this **likely** to happen, how sure do you feel that X will happen?

3. Finally, consider the balance of evidence. If you had to put a single figure on the probability of this outcome, what would it be?

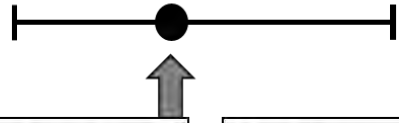


1. Realistically what do you think is the **lowest** plausible value?

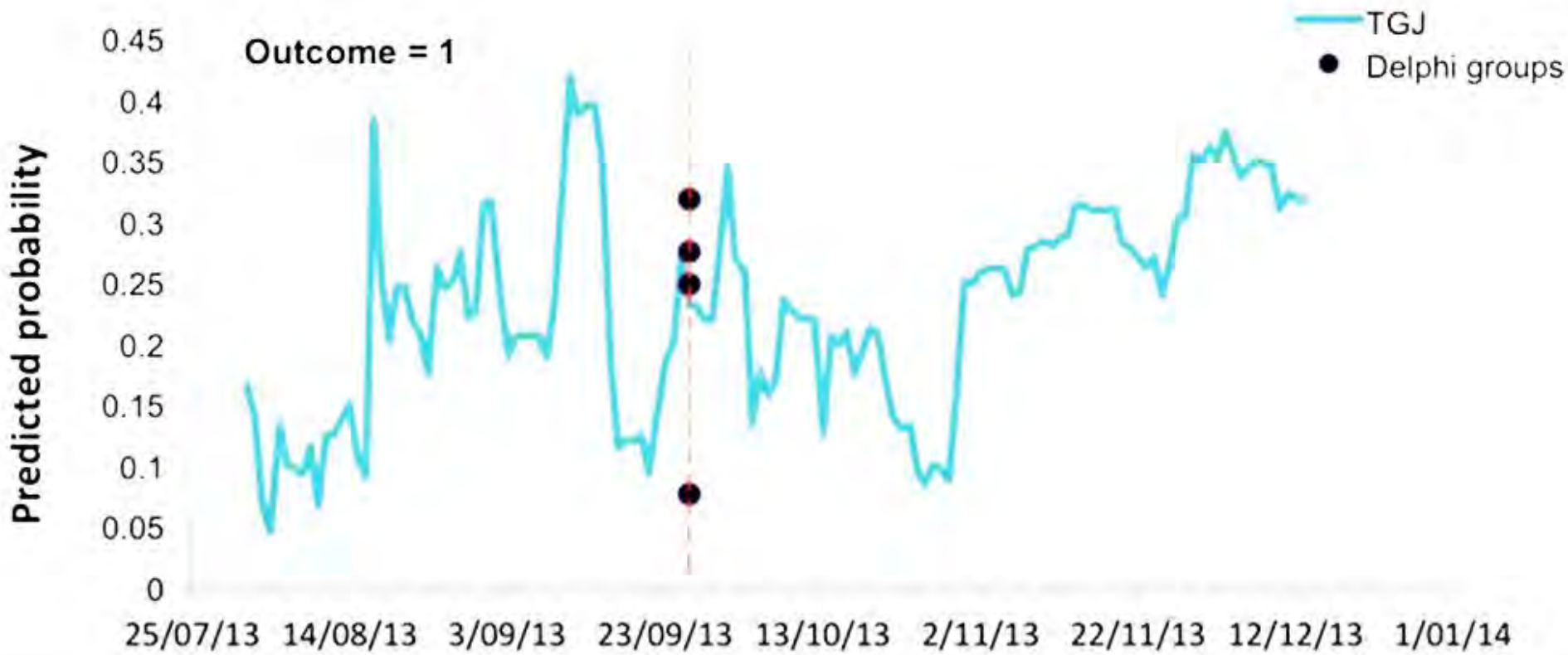
2. Realistically what do you think is the **highest** plausible value?

3. Realistically what is your best estimate?

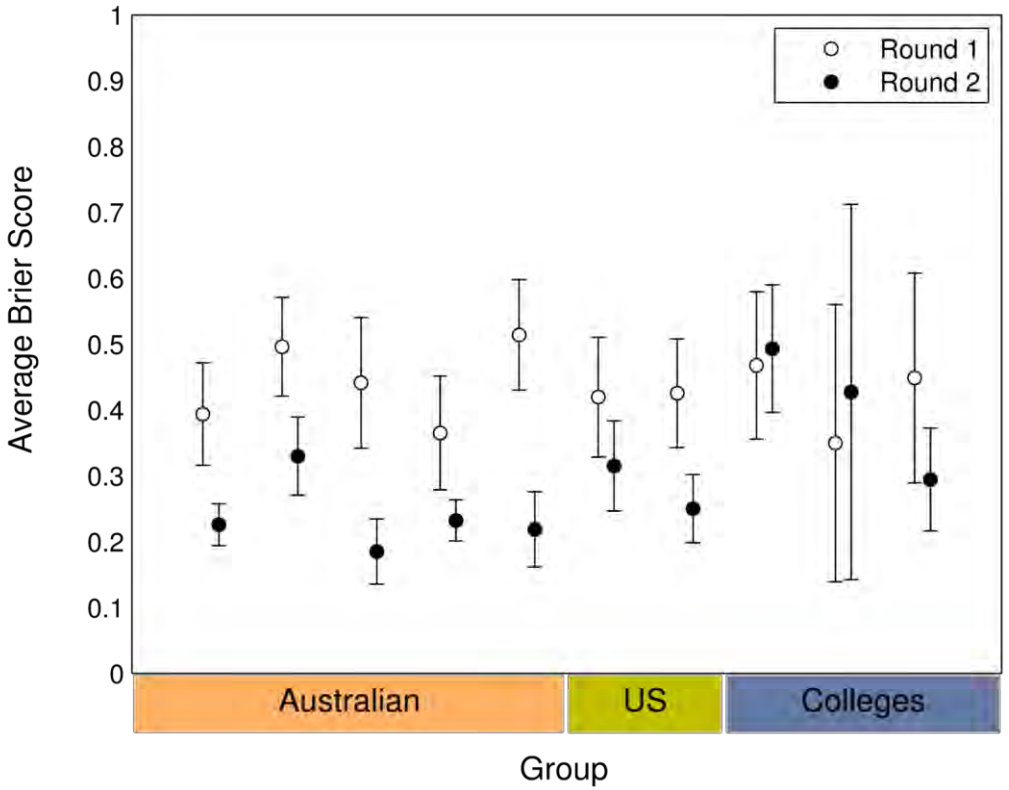
4. How confident are you that the interval you created from lowest to highest, captures the true value?

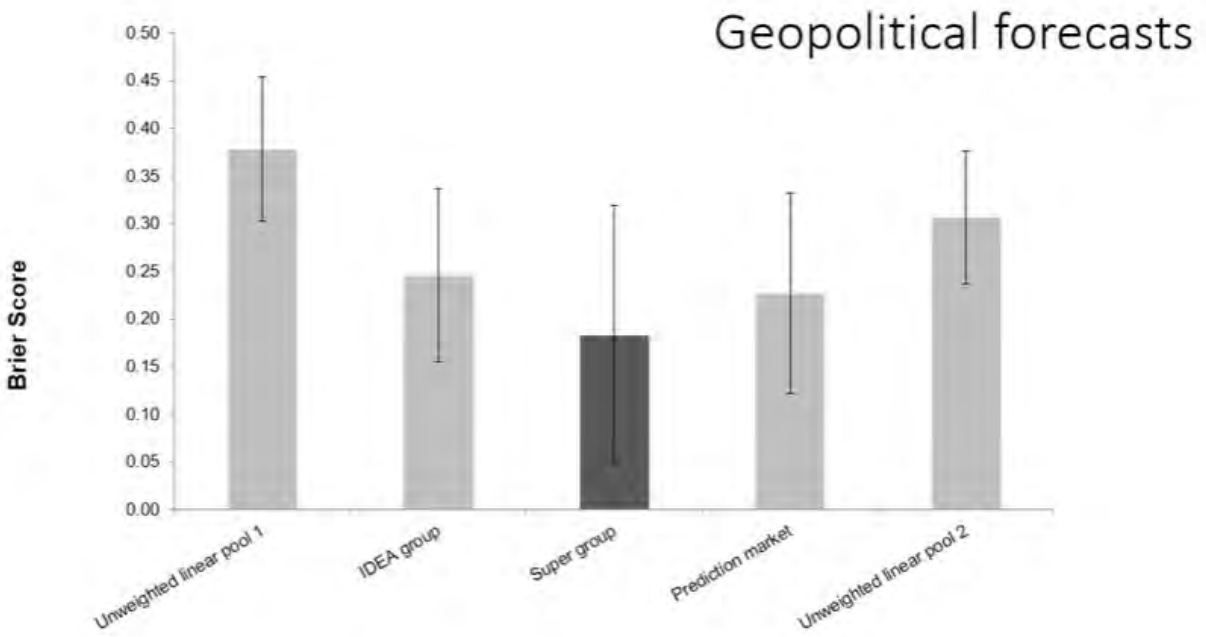


Will Chinese armed forces or maritime law enforcement forces attempt to interdict or make physical contact with at least one U.S. government naval vessel or airplane or Japanese government naval vessel or airplane that it claims is in its territorial waters or airspace, before 1 May 2014?

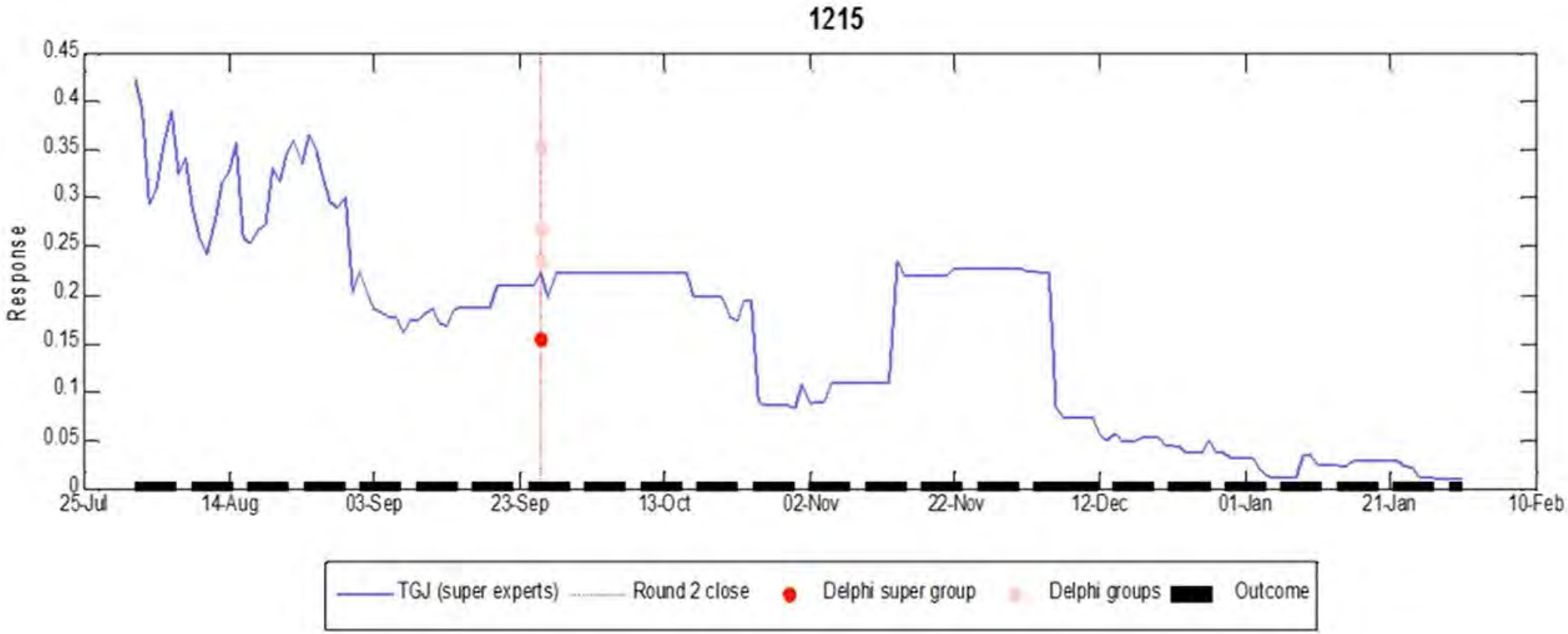


Results: Effect of Discussion

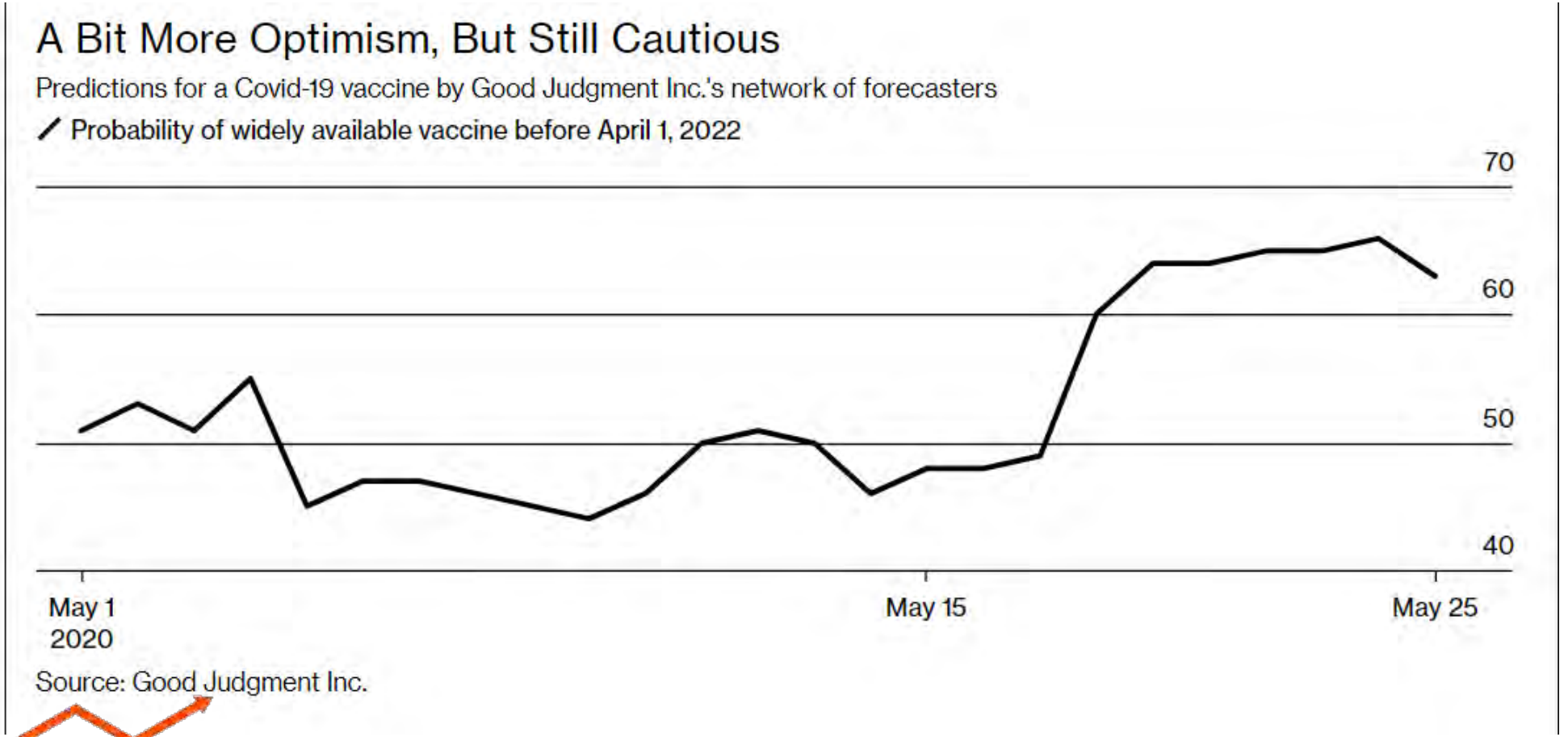




For those who revised their judgments, **second round estimates outperformed first round estimates.**


1215
Response
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
25-Jul
14-Aug
03-Sep
23-Sep
13-Oct
02-Nov
22-Nov
12-Dec
01-Jan
21-Jan
10-Feb
TGJ (super experts)
Round 2 close
Delphi super group
Delphi groups
Outcome

COVID Recovery metrics



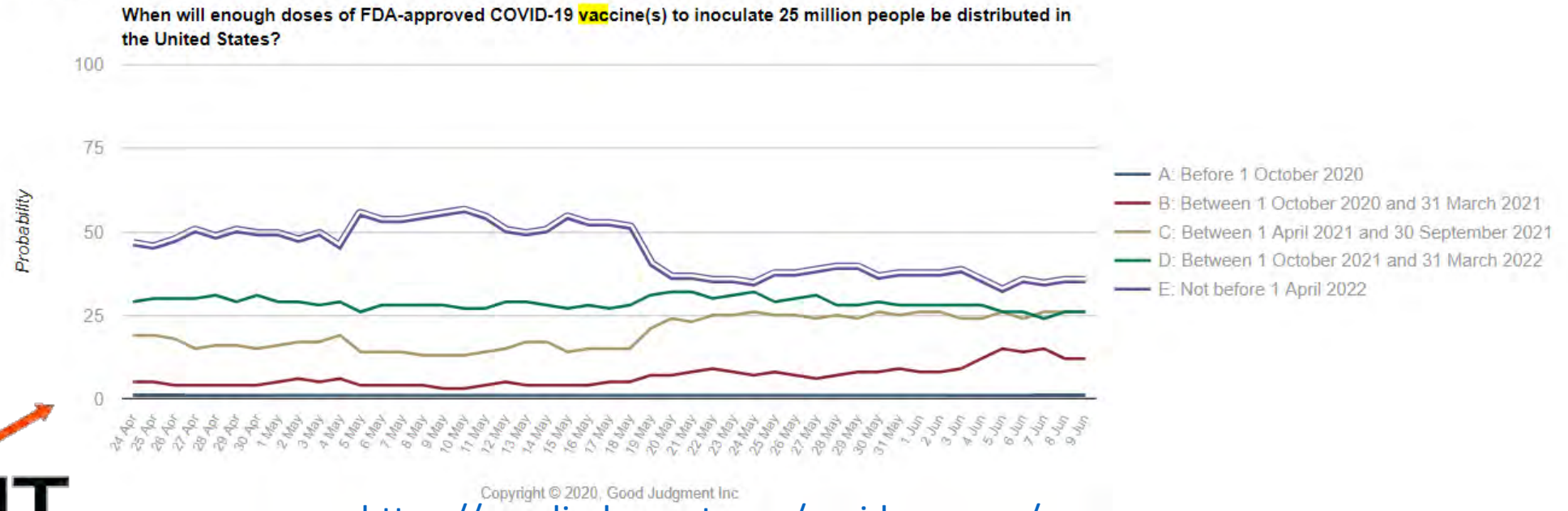
GOOD JUDGMENT

https://goodjudgment.com/covidrecovery/

COVID Recovery metrics

When will enough doses of FDA-approved COVID-19 vaccine(s) to inoculate 25 million people be distributed in the United States?	Today's Forecast
Today's Forecast:	
A Before 1 October 2020	1%
B Between 1 October 2020 and 31 March 2021	12%
C Between 1 April 2021 and 30 September 2021	26%
D Between 1 October 2021 and 31 March 2022	26%
E Not before 1 April 2022	35%

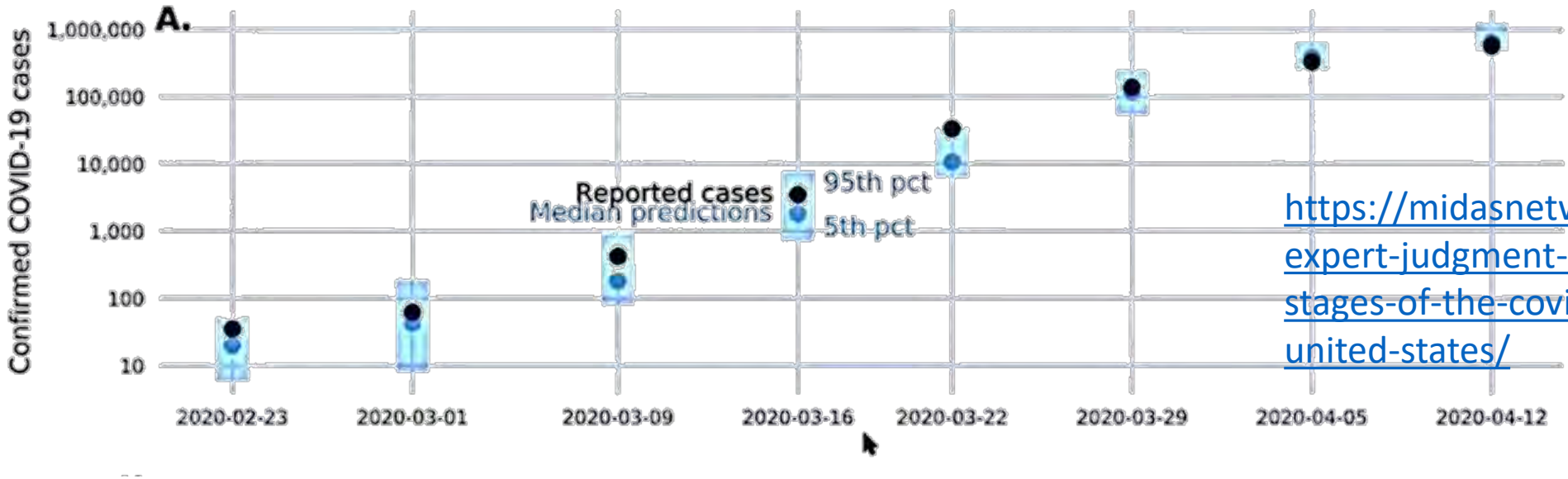
Forecast History:



GOOD JUDGMENT

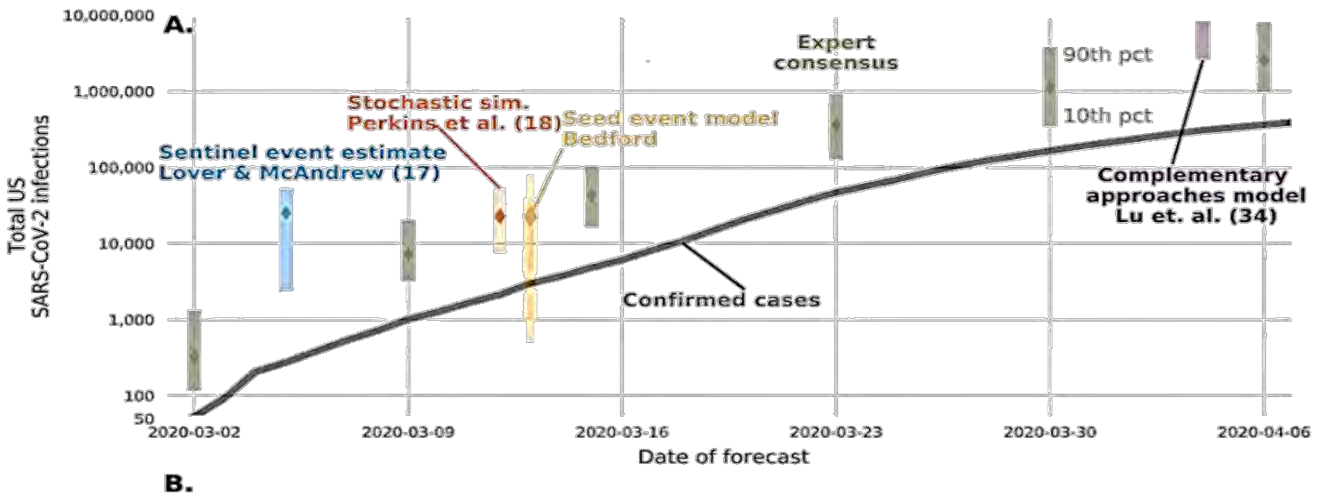
https://goodjudgment.com/covidrecovery/

Expert predictions of US COVID-19 cases



https://midasnetwork.us/midas-webinar-an-expert-judgment-model-to-predict-early-stages-of-the-covid-19-outbreak-in-the-united-states/

Predictions made 1 week in advance



Dr Thomas McAndrew, Reich lab
University of Amherst

http://reichlab.io/

http://www.thomasmcandrew.com/

Lessons learnt for better structured elicitations

Ask individuals to...

- Answer the same question in different ways (lowest, highest, most likely)
- Indicate confidence
- Examine estimates made by other people (feedback), consider counter-argument
- Revise original estimates after feedback
- Anticipate issues with conditional probabilities, base rates, …

Then, don't rely on individuals...

- Discuss questions to eliminate linguistic uncertainty
- Make groups diverse—age, gender, background and cognitive style, culture
- Aggregate mathematically – don't force a consensus
- Avoid group think— Delphi structures / anonymity in judgments

Acknowledgements

Mark Burgman
Anca Hanea
Victoria Hemming Bonnie Wintle
Terry Walshe
Fiona Fidler
Many more….

The experts involved

ceer
Centre for Environmental & Economic Research

Imperial College London

THE UNIVERSITY OF MELBOURNE

cebra
Centre of Excellence for Biosecurity Risk Analysis