Theory Comparisons for Generalized Quantifiers

Marco Ragni, Henrik Singmann, Eva-Maria Steinlein

University of Freiburg

July 21st, 2014

A Syllogism

Some frenchmen are wine drinkers

None of the wine drinkers are beer drinkers

Therefore, . . . ?

A Syllogism

Some frenchmen are wine drinkers

None of the wine drinkers are beer drinkers

Therefore, . . . ?

- Premises in classical syllogisms use one of four quantifiers: all (**A**), some (**I**), some not (**O**), or none (**E**)

A Syllogism

Some frenchmen are wine drinkers

None of the wine drinkers are beer drinkers

Therefore, . . . ?

- Premises in classical syllogisms use one of four quantifiers: all (**A**), some (**I**), some not (**O**), or none (**E**)
- *Therefore, some of the frenchmen are not beer drinkers.*

A Syllogism

Some frenchmen are wine drinkers

None of the wine drinkers are beer drinkers

Therefore, ...?

- Premises in classical syllogisms use one of four quantifiers: all (**A**), some (**I**), some not (**O**), or none (**E**)
- *Therefore, some of the frenchmen are not beer drinkers.*
- Everyday human reasoning is "based [...] on beliefs, in which there are varying degrees of confidence" (Evans, 2002, p.980)

A Syllogism

Some frenchmen are wine drinkers

None of the wine drinkers are beer drinkers

Therefore, . . . ?

- Premises in classical syllogisms use one of four quantifiers: all (**A**), some (**I**), some not (**O**), or none (**E**)
- *Therefore, some of the frenchmen are not beer drinkers.*
- Everyday human reasoning is "based [. . .] on beliefs, in which there are varying degrees of confidence" (Evans, 2002, p.980)
- We consider generalized quantifiers most (**M**) and few (**F**)

A Syllogism

Some frenchmen are wine drinkers

Few wine drinkers are beer drinkers

Therefore, . . . ?

A Syllogism

Some frenchmen are wine drinkers

Few wine drinkers are beer drinkers

Therefore, few frenchmen are beer drinkers.

Therefore, some frenchmen are beer drinkers.

Theory Predictions and Extensions

- Probability Heuristics Model (PHM) (Chater & Oaksford, 1999) use 3 heuristics and predict that conclusions can be ordered

$$A > M > F > I > E >> O$$

Theory Predictions and Extensions

- Probability Heuristics Model (PHM) (Chater & Oaksford, 1999) use 3 heuristics and predict that conclusions can be ordered

$$A > M > F > I > E >> O$$

- Extension 1: Matching Hypothesis Ordered from most to least conservative quantifier:

$$E > O = I >> A$$

Theory Predictions and Extensions

- Probability Heuristics Model (PHM) (Chater & Oaksford, 1999) use 3 heuristics and predict that conclusions can be ordered

$$A > M > F > I > E >> O$$

- Extension 1: Matching Hypothesis Ordered from most to least conservative quantifier:

$$E > O = I >> A$$

$$E > O = I > M = F >> A$$

Theory Predictions and Extensions

- Extension 2: Mental Models + Heuristics Construction of Mental Model and $E > I \geq F > O > M > A$

X	Y	Z
X		
	Y	
	Y	

- Extension 3: Preferred Mental Models Formalization of Mental Models as (minimal) spatial models $\varphi_1 : \Omega_1 \to \mathbb{N}^2$ satisfying premise P_1 and P_2

Example Item

Some brokers are waiters.
Few waiters are agents.

What follows?

$\boxed{}$ of the brokers are agents.

Quantifiers: All, Some, Some Not, Most, Few, None.
[*"Nothing follows" was not a provided option.*]

Experiment

- Online study (Amazon MT) with 25 participants.
- 40 items per participant
    - All items of Figure 1 (P1: X - Y, P2: Y - Z)
    - Conclusion: 20 trials X - Z, 20 trials Z - X
    - for each set of 20 items:
        - 6 syllogisms with *most* in P1
        - 6 syllogisms with *few* in P1
        - 4 syllogisms with *most* in P2
        - 4 syllogisms with *few* in P2
    - Different professions and hobbies constituted the content of the terms.

Predictions and Results

Observed responses and predictions of the four theories for selected syllogisms (X-Z conclusion).

Syll.	Data	PHM	Matching	PMM	Min. Models
MM	M(84%)	M, (I, O)	M	M	M
FF	F(84%)	F, (I, O)	F	F	F
IF	F(56%), I(32%)	I, (O)	I	F	F, I
FI	F(64%)	I, (O)	I	F	F, I
FO	F(44%), I(32%)	O, (I)	O	I	F
OF	F(48%), I(24%)	O, (I)	O	I	F
MO	I(56%)	O, (I)	O	I	O
OM	I(48%), F(36%)	O, (I)	O	I	O

Note. Predictions in parentheses indicate predictions by the non-preferred process, i.e., p-entailments for PHM.

Multinomial Processing Tree (MPT) models

- MPT models prominent class of measurement models for categorical data (Riefer & Batchelder, 1988).

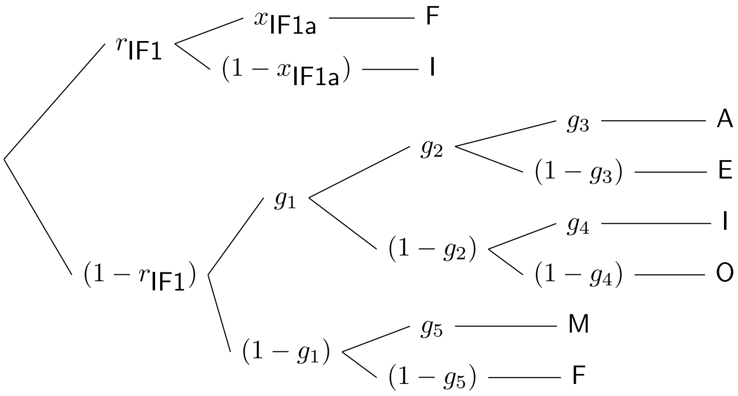
Multinomial Processing Tree (MPT) models

- MPT models prominent class of measurement models for categorical data (Riefer & Batchelder, 1988).
- Describe observed response frequencies as resulting from set of mutually exclusive latent cognitive states:
  - Reasoning state: response predicted by theories.
  - Uncertainty state: Any response can be guessed.

Multinomial Processing Tree (MPT) models

- MPT models prominent class of measurement models for categorical data (Riefer & Batchelder, 1988).
- Describe observed response frequencies as resulting from set of mutually exclusive latent cognitive states:
  - Reasoning state: response predicted by theories.
  - Uncertainty state: Any response can be guessed.
- Model parameters represent probability with which states are reached.

MPT model for IF1 (MMT)



MPT model comparison

- Model for each theory consisted of 40 different trees
  - For each theory only one guessing tree (constant across all items)
  - Full dataset $40 \times 5 = 200$ available degrees of freedom

MPT model comparison

- Model for each theory consisted of 40 different trees
  - For each theory only one guessing tree (constant across all items)
  - Full dataset $40 \times 5 = 200$ available degrees of freedom
- Each model fitted to data summed across individuals
  - Full dataset had $40 \times 25 = 1000$ observations

MPT model comparison

- Model for each theory consisted of 40 different trees
  - For each theory only one guessing tree (constant across all items)
  - Full dataset $40 \times 5 = 200$ available degrees of freedom
- Each model fitted to data summed across individuals
  - Full dataset had $40 \times 25 = 1000$ observations
- Model selection: Weighing model fit and model flexibility
  - AIC and BIC: Employ number of parameters as proxy for complexity
  - FIA: Estimates the functional complexity (third term below)

$$\text{FIA} = \frac{1}{2}G^2 + \frac{k}{2} ln \frac{N}{2\pi} + ln \int \sqrt{\det I(\Theta)}\, d\Theta$$

Model Comparison

Model Comparison

Theory	k	G^2	AIC	BIC	FIA
PMM	45	235.8	325.8	**546.6**	197.8
Min. M.	49	223.5	**321.5**	562.0	195.7
Matching	49	261.7	359.7	600.1	214.2
PHM	101	**187.2**	389.2	884.9	**182.4**

Model Comparison

Model Comparison

Theory	k	G^2	AIC	BIC	FIA
PMM	45	235.8	325.8	**546.6**	197.8
Min. M.	49	223.5	**321.5**	562.0	195.7
Matching	49	261.7	359.7	600.1	214.2
PHM	101	**187.2**	389.2	884.9	**182.4**

- Matching Hypothesis outperformed which contrasts with meta-analysis on classical syllogisms (Khemlani & Johnson-Laird, 2012)

Comparison of Reasoning Parameters

Comparison of Reasoning (r_i) Parameters

Theory	Mean	SD	Median	Min	Max
PMM	.44	.21	.46	.00	.82
Minimal Models	.46	.21	.48	.00	.82
Matching	.38	.26	.39	.00	.83
PHM	.51	.25	.53	.08	.93

Note. Although .00 is the smallest value for three theories, it does not occur at the same syllogism for all of them.

Comparison of Reasoning Parameters

Comparison of Reasoning (r_i) Parameters

Theory	Mean	SD	Median	Min	Max
PMM	.44	.21	.46	.00	.82
Minimal Models	.46	.21	.48	.00	.82
Matching	.38	.26	.39	.00	.83
PHM	.51	.25	.53	.08	.93

Note. Although .00 is the smallest value for three theories, it does not occur at the same syllogism for all of them.

- r_i parameters overall larger for **M** than for **F**.

Summary and Conclusion

- Everyday reasoning is based on degrees of belief rather than absolute certainty (Evans, 2002)
  $\Rightarrow$ generalized quantifiers "Most" and "Few"
- Only one theory so far
  - Probability Heuristics Model (Chater & Oaksford, 1999)
  - Extended Matching Hypothesis and two MM approaches
- Formalized as MPT models and empirically evaluated
- PHM and MM approaches outperform Matching Hypothesis
  - (which shows a good fit to the data on classical syllogistic reasoning; Khemlani & Johnson-Laird, 2012)
- MPT can be (even) used to build better theories!

The End

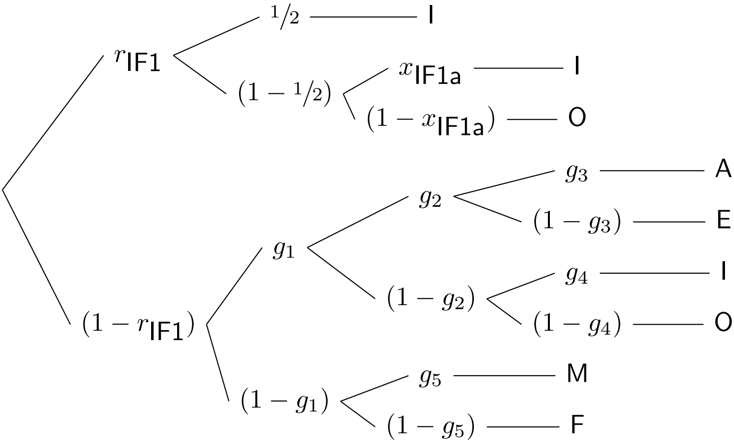
Thank You for Your Attention!

Special Thanks to

Franz Dietrich

Research partially funded by the DFG within
SFB/TR8 Project R8-[CSpace] and
the SPP “New Frameworks of Rationality”
project “Nonmonotonic logic”

MPT model for IF1 (PHM)



Model Comparison

Model Comparison

Theory	k	G^2	AIC	BIC	FIA	C_{FIA}
PMM	45	235.8	325.8	**546.6**	197.8	**79.9**
Min. M.	49	223.5	**321.5**	562.0	195.7	83.9
Matching	49	261.7	359.7	600.1	214.2	83.4
PHM	101	**187.2**	389.2	884.9	**182.4**	88.8

- Matching hypotheses outperformed which contrasts with large meta-analysis on classical syllogisms (Khemlani & Johnson-Laird, 2012)