Biases and [Ir]rationality

Informatics 1 CG: Lecture 18

Chris Lucas

clucas2@inf.ed.ac.uk

Why?

- Human failures and quirks are windows into cognition
  - Past examples: overregularisation, theory of mind tasks

- Maybe understanding bad biases can help us avoid them

  (or not…)

Some classic examples

1. Framing effects
2. Base-rate neglect
3. Representativeness
4. Availability
5. Wason's card selection task

* or apparent irrationality; we'll see

Framing effects

Which of the following options do you prefer?

A. Sure gain of £240.
B. 25% chance to win £1000, 75% chance to win nothing.

(Tversky & Kahneman (1981): The framing of decisions and the psychology of choice)

Framing effects

Which of the following options do you prefer?

C. Sure loss of £750.

D. 25% chance to lose nothing, 75% chance to lose £1000.

(Tversky & Kahneman (1981): The framing of decisions and the psychology of choice)

Framing effects

Which of the following options do you prefer?

A. Sure gain of £240.
B. 25% chance to win £1000, 75% chance to win nothing.

A vs B?

C. Sure loss of £750.
D. 25% chance to lose nothing, 75% chance to lose £1000.

C vs D?

(Tversky & Kahneman (1981): The framing of decisions and the psychology of choice)

Framing effects

Possible combined bets:

A&C, A&D, B&C, B&D.

Most popular choice: **A&D** (73% of participants)

25% chance to gain £240, 75% chance to lose £760

Compare to **B&C:**

25% chance to gain £250, 75% chance to lose £750.

(Tversky & Kahneman (1981): The framing of decisions and the psychology of choice)

Framing effects

People tend to choose a *dominated* option – strictly worse than an alternative!

Contra conventional wisdom that people maximise their expected rewards.

Also applies with real money, and questions about human lives.

Framing effects

What's going on?

Tversky & Kahneman: Prospect theory

- People have non-linear utility functions
  - U(+£120)-U(+£110) > U(+£20)-U(+£10)
- People treat gains and losses differently: losses more extreme.
- Other phenomena, e.g.,
  - weighting of extreme probabilities

Base rates

"Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run).

85% of the cabs in the city are Blue, **15%** are Green.

A cab was involved in a hit-and-run accident at night.
A witness later identified the cab as a **Green** cab.

The court tested the witness' ability to distinguish between Blue and Green cabs under nighttime visibility conditions.

It found that the witness was able to identify each color correctly about **80%** of the time, but confused it with the other color about **20%** of the time.

What do you think are the chances that the errant cab was indeed Green, as the witness claimed?"

Kahneman & Tversky (1972) via Maya Bar-Hillel (1980).

Base rates

Typical human judgment: Green cab

Bayes' theorem:

$$P(\text{green}|\text{witness}=g) = P(\text{green})P(\text{witness}=g|\text{green}) / (P(\text{green})P(\text{witness}=g|\text{green})+P(\text{blue})P(\text{witness}=g|\text{blue}))$$

That is: $0.15*0.80 / (0.15*0.80+0.85*0.20) = 0.12 / (0.12 + 0.17)$

$= 0.41$ (less than .5)

Kahneman & Tversky (1972) via Maya Bar-Hillel (1980).

Base rate neglect

What's going on?

Maya Bar-Hillel:

We have heuristics for determining the relevance of info

Kahneman:

The green cab is more *representative* of the witness's report

This phenomenon is important:

- Physicians are subject to base-rate neglect in evaluating diagnostic tests!
- Relevant in legal settings too -- See Bar-Hillel (1980) for more.

Kahneman & Tversky (1972) via Maya Bar-Hillel (1980): The Base-rate fallacy in probability judgments

Another example

"Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities."

Rank the following statements in terms of how likely they are to be true:

- Bill is a physician who plays poker for a hobby.
- Bill is an architect.
- Bill is an accountant.
- Bill plays jazz for a hobby.
- Bill surfs for a hobby.
- Bill is a reporter.
- Bill is an accountant who plays jazz for a hobby.
- Bill climbs mountains for a hobby.

(Tversky & Kahneman, 1983)

The conjunction fallacy

"Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities."

Rank the following statements in terms of how likely they are to be true:

- Bill is a physician who plays poker for a hobby.
- Bill is an architect.
- Bill is an accountant.
- Bill plays jazz for a hobby.
- Bill surfs for a hobby.
- Bill is a reporter.
- Bill is an accountant who plays jazz for a hobby.
- Bill climbs mountains for a hobby.

(Tversky & Kahneman, 1983)

The conjunction fallacy

- Bill is an accountant. (A)
- Bill plays jazz for a hobby. (J)
- Bill is an accountant who plays jazz for a hobby. (A&J)

Most people said A&J is more likely than J.

Since J includes both J&A and J&!A, this appears to be a fallacy.

(Tversky & Kahneman, 1983)

The conjunction fallacy

What's going on?

T&K: Bill being an accountant is *representative* of his description.

Akin to thinking of P(description|category).

(Tversky & Kahneman, 1983)

Availability

What's more dangerous?

Spending an hour on a large aircraft?

Spending an hour in a typical passenger car?

Availability

What's more dangerous to someone living in the US?

Terrorism?

Being hit by lightning?

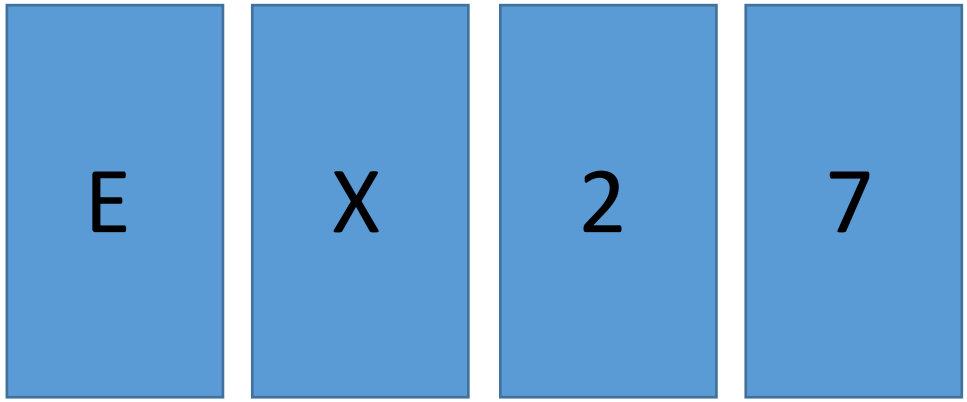
Availability

What's going on?

T&K: We use the ease with which examples come to mind as a proxy for probability.

Wason's card-selection task

The rule:
If there is a vowel on one side of a card,
there is an even number on the other side.

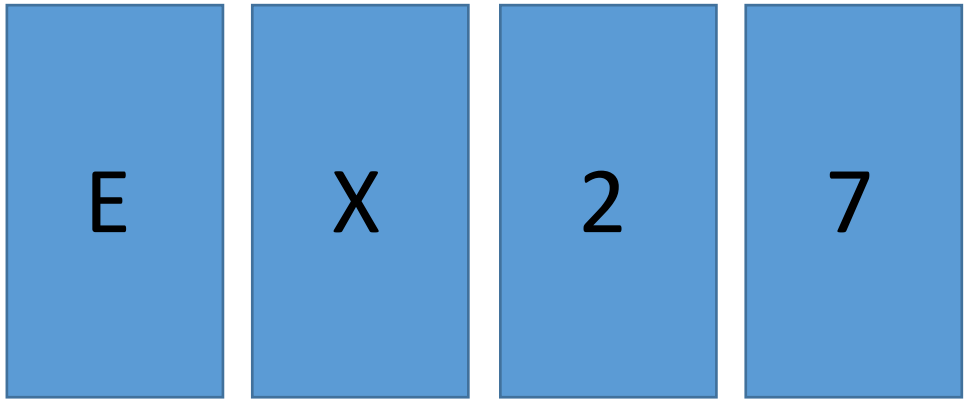


What cards should we reverse to evaluate the rule's truth,
assuming cards have letters on one side and number on the other?

(Oberauer et al. 1999; Wason, 1968; participants were Edinburgh first-year psychology undergraduates)

Wason's card-selection task

The rule:
If there is a vowel on one side of a card,
there is an even number on the other side.



Logically: E and 7.

(Oberauer et al. 1999; Wason, 1968; participants were Edinburgh first-year psychology undergraduates)

Wason’s card-selection task

What’s going on?

Wason: People are bad at logic (and “formal operations” in general).

Explanations

Explanations

1. People aren't solving the problem we think they are
    - Problem interpretation
    - "Ecologically appropriate" inductive biases

Explanations

2. Errors reflect rational trade-offs given resource limitations
    - “Resource-rational” models and theories
    - “Fast and frugal” Heuristics

Explanations

3. Inappropriate standards: participants are doing **better**
    - Economic decision-making
        - Nash equilibria and the traveller's dilemma

What's "rational"?

- Logic?
- Maximising reward/minimising loss?
- Probability theory?

Explanations

3. People are just bad at solving some problems

- No rational explanation
- Local optima
- Optic nerve
- The Spandrels of San Marco

Summary

- People deviate from certain standards for rational behaviour
  - Logic
  - Probability theory
- Revealed across a range of tasks, including ones with serious implications.
- Provoked investigations into “bounded rationality” and how human learners represent problems.

Next time

- A deeper look into what counts as rational, and different ways of examining some of the tasks we've seen.