The role of linguistic interpretation in human failures of reasoning

Salvador Mascarenhas
The University of Oxford
Ecole Normale Supérieure

ESSLLI 2016, week 2, lectures #2&#3

1 Recap

Some influential theories of reasoning [Slide 4]

- Heuristics and biases (Tversky and Kahneman, 1974)
- Bayesian reasoning (Oaksford and Chater, 1991)
- Mental logic (Rips, 1994): Our capacity for reasoning is underwritten by tacit natural deduction rules, but proofs are hard and we may be mistaken about what the right rules are.
- Mental models (Johnson-Laird, 1983; Koralus and Mascarenhas, 2013): Reasoning proceeds by manipulating representations of premises. A combination of the rules used and the nature of the representations is responsible for our successes and failures.

Reasoning and interpreting [Slide 5]

- Subjects in a reasoning experiment are performing two tasks: there is an *interpretive* step followed by a *reasoning* step.
- Psychologists tend to identify the processes of reasoning as the culprits of failures.
- But in principle subjects could be reasoning classically on non-obvious interpretations of the premises.

Semantically responsible psychology of reasoning
A theory of reasoning must rely on a comprehensive account of interpretive processes. Otherwise we risk misdiagnosing interesting but entirely *reasonable* interpretive quirks as fallacies.

2 Illusory inferences from disjunction

Illusory inference from disjunction [Slide 7]

(1) 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.

Does it follow that *Jane is looking at the TV*?

A fallacy [Slide 8]

(2) Illusory inference from disjunction, schematically:
P_1: $(a \wedge b) \vee (c \wedge d)$
P_2: a
Conclusion: b

- About 85% of subjects accept the conclusion (Walsh and Johnson-Laird, 2004)
- There is no significant effect of whether a, b, c, and d have distinct subjects

Falsified at a model where a, c, and d are true, but b is false.

Not a trivial issue of exclusive 'or'
$(a \wedge b \wedge \neg(c \wedge d)) \vee (c \wedge d \wedge \neg(a \wedge b))$

Mental models account [Slide 9]

Mental model theory account of the illusory inference from disjunction (combining elements from Johnson-Laird (1983) and Koralus and Mascarenhas (2013))

- Reasoners build mental representations (mental models) that verify each of the premises.
- Disjunctive premises are represented as sets of alternative mental models.
- P_1 gives rise to a set of two alternative models: a minimal model of $a \wedge b$ and a minimal model of $c \wedge d$.
- **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.
- The combined representation of the premises is therefore only one mental model: $a \wedge b$. From here, b follows.

3 A reasoning-based account: the erotetic theory of reasoning

The erotetic theory of reasoning [Slide 11]

The erotetic principle

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

Commitment on interpretation
Disjunctions raise alternatives and put pressure toward *choosing* an alternative — *disjunctions are like questions* in this regard (Inquisitive Semantics: Groenendijk, 2008, Mascarenhas, 2009)

Illusory inference on the erotetic theory [Slide 12]

(3) 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.

Evidence for the erotetic theory [Slide 13]

- Order effects if the premises are reversed: fewer people commit the fallacy if they see the categorical premise before the disjunctive premise (Mascarenhas and Koralus, 2016)

(4) P_2: John is watching TV.
P_1: John is watching TV and Mary is playing tennis, or Bill is doing homework.

Predicted if subjects are engaged in a question-answer task: the question must come first.

($p < .05$ for propositional case, $p < .01$ for indefinites case, insignificant for valid and invalid controls alike; controls had sentences of comparable length)

ETR's operations (simplified version) — 1 [Slide 14]

C(onjunctive)-Update

$$\begin{aligned}\Gamma[\Delta]^{\mathsf{C}} &= \Gamma \times \Delta \\ &= \{\gamma \sqcup \delta : \gamma \in \Gamma \;\&\; \delta \in \Delta\}\end{aligned}$$

C-Update pairwise combines each element of Γ with each element of Δ. It incorporates the new information in Δ into Γ.

ETR's operations (simplified version) — 2 [Slide 15]

Q(uestion)-Update

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

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.

ETR's operations (simplified version) — 3 [Slide 16]

Update

$$\Gamma[\Delta]^{\mathsf{Up}} = \begin{cases} \Gamma[\Delta]^{\mathsf{C}} & \text{if } \Gamma[\Delta]^{Q} = \emptyset \\ \Gamma[\Delta]^{\mathsf{Q}}[\Delta]^{\mathsf{C}} & \text{otherwise} \end{cases}$$

The complete Update procedure first *tests* whether Δ provides an answer to the question in Γ by attempting a Q-Update. If it *doesn't* (i.e. Q-update returns $\emptyset$), then Update performs a simple C-Update, incorporating the new information in Δ. If it *does*, then Update keeps the (possibly only partly) answered question and C-Updates with Δ, in case Δ provides some new information *beside* providing an answer to Γ.

ETR's operations (simplified version) — 4 [Slide 17]

Molecular Reduction

$$\Gamma[\alpha]^{\mathsf{MR}} = \begin{cases} (\Gamma - \{\gamma \in \Gamma : \alpha \sqsubseteq \gamma\}) \cup \{\alpha\} & \text{if } (\exists \gamma \in \Gamma)\alpha \sqsubseteq \gamma \\ \textit{undefined} & \text{otherwise} \end{cases}$$

Molecular Reduction of Γ on a mental molecule α reduces every alternative in Γ that contains α to α alone. It is undefined in case no alternative in Γ contains α. It amounts to *disjunct simplification* ($(\varphi \wedge \psi) \vee \theta \vdash \varphi \vee \theta$), and as a special case it allows for conjunction elimination.

ETR's operations (simplified version) — 5 [Slide 18]

Filter

$$\Gamma[\cdot]^{\mathsf{F}} = \{\textsc{dne}(\gamma) : \gamma \in \Gamma \;\&\; \neg\textsc{contr}(\gamma)\}$$

Filter eliminates all contradictory alternatives in Γ by testing for the presence, within an alternative, of a molecule α and its negation (this is the function $\textsc{contr}(\cdot)$). Further, it eliminates double negations from the surviving alternatives ($\textsc{dne}(\cdot)$).

ETR's operations (simplified version) — 6 [Slide 19]

Inquire

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

Inquire performs a simple conjunctive update (NB: no Q-Update) with a mental model Δ and its negation, followed by filtering out any contradictory alternatives and removing double negations.

ETR's accessory functions [Slide 20]

Mental Model Negation

For Γ a mental model, notice that $\Gamma = \{\alpha_0, \ldots, \alpha_n\}$ and for each $\alpha_i \in \Gamma$ we have that $\alpha_i = \bigsqcup\{a_{i0}, \ldots, a_{im_i}\}$, for $m_i + 1$ the number of mental model nuclei in α_i. Now,

$$\textsc{Neg}(\Gamma) \quad = \quad \textsc{Neg}(\{\alpha_0, \ldots, \alpha_n\}) \quad = \quad \{\neg a_{00}, \ldots, \neg a_{0m_0}\} \quad \times \quad \cdots \quad \times \quad \{\neg a_{n0}, \ldots, \neg a_{nm_n}\}$$

Double negation elimination

$$\textsc{dne}(a) = \begin{cases} b & \text{if } a = \neg\neg b \text{ for some } b \in \mathit{Atoms}(\mathscr{M}) \\ a & \text{otherwise} \end{cases}$$

$$\textsc{dne}(\alpha) = \bigsqcup \{\textsc{dne}(a) : a \in \mathit{Atoms}(\mathscr{M}) \;\&\; a \sqsubseteq \alpha\}$$

4 An interpretation-based account: scalar implicature

Interpretation-based accounts [Slide 22]

- On an **interpretation-based account,**
  1. there is nothing in principle non-classical about the human capacity for reasoning,
  2. but the **interpretive processes** are more complex that meets the eye. In other words: the premises do not mean what one might think they mean.
- Accounts in this spirit have been given to some classical fallacies within formal pragmatics (e.g. Horn, 2000, discusses affirming the consequent and denying the antecedent).

(5) P_1: If the card is long then the number is even.
P_1': *Only* if the card is long is the number even.
P_2: The number is even.
Conclusion: The card is long.

Preview: the illusory inference from disjunction in terms of scalar implicature [Slide 23]

- The illusory inference from disjunction follows **classically** if we assume that a classically-tuned reasoning module acts on suitably pragmatically strengthened premises.

(6) Illusory inference from disjunction, schematically:
P_1: $(a \wedge b) \vee c$
P_2: a
Conclusion: b

(7) Strengthened meaning of (6):
${P_1}^+$: $(a \wedge b \wedge \neg c) \vee (c \wedge \neg a \wedge \neg b)$
${P_2}^+$: a
Conclusion: b

Intuitively [Slide 24]

(8) 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.

Premise 1 of the illusory inference is interpreted as

(9) Either John is watching TV and Mary is playing tennis *and nothing else that is relevant is true* or Bill is doing homework *and nothing else that is relevant is true*.

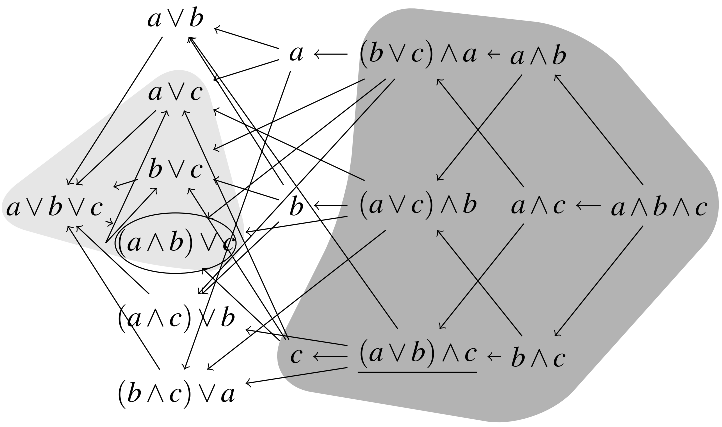
Calculating scalar implicatures [Slide 25]

1. Compute the alternatives to S that are at most as complex as S (Katzir, 2007).
2. Collect those alternatives S' that are (1) alternatives to S and (2) strictly stronger than S. Call this set A.
3. Compute primary implicatures: for each sentence $S' \in A$, "the speaker does not believe that S'."
4. Compute secondary implicatures: for each $S' \in A$ such that the negation of S' does not contradict the literal meaning of S or any of the primary implicatures of S, conclude (that the speaker believes) that S' is false.

5. Call the conjunction of the literal meaning of S together with all of its secondary implicatures the strengthened (exhaustive) meaning of S.

The alternative propositions [Slide 26]

These are all the alternatives for a sentence of the form $(a \wedge b) \vee c$:



Getting the implicature [Slide 27]

Alternatives that will give rise to secondary implicatures:
$\{\neg((a \vee b) \wedge c), \neg(a \wedge c), \neg(b \wedge c), \neg(a \wedge b \wedge c)\}$

Equivalently:
$(\neg a \wedge \neg b) \vee \neg c$

Conjoined with the literal meaning:
$((a \wedge b) \wedge \underline{((\neg a \wedge \neg b) \vee \neg c)}) \quad \vee \quad (c \wedge \underline{((\neg a \wedge \neg b) \vee \neg c)})$

Equivalently:
$(a \wedge b \wedge \neg c) \vee (c \wedge \neg a \wedge \neg b)$

Illusory inference explained [Slide 28]

(10) 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.

Premise 1 of the illusory inference is interpreted as

(11) Either John is watching TV and Mary is playing tennis *and nothing else that is relevant is true* or Bill is doing homework *and nothing else that is relevant is true*.

(12) Among these three possibilities, either it is *only* the case that John is watching TV and Mary is playing tennis, or it is *only* the case that Bill is doing homework.

From here the fallacious conclusion follows *classically*.

	2	3	4	n
1. Propositions	16	256	65,536	$2^{(2^n)}$
2. Positive propositions	4	18	166	Dedekind numbers: $M(n)-2$
3. Katzir (2007)	20	552	20,679	$\sum_{k<n}(2n-1)^{k+1}2^k-k$

Table 1: Number of alternatives by procedure, for a source with 2, 3, 4, and n atoms.

5 Excursus: too many alternatives...

Too many alternatives... [Slide 30]

- Every theory of scalar implicature needs to specify what the relevant alternatives are.
- But most proposals for alternative-set generation in the literature involve rapidly growing sets as a function of the number of atoms in the input.

6 Expanding the paradigm: enter quantifiers

Illusory inferences with quantifiers [Slide 32]

- When psychologists think about reasoning with quantifiers, they think about syllogisms.
- But syllogisms are only a small fragment of first order logic.
- *Universal* quantification relates to *conjunction* and *existential* quantification to *disjunction*.

(13) a. Every student snores.
b. Student *a* snores *and* student *b* snores *and* ...

(14) a. Some student snores.
b. Student *a* snores *or* student *b* snores *or* ...

Can we recast the illusory inference with quantifiers instead of propositional connectives?

Universals [Slide 33]

90% acceptance, significantly more than invalid controls at less than 10%

(15) a. Every boy or every girl is coming to the party.
John is coming to the party.
Does it follow that Bill is coming to the party?
b. Mary has met every king or every queen of Europe.
Mary has met the king of Spain.
Does it follow that Mary has met the k. of Belgium?

Mascarenhas (2014), Mascarenhas & Koralus (2016)

Indefinites [Slide 34]

- Indefinites are also like questions (Kratzer & Shimoyama, 2002, Mascarenhas, 2011)

(16) a. Some pilot writes poems.
b. Which pilot writes poems?

40% acceptance, significantly more than invalid controls ($p < .01$)

(17) a. Some pilot writes poems.
John is a pilot.
Does it follow that John writes poems?

(18) a. Some firmicute produces endospores.
Clostridium is a firmicute.
Does it follow that clostridium produces endospores?

b. Some thermotogum stains gram-negative.
Maritima is a thermotogum.
Does it follow that maritima stains gram-negative?

Interpretation *or* reasoning? [Slide 35]

- Can we decide between the erotetic (reasoning based) and pragmatic (interpretation based) accounts?

First attempt

Implicatures are much less likely to arise in downward entailing contexts. We could try to embed the crucial premise of the illusory inference in such a context. If the pragmatic theory is right, people's performance should *improve*.

If every boy and every girl is coming to the party, and moreover John is coming to the party, then Bill will come as well.

7 Synthesis: two sources of illusory inferences

Interpretation *and* reasoning [Slide 37]

- Propositional connectives and universals pattern alike: high acceptance rate (90%)
- Universals get an implicature rather like the propositional case:

(19) a. Every boy or every girl is coming to the party.

b. Implicatures:
Every boy *and no girl* or every girl *and no boy* is coming to the party.

- Indefinites induce fallacious reasoning, but the effect is significantly weaker (40%, between subjects $p < .01$)
- Indefinites *lack* the corresponding scalar implicature

(20) a. Some pilot writes poems.

b. *Not* an implicature: There is exactly one pilot and she writes poems.