Structured judgment and decision making

Budapest

European Commission and EKLIPSE

September, 2018

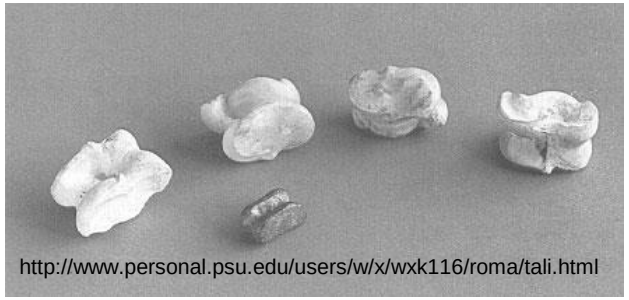
Mark Burgman
Imperial College London

Against the grain: in November last year, demonstrators took their grievances to the World Trade Organisation in Paris.

A Brief History of Risk

Greek, Hebrew and Roman systems used letters for numbers.
Lacked a numbering system that allowed calculations.

Hindu invention
of numerals



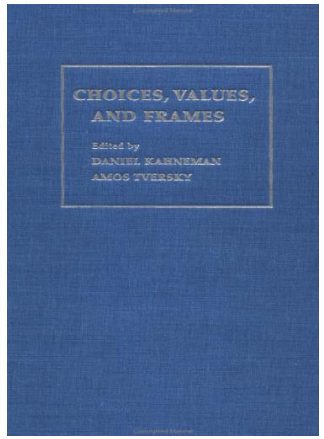
al-Khowârizmî, c. 825
Rules of arithmetic

Fibonacci, *Liber Abaci* (1202)
Fractions, roots, interest, profit, ...

Pascal *Port Royal Logic*, (1654), Fermat, Huygens,
Hobbs, … (probability as chance <u>and</u> belief)

Why people worry about the ‘wrong’ things

Judgements in uncertain situations are coloured by…

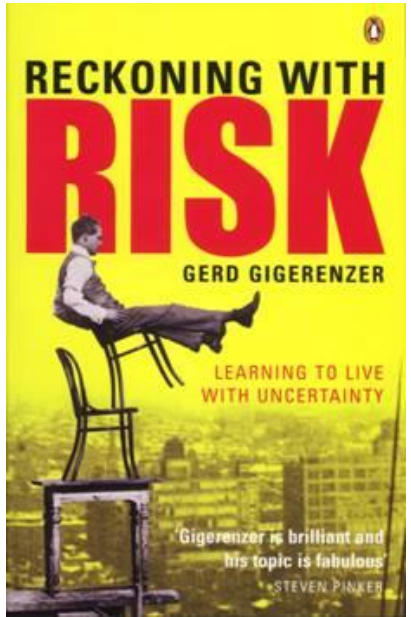


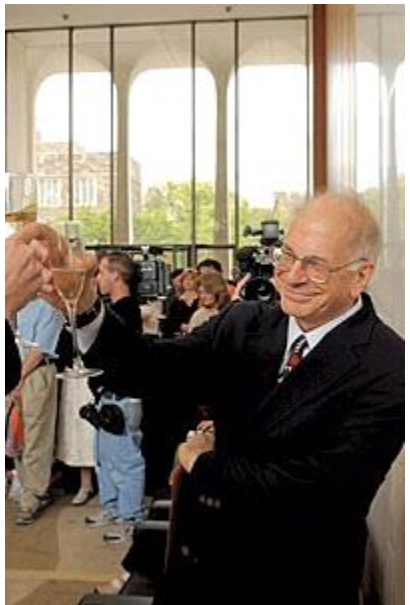
- overconfidence
- framing
- level of personal control
- understanding of the issues
- degree of personal experience
- dreadfulness of the outcome (kill size, outrage)
- equitability
- visibility
- status

Pathology of risk perception

leading to ...

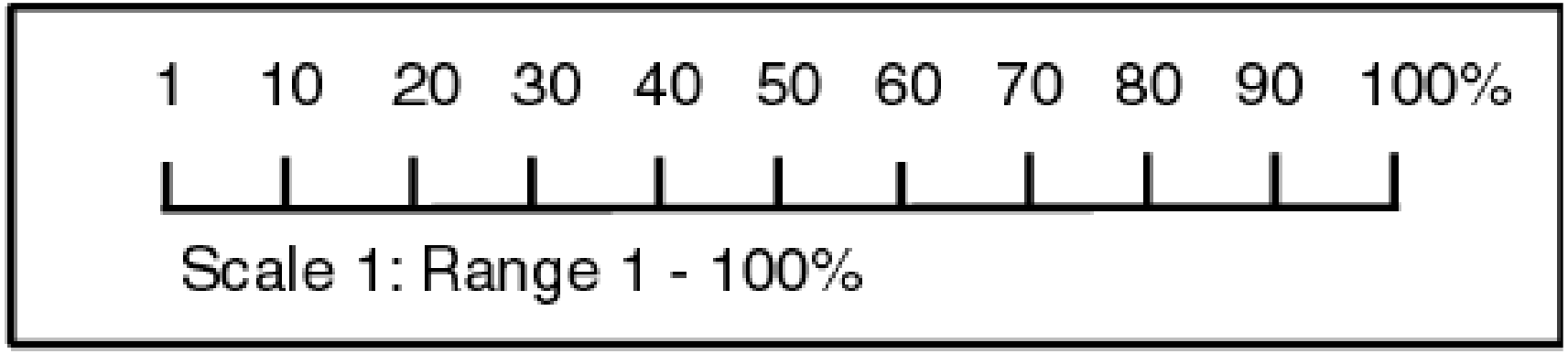
- insensitivity to sample size (Law of Small Numbers)
- framing effects
- overconfidence
- anchoring
- hindsight bias
- availability bias
- motivational bias
- risk aversion

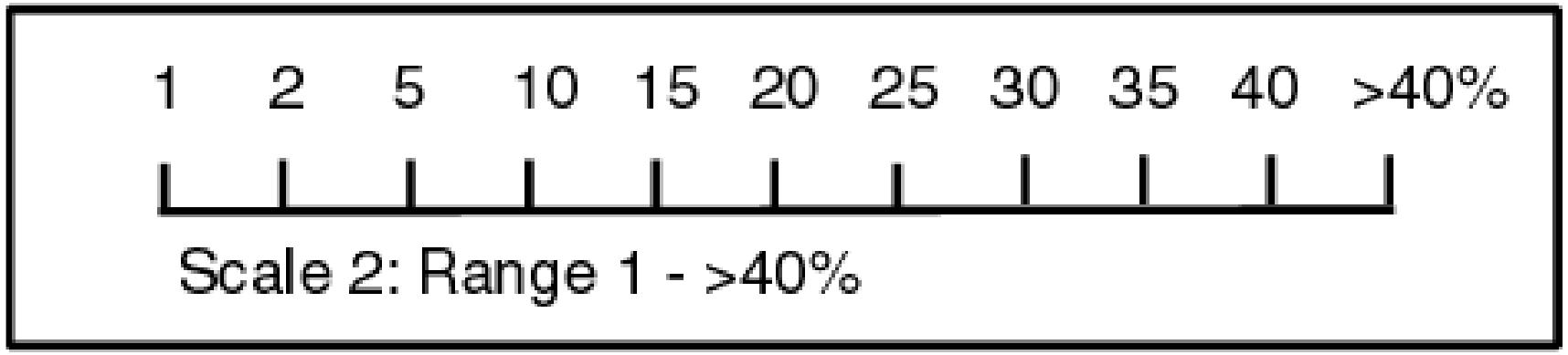


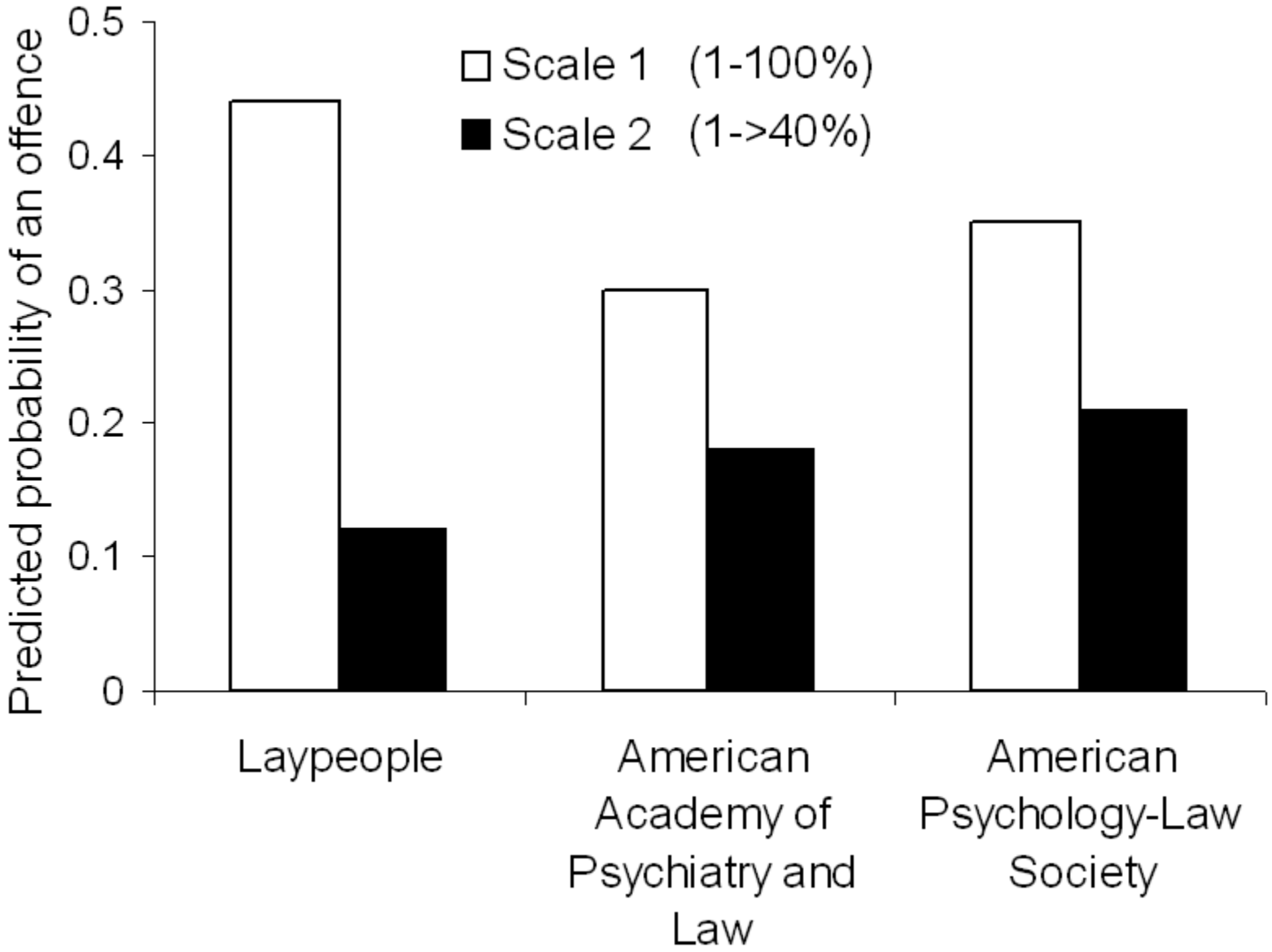


Framing

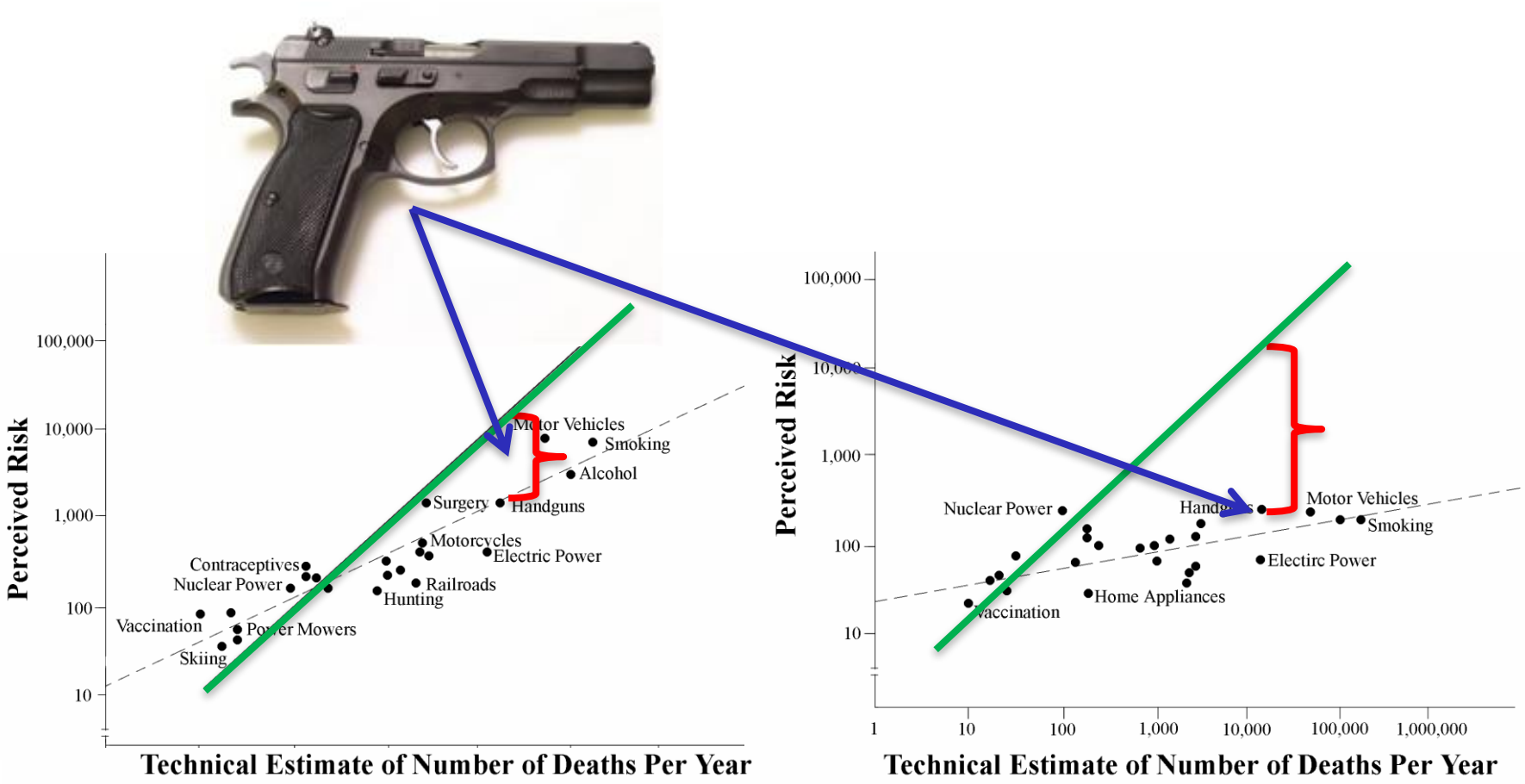
Two scales used to guide judgments about probabilities that violent criminals will re-offend (after Slovic et al. 2000)





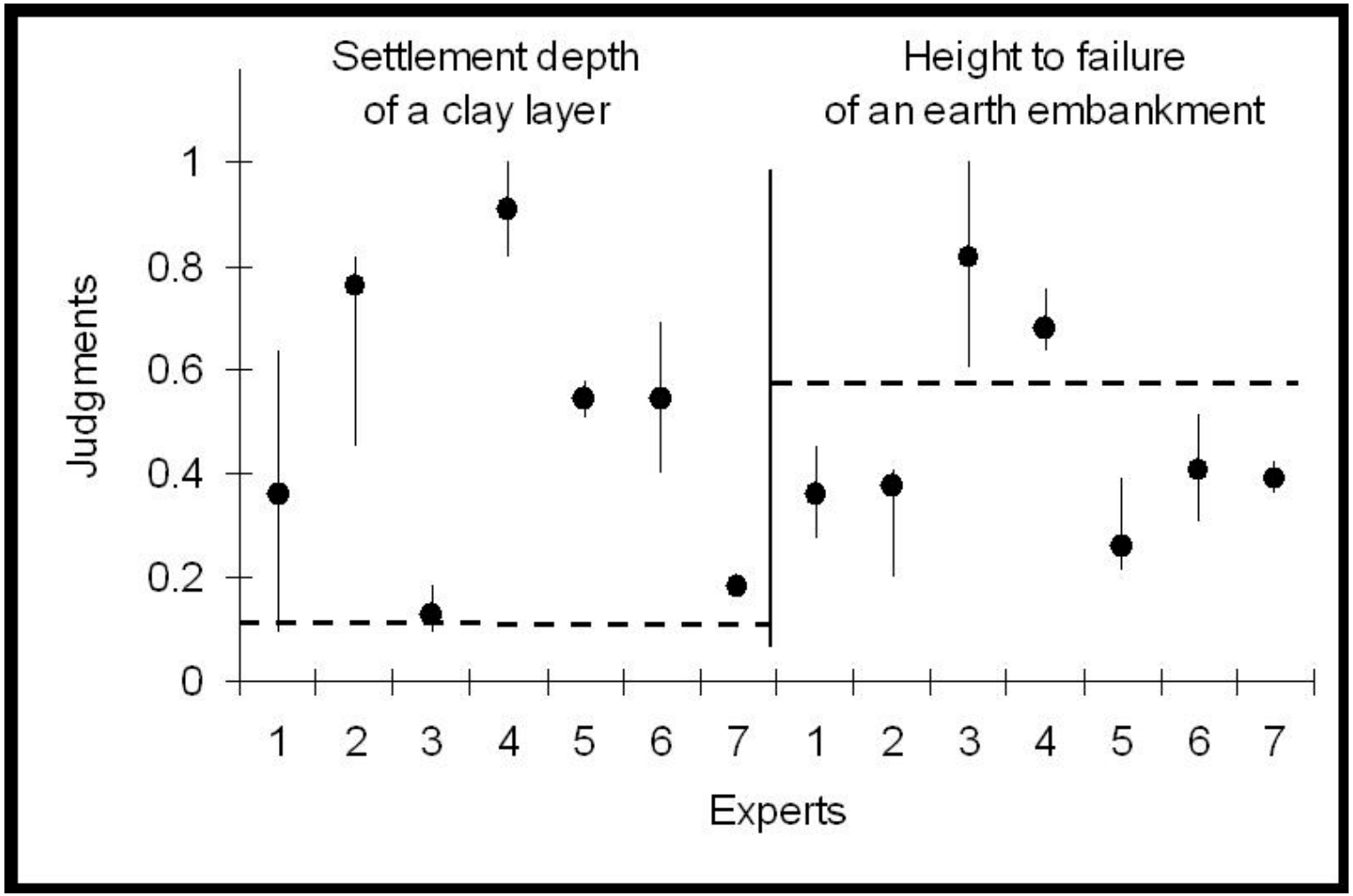

Predicted probability of an offence
0.5
0.4
0.3
0.2
0.1
0
Scale 1 (1-100%)
Scale 2 (1->40%)
Laypeople
American Academy of Psychiatry and Law
American Psychology-Law Society

Personal control, understanding…

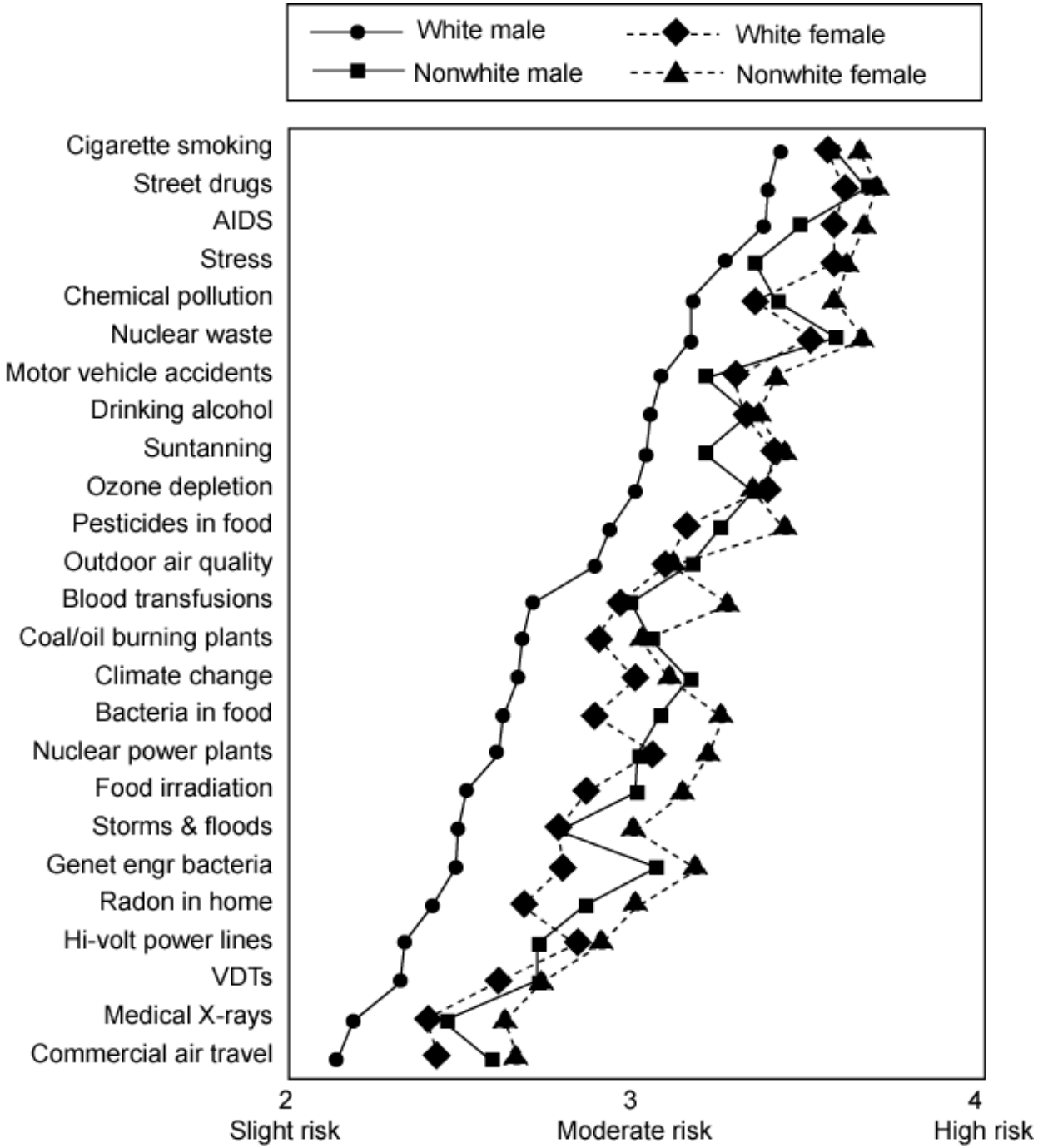


Slovic et al. (1979)

Overconfidence

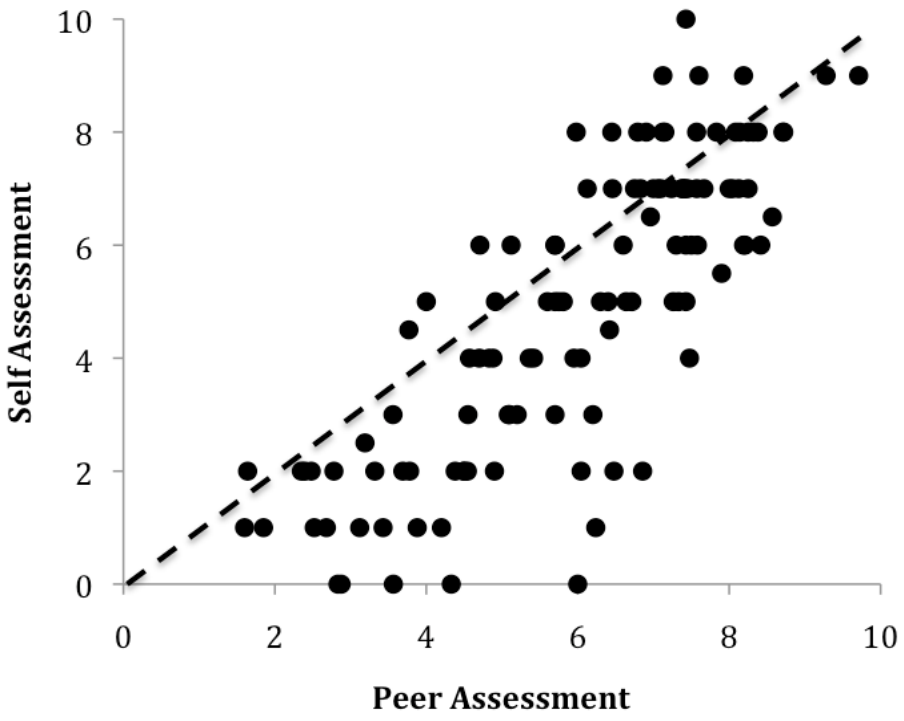


(Hynes and Vanmarche 1977)


White male
White female
Nonwhite male
Nonwhite female
Cigarette smoking
Street drugs
AIDS
Stress
Chemical pollution
Nuclear waste
Motor vehicle accidents
Drinking alcohol
Suntanning
Ozone depletion
Pesticides in food
Outdoor air quality
Blood transfusions
Coal/oil burning plants
Climate change
Bacteria in food
Nuclear power plants
Food irradiation
Storms & floods
Genet engr bacteria
Radon in home
Hi-volt power lines
VDTs
Medical X-rays
Commercial air travel
2
Slight risk
3
Moderate risk
4
High risk


DAN ARIELY
PREDICTABLY IRRATIONAL
FUTURE BABBLE
Why Expert Predictions Fail — and Why We Believe Them Anyway
Dan Gardner
Bestselling author of Risk
A NEW YORK TIMES BUSINESS BESTSELLER
THE WISDOM OF CROWDS
JAMES SUROWIECKI
WITH A NEW AFTERWORD BY THE AUTHOR
PHILIP E. TETLOCK
EXPERT POLITICAL JUDGMENT
How Good Is It? How Can We Know?
WRONG
Why Experts Keep Failing Us — And How to Know When Not to Trust Them
Scientists, finance wizards, doctors, relationship gurus, celebrity CEOs, high-powered consultants, health officials and more
DAVID H. FREEDMAN
author of THE UNDERCOVER ECONOMIST
ADAPT
WHY SUCCESS ALWAYS STARTS WITH FAILURE
TIM HARFORD
'Harford presents complex arguments with unfailing clarity and wit'
Sunday Times
Harnessing the Power of Counterintuition
THINK TWICE
MICHAEL J. MAUBOUSSIN
SUPER CRUNCHERS
IAN AYRES
SUPER CRUNCHERS
HOW ANYTHING CAN BE PREDICTED
IAN AYRES
CROWD SOURCING
WHY THE POWER OF THE CROWD IS DRIVING THE FUTURE OF BUSINESS
JEFF HOWE
Being Wrong
blink
By the author of THE TIPPING POINT
The Power of Thinking Without Thinking
Malcolm Gladwell
HERE COMES EVERYBODY
CLAY SHIRKY
Predictocracy
Market Mechanisms for Public and Private Decision Making
MICHAEL ABRAMOWICZ
SECOND EDITION
NEW YORK TIMES BESTSELLER
THE BLACK SWAN
The Impact of the HIGHLY IMPROBABLE
Nassim Nicholas Taleb
DOUGLAS W. HUBBARD
THE FAILURE OF RISK MANAGEMENT
Why It's Broken and How to Fix It
ON BEING CERTAIN
Believing You Are Right Even When You're Not
ROBERT A. BURTON, M.D.

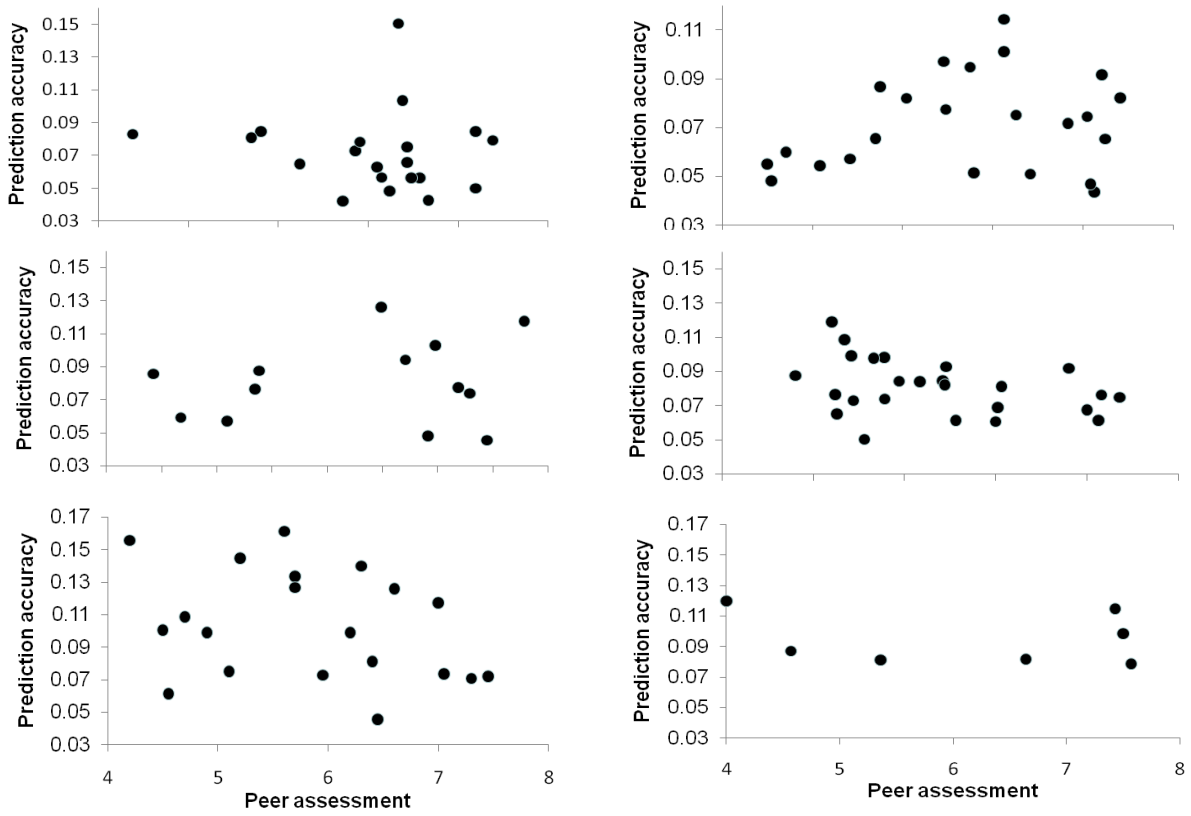
Peer versus self assessments



***r* = 0.85**

Range across workshops
[0.675 to 0.944]

Do peer assessments correlate with performance?



IUCN rules
‘Critically endangered’

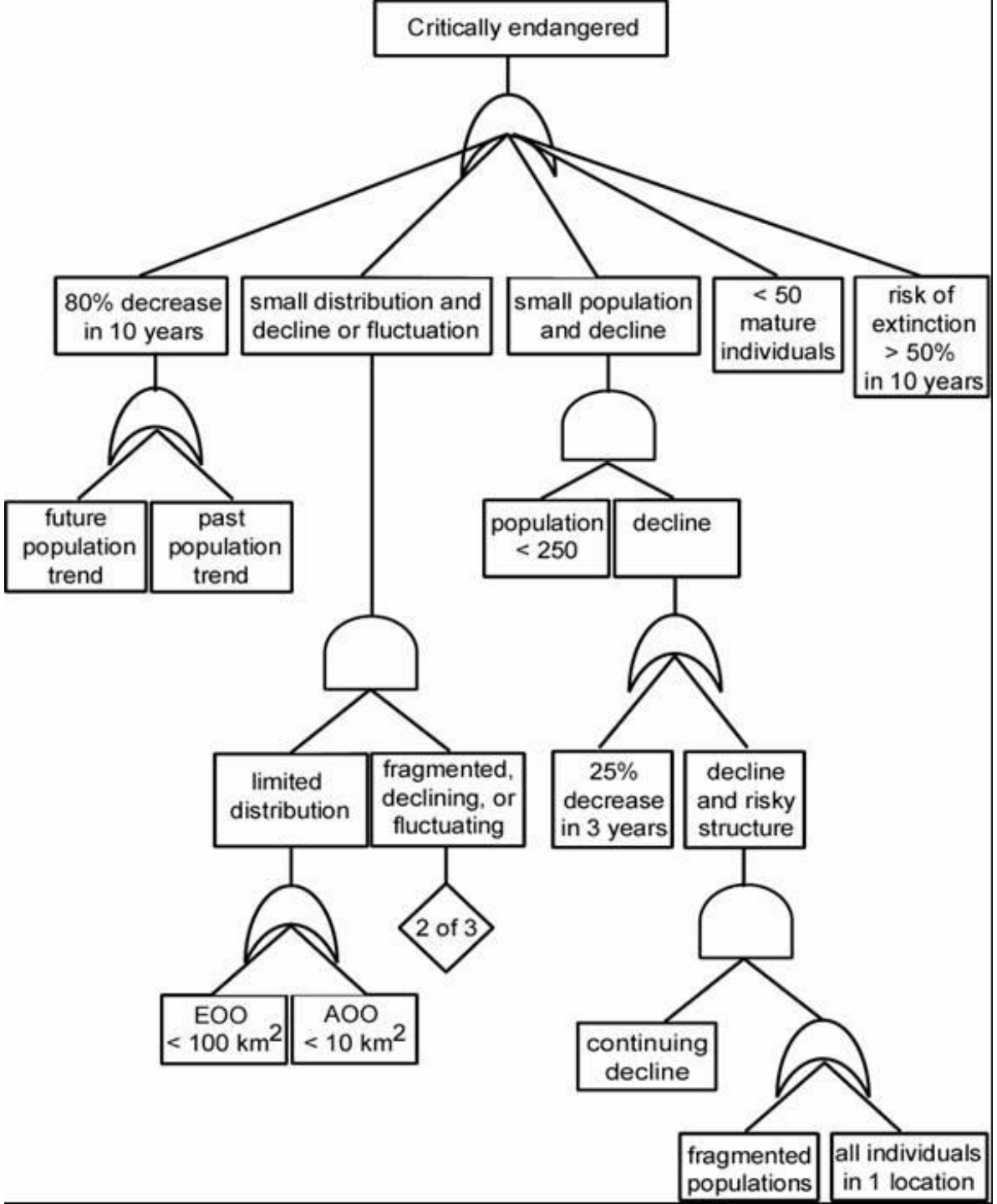
- ***IF*** *Decline of >= 80% in 10 years or 3 generations*
- ***OR*** *Range <100 km2 or occupied habitat <10 km^2*

  AND

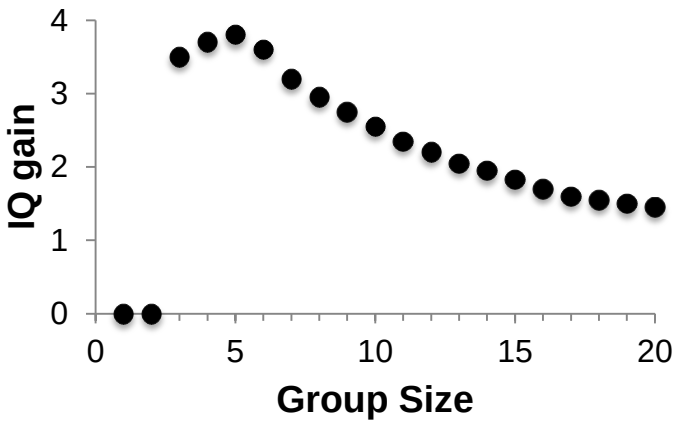
  at least 2 of the following 3 conditions are met:
  1) severely fragmented or in 1 subpopulation
  2) continuing to decline
  3) fluctuations > 1 order of magnitude
- **OR** *number of mature individuals < 250*

  AND

  at least 1 of the following 2 conditions are met:
  1) >=25% decline in 3 years / 1 generation
  2) continuing decline and 1 subpopulation or <=50 per subpopulation
- **OR** *< 50 mature individuals*
- **OR** *>= 50% risk of extinction in 10 years / 3 generations.*


Critically endangered
80% decrease in 10 years
small distribution and decline or fluctuation
small population and decline
< 50 mature individuals
risk of extinction > 50% in 10 years
future population trend
past population trend
population < 250
decline
limited distribution
fragmented, declining, or fluctuating
25% decrease in 3 years
decline and risky structure
2 of 3
EOO < 100 km2
AOO < 10 km2
continuing decline
fragmented populations
all individuals in 1 location

What can groups do?



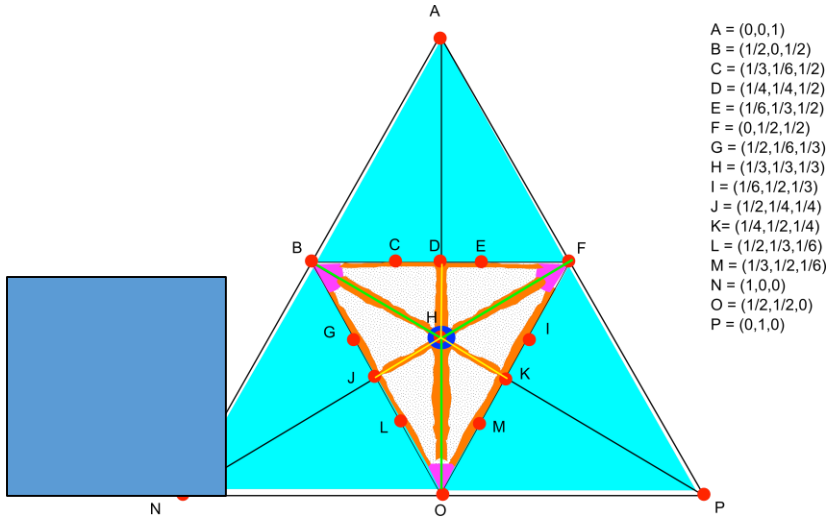
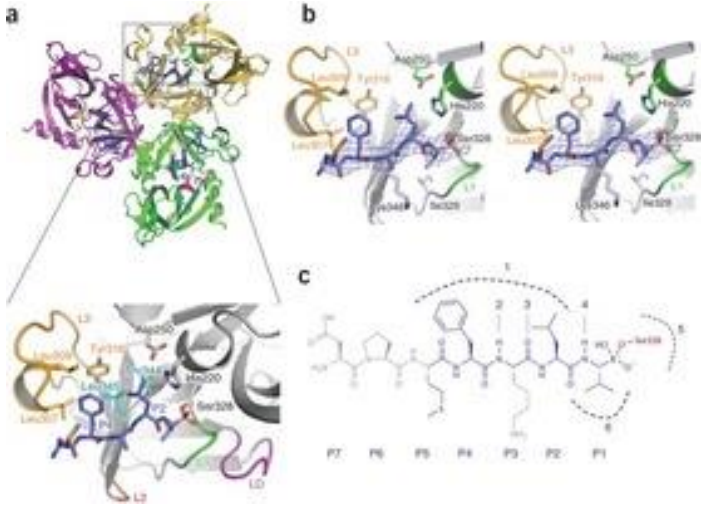


Photo # NH 70305 USS Scorpion comes alongside USS Tallahatchie County, April 1968



4-step question format…

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

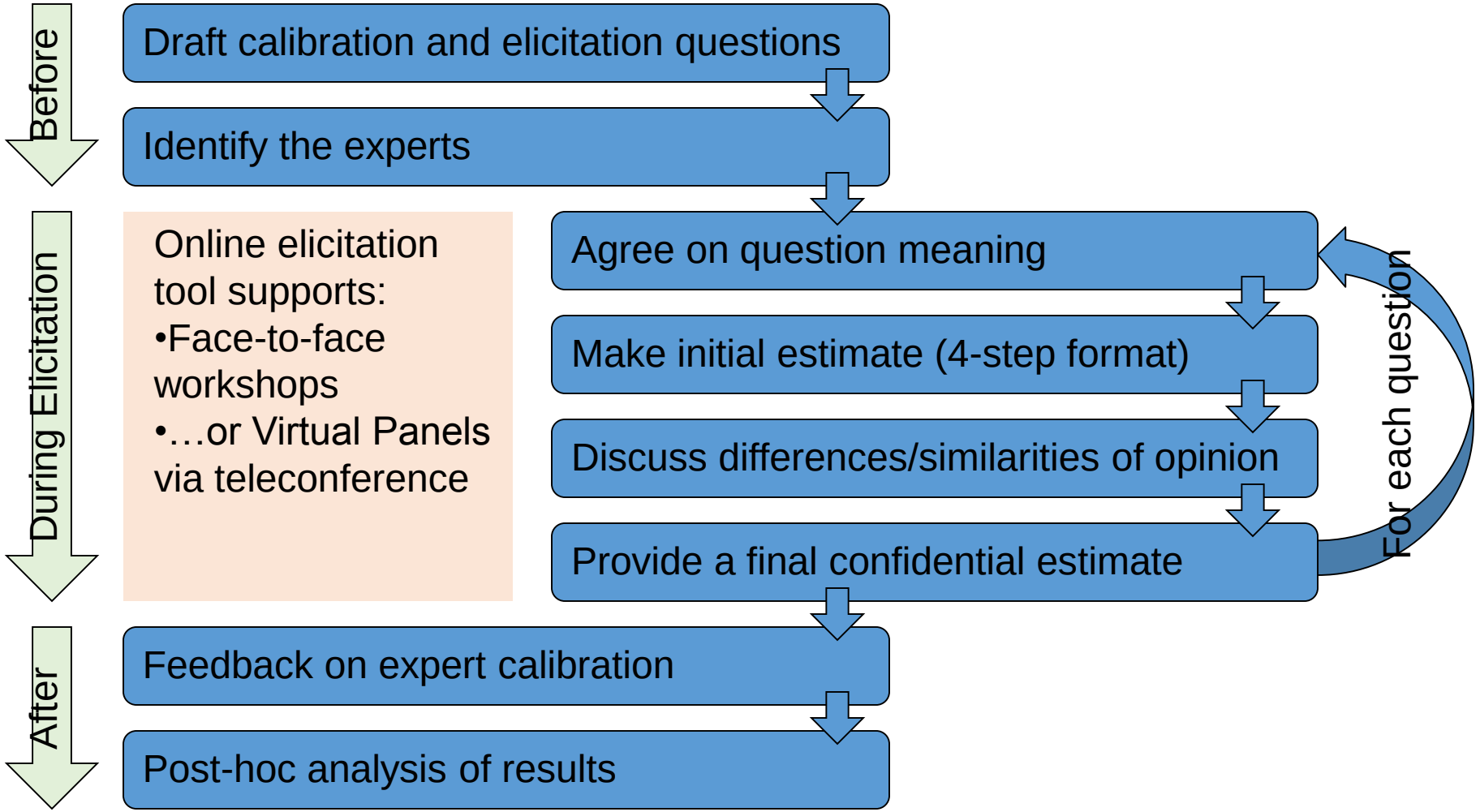
2. Realistically, what do you think the highest plausible value is?

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?

(Speirs-Bridge et al., 2010)

Integrated elicitation process



How many beans?: individual response

1. Realistically, what is the *lowest* number of jellybeans you think are in the jar?
2. Realistically, what is the *highest* number of jellybeans you think are in in the jar?
3. What's your *best guess* at number of jellybeans in the jar?
4. *How confident* are you that the interval you've created captures the true number of jellybeans? (for this question answer between 50-100%)

Trivial Pursuit

Group responses

1. In what year did the world’s population reach 2 billion?

Group responses

1. In what year did the world’s population reach 2 billion?

2. How long is a newborn Humpback whale?

Group responses

1. In what year did the world's population reach 2 billion?

1. How long is a newborn humpback whale?

1. What % of American presidents have been lawyers (asked in 2007)

Group responses

1. In what year did the world's population reach 2 billion?

2. How long is a newborn humpback whale?

3. What % of American presidents have been lawyers (asked in 2007)

4. In what year was the microchip invented?

Group responses

1. In what year did the world's population reach 2 billion?

2. How long is a newborn humpback whale?

3. What % of American presidents have been lawyers (asked in 2007)

4. In what year was the microchip invented?

1. What percent of a watermelon's weight is water?

Answers

1. In what year did the world’s population reach 2 billion?

Answers

1. In what year did the world’s population reach 2 billion?
   (1927)

Answers

1. In what year did the world’s population reach 2 billion? (1927)

2. How long is a newborn humpback whale?

Answers

1. In what year did the world's population reach 2 billion? (1927)

2. How long is a newborn humpback whale? (4.15m)

Answers

1. In what year did the world's population reach 2 billion? (1927)

2. How long is a newborn humpback whale? (4.15)

3. What % of American presidents have been lawyers (asked in 2007)

Answers

1. In what year did the world's population reach 2 billion? (1927)

2. How long is a newborn humpback whale? (4.15)

1. What % of American presidents have been lawyers (asked in 2007) (52.4%)

Answers

1. In what year did the world's population reach 2 billion? (1927)

2. How long is a newborn humpback whale? (4.15)

1. What % of American presidents have been lawyers (asked in 2007) (52.4%)

2. In what year was the microchip invented?

Answers

1. In what year did the world's population reach 2 billion? (1927)

2. How long is a newborn humpback whale? (4.15)

1. What % of American presidents have been lawyers (asked in 2007) (52.4%)

2. In what year was the microchip invented? (1959)

Answers

1. In what year did the world's population reach 2 billion? (1927)

2. How long is a newborn humpback whale? (4.15)

1. What % of American presidents have been lawyers (asked in 2007) (52.4%)

2. In what year was the microchip invented? (1959)

1. What percent of a watermelon's weight is water?

Answers

1. In what year did the world's population reach 2 billion? (1927)

2. How long is a newborn humpback whale? (4.15)

1. What % of American presidents have been lawyers (asked in 2007) (52.4%)

2. In what year was the microchip invented? (1959)

1. What percent of a watermelon's weight is water? (92%)

The 3-step Procedure

Columbia University admitted women before 1980
(Yes / No)?

Some facts

1. Columbia University is located on Manhattan in New York, USA (a progressive city and state).

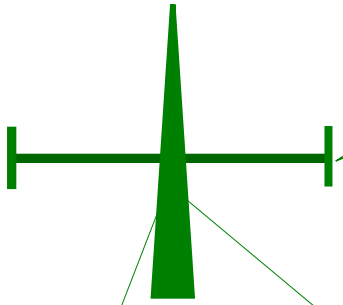
2. Columbia University was the last ‘ivy league’ school in the USA to admit women.

The 3-step Procedure

Columbia University admitted women before 1980 (Yes / No)?

1. What do you think is the lowest probability that this statement is true?

2. What do you think is the highest probability that this statement is true?



3. Realistically, what is your best estimate?

Answer

Columbia University admitted women before 1980
(Yes / No)?

Answer

Columbia University admitted women before 1980 (Yes / No)?

NO! They were admitted in 1983.

Prizes go to .....

What doesn't work

Relying on individuals...

- overconfidence, hindsight bias
- framing
- availability bias
- reference group, base rate neglect
- using the person who (everyone believes) knows the most: the status effect
- undetected linguistic uncertainty

Relying on naïve groups...

- naïve question formulation: linguistic uncertainty
- unstructured discussion
- dominance, group-think
- common data sources / lack of independence
- uniformity in context, culture, styles of reasoning

To fix the problems

Ask individuals to...

- Consider counter-arguments
- Answer the same question in different ways (lowest, highest, most likely)
- Indicate confidence
- Examine estimates made by other people (feedback)
- 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
- Encourage groups to discuss and revise question meaning
- Use psychologically, culturally and contextually diverse, independent people
- Avoid group think— Delphi / independent data / anonymity in judgments

Office of the Director of National Intelligence

IARPA

BE THE FUTURE

RESEARCH PROGRAMS | OUR PROGRAM MANAGERS | WORKING WITH IARPA | CAREERS

Home ▸ Research Programs ▸ ACE

Aggregative Contingent Estimation (ACE)

The goal of the ACE Program is to dramatically enhance the accuracy, precision, and timeliness of intelligence forecasts for a broad range of event types, through the development of advanced techniques that elicit, weight, and combine the judgments of many intelligence analysts. The ACE Program seeks technical innovations in the following areas: (a) efficient elicitation of probabilistic judgments, including conditional probabilities for contingent events; (b) mathematical aggregation of judgments by many individuals, based on factors that may include: past performance, expertise, cognitive style, metaknowledge, and other attributes predictive of accuracy; and (c) effective representation of aggregated probabilistic forecasts and their distributions. The ACE Program will build upon technical achievements of past research and on state-of-the-art systems used today for generating probabilistic forecasts from widely-dispersed experts. The program will involve empirical testing of forecasting accuracy against real events.

Program Manager

For information contact: dni-iarpa-info@iarpa.gov

Program Information

IARPA-BAA-10-05

IARPA Day Poster

http://daggre.org

George Mason University (Mason)

- Charles Twardy (PI), Kathryn Laskey (Co-PI), Robin Hanson

Australian Center of Excellence for Risk Analysis

And a supporting cast of thousands….

Intelligence Advanced Research Projects Activity (IARPA) invests in high-risk/high-payoff research programs that have the potential to provide our nation with an overwhelming intelligence advantage over future adversaries

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

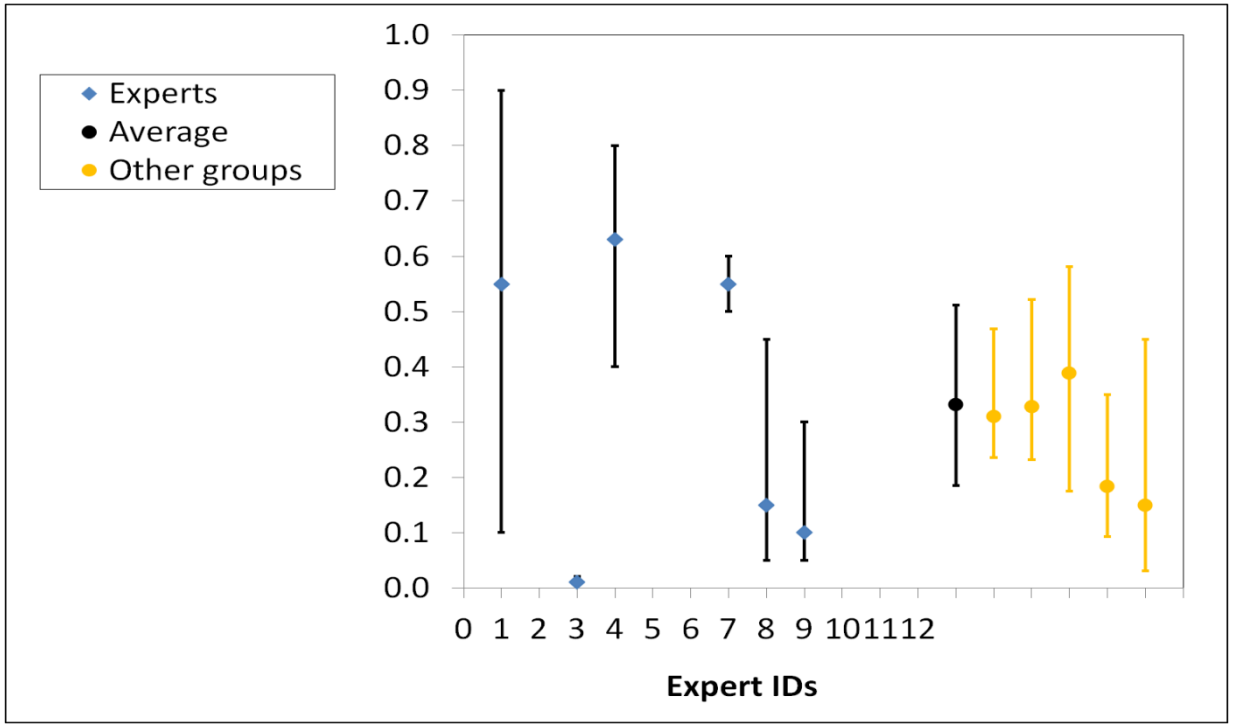
Print this Month's Questions

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?

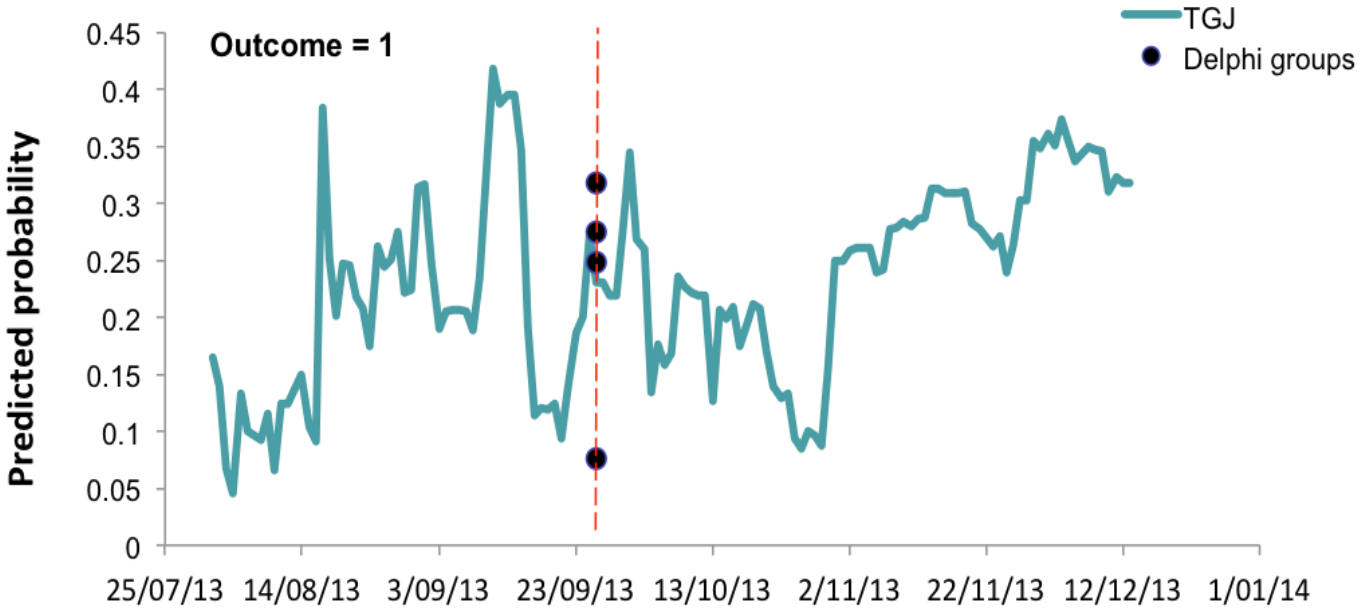
Status: -

Will Australia sell uranium…?

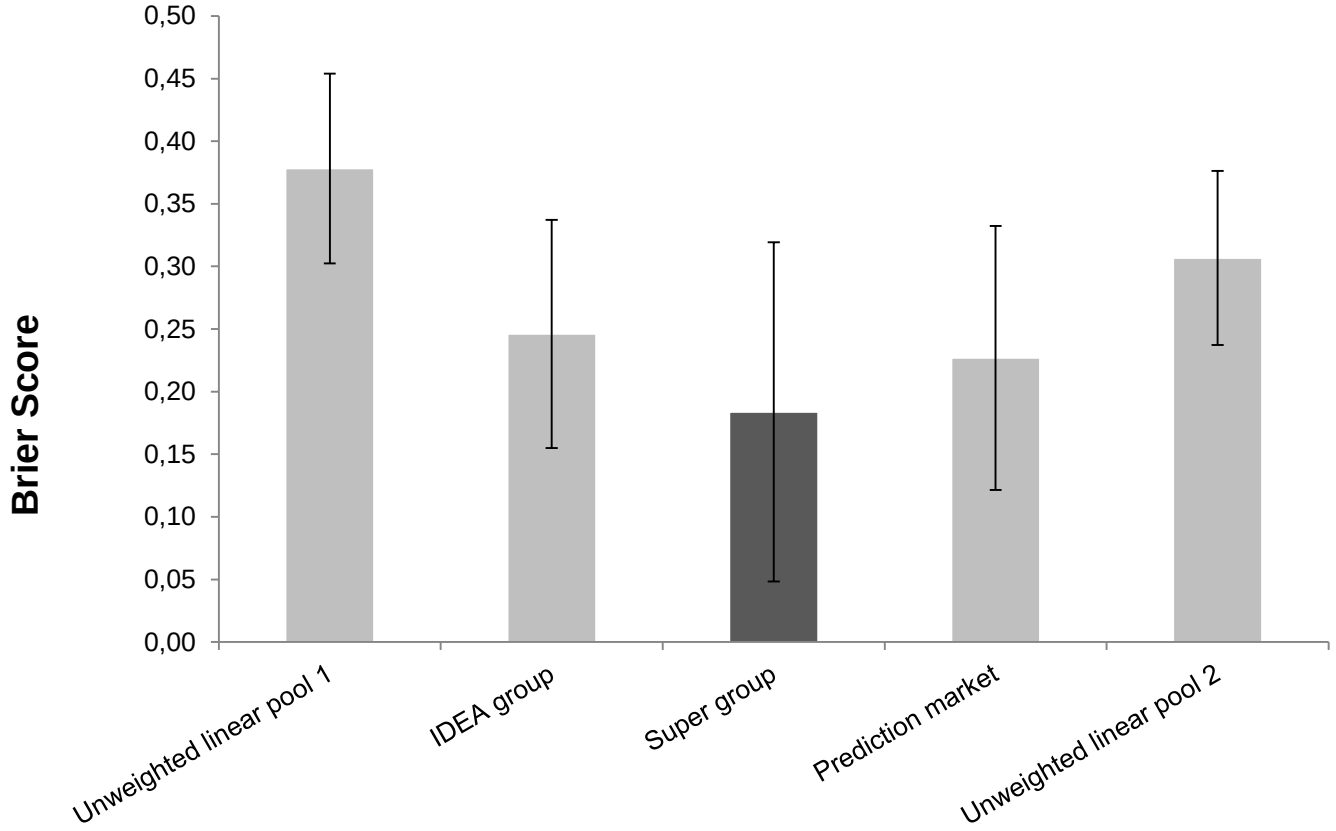


1220

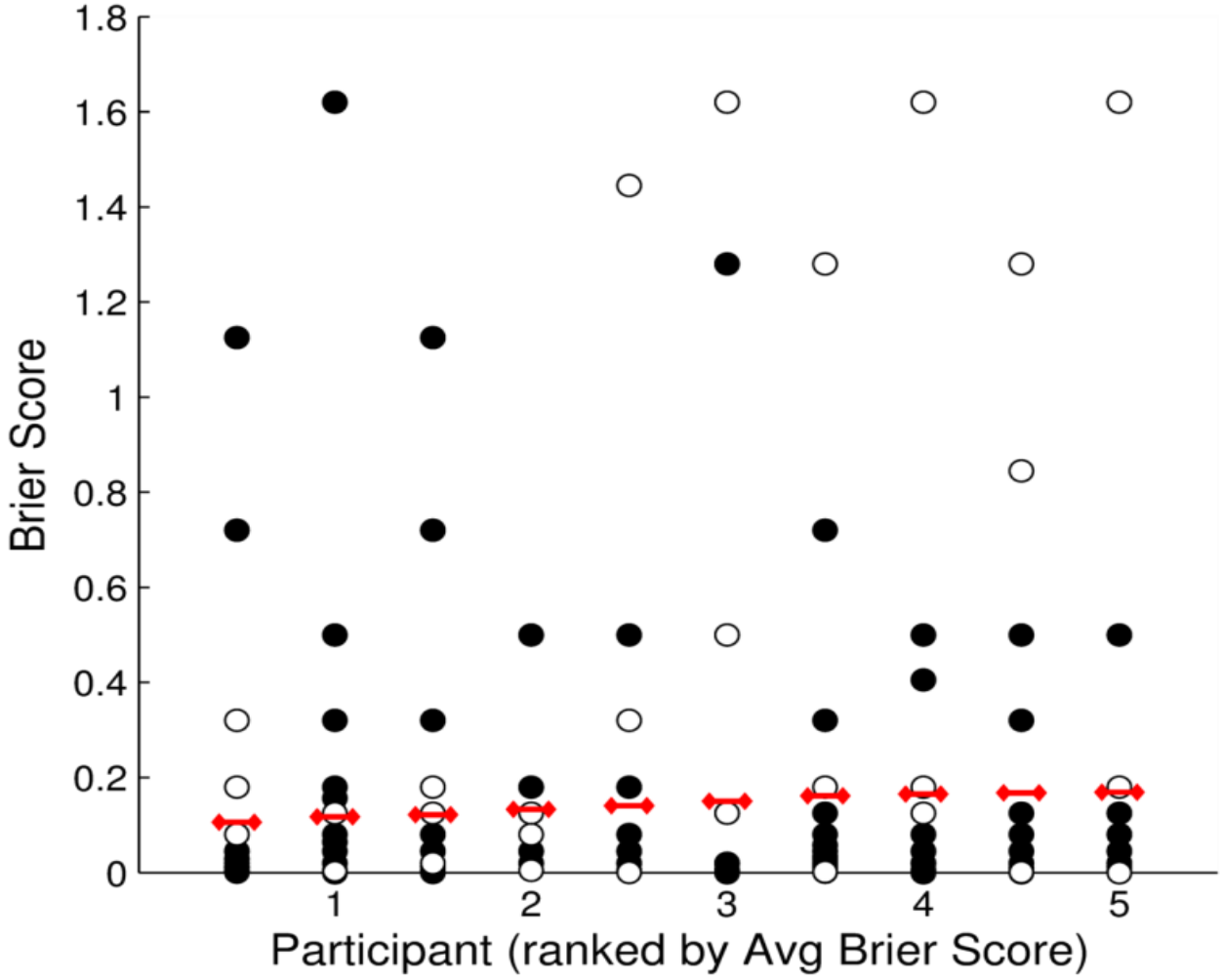
Will Chinese armed forces or maritime law enforcement forces attempt to interdict or make physical contact with at least one U.S. government naval vessel or airplane or Japanese government naval vessel or airplane that it claims is in its territorial waters or airspace, before 1 May 2014?



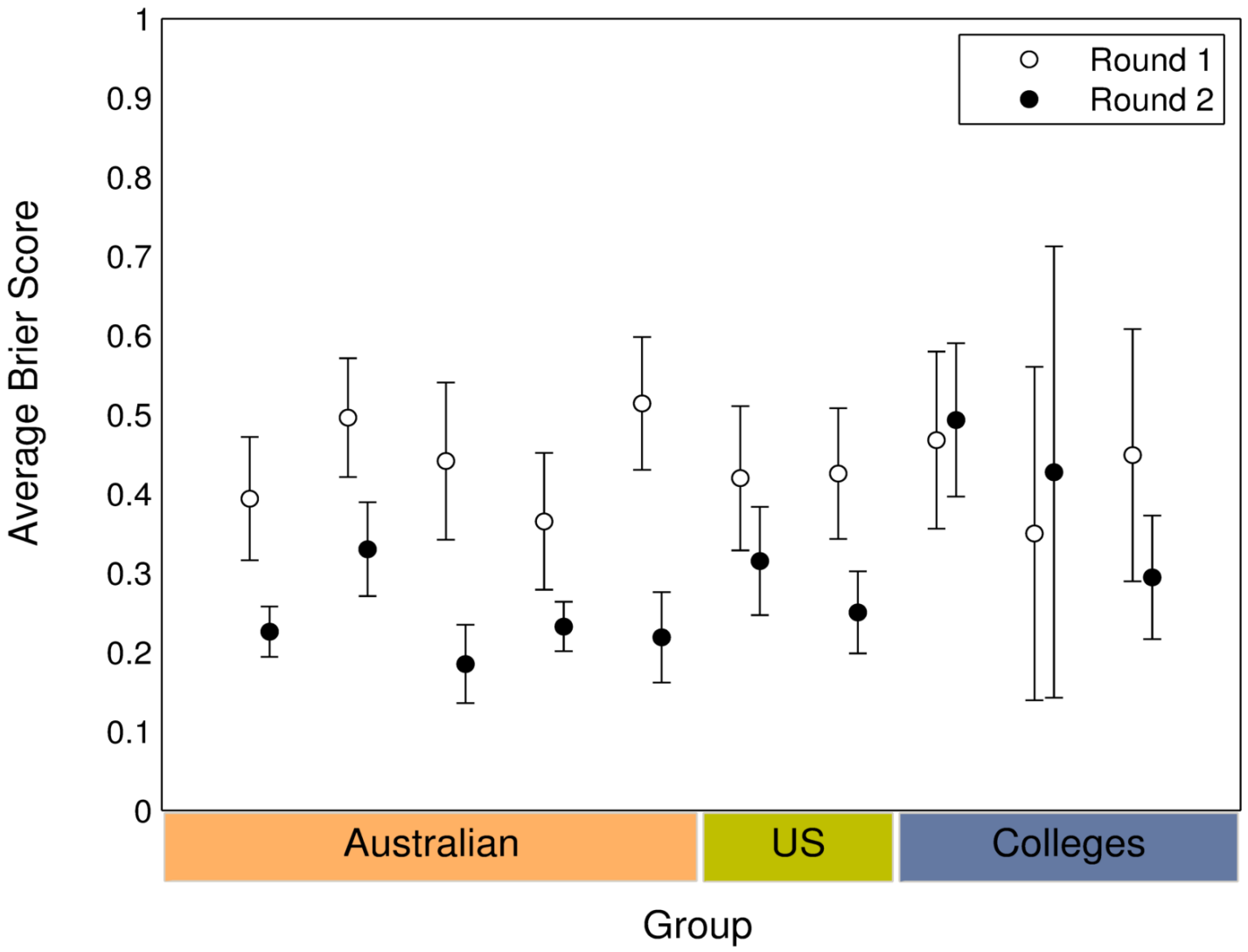
Geopolitical forecasts



Best performing individuals



Feedback helps



Office of the Director of National Intelligence

IARPA

BE THE FUTURE

home contact faqs

Search ... GO

RESEARCH PROGRAMS | OUR PROGRAM MANAGERS | WORKING WITH IARPA | CAREERS | NEWSROOM | ABOUT IARPA

Home › Research Programs › CREATE

Crowdsourcing Evidence, Argumentation, Thinking and Evaluation (CREATE)

The CREATE program seeks proposals to develop, and experimentally test, systems that use crowdsourcing and structured analytic techniques (STs) to improve analytic reasoning. These systems will help people better understand the evidence and assumptions that support—or conflict with—conclusions. Secondarily, they will also help users better communicate their reasoning and conclusions.

Interested offerors are required to submit full proposals in order to receive consideration for funding. Proposals must be received by May 9, 2016 in order to be assured of consideration during the initial round of selections.

The CREATE program is envisioned as a 4.5-year effort that is intended to begin in September 2016. Phase 1 of the program will last 20 months, Phase 2 will last 17 months and Phase 3 will last 17 months. Multiple Phase 1 awards are anticipated.

Program Manager

Steve Rieber

Program Information

IARPA-BAA-15-11

Research Area(s)

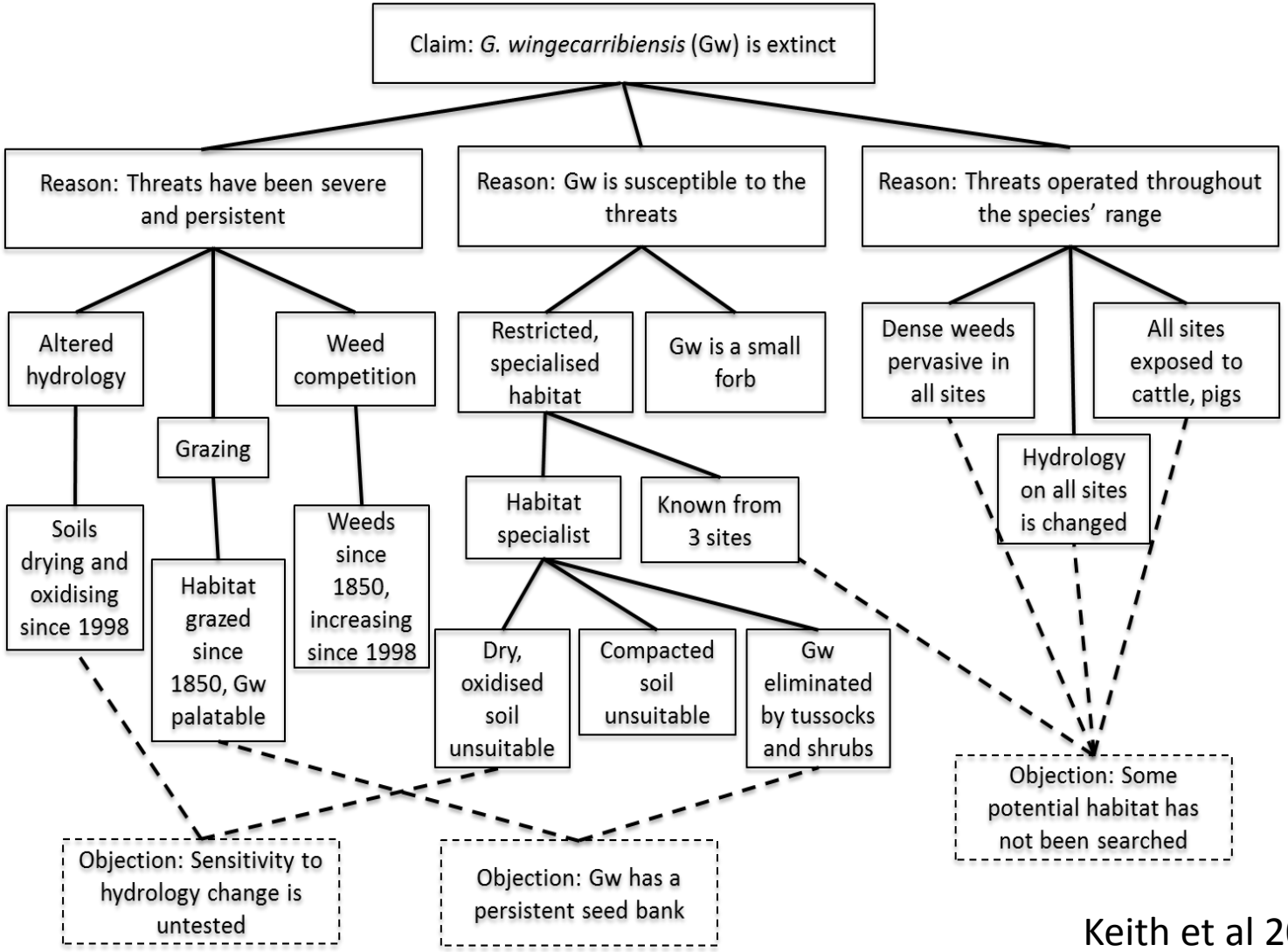
- Social and behavioral sciences
- Informal reasoning
- Computer science
- Structured analytic techniques
- Crowdsourcing

Related Article(s)

Dr. Steve Rieber: Program aims to improve human reasoning

The future of crowdsourcing: Integrating humans with machines

Argument map



Keith et al 2017

Tradeoffs

Decisions involve tradeoffs

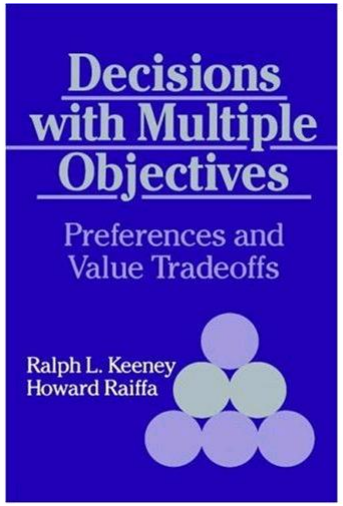
Tradeoffs involve weights (but whose?)

Weights are a function of two things:

i. How important the attribute (the criterion) is to the decision-maker

ii. The range of the attributes, that is, how much the criterion 'swings' in value across the range of the alternatives on offer

Von Neumann and Morgenstern



Utility and rights (after Morgan and Henrion 1990)

Utility-based criteria	**Description**
Probabilistic benefit-cost	Estimate benefits and costs of alternatives in economic terms, and use expected value (weighted by risk) to find the option with the greatest net benefit
Maximise multi-attribute utility	In place of economic value, use a utility function that incorporates the utility in terms of all important attributes
Maximise/minimise extreme outcomes	Minimise the chance of the worst outcome, or maximise the chance of the best outcome, usually dictated by social or political context
Rights-based criteria	**Description**
Zero risk	Eliminate risks entirely, irrespective of benefits or costs
Constrained risk	Constrain risk so that it does not exceed a specified threshold
Approval/ compensation	Impose risks on only those parts of the population that have given consent, perhaps after compensation
Hybrid	Maximise probabilistic benefit-cost within a constraint of an upper bound on risk to an element(s) of the system

Performance matrix

- One of the simplest ways to avoid misusing MCDA is to avoid numerical solutions to trade-offs.
- Wherever possible, use the performance matrix to support decision-making.
- Look for dominated alternatives and redundant criteria

Simple example: Pick a flight ticket to Paris

(informal) problem formulation

Issues	Objectives	Indicators
•I don't want to spend much money •I don't want hidden fees	→ Minimise cost	$ total
•I want a direct flight •I want easy check-ins	→ Minimise travel time	Hours
•I want decent leg room •I want an aisle seat •I want friendly service	→ Maximise comfort	Scale (5 = best, 0 = worst)
•I am concerned about recent airline safety incidents •I'm uncomfortable flying with a new airline	→ Maximise safety	# accidents / 1 million take-offs (5 yr average)

Performance matrix

Objectives (criteria)	Indicators	Preferred direction	A Air Aussie	B Westjet	
Minimise cost	$	Lower is better	$2000	$1500	
Minimise travel time	Hours	Lower is better	8 - 9	13 – 15	