Reasoning with alternatives as Bayesian confirmation
Revisiting the lawyers and engineers problem

Mathias Sablé-Meyer[1]; Janek Guerrini[2]; Salvador Mascarenhas[2]

[1]Université PSL; Collège de France; UniCog
[2]Ecole Normale Supérieure; Department of Cognitive Studies; Institut Jean-Nicod

June 22, 2021

Introduction

Lawyers and engineers (Kahneman & Tversky, 1973, *Psychology of prediction*)

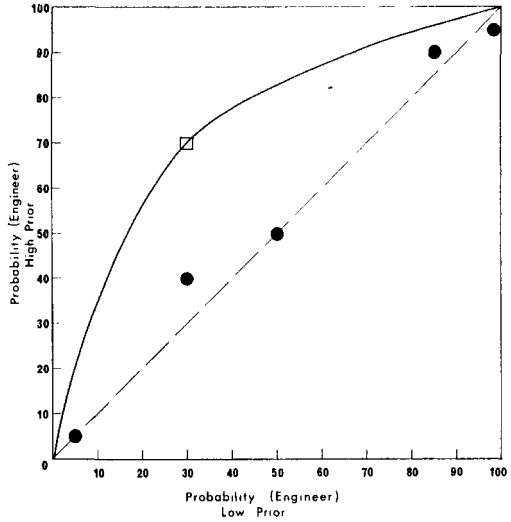
- Subjects were shown personality descriptions of individuals chosen at random from a group of lawyers and engineers. They were told descriptions had been composed by psychologists based on personality tests.

  Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

- Two between-subjects conditions:
  - 70% lawyers and 30% engineers
  - 30% lawyers and 70% engineers

- Question: what is the probability that Jack is an engineer?

Base-rate neglect and representativeness



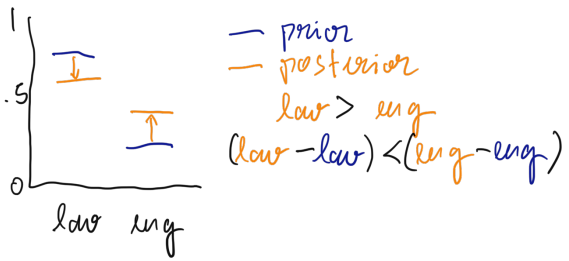
- Responses between the two conditions did not (seem to) vary according to the rational norm
- Instead, responses (seemed to) altogether **ignore the prior probabilities**
- Theory: **representativeness** (plus judgment by substitution heuristic)
  Probabilities are too hard. When faced with such questions, and in the presence of **familiar stereotypes** and **individuating information**, human reasoners substitute for the probabilities question an easier question in terms of typicality. How typical an example of the familiar stereotype for engineer is Jack?

Confirmation theory

- Recent approaches to **conjunction effects** from the representativeness literature have used **confirmation theory** (Tentori et al., 2013, *Determinants of the conjunction fallacy*)

Representativeness effects as confirmation

In the presence of competing **hypotheses** and some **evidence**, reasoners don't calculate posterior probabilities, but instead ask themselves about the extent to which the hypotheses under consideration are supported by the available evidence.



- A piece of evidence might **support a hypothesis** (engineer), while not raising its posterior probability enough to overtake its competitor (lawyer)
- Subtracting prior from posterior is one way to quantify evidential support

Zooming out: breadth and depth of confirmation-theoretical accounts

1. Multiple variants of the conjunction fallacy (Tentori et al., 2013, *Determinants of the conjunction fallacy*)
2. Medical decision making (Crupi et al., 2018, *Physicians' probability judgment*)
3. Lawyers and engineers (this talk)
4. Deductive reasoning with disjunctions (Sablé-Meyer & Mascarenhas, 2021, *Indirect illusory inferences*)

Question dynamics in reasoning

Humans find **questions** in many if not most reasoning problems (Koralus & Mascarenhas, 2013, 2018, *Erotetic theory of reasoning*). These questions can be straightforward (1., 2., and 3.) or hidden (4. and a wealth of other deductive inferences). As they find potential **answers** to those questions elsewhere in the reasoning problem, they would do well to apply pragmatic principles of **relevance**. Probabilistic theories of relevance will leverage confirmation-theoretic quantities rather than posterior probabilities.

Testing the theory and understanding lawyers and engineers

Better lawyers and engineers

- K&T (1973) could not exclude that probabilities besides posteriors were at play in responses to the lawyers and engineers problem
- Their much maligned setup was long winded and stacked mysterious elements that pushed towards the irrational response
  - "Psychologists wrote personality descriptions of individuals"
  - **The experimenter** "randomly" selected a description
  - Descriptions were long and piled on diagnostic information
- We came up with seven **minimal** triplets of h_1, h_2, e and normed them for each term in Bayes' formula: priors on the hypotheses, posteriors, and likelihoods
- Then we ran a lawyers-and-engineers task with the normed items

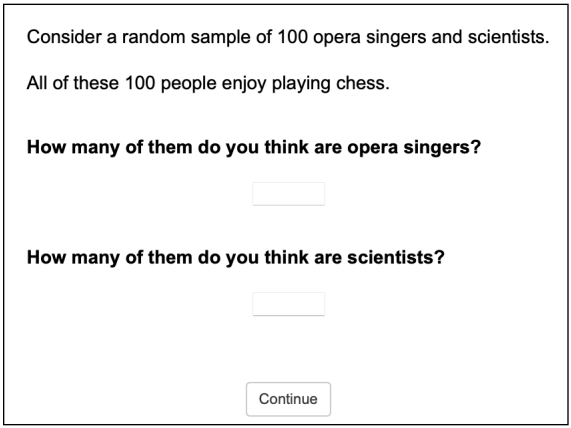
Norming study

- We collected probabilities in **frequency format** ("how many out of n" rather than "what percentage")
  - Known to make people more rational (Gigerenzer & Hoffrage, 1995)
- Trials were blocked by Bayes' term, the order of the blocks was randomized, and the order of the items was randomized internally to each block

- We collected
  - Prior on hypotheses: $P(h)$
  - Posteriors: $P(h|e)$
  - Likelihoods: $P(e|h)$
- We inferred:
  - $P(e) = P(e|h_1) \cdot P(h_1) + P(e|h_2) \cdot P(h_2)$

Norming study

Task example: norming of the posteriors ($P(h|e)$)



- We excluded any participants (40/120)
  - whose responses to questions that ought to sum up to 1 did not ($\pm.05$);
  - who responded to each question with one of 0, .5, or 1;
  - who responded with a non-number to **any** question, or who skipped **any** question

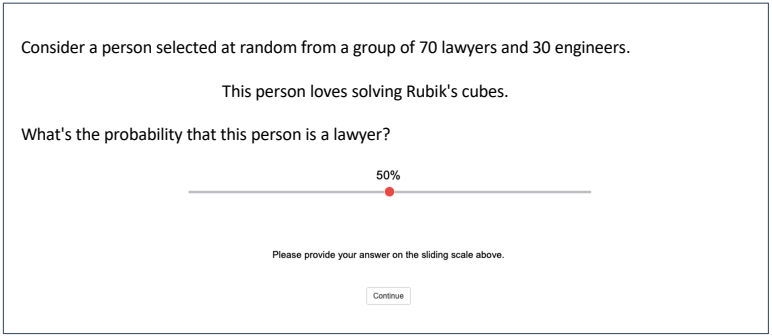
Norming study: internal rationality

- Participants' responses were **very coherent**:
  - their responses to the posteriors question was significantly predicted by their responses to likelihoods and priors via Bayes' theorem
  - LME model intercept .3, Bayes predictor estimate 0.45; model significantly better than null model (likelihood ratio test); all $p < 10^{-5}$

- This **confirms the effectiveness of our norming strategy**: the means we will use are meaningful as components of subjects' reasoning.
  - The consistency of the aggregate suggests that the means of the different factors of Bayes' laws are reasonable predictive elements

L&E study

- Using the normed materials, we ran a lawyers-and-engineers study with **minimal instructions** and no story about the source of the descriptions or the nature of the random selection
- We tested 120 subjects on all seven items times the two hypotheses for each item, blocked by hypothesis, with order of blocks and internal order of items randomized
- We tested five prior-probability conditions between subjects: .9–.1, .7–.3, .5–.5, .3–.7, .1–.9



Theories (1)

Posteriors (null theory)

$$\frac{target\_\mathbf{normed}\_likelihood \cdot target\_\mathbf{displayed}\_prior}{h1\_\mathbf{normed}\_likelihood \cdot h1\_\mathbf{displayed}\_prior + h2\_\mathbf{normed}\_likelihood \cdot h2\_\mathbf{displayed}\_prior}$$

An example

$$\frac{\mathbf{normed}\_P(Rubik's\_cube\_lover|lawyer) \cdot (.7)}{\mathbf{normed}\_P(Rubik's\_cube\_lover|engineer) \cdot (.3) + \mathbf{normed}\_P(Rubik's\_cube\_lover|lawyer) \cdot (.7)}$$

Theories (2)

Likelihoods

posteriors & normed_likelihood

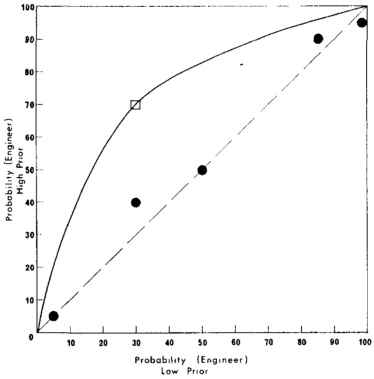
Difference

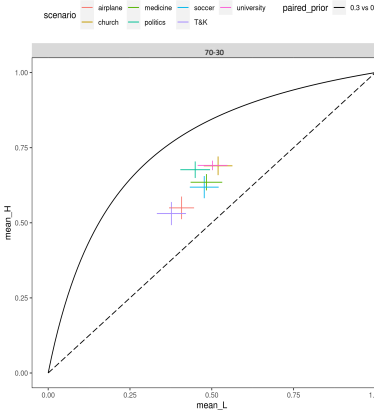
posteriors & (posteriors – displayed_prior)

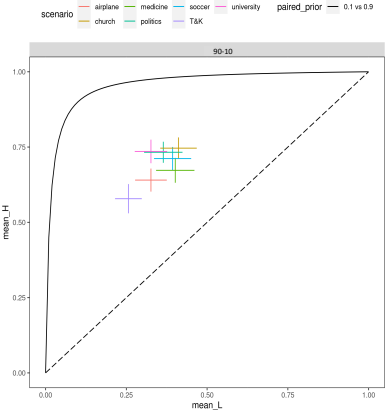
Ratio

posteriors & $\left(\frac{\text{posteriors}}{\text{displayed\_prior}}\right)$

Shifted rationality: likelihoods explain incomplete integration of priors







Results (1)

- We built LME models for the theories of interest and compared them
- A likelihoods ratio test shows that the posteriors & normed_likelihood performs significantly better than a posteriors alone models ($\chi^2 = 12.951$, $p < .0005$)

Rational model

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.38641	0.02233	151.41035	17.30	<2e-16	***
normed_posterior	0.37719	0.03624	138.54125	10.41	<2e-16	***

Rational + likelihoods model

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	3.313e-01	2.698e-02	3.221e+02	12.283	< 2e-16	***
normed_posterior	3.666e-01	3.615e-02	1.384e+02	10.144	< 2e-16	***
normed_likelihood	9.402e-02	2.610e-02	2.075e+03	3.602	0.000323	***

Results (2)

- We compared the predictions of three confirmation measures, looking at the Akaike information criterion and the Bayesian information criterion
- The model with posteriors and target likelihoods minimizes information loss the best.
  - E.g. the posteriors+ratio model has a relative likelihood less than 0.007 compared to the posteriors+target likelihoods model

	df	AIC
lme_pos_lik	10	-593.2680
lme_pos_ratio	10	-583.1611
lme_pos_difference	10	-581.6118

Lawyers and engineers distilled

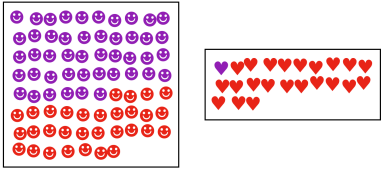
We set out to

1. remove stereotypes and individuating information as much as possible;
2. reduce language to a minimum;
3. provide participants with all they needed to give fully rational responses;
4. test the prediction that salient **alternatives** are essential

Design wise, this was a

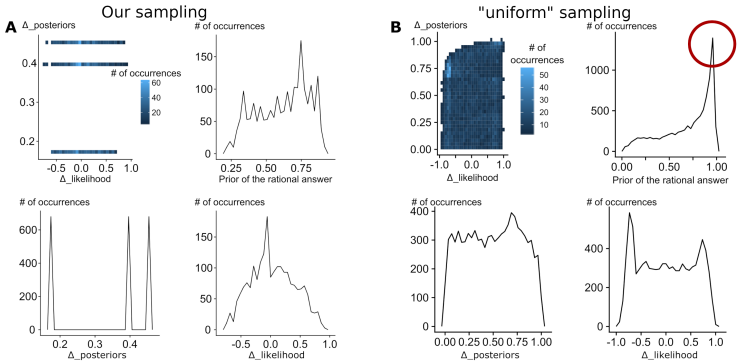
- Two between-subjects conditions: flat (all in one box) vs. structured (two boxes)
- Forced choice: is this a S_1 or a S_2? NB: not a question about probabilities

We used colors and symbols to convey the entire distribution for each trial:



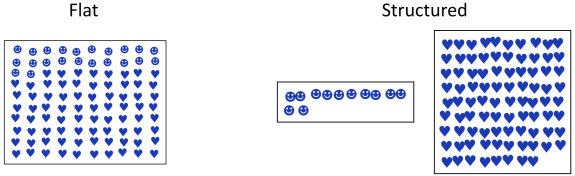
Lawyers and engineers distilled

- Fully controlled materials allowing for **broad sampling**
- Still, of course, constrained by Bayes' theorem.

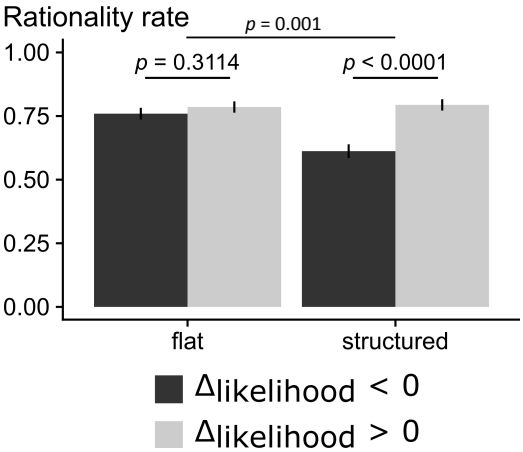


Possible confounds

- Potential confound in structured condition: participants might think there are **two random events**: pick a box at random, then pick an element from that box. Effectively this would make the prior probabilities on the categories idle.
- Control for confound: we tested subjects at the end of the experiment on evidenceless stimuli: the confounding interpretation would predict chance responses in these items. (Further control: "equally likely" option)



Results and theory



- We tested a more sophisticated theory: rationality would be predicted by an mixture of the salience of posteriors and the salience of comparative likelihoods
  - rational strategy: rat.l posterior − irrat.l posterior
  - erotetic strategy: rat.l likelihood − irrat.l likelihood
- We saw a main effect of condition & an interaction with the erotetic strategy, as predicted
- The rational + erotetic LME was highly significantly better than the rational alone model

Discussion

Confirmation theory and open questions

Confirmation-theoretic reasoning

When faced with candidate **hypotheses** and some **evidence**, one strategy humans apply, irrespective of the exact phrasing of the task at hand, is to consider the extent to which **the evidence confirms a target hypothesis**.

Confirmation theory and open questions

Confirmation-theoretic reasoning

When faced with candidate **hypotheses** and some **evidence**, one strategy humans apply, irrespective of the exact phrasing of the task at hand, is to consider the extent to which **the evidence confirms a target hypothesis**.

1. *How* do reasoners find **hypotheses** and **evidence**?
2. *Why* do reasoners engage in confirmation-theoretic reasoning to begin with?

A theory of questions

Erotetic theory of reasoning (Koralus & Mascarenhas, 2013, 2018)

- Representing **candidate alternatives** is the same as representing **a question**
- When humans represent a question, they seek to resolve it as soon as possible, by trying to find something to interpret as **an answer** elsewhere
- Reasoners find **questions** in entirely non-obvious places such as purely declarative sentences, that are however **predictable** with the right theories of meaning

Indirect fallacies with disjunction (Sablé-Meyer & Mascarenhas, 2021, *Indirect illusory inferences*)

The gun fired and the guitar was out of tune, or else someone was in the attic.
The trigger was pulled.
Does it follow that *the guitar was out of tune*?

A theory of questions

Erotetic theory of reasoning (Koralus & Mascarenhas, 2013, 2018)

- Representing **candidate alternatives** is the same as representing **a question**
- When humans represent a question, they seek to resolve it as soon as possible, by trying to find something to interpret as **an answer** elsewhere
- Reasoners find **questions** in entirely non-obvious places such as purely declarative sentences, that are however **predictable** with the right theories of meaning

Indirect fallacies with disjunction (Sablé-Meyer & Mascarenhas, 2021, *Indirect illusory inferences*)

The gun fired and the guitar was out of tune, or else someone was in the attic.
The trigger was pulled.
Does it follow that *the guitar was out of tune*?

(gun-fire AND guitar)$_h$ OR (attic)$_{h'}$
(trigger)$_e$

A theory of answers

The answers to all your questions

Under an assumption of **cooperativeness** and **informativeness**, we should interpret an **answer** with mitigated regard for prior probabilities.

A theory of answers

The answers to all your questions

Under an assumption of **cooperativeness** and **informativeness**, we should interpret an **answer** with mitigated regard for prior probabilities.

- Gricean relevance, Relevance Theory, the Rational Speech Act model, all provide **justifications** for an oversize role for confirmation-theoretic reasoning when answering questions
- Reasoners are far more rational in lawyers-and-engineers when **a computer** randomly selects Jack's description (Schwarz et al., 1991, *Relevance of "irrelevant" information*)

Open question

Is this an entirely productive, on-the-fly pragmatic process? Or has it become somewhat crystallized in the human faculty for reasoning?

Conclusions

- Confirmation-theoretic reasoning accounts for a broad range of data, starting with the conjunction fallacy, but extending to superficially unrelated deductive problems
- We've shown that aspects of base-rate neglect can be analyzed in the same way, and there isn't enough evidence in favor of a central role for **stereotypes** and **individuating information**
- We need a theory that can generate hypotheses and evidence from reasoning problems,
- and a theory of *why* confirmation-theoretic reasoning happens altogether
- We propose that the connection with independent experimental results and theories on question-answer dynamics provide a new and promising avenue to answer these central theoretical questions

Thank you!

The LANG-REASON team, Vincenzo Crupi, Katya Tentori,
Benjamin Spector, Ulrike Hahn, Emmanuel Chemla, Alina Dracheva

Agence Nationale de la Recherche grants
ANR-17-EURE-0017 (FrontCog: DEC) and
ANR-18-CE28-0008 (LANG-REASON, PI: Mascarenhas)