Toward probabilistic mental logic

Jakub Szymanik
jakub.szymanik@gmail.com

Universiteit van Amsterdam

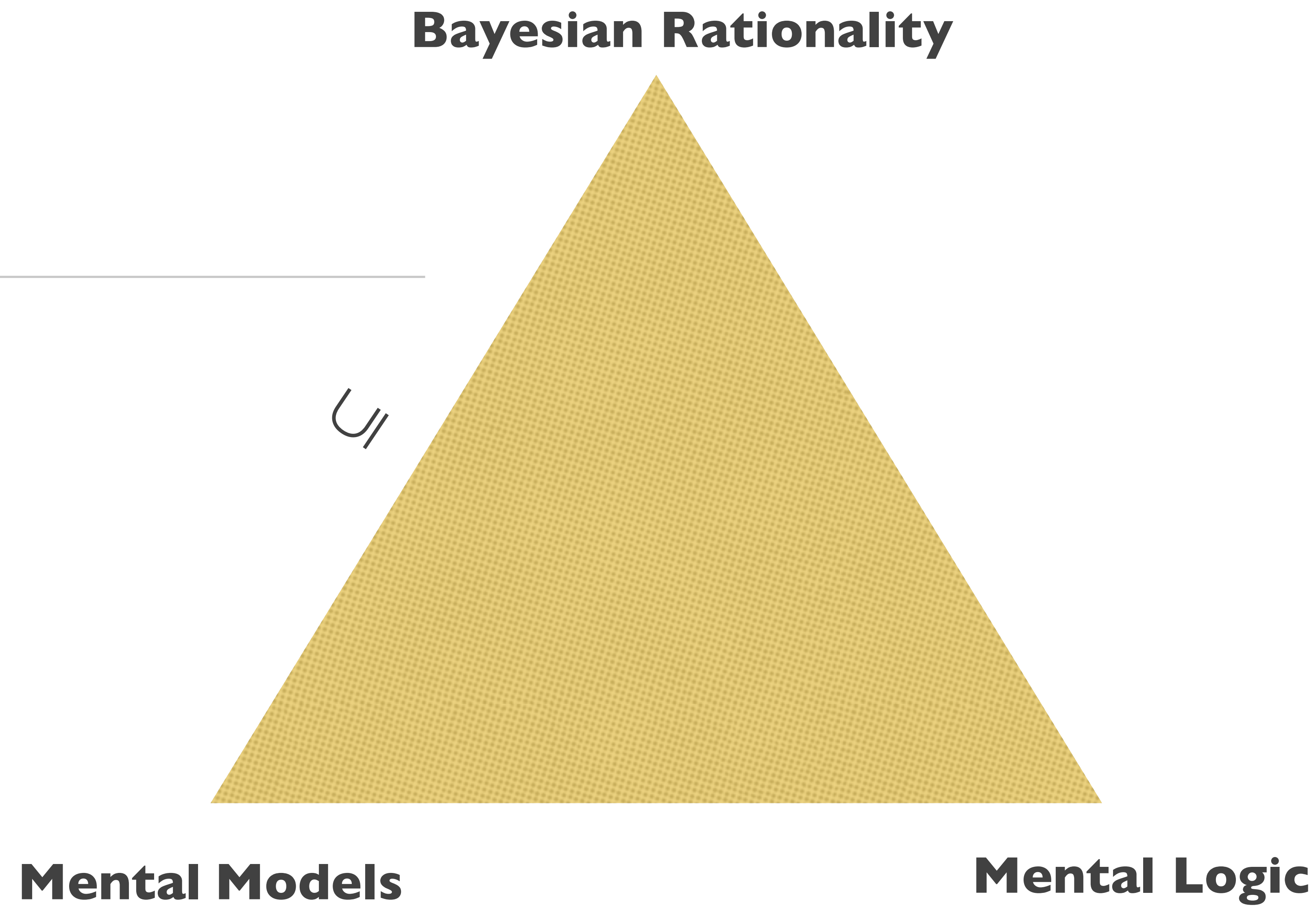
Plan

- Revive the project of mental logic
- Probabilistic natural logic for syllogistic reasoning
- Weights based in empirical data
- Reflecting `complexity / preferability' of single reasoning rules
- Proof-of-concept providing guidelines for further work

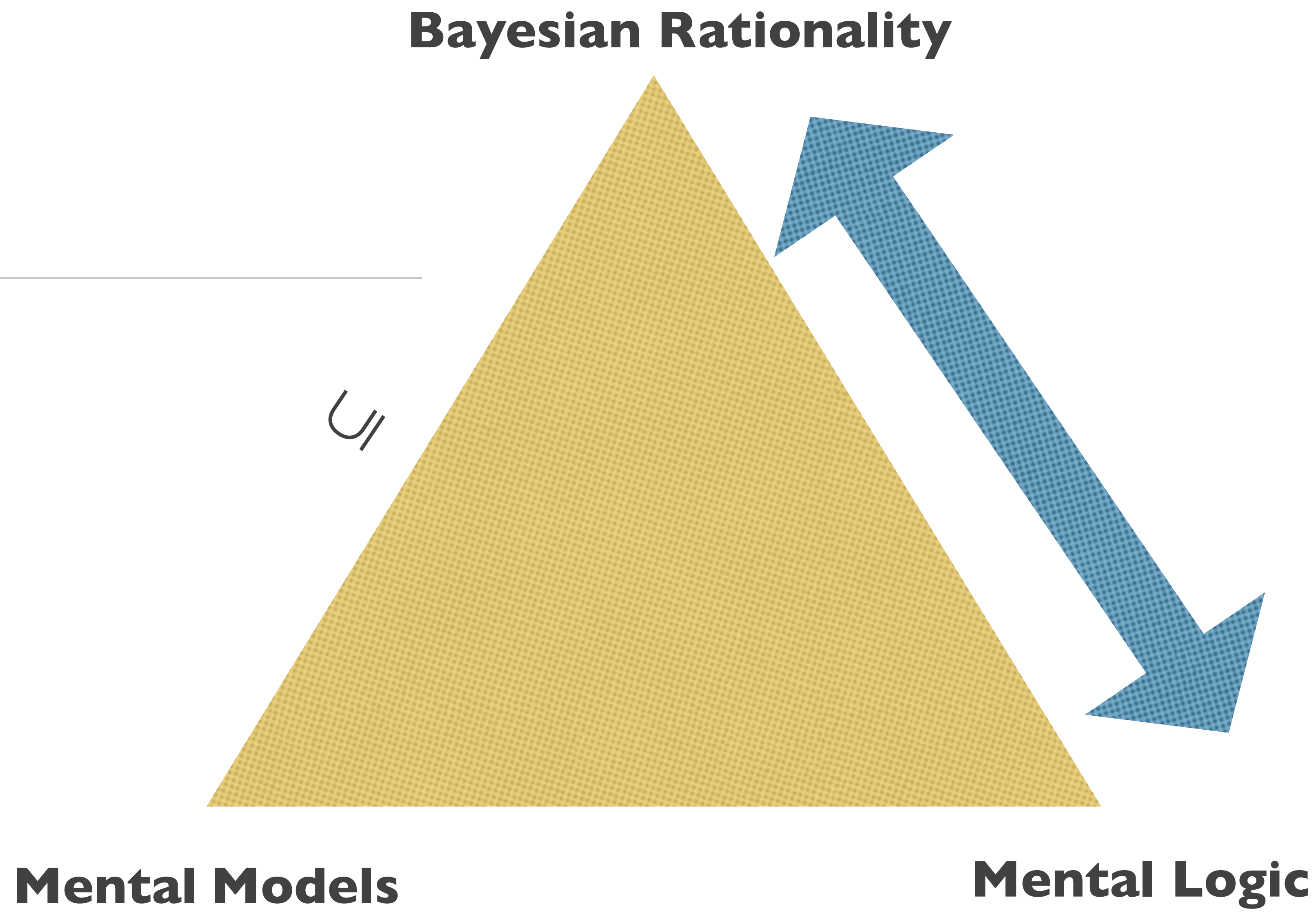
Logic as the theory of reasoning & its challenges

- Logical Omniscience
- Conjunction Fallacy
- Wason Selection Task
- Suppression Task
- etc.

Reaction:



Reaction:



Mental Logic

- Rips (1994):
- Formulas as the underlying mental representations
- Inference rules are the basic operations
- PSYCOP based on Natural Deduction
- You can think about proofs as computations.

ML's shortcomings

- Abstract rules and formal representations
- Based in natural deduction for FOL
- Ad hoc `psychological completness'
- Explains only validities, no story on mistakes
- No learning or individual differences

Natural Logic Program

- van Benthem 1986, Sánchez-Valencia 1991:
- Computationally minimal systems
- Following `the surface structure of NL'
- Traditionally monotonicity and semantic containment
- Recently intensively studied, extended, and applied, e.g., by Stanford NLP group
- So, why not build MLs based on these ideas?

Natural Logic Program

- van Benthem 1986, Sánchez-Valencia 1991:
- Computationally minimal systems
- Following `the surface structure of NL'
- Traditionally monotonicity and semantic containment
- Recently intensively studied, extended, and applied, e.g., by Stanford NLP group
- So, why not build MLs based on these ideas?

IF No aardvark without a keen sense of smell can find food.
THEN No aardvark without a sense of smell can find food.

Benchmark Task: arena of syllogistic reasoning

- All A are B : universal affirmative (A)
- Some A are B: particular affirmative (I)
- No A are B: universal negative (E)
- Some A are not B: particular negative (O)

Figure 1	*Figure 2*	*Figure 3*	*Figure 4*
B C	C B	B C	C B
A B	A B	B A	B A
A C	A C	A C	A C

AE4O:

All C are B
No B are A

Some A are not C

Benchmark Task: arena of syllogistic reasoning

- All A are B : universal affirmative (A)
- S
- N
- S

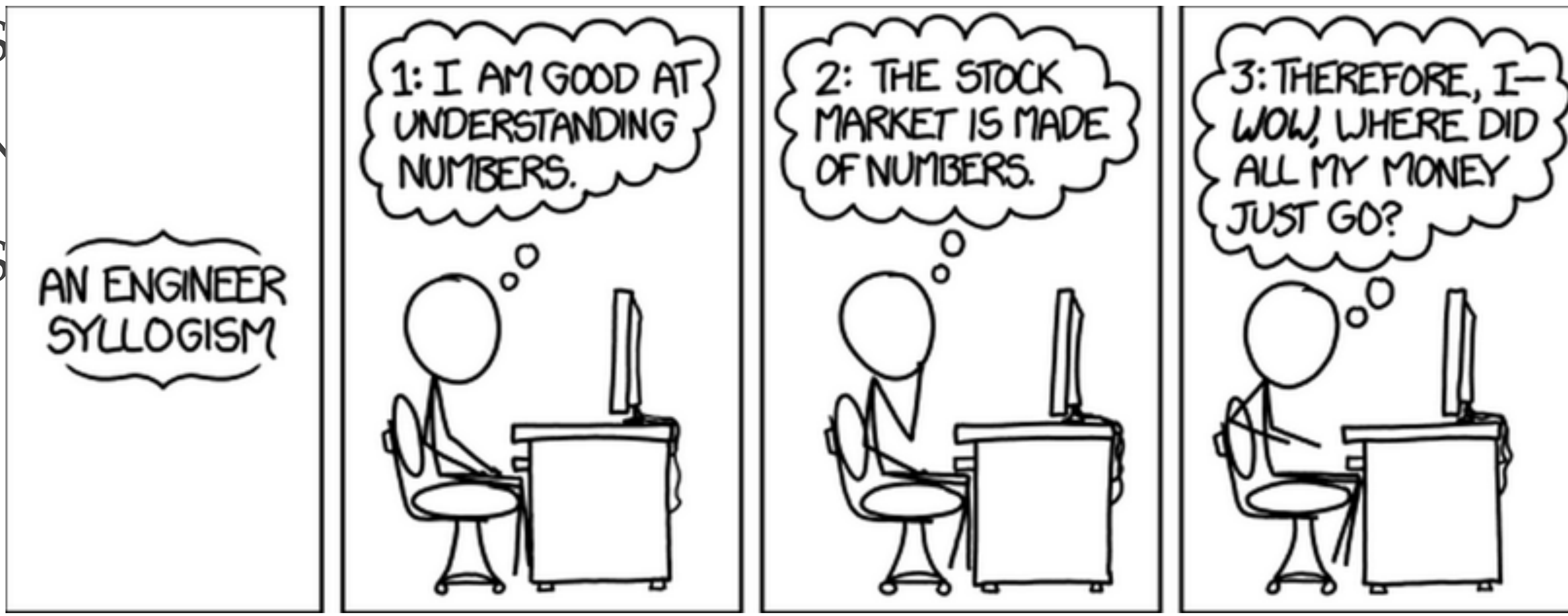


Figure 4
C B
B A

A C

AE4O: All C are B
No B are A

Some A are not C

Syllogistic reasoning

Syllogism	Conclusion					Syllogism	Conclusion				
	A	I	E	O	NVC		A	I	E	O	NVC
AA1	**90**	**5**	0	0	5	AO1	1	6	1	57	35
AA2	58	8	1	1	32	AO2	0	6	3	**67**	24
AA3	57	**29**	0	0	14	AO3	0	10	0	66	24
AA4	75	**16**	1	1	7	AO4	0	5	3	72	20
AI1	0	92	3	3	2	OA1	0	3	3	68	26
AI2	0	57	3	11	29	OA2	0	11	5	56	28
AI3	1	**89**	1	3	7	OA3	0	15	3	**69**	13
AI4	0	71	0	1	28	OA4	1	3	6	27	63
IA1	0	72	0	6	22	II1	0	41	3	4	52
IA2	13	49	3	12	23	II2	1	42	3	3	51
IA3	2	**85**	1	4	8	II3	0	24	3	1	72
IA4	0	**91**	1	1	7	II4	0	42	0	1	57
AE1	0	3	59	6	32	IE1	1	1	22	16	60
AE2	0	0	88	**1**	11	IE2	0	0	39	30	31
AE3	0	1	61	13	25	IE3	0	1	30	33	36
AE4	0	3	**87**	**2**	8	IE4	0	42	0	1	57
EA1	0	1	**87**	**3**	9	EI1	0	5	15	**66**	14
EA2	0	0	**89**	**3**	8	EI2	1	1	21	**52**	25
EA3	0	0	64	**22**	14	EI3	0	6	15	**48**	31
EA4	1	3	61	**8**	28	EI4	0	2	32	**27**	39
OE1	1	0	14	5	80	OO1	1	8	1	12	78
OE2	0	8	11	16	65	OO2	0	16	5	10	69
OE3	0	5	12	18	65	OO3	1	6	0	15	78
OE4	0	19	9	14	58	OO4	1	4	1	25	69
IO1	3	4	1	30	62	OI1	4	6	0	35	55
IO2	1	5	4	37	53	OI2	0	8	3	35	54
IO3	0	9	1	29	61	OI3	1	9	1	31	58
IO4	0	5	1	44	50	OI4	3	8	2	29	58
EE1	0	1	34	1	64	EO1	1	8	8	23	60
EE2	3	3	14	3	77	EO2	0	13	7	11	69
EE3	0	0	18	3	78	EO3	0	0	9	28	63
EE4	0	3	31	1	65	EO4	0	5	8	12	75

Table 2.1: Percentage of times each syllogistic conclusions was endorsed. The data is from a meta-analysis by Chater and Oaksford (1999). "NVC" stands for "No Valid Conclusion", all numbers have been rounded to the closest integer. A bold number indicates that the corresponding conclusion is valid.

Chater and Oaksford, 1999

Geurts (2003)'s model

- Logic including syllogistics and pivoting on monotonicity with rules:
- *All-Some:* `All A are B′ implies `Some A are B′.
- *No-Some not:* `No A are B′ implies `Some A are not B′.
- *Conversion1:* `Some A are B′ implies `Some B are A′;
- *Conversion2:* `No A are B′ implies `No B are A".
- *Monotonicity:* If A entails B, then the A in any upward entailing position can be substituted by a B, and the B in any downward entailing position can be substituted by an A.
- *Extra rule:* `No A are B′ and `Some C are A′ implies `Some C are not B′.

Example for EA2E

$$
\begin{array}{lll}
\text{No C are B} & (1) & \\
\text{All A are B} & (2) & \\
\hline
\text{No B are C} & (3) & \text{Conversion}(1) \\
\text{No A are C} & (4) & \text{Monotonicity}(2,3)
\end{array}
$$

Geurts' (2003) model c'td

- The shorter the proof the easier the syllogism.
- Initial budget of 100 units. Each use of the monotonicity rule costs 20, the extra rule costs 30; a proof containing a "Some Not" proposition costs an additional 10 units. Take the remaining budget as an evaluation of the difficulty.
- It gives a good fit with data.

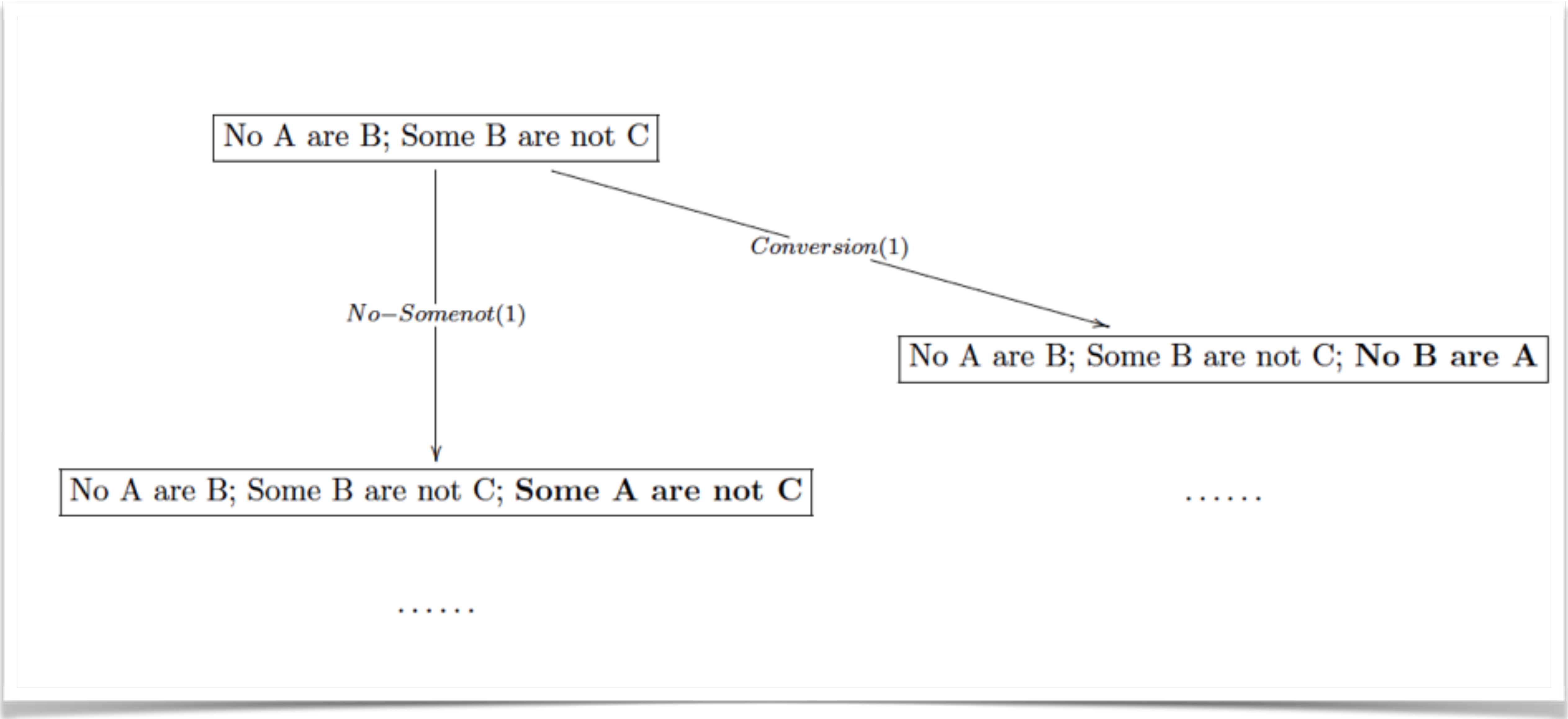
Table 4
Predicted difficulty of valid syllogisms according to the model described in the text, compared with Chater and Oaksford's scores (in parentheses)

AA1A	80	(90)	OA3O	70	(69)	EA1O	40	(3)
EA1E	80	(87)	AO2O	70	(67)	EA2O	40	(3)
EA2E	80	(89)	EI1O	60	(66)	EA3O	40	(22)
AE2E	80	(88)	EI2O	60	(52)	EA4O	40	(8)
AE4E	80	(87)	EI3O	60	(48)	AE2O	40	(1)
IA3I	80	(85)	EI4O	60	(27)	AE4O	40	(2)
IA4I	80	(91)	AA1I	60	(5)			
AI1I	80	(92)	AA3I	60	(29)			
AI3I	80	(89)	AA4I	60	(16)			

- Similar strategy works for other cognitive tasks, see Gierasimczuk et al. 2014.

Learning the inference rules from the data

Joint work with Fangzhou Zhai and Ivan Titov



Vanilla version

- Geurts' logic
- Tree representation: states linked by reasoning events
- No vapid transitions

Probabilities

- Tendency value: an easier rule is adopted with higher probability, while a more difficult one is adopted with lower probability.
- Let T_r any rule and c_r the number of ways that it can be adopted at S:

$$p_0(S_r|S,\theta_0) = \frac{T_r}{\sum_{r\in R} c_r \cdot T_r}$$

The output of the model

- A probability with which a syllogism is endorsed.
- 5 possible conclusions: A, I, E, O, NVC.
- Each leaf uniquely determines a path from the root.
- We can compute the probability that a given conclusion is drawn.

$$p'_0(y|R,\theta_0) = \sum_{\text{S is a leaf consistent with y}} p_0(S|R,\theta_0)$$

The output of the model

- A probability with which a syllogism is endorsed.
- 5 possible conclusions: A, I, E, O, NVC.
- Each leaf uniquely determines a path from the root.
- We can compute the probability that a given conclusion is drawn.

$$p_0'(y|R,\theta_0) = \sum_{\text{S is a leaf consistent with y}} \prod_{0 \le i < n} p_0(S_{i+1}|S_i,\theta_0)$$

Training

- Subset of the data from Chater and Oaksford (1999)
- We use the Expectation-Maximization algorithm
- Compute:

$$\arg\max_{\theta_0} \; p_0(\{(X_i, y_i)\}_{i \leq n} | \theta_0)$$

Evaluation

- The Khemlani and Johnson-Laird (2012) method
- Detection theory

Predictions \ Exp. Data	$< 30\%$	$\geq 30\%$
$< 30\%$	Correct Rejection	Miss
$\geq 30\%$	False Alarm	Hit

Performance of Vanilla Version

- 95,8% correct predictions on syllogisms with at least one conclusion.
- 81,6% correct predictions on all syllogisms.
- But no mechanism to explain the errors.
- The models always returns NVC for invalid syllogisms.

Adding illicit conversions

- *Conversion:* For every Q,
  `Q A are B' implies `Q B are A'.
- Half the number of misses.
- 91,9% correct predictions on all syllogisms.
- For II, IO, EE, OI, OE, OO always returns NVC.

Let's guess

- Probability of guessing NVC is negatively related to the informativeness of the premises.
- Atmosphere hypothesis when there is a negation in the premises, individuals are likely to draw a negative conclusion; when there is `some' in the premises it will be likely in the conclusion; when neither is the case, the conclusion is often affirmative.

Performance

- 95% correct predictions on all syllogisms
- The training gives the informativeness order as assumed by Chater & Oaksford

A(1.11) > E(0.33) > I(0.199) > O(-0.78)

- And data yields the complexity order:

Conversion<Monotonicity<All-Some<No-SomeNot

Theory	Hit	Miss	False Alarm	Correct Rejection	Correct Predictions
Atmosphere	44	41	20	215	259 /80.9%
Matching	41	44	55	180	221 /69.1%
Conversion	52	33	12	223	275 /85.9%
PHM*	40	45	63	172	212 /66.3%
PSYCOP	45	40	26	209	254 /79.4%
Verbal Models*	54	31	29	206	260 /81.2%
Mental Models*	85	0	55	180	265 /82.8%
Generative Model Ver. 1	51	33	26	210	261/81.6%
Generative Model Ver. 2	67	17	9	227	294/91.9%
Generative Model Ver. 3	74	10	6	230	304/95.0%
Experimental Data	85	0	0	235	320/100%

Comparing with other theories

Khemlani and Johnson-Laird (2012)

Theory	Hit	Miss	False Alarm	Correct Rejection	Correct Predictions
Atmosphere	44	41	20	215	259 /80.9%
Matching	41	44	55	180	221 /69.1%
Conversion	52	33	12	223	275 /85.9%
PHM*	40	45	63	172	212 /66.3%
PSYCOP	45	40	26	209	254 /79.4%
Verbal Models*	54	31	29	206	260 /81.2%
Mental Models*	85	0	55	180	265 /82.8%
Generative Model Ver. 1	51	33	26	210	261/81.6%
Generative Model Ver. 2	67	17	9	227	294/91.9%
Generative Model Ver. 3	74	10	6	230	304/95.0%
Experimental Data	85	0	0	235	320/100%

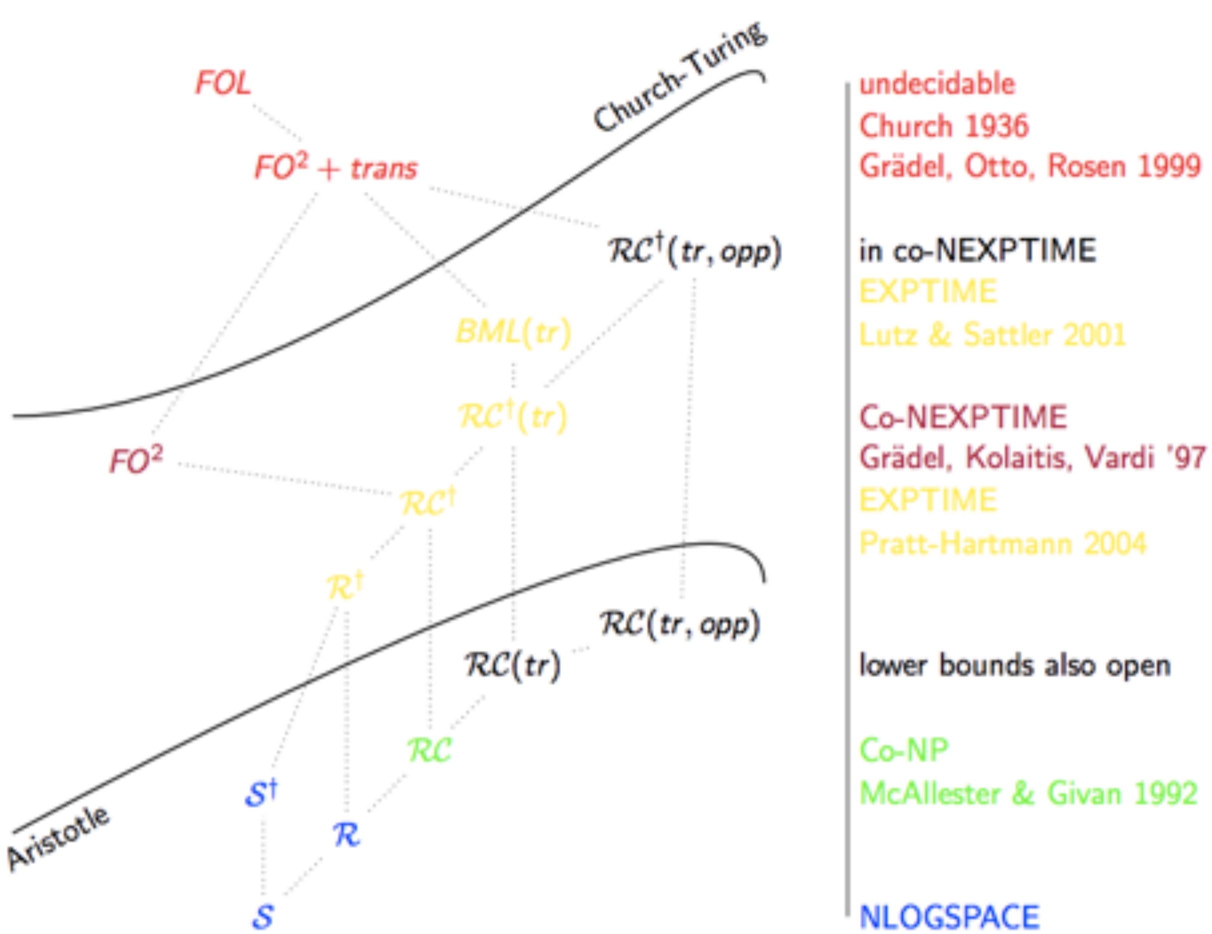
Comparing with other theories

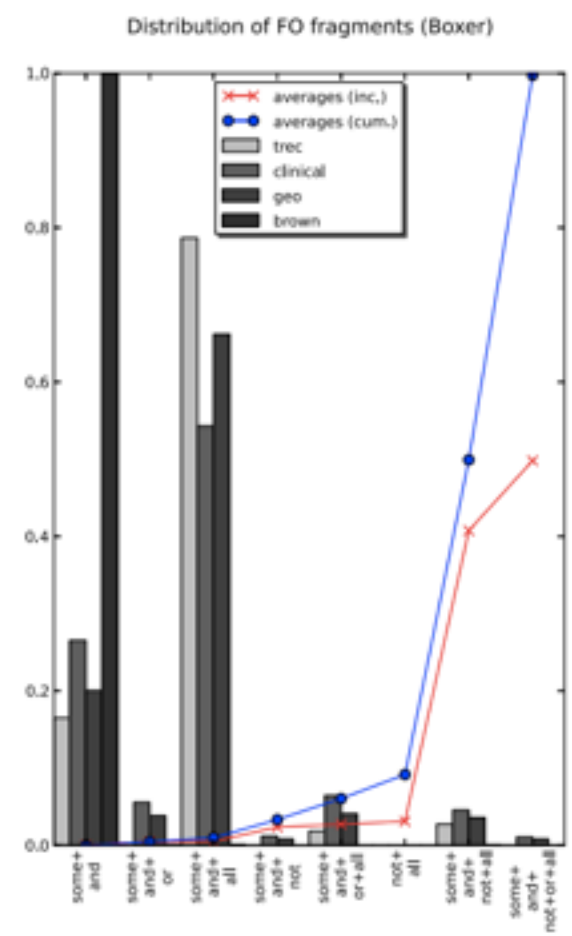
Khemlani and Johnson-Laird (2012)

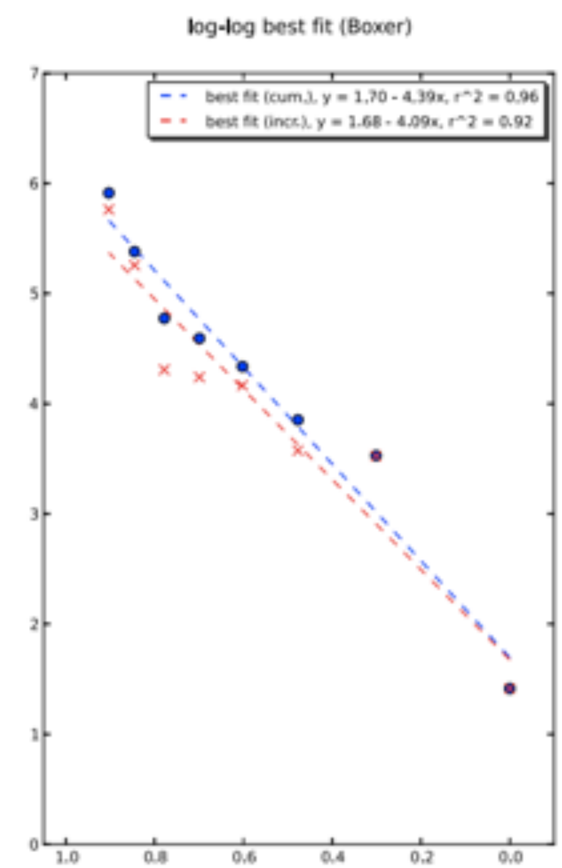
Summary

- Abstract ND rules of ML can be replaced by NL.
- Ad hoc `psychological completeness' can be derived from data, some rules are unlikely to fire.
- It can give a more systematic take on reasoning errors.
- A way to classify inferences steps wrt cognitive difficulty.
- Yields computationally friendlier systems.
- Modular approach.

How much logic do we need?







(Pratt-Hartmann 2010; Thorne, 2010; Larry Moss, 2010)

(Thorne, 2010)

Further work

- Extend to wider fragments of language.
- But also other types of reasoning
  (see, e.g. Gierasimczuk et. al. 2013, Braüner 2013).
- Run experiments / train model on better data.
- Understand learning and individual differences

  (joint work with N. Gierasimczuk & A.L. Vargas Sandoval).
- Think about processing model and its complexity.
- …

Thank you!

Amsterdam Colloquium 2015

Workshop `Reasoning in Natural Language: Symbolic and Sub-symbolic Approaches'