Biases and [Ir]rationality 2

Informatics 1 CG: Lecture 19

Chris Lucas
clucas2@inf.ed.ac.uk

Last time

Some classic experiments suggesting we're irrational.

Today

(1) What is it to be "rational"?

(2) A deeper look at some of last lecture's experiments

What's a rational inference?

One view: **logic**

- The standard in the Wason card selection task
- Logical approaches were dominant in early cognitive science and artificial intelligence.
  - Logic theorist: proved theorems (Newell & Simon, 1955)
  - Offered as an account of creative problem solving
  - Deductive reasoning a marker of Piaget's "formal operational" stage of development.

What's a rational inference?

Is classical logic* rational?

- Yes, according to Aristotle and others.
- If one accepts a set of premises and rules of inference, inferences that follow are valid.

Are human judgments consistent with classical logic?

- Sometimes, but consider categories and category-based induction.
- Humans make inductive inferences that aren't logically valid.
- Human inferences are non-monotonic.

E.g., propositional or first-order logic.

What's a rational inference?

One answer:
- Defeasible/non-monotonic/fuzzy logics/predicate invention [not covered here].

Another:
- **Probability**.

Logical and probabilistic views are compatible
- Classical logics applicable when P is 0 or 1.
- Can express distributions over logical statements.

What's a rational inference?

Another view: **probability**

(1) Subjectivist/Bayesian probability:
P(*h*|*d*) is about subjective belief about *h* in light of *d*.

(2) Frequentist probability:
P(*h*|*d*) is about frequencies in the world.

We're talking about (1).

What's a rational inference?

Is probability theory rational?

Cox's theorem: Desiderata for talking about plausibility require that we use probability.

(1) The plausibility of a statement is a real number.
(2) "Common sense" [a handful of constraints on how plausibilities vary together].
(3) If we can derive the plausibility of a statement in two ways, they should agree.

Given particular mathematical expressions of these ideas, plausibility = probability.

You needn't know how (1-3) are expressed or the details of the theorem, but the curious might have a look at [1].

[1] Van Horn, "Constructing a logic of plausible inference: a guide to Cox's theorem"

What's a rational inference?

Is probability theory rational?

Dutch book: If you accept bets that are inconsistent with probability theory, someone can take your money.

Example:
You're betting whether a toss of a weighted coin will be heads (H).

http://plato.stanford.edu/entries/dutch-book/

Dutch book example

Not this kind of book

S: The stake; the total amount involved in the wager.
q: The betting quotient.

If the coin is heads, buyer **gets** S-qS.
If it's tails, buyer **loses** qS (the wager).

E.g., If the stake is £1 and q=0.1, heads gets you £9. Tails loses you £1.

The better chooses q based on beliefs about the coin. The bookie decides whether to buy or sell. Suppose S=1 for simplicity.

http://plato.stanford.edu/entries/dutch-book/

Dutch book example

Axioms of probability theory:

1. P(H) >= 0
2. The probability of at least one event occurring (a tautology) is 1
3. For a collection of disjoint (non-overlapping events), the probability of any happening is the sum of their separate probabilities.

If your policy for choosing q doesn't respect these rules, you lose.

http://plato.stanford.edu/entries/dutch-book/

Dutch book example

Consider:
1. P(H) >= 0

If q < 0, the bookie can buy the bet, thereby making
£-q on tails and
£(1-q) on heads.

Both are positive; you lose no matter what.

For the rest, see http://plato.stanford.edu/entries/dutch-book/

What's a rational inference?

Are human judgments consistent with probability theory?

- To test this, we can give problems where P(h|d) is computable and unambiguous and compare to human judgments.
- Last time, we saw evidence that they disagree.

What's a rational inference?

Base-rate neglect in the cab problem:

Probability says:

P(green|witness) = P(h)P(d|h)/P(d) =

0.15*0.80/(0.15*0.80+0.85*0.20) = 0.41

People say

>0.5.

What's a rational inference?

Conjunction fallacy:

Probability says:

P(accountant & jazz-player) =< P(jazz-player)

People say

accountant & jazz-player is more likely than jazz-player.

What's a rational inference?

So, people are irrational?

That depends on how strict our standard is.

What's a rational inference?

One extreme: We don't call human beings rational unless they always act in accordance with the prescriptions of probability theory.

What's a rational inference?

One extreme:

We aren't rational unless we always act in accordance with probability theory.

A test: For all of the slides you've seen so far, what's the probability that a randomly-selected one has fewer words than this one?

What's a rational inference?

Some alternatives:

(1) Bounded rationality: "…theories that incorporate constraints on the information-processing capacities of the actor…" (Simon, 1972)

(2) We answer different questions than the experimenters intended.

Bounded rationality

(1) Bounded rationality: "…theories that incorporate constraints on the information-processing capacities of the actor…" (Simon, 1972)

- We trade off between the value of an inference and the cost of obtaining it.
- We use heuristics that are usually indistinguishable from rational behaviour.
- We are as close to rational as our limited resources allow.

Task interpretation

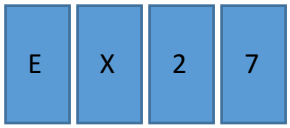
(2) Answering different questions than the experimenters intended

Claim: The conjunction fallacy is the result of not answering
"How likely is it that Bill is a jazz-playing accountant?"
but rather:
"How representative is Bill's description of a jazz-playing accountant?

(For a formal account, see "The rational basis of representativeness" by Tenenbaum & Griffiths)

Task interpretation

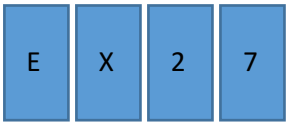
The rule: If there is a vowel on one side of a card, there is an even number on the other side.



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

Uncertainty reduction

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



What if we want to reduce our uncertainty about P(R=true)?
If the antecedents and consequents are rare, then we might want to flip the 2.

(Oaksford & Chater, 2009; 2003)

Uncertainty reduction

The rule: "If a person is a time-traveller, they will wear anachronistic clothing."

Do we:

1. Inspect the clothing of known time-travellers?
2. Inspect the clothing of known non-travellers?
3. Ask people wearing normal clothing if they are from the future?
4. Ask people wearing anachronistic clothing if they are from the future?

(Oaksford & Chater, 2009; 2003)

Uncertainty reduction

The rule: "If a person is a time-traveller, they will wear anachronistic clothing."

Do we:

1. Inspect the clothing of known time-travellers?
2. Inspect the clothing of known non-travellers?
3. Ask people wearing normal clothing if they are from the future?
4. Ask people wearing anachronistic clothing if they are from the future?

(Oaksford & Chater, 2009; 2003)

Summary

- Human rationality is a matter of one's standards
  - What model do we use – classical logic? Probability?
  - Can we justify our model as a rational one?
  - Even given a prescriptive model, there are different kinds of rationality.
- Classical decision-making tasks are open to multiple interpretations.
  - Wason's card-selection task.
  - Conjunction fallacy
  - (as well as base-rate neglect and others)