Structured judgment and decision making

Budapest
European Commission and EKLIPSE
September, 2018

Mark Burgman
Imperial College London
WHY WE WORRY ABOUT THE WRONG THINGS
The Psychology of Risk
BY JEFFREY KLUGER

PLUS:
The Year in Medicine A to Z

Against the grain: in November last year, demonstrators took their grievances to the World Trade Organization 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 and 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)

Scale 1: Range 1 - 100%

Scale 2: Range 1 - >40%
Personal control, understanding…

Slovic et al. (1979)
Overconfidence

(Hynes and Vanmarche 1977)
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 $\geq 80\%$ in 10 years or 3 generations
- **OR**  Range $<$100 km$^2$ or occupied habitat $<$10 km$^2$
  
  \textbf{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
  
  \textbf{AND}
  
  at least 1 of the following 2 conditions are met:
  1) $\geq$25\% decline in 3 years / 1 generation
  2) continuing decline and 1 subpopulation or $\leq$50 per subpopulation

- **OR**  $<$ 50 mature individuals
- **OR**  $\geq 50\%$ risk of extinction in 10 years / 3 generations.
What can groups do?

Photo # NH 70365  USS Scorpion comes alongside USS Tallahassee County, April 1968
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

Before
- Identify the experts

During Elicitation
- Draft calibration and elicitation questions
- Online elicitation tool supports:
  • Face-to-face workshops
  • …or Virtual Panels via teleconference
- Agree on question meaning
- Make initial estimate (4-step format)
- Discuss differences/similarities of opinion
- Provide a final confidential estimate

After
- Feedback on expert calibration
- Post-hoc analysis of results

For each question
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 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
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?
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)
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. 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?
1. In what year did the world’s population reach 2 billion?
1. In what year did the world’s population reach 2 billion? (1927)
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)
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)
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?
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?
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%)
Columbia University admitted women before 1980 (Yes / No)?
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

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?

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

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?
Will Australia sell uranium...?
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

![Graph showing Brier Score against Participant (ranked by Avg Brier Score).]
Feedback helps

![Graph showing average Brier score for different groups. The graph compares Round 1 (open circles) and Round 2 (solid circles) across Australian, US, and Colleges groups.]
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 (STTs) to improve analytic reasoning. These systems will help people better understand the evidence and assumptions that support—or conflict with—conclusions. Secondly, 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.
Argument map

Claim: *G. wingecarribiensis* (Gw) is extinct

Reason: Threats have been severe and persistent
- Altered hydrology
  - Grazing
    - Soils drying and oxidising since 1998
- Weed competition
  - Weeds since 1850, increasing since 1998
- Habitat grazed since 1850, Gw palatable

Reason: Gw is susceptible to the threats
- Restricted, specialised habitat
- Gw is a small forb
- Habitat specialist
- Known from 3 sites
- Dry, oxidised soil unsuitable
- Compacted soil unsuitable
- Gw eliminated by tussocks and shrubs

Reason: Threats operated throughout the species’ range
- Dense weeds pervasive in all sites
- All sites exposed to cattle, pigs

Objection: Sensitivity to hydrology change is untested
- Objection: Gw has a persistent seed bank
- Objection: Some potential habitat has not been searched

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
### Utility and rights (after Morgan and Henrion 1990)

<table>
<thead>
<tr>
<th>Utility-based criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic benefit-cost</td>
<td>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</td>
</tr>
<tr>
<td>Maximise multi-attribute utility</td>
<td>In place of economic value, use a utility function that incorporates the utility in terms of all important attributes</td>
</tr>
<tr>
<td>Maximise/minimise extreme outcomes</td>
<td>Minimise the chance of the worst outcome, or maximise the chance of the best outcome, usually dictated by social or political context</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rights-based criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero risk</td>
<td>Eliminate risks entirely, irrespective of benefits or costs</td>
</tr>
<tr>
<td>Constrained risk</td>
<td>Constrain risk so that it does not exceed a specified threshold</td>
</tr>
<tr>
<td>Approval/compensation</td>
<td>Impose risks on only those parts of the population that have given consent, perhaps after compensation</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Maximise probabilistic benefit-cost within a constraint of an upper bound on risk to an element(s) of the system</td>
</tr>
</tbody>
</table>
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

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

The table above shows the performance matrix for different airlines based on various criteria. For example, under the objective of minimizing cost, Air Aussie has a preferred cost of $2000, while Westjet has a preferred cost of $1500. Similarly, for minimizing travel time, Air Aussie has a preferred travel time of 8 - 9 hours, while Westjet has a preferred travel time of 13 – 15 hours.