Theory Comparison for Generalized Quantifiers

Marco Ragni[a] **(marco.ragni@cognition.uni-freiburg.de)**
Henrik Singmann[b] **(henrik.singmann@psychologie.uni-freiburg.de)**
Eva-Maria Steinlein[a] **(eva.steinlein@jupiter.uni-freiburg.de)**
[a] Center for Cognitive Science, [b]Department of Psychology
University of Freiburg

Abstract

Premises and conclusions in classical syllogistic reasoning are formed using one of four quantifiers (All, Some, Some not, None). In everyday communication and reasoning, however, statements such as "most" and "few" are formed as well. So far only Chater and Oaksford's (1999) Probability Heuristics Model (PHM) makes predictions for these so-called generalized quantifiers. In this article we (i) extend existing and develop new theories, (ii) develop multinomial processing tree (MPT) models for these theories, and (iii) conduct an experiment to test the models. The models are evaluated with G^2, Akaike's (AIC) and Bayesian Information Criteria (BIC), and Fisher's Information Approximation (FIA). Mental model-based accounts and PHM provide an equal account to the data.

Keywords: Syllogistic Reasoning; Generalized Quantifiers; MPTs; Model Selection

Introduction

Syllogistic reasoning derives a conclusion from two quantified statements. Each statement is formed using one of the four quantifiers: all (**A**), some (**I**), some not (**O**), or none (**E**). Consider, a simple syllogism, e.g.,

(1) All cognitive scientists are intelligent.
Some intelligent people are rich.

Most individuals erroneously infer from these statements that "Some cognitive scientists are rich" (Khemlani & Johnson-Laird, 2012). The only *logically valid response* is, however, that nothing follows. A meta-analysis by Khemlani and Johnson-Laird (2012) investigating twelve theories shows that there is still no psychological theory that provides a comprehensive account for human syllogistic reasoning. The assessed theories can be categorized into one of the following three classes: *heuristic theories*, *logical theories*, and *model-based theories*. The findings indicate that heuristic-based theories (such as the Matching Hypothesis by Wetherick & Gilhooly, 1990) and the Probabilistic Heuristics Model (PHM by Chater & Oaksford, 1999) and model-based theories (such as the Theory of Mental Models, see Johnson-Laird, 2006) performed better overall than theories of mental logic.

It has been criticized that classical quantifiers are too limited with respect to everyday contexts. The quantifiers A and E are too strict, not allowing any exceptions, whereas I and O are considered too weak requiring only a single individual (Pfeifer, 2006). In contrast, everyday human reasoning is "based [...] on beliefs, in which there are varying degrees of confidence" (Evans, 2002, p. 980). In accordance with that, we consider the generalized quantifiers most (**M**) and few (**F**), e.g.,

(2) Some X are not Y.
Few Y are Z.

From these premises a reasoner could conclude "Few X are Z" although again (logically) nothing follows. An additional advantage of using generalized quantifiers is that they can provide a new benchmark for theories of classical syllogistic reasoning.

The remainder of this article is structured as follows: We first review the only existing theory for the generalized quantifiers most and few – the PHM (Chater & Oaksford, 1999). In a second step we extend successful approaches for classical syllogistic reasoning. Specifically, we extend the Matching Hypothesis and develop two additional mental model-based accounts – one based on minimal models and the other on preferred mental models. Finally, all theories are formalized via multinomial processing tree (MPT) models and their empirical performance is assessed on data from a new experiment.

Theories for Generalized Quantifiers

Established Theory: Probability Heuristics Model

Chater and Oaksford (1999) proposed a *Probability Heuristics Model* (PHM) that accounts for the four classical and two generalized quantifiers M and F. It predicts that conclusions are chosen by their likelihood. Chater and Oaksford (1999) argue that quantifiers can be ordered according to their respective informativeness:

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

with A being the most informative quantifier as no exceptions are allowed. Simple heuristics like the *min-heuristic* and *p-entailments* describe the conclusion drawn by individual reasoners. The min-heuristic postulates that the quantifier of the conclusion is the same as the one in the least informative premise. Hence, for Example (2) above, the min-heuristic predicts reasoners conclude that "Some X are not Z" as the first premise (containing O) is less informative than the second one (containing F). The conclusion drawn by this heuristic is called min-conclusion. If more than one conclusion can be derived from a given problem, the following conclusions are called p-entailments (i.e., they are probabilistically entailed) of the min-conclusion. For Example (2) the min-conclusion "Some X are not Z" entails that "Some X are Z."

The predictions were evaluated in two experiments (Chater & Oaksford, 1999). Participants received two premises and had to choose one, multiple, or none of four response options. Each answer option contained a different quantifier and related the third term Z to the first term X. The first experiment investigated the quantifiers A, M, F, and O, while a second experiment considered the quantifiers M, F, I, and E. Forcing participants to draw $Z-X$ inferences (though in good tradition as they argue in their paper) influences the results; experiments in which both orders are permitted (e.g., Bucciarelli & Johnson-Laird, 1999) have shown that participants (sometimes) prefer the $X-Z$ order. An additional point is that multiple responses are more difficult to interpret.

Extending Classical Theories

Matching Hypothesis. Another heuristic approach to syllogistic reasoning is the *Matching Hypothesis* (Wetherick & Gilhooly, 1990). This theory claims that the quantifier of the conclusion is selected according to the quantifier of the most conservative premise, i.e., referring to the smallest number of individuals. Ordered from most to least conservative the quantifiers are:

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

A strength of this theory is that it can easily be generalized to additional quantifiers. By assuming that M and F are less conservative than I and O – as they not only make statements about an arbitrary number, but also impose quantitative restrictions – we get the order:

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

For Example (2) above, the predicted responses by the Matching Hypothesis are "Some X are not Z", and "Some Z are not X". This approach, however, is weak in that it cannot explain instances in which *nothing follows* from the premises since a least conservative quantifier always exists. Although the postulated order of quantifiers differs from the order predicted by the PHM, the underlying mechanisms are quite equivalent.

Mental Models and Heuristics. An important non-heuristic approach is the *Mental Model Theory* (MMT) (Bucciarelli & Johnson-Laird, 1999; Khemlani & Johnson-Laird, 2012). This theory claims that deductive reasoning consists of the construction and manipulation of mental models.

While the classical MMT (Bucciarelli & Johnson-Laird, 1999) does not provide predictions for reasoning with generalized quantifiers (but, see Neth & Johnson-Laird, 1999) a new approach has been proposed by Johnson-Laird and Khemlani (2014) – a Unified Theory – which integrates heuristics into the MMT. This theory is implemented in a computer program called *mReasoner*[1]. It contains a semantic for the generalized quantifier most (M) in addition to the classical quantifiers.

We extend the classical theory with the principle of parsimony, i.e., we assume that reasoners construct *Minimal Models*. This approach receives support from the limited working memory capacity. Additionally, heuristics guide the reasoning process: Reasoners first construct a model in which they try to verify a conclusion with a quantifier from one of the premises. The minimal mental model for Example (2) is:

X		
X	Y	Z
	Y	
	Y	

Consequently the reasoner chooses a quantifier from the premises and verifies whether it holds in the initial model. Empirical data suggests that this choice is based on the following ordering of quantifiers:

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

As F appears before O in this order, F is checked and the initial model of Example (2) supports the conclusion that "Few X are Z." Note that during the construction of the *initial model*, the preferred strategy is to minimize the number of individuals, which implies a maximization of the overlap between individuals.

Although this might in some ways resemble the heuristics of the Unified Theory (for a detailed description, see Johnson-Laird & Khemlani, 2014), there is one important difference: The Unified Theory assumes that an initial model is constructed first and only with this model in mind do heuristics affect the derivation of a conclusion. In contrast to this, our account of Minimal Models predicts that heuristics play a role with regard to the construction of the initial model, which is governed by the quantifiers.

Preferred Mental Models. In a current analysis (Ragni, Schrögendorfer, & Nebel, in prep) we formalized the MMT as spatial models, i.e., a mental model can be seen as a mapping from the set of premises $\mathbb{P}$ onto a discrete space $\mathbb{N}^2$. The preferred model is the one which has minimal extension. To provide an example:

(3) Few of the Architects are Beekeepers.
Some of the Beekeepers are Chemists.

We identify all the architects with a set of instances A, the beekeepers with a set of instances B, and the set of instances of chemists with C for each premise. The syllogism above is $P_1 := (\textbf{Few}, A, B)$ and $P_2 := (\textbf{Some}, B, C)$. The preferred mental model for Example (3) is constructed as the minimal interpretation Ω_1 and a map $\varphi_1 : \Omega_1 \to \mathbb{N}^2$ satisfying premise P_1.

$$\Omega_1 := A \dot{\cup} B$$
$$A_1 := \{a, a', a''\}, \quad B_1 := \{b\}$$
$$\varphi_1 : \quad a \mapsto (1,1) \quad a' \mapsto (1,2)$$
$$a'' \mapsto (1,3) \quad b \mapsto (2,1)$$

The graph of φ_1 is shown in Figure 1 on the upper left and consists of the instances $\{a, a', a'', b\}$. The instances a and b share the same row. Therefore, exactly one A-instance

[1] http://mentalmodels.princeton.edu/models/mreasoner/

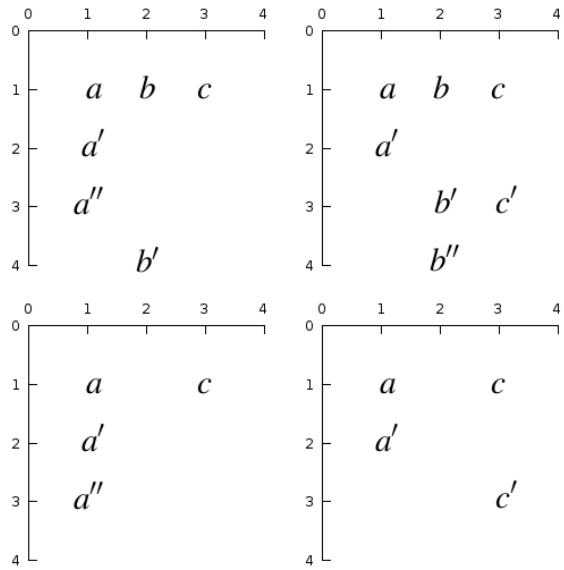

Figure 1: Graphs of preferred mental models to syllogisms of the form **F**-**O** (ul), and **I**-**M** (ur), the conclusion models in lower row implying the conclusion **F** (ll), and **I** (lr)

a is a B-instance, whereas the A-instances a' and a'' are not B-instances. Thus, the premise P_1 is satisfied, and it easily follows that Ω_1 is chosen with minimal cardinality.

We now choose a proper extension of the model φ_1 to get the preferred mental model φ of Example (3) shown in Figure 1 on the upper left. The formal model φ_2 is minimal. Thus, to satisfy $P_2 := (\textbf{Some}, B, C)$ exactly the B-instance b is row equivalent to $c \in C$. Therefore, $A := \{a, a', a''\}$, $B := \{b, b'\}$, and $C := \{c\}$. This model is constructed to be minimal, parsimonious, and to satisfy premise P_2. Consider Example (4):

(4) Some of the Artists are not Bakers.
Most of the Bakers are Cheerleaders.

On the upper right side of Figure 1 the preferred mental model φ' of the syllogism

$$((\textbf{Some Not}, A, B), (\textbf{Most}, B, C))$$

can be found. The model φ'_1 (that is restricted to a, a', b or formally, $\varphi' \mid_{\{a,a',b\}}$) is minimal and satisfies the first premise $P'_1 := (\textbf{Some Not}, A, B)$, and is a proper submodel of φ. Through extension of the map φ'_1 to a minimal and parsimonious model which satisfies P'_2 the preferred mental model φ' is constructed.

Examine the A-C-submodels of φ and φ' to draw conclusions (from these models). The models for Example (3) and Example (4) are given in the lower row of Figure 1. The submodel φ_{AC} to example (3) uniquely satisfies the premise $P := (\textbf{Few}, A, C)$, and the submodel φ'_{AC} to example (4) holds $P' := (\textbf{Some}, A, C)$. As nothing else holds, we conclude from the constructed preferred mental models for problem (3) that "Few of the Architects are Chemists." and for problem (4) that "Some of the Artists are Cheerleaders".

Besides heuristics and mental models, another important approach in syllogistic reasoning are theories based on formal rules, like for example the PSYCOP model (Rips, 1994). According to this theory, reasoners rely on formal rules of inference which they apply to propositional representations of the premises. So far this approach is limited to the classical quantifiers and thus not included in our analysis.

Model Comparison and MPT-Analysis

One of our main goals was to assess the empirical adequacy of the presented theories. To achieve this in a statistical sophisticated manner, we resorted to formalizing the theories as multinomial processing tree (MPT) models (Riefer & Batchelder, 1988). MPT models are a class of cognitive measurement models for multinomial (i.e., categorical) data that describe observed response frequencies as resulting from latent cognitive states. The probabilities that the cognitive states are reached are estimated from the data and provide a measure of the contribution of said states to the observed responses. To formalize all models in a consistent manner, we assumed that responses can be produced by one of two (mutually exclusive) cognitive states: A response is either (a) produced by the reasoning processes assumed by the theories, in which case a response predicted by the theories is given, or (b) a response is guessed in which case any possible response can be given (see Oberauer, 2006, for a similar approach for conditional reasoning).

More specifically, regarding (a) we assumed that for each syllogism with probability r_i (where i represents the specific syllogism) a reasoning state is reached. In the reasoning state, one of the predicted responses is invariantly given (i.e., we assume the absence of post-reasoning response processes such as motor errors). If a theory predicts more than one response for a given syllogism, the probability with which each of the predicted responses is given is estimated freely from the data (if only one response is predicted, this response is given). One exception existed for PHM which assumes that two distinct processes, min-heuristic and p-entailment, can generate conclusions and min-heuristic is the preferred process. Consequently, we imposed a weak-order on the multinomial distribution predicted by the reasoning state of PHM such that the probability of responses by min-heuristic was equal or greater than the probability of responses by p-entailments.[2]

Regarding (b), we assumed that in cases where the reasoning state is not reached a guessing state is entered with probability $(1 - r_i)$. Further, we assumed that the guessing state was identical across all syllogisms (i.e., across all i). Within the guessing state, the probability with which each response is given is estimated freely from the data. In other words, whereas the guessing tree can in principle account for any observed data pattern, the models are restricted in such a way that the predicted guessing pattern needs to be identical for all

[2] In cases where one of the two processes predicted more than one response (which occurred several times), this restriction pertained to the sum of the predictions from one of the processes.

Table 1: Reliable responses for all 40 syllogisms and the corresponding response proportions.

		2. Premise					
1. Prem.	Concl. Type	**A**	**M**	**I**	**F**	**E**	**O**
A	X-Z		**M(72%)**, I(24%)		F(60%)		
	Z-X		M(56 %), I(32%)		F(68%)		
M	X-Z	M(84%)	M(84%)	I(68%)	F(60%)	E(36%), F(36%)	I(56%)
	Z-X	M(68%)	M(56%), I(36%)	I(60%)	F(56%), I(36%)	E(44%), F(28%)	I(52%)
I	X-Z		I(76%)		F(56%), I(32%)		
	Z-X		I(68%)		F(52%)		
F	X-Z	F(76%)	F(64%)	F(64%)	F(84%)	E(40%), I(28%)	F(44%), I(32%)
	Z-X	F(72%)	F(56%), I(32%)	I(40%), F(36%)	F(68%)	E(44%)	F(32%), I(28%)
E	X-Z		E(68%)		E(64%)		
	Z-X		E(72%)		E(56%)		
O	X-Z		I(48%), F(36%)		F(48%), I(24%)		
	Z-X		I(56%)		F(56%), I(24%)		

Note. Responses which occurred significantly often (>22%). For one item the modal response (*M*, in bold) was given significantly more often than a second reliable response (*I*).

syllogisms for a given theory (i.e., guessing is constant across syllogisms).

The advantages of formalizing the theories in such a way are basically threefold. First, formalizing all theories within one model class allows a consistent statistical treatment across theories. Second, MPT models are a statistically well developed model class (see Singmann & Kellen, 2012, for an overview). Specifically, in addition to assessing model fit via the G^2 statistics we can employ model selection indices that provide a sort of automatic Occam's razor by taking both model fit and model complexity into account. Here we use the well-known Akaike information criterion (AIC), the Bayesian information criterion (BIC), as well as the Fisher Information Approximation (FIA Wu, Myung, & Batchelder, 2010) a measure that estimates the flexibility of a model. Third, by separating and estimating the contribution of reasoning versus guessing, we can evaluate the contribution of both types of processes across theories, but also across syllogisms. We now describe the experimental data obtained to compare the formalized theories.

Method

Participants.

Twenty-five English native speakers (M = 30.8 years) participated in this experiment. They were recruited by Amazon Mechanical Turk and paid for their participation.

Materials, Procedure, and Design.

The experiment was conducted as an online experiment via Amazon Mechanical Turk. Each participant completed 40 syllogistic reasoning tasks and additionally one easy test problem which was excluded from the analysis. The tasks consisted of two premises at least one of which contained a generalized quantifier (F or M). Simultaneously with the premises, the question "what follows?" as well as a response field for the conclusion appeared. The participants had to generate a conclusion using one of the six quantifiers (A, E, F, I, M, O), which were displayed at the bottom of the screen, and type their answer (i.e., only the quantifier) into a provided response field (see example below). Participants could only proceed to the next syllogism if they entered one of the six quantifiers into the response field. In order to enhance reasoning and avoid fast responses, "nothing follows" was not provided as a response option. An example item follows:

All brokers are waiters.
Few waiters are agents.

What follows?

[ ] of the brokers are agents.

Quantifiers: All, Some, Some Not, Most, Few, None.

In half of the trials the conclusion related the subject of the first premise (i.e., X) to the end-term of the second premise (i.e., Z), and vice versa in the remaining 20 trials; all 40 tasks were of the same form as Example (2) above (i.e., $X-Y$ $Y-Z$, which is also known as Figure I). Each set of 20 syllogisms consisted of the following items: 6 syllogisms in which *most* appeared as first quantifier and each quantifier appeared as second quantifier once, 6 syllogisms in which *few* appeared as first quantifier and each quantifier appeared as second quantifier once, 4 syllogisms in which each of the four standard quantifiers (i.e., A, E, F, I) appeared as first quantifier once and *most* appeared as second quantifier, and 4 syllogisms in which each of the four standard quantifiers appeared as first quantifier once and *few* appeared as second quantifier. The syllogisms were presented in a individually randomized order. Different professions and hobbies constituted the content of the terms (e.g., Chater & Oaksford, 1999).

Results

Descriptive Results

Table 1 presents the reliable responses (i.e., responses which occurred significantly often, >22%) and the corresponding response proportions for all 40 syllogisms. For one item (printed in bold) the modal response (M) was given significantly more often than a second reliable response (I).

Table 2: Predictions of the four theories for selected syllogisms (X-Z conclusion) and significant choices.

Syll.	Data	PHM	Matching	PMM	Min. Models
IF	F, I	I, (O)	I	F	F, I
FI	F	I, (O)	I	F	F, I
FO	F, I	O, (I)	O	I	F
OF	F, I	O, (I)	O	I	F
MO	I	O, (I)	O	I	O
OM	I, F	O, (I)	O	I	O

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

Table 3: 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

Note. k is the number of model parameters, the total number of available df is 200 (i.e., model $df = 200 - k$). All $p < .001$. Theories are ordered by the number of parameters. The smallest value per column is printed in bold. FIA gives the value of the Fisher Information Approximation and C_{FIA} is the penalty term for FIA representing the flexibility which is estimated using 1 million Markov chain Monte Carlo (MCMC) samples.

MPT analysis

For each of the 40 syllogisms in our data we created predictions from the four theories. A few examples in which the theories give specifically diverging predictions are presented in Table 2 along with the reliable responses given by the participants. From the predictions we constructed MPT models as described above. The full model for each theory consisted of 40 trees, each representing one syllogism. Each tree contained an individual r_i parameter estimating the probability with which a response was produced by a reasoning or guessing state. As described above, the probabilities of responses within the reasoning branch of the trees were freely estimated (with the restriction regarding the min-heuristic for PHM) and the parameters in the guessing tree were constant across all trees (i.e., each model only had one set of five guessing parameters for all 40 syllogisms).

We fitted each model to the aggregated data via the maximum likelihood method using `MPTinR` (Singmann & Kellen, 2012). The full dataset had 5 $\times$ 40 = 200 available degrees of freedom for a total of 25 $\times$ 40 = 1000 responses. Model comparison data is presented in Table 3. In terms of model fit as measured by the G^2 statistic the first important observation is that all models provide an inadequate account (i.e., all $p < .001$). Furthermore, PHM provided the best account (smallest $\Delta G^2 = 36.30$). This is not too surprising given that it makes the most predictions and has the most parameters.

When considering both model fit and model flexibility the picture somewhat changes. The two "naive" measures AIC and BIC (which operationalize model flexibility only via numbers of parameters) favour the two Mental Model based theories (AIC: Minimal Models; BIC: PMM). But it should be noted that AIC and BIC are somewhat inappropriate for models where parameters impose order restrictions as is the case for PHM. For such a case, FIA is a more appropriate measure as it estimates the flexibility of a model, taking order restrictions into account (see Wu et al., 2010, for a discussion). According to FIA, PHM provides the best account. However, when inspecting the FIA penalty terms (which are added to $\frac{1}{2}G^2$) estimated model flexibility only minimally differs (maximal difference is 8.9). Given the rather huge differences in number of parameters and predictions between e.g., PMM and PHM, this is odd. It most likely reflects the asymptotic nature of FIA, which only perfectly captures flexibility differences for $n \to \infty$, and our small sample size (only $n = 25$ per tree; where n in both cases is the number of items). Hence, we interpret the results that the model that strikes the best balance between model fit and flexibility is either one of the two mental model-based theories or PHM.

In the next step of the analysis we looked at the r_i parameters. As shown in Table 4, the theories differ wrt. the amount of reasoning compared to guessing they assume. In line with the G^2 values, PHM estimated the greatest probability for reasoning based responses. However, even for PHM the mean probability for a reasoning based response was .5. In other words, even for the best fitting theory (i.e., ignoring the issue of model flexibility), half of the responses were produced by guessing processes and only the other half of the responses by the reasoning processes the theories assume. This finding may well be responsible for the overall unsatisfactory model fit given that the guessing process which was responsible for at least 50% of the responses was assumed to be constant across all 40 syllogisms (i.e., only 5 parameters for 50% of the responses across 40 trees).

Finally, we analyzed the r_i parameters with a mixed ANOVA (i.e., an item-wise analysis) with *theory* as within-

Table 4: 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.

subject factor and *generalized quantifier* (M vs. F), *position* of generalized quantifier (first vs. second premise), and *conclusion* (X-Z vs. Z-X) as between-subjects factors. The analysis revealed, besides a main effect of theory ($F(2.59, 83.00) = 10.36$, $p < .001$), a significant theory $\times$ generalized quantifier interaction ($F(2.59, 83.00) = 11.22$, $p < .001$).[3] With the exception of the Minimal Model theory were the reasoning parameters consistently larger for M than for F (differences are .10 for PMM, -.03 for Minimal Models, .13 for Matching, and .21 for PHM). This indicates that most theories predicted more reasoning when M was the generalized quantifier compared to F.

Discussion

Everyday reasoning is based on degrees of belief rather than absolute certainty. To this end, we investigated syllogisms with generalized quantifiers such as "Most" (M) and "Few" (F). There is currently only one approach – the Probability Heuristics Model (Chater & Oaksford, 1999) – that makes any predictions about possible inferences. We extended the Matching Hypothesis and developed two model-based approaches for generalized quantifiers, formalized them as MPT models and evaluated them on a new dataset.

Our results show that MPT models can be used to successfully compare reasoning theories. In terms of model fit, however, all discussed theories provide an inadequate account to the data. Nevertheless, differences between these approaches exist in their predictive power. PHM and the mental model-based approaches clearly outperform the Matching Hypothesis (which shows a good fit to the data on classical syllogistic reasoning; Khemlani & Johnson-Laird, 2012). Deciding between the remaining three theories seems to require a larger number of responses as indicated by the inconclusive results from the different model selection indices employed. Specifically the small differences in FIA penalty given the rather dramatic differences in the predictions between the theories point to this conclusion.

While the results indicate that no current theoretical approach captures the full set of responses, one further take home message can be distilled from our analysis: Disentangling the contribution of reasoning and guessing based responses reveals that almost all theories are more successful in predicting participants' responses when reasoning with "Most" compared to "Few". This result shows that our MPT modelling can help in providing specific insight for further theoretical development.

In contrast to other analysis strategies, MPT modeling allows to test specific, possibly ordered, predictions and uses raw response frequencies instead of transformed data. Furthermore, the ability to disentangle latent cognitive processes such as reasoning and guessing provides important theoretical insights. From this perspective, MPT models are an excellent cognitive modeling methodology for assessing reasoning theories on a new and more precise level and allow for distinguishing between competing theories in a unified framework.

[3] *df* are Greenhouse-Geisser corrected.

References

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science: A Multidisciplinary Journal*, *23*(3), 247–303.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive psychology*, *38*(2), 191–258.

Evans, J. S. B. T. (2002). Logic and Human Reasoning: An Assessment of the Deduction Paradigm. *Psychological Bulletin*, *128*(6), 978–996.

Johnson-Laird, P. N. (2006). *How we reason*. New York: Oxford University Press.

Johnson-Laird, P., & Khemlani, S. (2014). Toward a unified theory of reasoning. *Psychology of Learning and Motivation*, *59*, 1 - 42.

Khemlani, S., & Johnson-Laird, P. N. (2012). The processes of inference. *Argument and Computation*, *1*(1), 1–17.

Neth, H., & Johnson-Laird, P. N. (1999). The search for counterexamples in human reasoning. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty-first annual conference of the cognitive science society* (p. 806). Mahwah: Lawrence Erlbaum.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology*, *53*, 238–283.

Pfeifer, N. (2006). Contemporary syllogistics: Comparative and quantitative syllogisms. In G. Kreuzbauer & G. J. W. Dorn (Eds.), *Argumentation in Theorie und Praxis: Philosophie und Didaktik des Argumentierens* (pp. 57–71). Wien: LIT.

Ragni, M., Schrögendorfer, D., & Nebel, B. (in prep). Formal Mental Model Theory in Syllogistic Reasoning.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: The MIT Press.

Singmann, H., & Kellen, D. (2012). MPTinR: Analysis of multinomial processing tree models in R. *Behavioral Research Methods*, *45*, 560–575.

Wetherick, N. E., & Gilhooly, K. J. (1990). Syllogistic reasoning: Effects of premise order. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought, vol. 1* (pp. 99–108). Chichester: Wiley.

Wu, H., Myung, J. I., & Batchelder, W. H. (2010). Minimum description length model selection of multinomial processing tree models. *Psychonomic bulletin & review*, *17*(3), 275–286.