Reasoning with disjunctions as a form of hypothesis testing

Salvador Mascarenhas & Mathias Sablé Meyer
Ecole Normale Supérieure, Department of Cognitive Studies, Institut Jean Nicod

LPPRD Seminar, February 6th, 2018

GOALS FOR TODAY & TAKE-HOME MESSAGE

- There is a relatively diverse paradigm of compelling fallacies involving disjunction, of a kind unfamiliar to most philosophers and psychologists of reasoning.
- Call them **illusory inferences from disjunction**:

(1) John speaks English and Mary speaks French, or else Bill speaks German.
John speaks English.
Does it follow that Mary speaks French? (Walsh and Johnson-Laird, 2004)

(2) The card in my hand is an ace and a heart, or else it's a black card.
The card is an ace.
Does it follow that the card is a heart? (Sablé Meyer and Masc., *in prep.*)

(3) Mary has met every king or every queen of Europe.
Mary has met the king of the Netherlands.
Does it follow that Mary has met the king of Spain? (Masc. and Koralus, 2017)

(4) Some pilot writes poems.
John is a pilot.
Does it follow that John writes poems? (Mascarenhas and Koralus, 2017)

(5) The gun fired & 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? (Sablé M. and Masc., *in prep.*)

(6) Each person at this party is either French or a linguist.
That guy is European.
Does it follow that he is French? (Sablé Meyer and Mascarenhas, *in prep.*)

(7) This party is only for French people and linguists.
That guy is European.
Does it follow that he is French? (Sablé Meyer and Mascarenhas, *in prep.*)

- The patterns in (1)–(4) are explained within mental-models approaches (Johnson-Laird, 1983) with resort to **a special semantics for disjunction** (and relevantly disjunction-like elements)[1] and a **syntactic matching procedure**.
- In a nutshell: the second premise **matches** part of the first disjunct and not the second disjunct. The latter drops from attention, leaving us the first disjunct, whence the fallacious conclusion follows by (some analog of) conjunction elimination.
- These same data points are very elusive from the perspective of extant probabilistic approaches to reasoning, which are the most well-subscribed approaches to many other compelling fallacious patterns in the literature.
- The class of matching procedures just sketched won't take us far enough: the data in (5)–(7) do not display the necessary configuration for any of the matching proposals in the literature.
- In addition, despite the difficulties of probabilistic approaches in accounting for any of these inference patterns, there are strong independent arguments for the role of credences, probabilities, and rational update procedures in naive reasoning.
- The work we present today
  - takes what is good about existing accounts of (part of) the illusory inference from disjunction paradigm — a special semantics for disjunction and a question-answer dynamics
  - replaces syntactic matching with a probabilistic procedure,
  - offering a full account of the paradigm — the only extant full account,
  - unifying essentially non-probabilistic ingredients with probabilistic ones, and in the process
  - elucidating the nature of (part of) the probabilistic component of human reasoning.
- Tentative bonus: a possible contribution to the question of what ought to characterize a competence theory of semantics.

[1] We can think of "Some pilot writes poems" as an infinitary disjunction of conjunctions "x_0 is a pilot and writes poems OR x_1 is a pilot and writes poems, …"

1. Background: illusory inferences from disjunction and syntactic matching

- The class of **illusory inferences from disjunction** this talk is about was discovered by Walsh and Johnson-Laird (2004).
- About 85% of subjects accepted the inference in (8).

(8)
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.

(9)
P_1: $(a \wedge b) \vee (c \wedge d)$
P_2: a
Concl.: b

- **Fallacy**: falsified at a model where a, c, and d are true, but b is false.
- Caveat: there is an interpretation-based account of the pattern in (9), in terms of scalar implicature (Mascarenhas, 2013, 2014). We have good reason to think that the phenomenon in (9) is due both to a scalar implicature and to the kinds of mechanisms this talk is about. But the jury is still out.
- Walsh and Johnson-Laird (2004) found no effect of dependencies between the atomic propositions.
- In our studies (Mascarenhas and Koralus, 2017), we found no effect of the way disjunction is expressed in inferences like (8).
- A class of inferences of the same name had been identified by Johnson-Laird and Savary (1999):

(10)
P_1: If there is a king in the hand then there is an ace in the hand, or else if there is a queen in the hand then there is an ace in the hand.
P_2: There is a king in the hand.
Concl.: There is an ace in the hand.

- A comparable proportion of subjects judged that (10) was valid.
- But (10) is very plausibly about something else: there is a "conjunctive feel" to P_1 of (10) that is absent from P_1 of (8). We take it that (10) illustrates a phenomenon largely orthogonal to what's behind (8).
- On **mental models theory**, illusory inferences like (8) are analyzed as follows.

Mental models account (Johnson-Laird, 1983; Johnson-Laird and Savary, 1999; Walsh and Johnson-Laird, 2004)

1. Reasoners build mental representations (mental models) that verify each of the premises.
2. Disjunctive premises are represented as sets of alternative mental models.
3. P_1 gives rise to a set of two alternative models: a minimal model of $a \wedge b$ and a minimal model of $c \wedge d$.
4. **Upon hearing P_2, a, reasoners notice that it is related to the first alternative model for P_1, but not the second.** This makes them ignore the second model.
5. The combined representation of the premises is therefore only one mental model: $a \wedge b$. From here, b follows.

- Step 4. in the MMs account is dubbed the **matching procedure**.
- The original MMs account lacks (among other things . . .)
  1. a precise regimentation of the representational system: what are mental models?
  2. an explanation of the workings of the matching procedure
  3. an explanation for the peculiar interpretation of disjunctions
- In part to address these questions, Koralus and Mascarenhas (2013) give a variant of mental models theory called the **erotetic theory of reasoning** (ETR).
- On ETR, mental models are sets of **exact verifiers**.
- We build on van Fraassen (1969) and Fine's 2012 **truth-maker semantics** — closely related to inquisitive semantics

(11) $\varphi \vee (\varphi \wedge \psi)$
  a. You're allowed to have a cookie or a cookie and some ice-cream.
  b. You're allowed to have a cookie.

- Convergence with inquisitive semantics (Groenendijk, 2008b; **?**): disjunctive premises give rise to mental models with multiple alternatives. Multiple alternatives are interpreted as questions.
- We get representations like the following in (12)

(12) $\{a \sqcup b, c \sqcup d\}$
$\{a\}$

- ETR postulates a novel goal for the human faculty for reasoning: to **answer questions** by reducing the number of alternatives under consideration at any given point.
- The operation analogous to MMs matching is Q(uestion)-Update.

Q-Update (from Koralus and Mascarenhas, 2013)

$$\Gamma[\Delta]^{Q} = \Gamma - \{\gamma \in \Gamma : (\bigsqcap \Delta) \sqcap \gamma = 0\}$$

Q-Update eliminates from Γ (the "question") all alternatives that have nothing in common with the intersection of all alternatives in Δ. In other words: take the information in Δ, that is the intersection of all alternatives in Δ. Keep in Γ only those alternatives that share some mental molecule with the information in Δ.

- **So**, the matching procedure looks at the **syntax** of mental representations (mental models) and requires **partial overlap**.
- This works pretty well for the patterns in (1)–(4) (classical illusory inferences and versions with quantifiers), but it fails for the rest of the paradigm.

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

(14) $(a \wedge b) \vee c$
d
Does it follow that b?
(where independently $d \nearrow a$)

- Whatever is going on with (13), it doesn't look like matching. If anything, it looks a lot like the result of a probabilistic procedure ...
- Even with additional assumptions about what is stored in the background knowledge, the Q-Update is a matching procedure

$$\{a \sqcup b, c\}[\{d\}]^{Q} = \{a \sqcup b, c\} - \{\gamma \in \{a \sqcup b, c\} : d \sqcap \gamma = 0\}$$

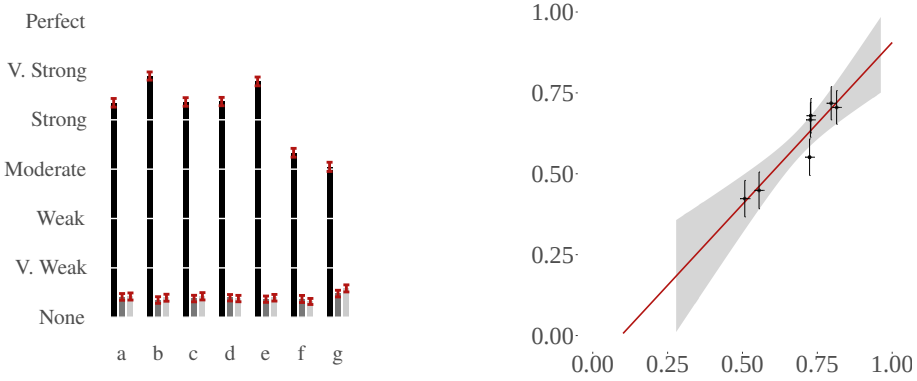

Figure 1: Left: Results from the rating task. Black is $d \to a$, dark gray is $d \to b$ and light gray is $a \to b$. Right: Correlation between the rating task (x axis, ordinal responses) and the inference task (y axis, frequencies of binary responses).

- Sablé Meyer and Mascarenhas (*in prep.*) recently showed that the pattern in (14) is very attractive, and that the extent to which subjects consider that b follows is closely correlated with the extent to which they endorse the conditional sentence $d \to a$.
- We took inspiration from work by Cummins et al. (1991); Cummins (1995) on inferences with conditionals.
- **Rating task:** subjects saw multiple $d \to a$ conditionals and were asked to rate the strength of the connection expressed by those conditionals on a 7-point scale.

(15)
a. If the brake was depressed, then the car slowed down.
b. If Mary jumped into the swimming pool, then Mary got wet.
c. If the trigger was pulled, then the gun fired.
d. If Larry grasped the glass with his bare hands, then Larry left fingerprints on his glass.
e. If the gong was struck, then the gong sounded.
f. If John studied hard, then John did well on the test.
g. If the apples were ripe, then the apples fell from the tree.

- **Inference task:** a different group of subjects performed a yes/no "does b follow" task on the pattern in (14).

- The possibility of these **indirect illusory inferences**, along with the correlation between endorsement of the hidden conditional and the attractiveness of the corresponding inference, suggest an approach that distinguishes between different levels of credence.
- Same goes for (6) and (7) with sets and subsets — plus scalar implicature stories have no good story about that.

2. Probabilistic approaches to reasoning with disjunctions

- Psychologists have always been interested in degrees of belief, and have naturally modeled that notion with recourse to probability measures, but the issue is fraught within philosophy and linguistic semantics.
- There is a longstanding analytic intuition that probability talk (Hamblin, 1959; Kratzer, 1991) and reasoning with degrees of belief (Harman, 1986) are about something *less* than probability measures.
- But more recent work has shown the benefits of going full probabilistic with probability talk (Lassiter, 2011), pragmatic reasoning (Frank and Goodman, 2012), and general-purpose reasoning (Oaksford and Chater, 2007).
- We take a different route in this talk: **Bayesian confirmation theory**.

An illustration of confirmation theory in reasoning

d = "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." (Tversky and Kahneman, 1983)

Which is more likely?
(A) Linda is a bank teller. b
(B) Linda is a b. t. who is active in the feminist movement. $b \wedge f$

(16) What T&K should people were saying
$P(b \wedge f|d) > P(b|d)$

(17) What people might in fact be saying
$P(d|b \wedge f) > P(d|b)$

(18) Erotetic formulation of the problem
P_1: $b \vee (b \wedge f)$
P_2: d
$d \nearrow f$ thus d answers the question in the $b \wedge f$ direction rather than b

- Confirmation theory is about hypothesis testing. Its goal is to characterize the extent to which a piece of evidence e **confirms** a hypothesis h.
- Confirmation theorists have proposed purely comparative measures that decide between candidate hypotheses satisfying certain conditions (e.g. mutual exclusivity).
- But most researchers endorse non-comparative confirmation measures.
- There are three main confirmation measures used in the literature. The Linda example illustrate the most pedestrian and inadequate of the three: likelihoodism.
- Crupi et al. (2008); Tentori et al. (2013) show that the account of the conjunction fallacy just sketched works for more sophisticated confirmation measures. In fact it works for **all** confirmation measures that have received some independent attention in this literature.

$$
\begin{aligned}
D(h,e) &= P(h|e) - P(h) \\
R(h,e) &= \log\left(\frac{P(h|e)}{P(h)}\right) \\
L(h,e) &= \log\left(\frac{P(e|h)}{P(e|\neg h)}\right) \\
C(h,e) &= P(e \wedge h) - P(h) \times P(e) \\
S(h,e) &= P(h|e) - P(h|\neg e) \\
Z(h,e) &= \begin{cases} \dfrac{P(h|e) - P(h)}{1 - P(h)} & \text{if } P(h|e) \geqslant P(h) \\ \dfrac{P(h|e) - P(h)}{P(h)} & \text{otherwise} \end{cases}
\end{aligned}
$$

- **The idea**: Substitute for the matching component of mental models approaches (including ETR) a confirmation-based hypothesis-testing mechanism.

Toward an erotetic-confirmation theory of reasoning
Look at the alternatives under consideration $h_0, \ldots, h_n$ and the answer provided by the second premise e. Keep only those alternatives h such that there is no h' with $c(h,e) < c(h',e)$, for c "the right" confirmation measure.

- ECTR as outlined above has three moving parts that deserve some attention.
  1. the specific measure of confirmation that can/should be plugged in for c,
  2. the *question* component: disjunctive premises induce representations with some structure, whose alternatives determine which hypotheses will be compared,
  3. the *answer* component: the second premise alone of illusory inferences from disjunction offers the evidence with respect to which competing hypotheses are compared.
- Point 1. seems not to matter: any of the confirmation measures in the literature will do the trick provided that the other two components are there.
- But points 2. and 3. are open questions. More concretely: do we *need* erotetics to account for this class of inferences, or is the confirmation-theoretic approach in itself sufficient?
- There are (at least) three interesting possibilities to consider.

(19) $P_1 = \varphi \vee \psi$
$P_2 = \theta$
proposed conclusions: c_1, c_2

Lean posteriors

$$P(c_1|P_1 \ \& \ P_2) > P(c_2|P_1 \ \& \ P_2)$$

Lean confirmation

$$c(c_1, P_1 \ \& \ P_2) > c(c_2, P_1 \ \& \ P_2)$$

Dynamic (lean?) confirmation

$$c(c_1, P_1 \ \& \ P_2) - c(c_1, P_1) > c(c_2, P_1 \ \& \ P_2) - c(c_2, P_1)$$

- We investigated this using simple illusory inferences from disjunction about standard decks of playing cards — easy way of controlling the actual probabilities and know exact predictions from the competing theories.

	Balanced priors		**Unbalanced priors**	
	Canonical	**Flat**	**Canonical**	**Flat**
Exp. design				
P_1	$(a \wedge h) \vee b$	$(a \wedge h) \vee (a \wedge b)$	$(a \wedge h) \vee b$	$(a \wedge h) \vee (a \wedge b)$
P_2	a		a	
Concl.	$h, c, =$	$h, c, =$	$h, b, =$	$h, b, =$
	Predictions			
Erotetic	h	$\varnothing$	h	$\varnothing$
Lean posteriors	$\varnothing$	$\varnothing$	b	b
Lean confirmation				
Measure: D	$\varnothing$	$\varnothing$	b	b
Measure: R	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
Dynamic confirmation	h	$\varnothing$	h	$\varnothing$

Table 1: Predictions of the different theories for the cards inference.

(20) The card is an ace and a heart, or else it's a black card.
The card is an ace.
Which is most likely?
  a. The card is a heart.
  b. The card is a club / The card is black.
  c. Heart and club/black are equally likely.

(21) The card is an ace and a heart, or else it's an ace and a black card.
Which is most likely?
  a. The card is a heart.
  b. The card is a club / The card is black.
  c. Heart and club/black are equally likely.

- We tested for a difference between (20) (predicted by ETR and dynamic confirmation to give rise to an illusory inference) and (21) (predicted not to give rise to an illusion).
- We found a significant difference between (20) and (21), showing that only the erotetic theory and the dynamic confirmation theory hold any promise.
- **NB**: The dynamic confirmation theory is properly *non-erotetic* with respect to the question component. It's unclear whether it's erotetic with respect to the answer component: dynamic procedure treats the second premise in a special way.

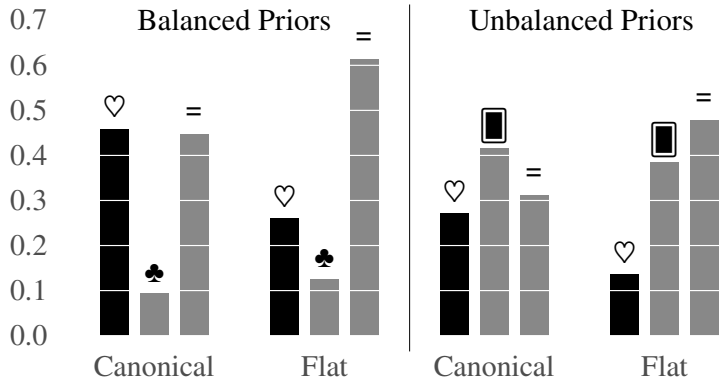

Figure 2: Results from the cards experiment. Legend: ♡ hearts are more likely; ♣ clubs are more likely; = options are equally likely; ▮ black card is more likely.

- The lean confirmation theory can give the same results as the erotetic theory, as long as we incorporate some elements of the erotetic confirmation account.
- We have the following
  - Measures *R* and *D* will make the same predictions as the erotetic confirmation theory (on any measure) if the question component is adopted: that is if the candidate conclusions are determined by the disjunctive premise rather than by the options given in the experiment.
  - Measure *R* does not care about the answer component: whether the first premise is considered part of the evidence or done away with, the predictions are the same.
- These results are in line with considerations in the confirmation literature on how *R* is better at factoring out priors than *D*. (Chandler, 2013).

4. Conclusions

- Mental models theory of reasoning and ETR were right that disjunctions are interpreted in a (mildly) structured way: this provides an account of the classical illusory inferences from disjunction as well as the ones with indefinites, where extant probabilistic approaches fail.
- But indirect illusory inferences with conditionals (and with properties, only tentative results for now) show that the matching procedures of these theories were only scratching at the surface of a more nuanced phenomenon.
- Incorporating probabilities in some fashion or other seems like an important ingredient of an account of these new data.
- One attractive and independently motivated way of cashing out an erotetic theory of reasoning with probabilities is to reframe the question-answer process in terms of Bayesian confirmation theory.
- But confirmation theory in itself doesn't seem enough: we still need properly erotetic elements in the account, and in particular the idea that disjunctions raise questions.
- Insofar as we're right about these erotetic processes, we have a novel perspective on what a competence theory of human reasoning should look like: **reasoning is partly about alternatives**, **reasoning involves representations with more structure that mere truth conditions**.
- This is about competence in the same sense that Oaksford and Chater's (2007) probabilistic approach is a theory of competence: O&C have a different theory of what the *functional aim* of human reasoning is. So does the erotetic theory.
- If this is right, then natural language semantics (in its guise as cognitive science), even at a computational level of analysis, should directly avail itself of these richer, non-lean representations.

References

Chandler, Jake (2013). Contrastive confirmation: some competing accounts. *Synthese*, 190:129–138.

Crupi, Vincenzo, Branden Fitelson and Katya Tentori (2008). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, 14(2):182–199.

Cummins, Denise D. (1995). Naive theories and causal deduction. *Memory and Cognition*, 23(5):646–658.

Cummins, Denise D., Todd Lubart, Olaf Alksnis and Robert Rist (1991). Conditional reasoning and causation. *Memory and Cognition*, 19(3):274–282.

Fine, Kit (2012). A difficulty for the possible world analysis of counterfactuals. *Synthese*.

van Fraassen, Bas (1969). Facts and tautological entailments. *The Journal of Philosophy*, 66(15):477–487.

Frank, Michael C and Noah D Goodman (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Groenendijk, Jeroen (2008a). Inquisitive semantics and dialogue management. Course given at ESSLLI 2008, Hamburg.

Groenendijk, Jeroen (2008b). Inquisitive Semantics: Two possibilities for disjunction. ILLC Prepublications PP-2008-26, ILLC.

Hamblin, Charles L. (1959). The modal "probably". *Mind*, 68(270):234–240.

Harman, Gilbert (1986). *Change in view: Principles of reasoning*. MIT press Cambridge, MA.

Johnson-Laird, Philip and Fabien Savary (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, 71(3):191–229.

Johnson-Laird, Philip N. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Koralus, Philipp and Salvador Mascarenhas (2013). The erotetic theory of reasoning: bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives*, 27:312–365.

Kratzer, Angelika (1991). Modality. In Arnim von Stechow and Dieter Wunderlich, editors, *Semantics: An International Handbook of Contemporary Research*. Berlin: Walter de Gruyter.

Lassiter, Daniel (2011). *Measurement and Modality: The Scalar Basis of Modal Semantics*. Ph.D. thesis, New York University.

Mascarenhas, Salvador (2013). An interpretation-based account of illusory inferences from disjunction. Ms. NYU.

Mascarenhas, Salvador (2014). *Formal Semantics and the Psychology of Reasoning: Building new bridges and investigating interactions*. Ph.D. thesis, New York University.

Mascarenhas, Salvador and Philipp Koralus (2017). Illusory inferences with quantifiers. *Thinking and Reasoning*, 23(1):33–48.

Oaksford, Michael and Nicholas Chater (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Tentori, Katya, Vincenzo Crupi and Selena Russo (2013). On the determinants of the conjunction fallacy: probability versus inductive confirmation. *Journal of Experimental Psychology: General*, 142(1):235–255.

Tversky, Amos and Daniel Kahneman (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, 90:293–315.

Walsh, Clare and Philip N. Johnson-Laird (2004). Coreference and reasoning. *Memory and Cognition*, 32:96–106.