# Reasoning with alternatives as a confirmation-theoretic process

Salvador Mascarenhas

Ecole Normale Supérieure Department of Cognitive Studies Institut Jean-Nicod

> LLC Webinar May 27, 2021

- Deductive fallacies involving *reasoning with alternatives* require a theory of mental representations that goes beyond classical logic.
- These fallacies are closely connected to superficially very dissimilar fallacies in probabilistic reasoning.
- A *confirmation theoretic* account that builds on *question-answer dynamics* can bridge this gap, thereby extending the empirical scope of confirmation-theoretic reasoning, and providing a new promising component of the explanatory dimension of confirmation theory in reasoning.

## Deduction from alternatives

# Illusory inferences from disjunction

- $P_1$ : Either Jane is kneeling by the fire and she is looking at the TV or otherwise Mark is standing at the window and he is peering into the garden.
- $P_2$ : Jane is kneeling by the fire.
- Concl.: Jane is looking at the TV.

$$P_1: (a \land b) \lor (c \land d)$$
$$P_2: a$$
$$Concl: b$$

- discovered by Walsh & Johnson-Laird (2004)
- 85%–90% acceptance, replicated multiple times
- Independent of how disjunction is expressed in English
- not a simple matter of exclusive 'or'

# Illusory inferences from disjunction(-like elements)

 John speaks English and Mary speaks French, or else Bill speaks German. John speaks English. Mary speaks French?

# Illusory inferences from disjunction(-like elements)

- John speaks English and Mary speaks French, or else Bill speaks German. John speaks English. Mary speaks French?
- 2 Mary has met every king or every queen of Europe. Mary has met the king of the Netherlands. Mary has met the king of Spain?
- 3 Some pilot writes poems. John is pilot. John writes poems?
- The gun fired and the guitar was out of tune, or else someone was in the attic.
   The trigger was pulled.
   The guitar was out of tune?

- 5 Each person at this party is either French or a linguist.That guy is European.That guy is French?
- 6 This party is only for French people and linguists. That guy is European. *That guy is French?*
- 7 John might speak English and French. John speaks English. John speaks French?

#### Mental models

My mental-models-inspired account to come

#### Mental models

My mental-models-inspired account to come

#### Not Xor

 $(a \wedge b \wedge \neg c) \lor (c \wedge \neg (a \wedge b))$ 

#### Mental models

My mental-models-inspired account to come

#### Not Xor

$$(a \wedge b \wedge \neg c) \lor (c \wedge \neg (a \wedge b))$$

# Formal pragmatics (but it's not the whole story, Picat, 2019) $(a \land b \land \neg c) \lor (c \land \neg a \land \neg b)$

#### Mental models

My mental-models-inspired account to come

#### Not Xor

$$(a \wedge b \wedge \neg c) \lor (c \wedge \neg (a \wedge b))$$

Formal pragmatics (but it's not the whole story, Picat, 2019)  $(a \land b \land \neg c) \lor (c \land \neg a \land \neg b)$ 

Not posterior probabilities (New Paradigm)  $P((a \land b) \lor c, a) = P((a \land b) \lor (a \land c))$ 

# The erotetic theory of reasoning (Koralus & Mascarenhas, 2013)

#### The erotetic principle

- Part I Our natural capacity for reasoning proceeds by treating successive premises as questions and maximally strong answers to them. (problem of failure)
- Part II Systematically asking a certain type of question as we interpret each new premise allows us to reason in a classically valid way. (problem of success)

#### Commitment on interpretation

Disjunctions propose alternatives and put pressure toward choosing an alternative — disjunctions are like questions in this regard (Inquisitive Semantics: Mascarenhas, 2009, Groenendijk, 2008; Alternative semantics: Kratzer & Shimoyama, 2002, Alonso-Ovalle, 2006)

- $P_1$ : John is watching TV and Mary is playing tennis, or Bill is doing homework.
- $P_2$ : John is watching TV.
- C: Mary is playing tennis.

- $P_1$ : John is watching TV and Mary is playing tennis, or Bill is doing homework.
- $P_2$ : John is watching TV.
- C: Mary is playing tennis.

#### Question

Are we in a John-watching-TV and Mary-playing-tennis situation, or in a Bill-doing-homework situation?

- $P_1$ : John is watching TV and Mary is playing tennis, or Bill is doing homework.
- $P_2$ : John is watching TV.
- C: Mary is playing tennis.

#### Question

Are we in a John-watching-TV and Mary-playing-tennis situation, or in a Bill-doing-homework situation?

Incomplete answer

We are in a John-watching-TV situation.

- $P_1$ : John is watching TV and Mary is playing tennis, or Bill is doing homework.
- $P_2$ : John is watching TV.
- C: Mary is playing tennis.

#### Question

Are we in a John-watching-TV and Mary-playing-tennis situation, or in a Bill-doing-homework situation?

Incomplete answer

We are in a John-watching-TV situation.

#### Jumping to conclusions

I see, so the first answer to the question is the true answer — content overlap

Salvador Mascarenhas (ENS-DEC-IJN)

# Disjunctions are like questions?

### A(n in)famous fact

Very many natural languages have the same morphemes for the interrogative complementizer and disjunction operator (and indefinites, more on which later)

Malayalam is a good example (Jayaseelan, 2004)

John-oo	Bill-oo	wannu.	Mary	wannu-oo?
John-or	Bill-or	came.	Mary	came-or
"John or	Bill came	e"	"Did I	Mary come"

(cf. also Japanese 'ka,' Korean 'na,' several variations of Slavic 'li,' Polish 'czy,' and so on)

# The crucial update rule

# Q(uestion)-Update

$$\mathsf{F}[\Delta]^{\mathsf{Q}} = \mathsf{F} - \{\gamma \in \mathsf{F} : (\square \Delta) \sqcap \gamma = 0\}$$

Let  $\Gamma$  be the first premise after it has been processed, and  $\Delta$  the second premise. Q-Update eliminates from  $\Gamma$  (the "question") all alternatives that have nothing in common with the intersection of all alternatives in  $\Delta$ . In other words: take the information in  $\Delta$ , that is the intersection of all alternatives in  $\Delta$ . Keep in  $\Gamma$  only those alternatives that share some mental molecule with the information in  $\Delta$ .

# Success on the erotetic theory

- Sound reasoning is possible for humans, so there must be some strategy using our natural faculties that guarantees it.
- This strategy has to be innate or learnable, but it should be costly, or we won't explain why the mind doesn't use it all the time.

#### Inquire, or Part 2 of the erotetic principle

 Systematically asking a certain type of question as we interpret each new premise allows us to reason in a classically valid way.

$$\mathsf{\Gamma}[\Delta]^{\mathsf{Inq}} = \mathsf{\Gamma}[\Delta \cup \operatorname{neg}(\Delta)]^{\mathsf{C}}[\cdot]^{\mathsf{F}}$$

# Soundness theorem

#### Soundness theorem

The ETR derivation strategy where an update with  $\Delta$  is immediately preceded by a sequence of Inquire steps for each atom p that occurs somewhere in  $\Delta$  is sound for classical propositional models.

#### Proof sketch

We need to show by induction that, for this class of derivations, every move will preserve classical validity. All of the operations not discussed today happen to be immediately sound, so the interesting step is Q-Update, which we looked at briefly. Q-Update eliminates alternatives not containing the intersection of  $\Delta$ . But if Inquire on all atoms in  $\Delta$  is guaranteed, then every alternative that Q-Update with  $\Delta$  eliminates is going to contain the negation of something in the intersection of  $\Delta$ , and can therefore be soundly eliminated.

Koralus & Mascarenhas (2013)

# Indirect illusory inferences from disjunction (with Sablé-Meyer)

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?

- $\begin{array}{cc} \mathsf{P}_1\colon & (a\wedge b)\vee c\\ \mathsf{P}_2\colon & d\\ \mathsf{Conclusion}\colon & b \end{array}$
- A norming group rated the strength of the inference from d to a (controlling for  $d \rightarrow b$  and  $a \rightarrow b$ ). Another group performed an inference task on the schema above
- Linear regression showed strong correlation: Group 2 accepted the inference with a frequency proportional to the extent to which Group 1 found the *d* → *a* conditional attractive.



	Estimate	Std. err	<i>t</i> -value	<i>p</i> -value
intercept	-0.095	0.110	-0.861	0.429
means	1.001	0.157	6.371	0.001

# The conjunction fallacy and indirect illusory inferences

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable? (Tversky & Kahneman, 1983)

- 1 Linda is a bank teller.
- 2 Linda is a bank teller and she is active in the feminist movement.

# The conjunction fallacy and indirect illusory inferences

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable? (Tversky & Kahneman, 1983)

- **1** Linda is a bank teller.
- 2 Linda is a bank teller and she is active in the feminist movement.

The gun fired and the guitar was out of tune, or else someone was in the attic.The trigger was pulled.The guitar was out of tune?(Sablé-Meyer & Mascarenhas, 2021)

# The conjunction fallacy and indirect illusory inferences

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable? (Tversky & Kahneman, 1983)

- **1** Linda is a bank teller.
- 2 Linda is a bank teller and she is active in the feminist movement.

The gun fired and the guitar was out of tune, or else someone was in the attic. The trigger was pulled. *The guitar was out of tune?* (Sablé-Meyer & Mascarenhas, 2021)

(gun and guitar) or attic trigger gun and guitar? (feminist and bank) or bank socially-aware bank and feminist?

## Reasoners as mini-theorists

#### Question for the erotetic theory

How exactly does "the trigger was pulled" answer the question in the "gun fired and guitar was out of tune" direction rather than "someone was in the attic?"

#### Proposal

When answering questions, reasoners attempt to weave a story with predictive and explanatory power relating the answer to the question. They engage in a kind of **confirmation reasoning**.

# Reasoners as mini-theorists

#### Question for the erotetic theory

How exactly does "the trigger was pulled" answer the question in the "gun fired and guitar was out of tune" direction rather than "someone was in the attic?"

#### Proposal

When answering questions, reasoners attempt to weave a story with predictive and explanatory power relating the answer to the question. They engage in a kind of **confirmation reasoning**.

- *h*<sub>1</sub>: The gun fired and the guitar was out of tune.
- $h_2$ : Someone was in the attic.
- e: The trigger was pulled.

*e* is (much) better evidence for  $h_1$  than for  $h_2$ .  $h_1$  **explains** the evidence *e* (much) better than  $h_2$  does.

# Confirmation theory

- Formal epistemologists investigate various criteria for evidential support. These notions have shown great power as accounts of reasoning in probabilistic problems such as the conjunction fallacy (Crupi et al., 2008, Tentori et al., 2013)
- One pedestrian but surprisingly efficacious strategy is to consider the likelihoods

P(trigger|gun) > P(trigger|attic)

A more sophisticated one is to look at the difference between posterior and prior probabilities

P(gun|trigger) - P(gun) > P(attic|trigger) - P(attic)

# Confirmation theory

- Formal epistemologists investigate various criteria for **evidential support**. These notions have shown great power as accounts of reasoning in probabilistic problems such as the conjunction fallacy (Crupi et al., 2008, Tentori et al., 2013)
- One pedestrian but surprisingly efficacious strategy is to consider the likelihoods

P(trigger|gun) > P(trigger|attic)

A more sophisticated one is to look at the difference between posterior and prior probabilities

$$P(gun|trigger) - P(gun) > P(attic|trigger) - P(attic)$$

#### Erotetic confirmation theory

Reasoners decide between **alternative hypotheses** (the question) by comparing their likelihoods with respect to the **evidence** (the answer)

# Reasoning with probabilities

#### Lawyers and engineers (Kahneman & Tversky, 1973)

• 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 lawyer is Jack?

#### Issues

#### Experimental

- The presence of base-rate neglect in the responses is questionable Only one item (one description) clearly displays the effect
- Design did not allow for a complete analysis of fallacious behavior The role that other relevant probabilities might play is inscrutable in the experiment

#### Theoretical

- Competing accounts of the phenomenon were dismissed too quickly Subjects might have relied on the *inverse probabilities* (likelihoods, i.e. probability of the description conditional on the category)
- Central elements of the representativeness account were not tested Reliance on familiar stereotypes and individuating information (i.e. a complete description of "John")

# Alternative account

- Participants aren't simply ignoring the prior probabilities and reasoning based on representativeness.
- They are engaged in an erotetic confirmation process where the professions are the two hypotheses and the description the evidence.
- Instead of being poor probabilistic reasoners, participants are doing something even more sophisticated with probabilities than merely calculating posteriors.

# Three birds, one stone

# Indirect illusory inferences from disjunction Slides 14–15

#### The conjunction fallacy

- What T&K thought people were saying: P(b ∧ f|d) > P(b|d)
- What people might in fact be saying:  $P(d|b \wedge f) > P(d|b)$

Erotetic formulation of the problem:
 P<sub>1</sub>: b ∨ (b ∧ f)
 P<sub>2</sub>: d
 d raises the probability of f and so it answers the question in the b ∧ f direction rather than b

#### Binary lawyers and engineers

P(d|engineer) > P(d|lawyer)

# Testing the theory and understanding lawyers and engineers

### Better lawyers and engineers (with Guerrini & Sablé-Meyer)

- 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*<sub>1</sub>, *h*<sub>2</sub>, *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")
- 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 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
- 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<sup>-5</sup>

# 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

Consider a person selected at random from a group of 70 lawyers and 30 engineers.									
This person loves solving Rubik's cubes.									
What's the probability that this person is a lawyer?									
50%									
Please provide your answer on the sliding scale above.									
Continue									

### Theories

### Posteriors (null theory)

normed\_likelihood · displayed\_prior normed\_evidence\_prior

Likelihoods

posteriors & normed\_likelihood

Difference

posteriors & (posteriors - displayed\_prior)

# Results (1)

We built LME models for the theories of interest and compared them Model convergence was extremely elusive with items as random effects (only 7 groups). We included them as fixed effects in the first instance, observed that they displayed no main effect or interactions, and then removed them from the fixed effects

#### 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 cum confirmation 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)

#### Model comparison

	$\mathtt{npar}$	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)	
rational	6	-590.69	-556.41	301.35	-602.69				
rational_erotetic	7	-601.64	-561.64	307.82	-615.64	12.951	1	0.0003197	***

# Results (2)

#### Model comparison

npar AIC BIC logLik deviance Chisq Df Pr(>Chisq) rational 6 -590.69 -556.41 301.35 -602.69 rational\_erotetic 7 -601.64 -561.64 307.82 -615.64 12.951 1 0.0003197 \*\*\*

#### Bonus: base-rate neglect?

	Estimate	Std. Error	df	t value	$\Pr(> t )$	
(Intercept)	0.38569	0.02355	145.21948	16.380	<2e-16	***
priors	0.37864	0.03887	127.17161	9.742	<2e-16	***

#### Conclusions

- An erotetic theory in terms of likelihoods explains the non-rational part of participants' behavior better than the null hypothesis (noise)
- We found no evidence of base-rate neglect as a strategy in itself (WIP)

# Results (3)

• We compared the predictions of three confirmation measures, looking at the Akaike information criterion and the Bayesian information criterion

 df
 AIC

 lme\_pos\_lik
 10
 -593.2680

 lme\_pos\_ratio
 10
 -583.1611

 lme\_pos\_difference
 10
 -581.6118

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

#### For discussion

Likelihoods alone are not a *Bayesian* measure of confirmation, many consider them to be a poor normative model, and they pose some conceptual puzzles. Yet they perform best in our study.

### Lawyers and engineers distilled (with Sablé-Meyer & Guerrini)

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 erotetic prediction that alternatives (the question) are essential

(This study is close to a recently published article from Tentori's lab [Mangiarulo et al., 2021]. Point 4. is of course entirely new to our study, and we had a different approach to point 3.)

# Lawyers and engineers distilled (with Sablé-Meyer & Guerrini)

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 erotetic prediction that **alternatives** (the question) are essential



# Design

- 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
- Symbols were randomly generated
- Fully controlled materials allowing for far broader sampling along the terms of Bayes' theorem
- Participants (120 per experiment) saw 15 items
- 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 evidencelss stimuli: the confounding interpretation would predict chance responses in these items. (Further control: "equally likely" option)

#### Visual L&E

# Sampling pains



# Familiarity and stereotypes



"I picked the shape that was more typically the specified color in real life, e.g. a red heart"

About 10% explicitly answered this in a bunch of control questions: data confirms this.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
color	6	0.41	0.07	0.44	0.85
eSymbolName	2	0.18	0.09	0.58	0.56
color:eSymbolName	12	3.58	0.3	1.94	0.03
Residuals	11205	1724	0.15		

# Results and theory

Rationality rate 1.00 p = 0.3114p < 0.00010.75 0.50 0.25 0.00 flat structured ■ ∆likelihood < 0</p>  $\Delta$ likelihood > 0

- We tested a more sophisticated theory: rationality would be predicted by an amalgam 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, and an interaction with the erotetic strategy, as predicted
- The rational + erotetic LME was highly significantly better than the null (rational alone) model

# Concluding remarks

# A reason for confirmation-theoretic reasoning

- There isn't a consensus picture yet on why humans would ever engage in confirmation-theoretic reasoning, as opposed to maximizing posterior probabilities. Some possibilities:
  - Confirmation measures connect substantially to causal relations, simple posteriors do not
  - Confirmation measures connect to categorization, as a measure of diagnosticity
- If today's theory is on the right track, erotetics can provide a reason for confirmation-theoretic reasoning: when interpreting an indirect answer to a question, focusing on confirmation is rational

# Erotetic confirmation and the RSA

#### Erotetic confirmation

The rational strategy when interpreting something meant as an answer to a question is to consider how that answer changes our beliefs about the possible answers, i.e. confirmation.

- A version of this falls out of the Rational Speech Act (Goodman & collaborators), which has an amplification effect on likelihoods
- This is not to say that "this is all just rational pragmatics!" Other elements of rational pragmatics have clearly been crystalized in language already (scalar implicature), something similar could have happened in reasoning.
- But it does mean that the connection with question-answering needs to be explored more carefully, and the confounds of communication need to be better controlled for.

# Conclusions

- Confirmation-theoretic reasoning accounts for a broad range of data, starting with the conjunction fallacy.
- I've shown that aspects of base-rate neglect can be analyzed in the same way, as well as some deductive problems involving reasoning with alternatives.
- Something like the erotetic theory is needed to understand where non-interrogative alternatives come from.
- The *explanatory* power of confirmation-theoretic reasoning is still somewhat elusive, but the connection with question-answer dynamics is a new and promising avenue.

# Thank you!

Mathias Sablé-Meyer, Janek Guerrini, the LANG-REASON team, Vincenzo Crupi, Benjamin Spector, Ulrike Hahn, Emmanuel Chemla, Alina Dracheva

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

