Proceedings of ICCM 2023

21st International Conference on Cognitive Modelling[1]

Edited by

Catherine Sibert

1 Co-located with the 56th Annual Meeting of the Society for Mathematical Psychology

Preface

The International Conference on Cognitive Modelling (ICCM) is the premier conference for research on computational models and computation-based theories of human cognition. ICCM is a forum for presenting and discussing the complete spectrum of cognitive modelling approaches, including connectionism, symbolic modeling, dynamical systems, Bayesian modeling, and cognitive architectures. Research topics can range from low-level perception to high-level reasoning. In 2023, ICCM was jointly held with MathPsych – the annual meeting of the Society for Mathematical Psychology. The conference was held at the University of Amsterdam from July 18th to July 21st. An additional, virtual conference was held online from June 19th to June 23rd. Submissions from both the in-person and virtual conferences are included in these proceedings.

Acknowledgements

We would like to thank the Society for Mathematical Psychology (SMP) for their ongoing commitment to the collaboration between our societies. In particular, the MathPsych Conference Chair (Joachim Vandekerckhove) maintained the conference website, and the officers of the SMP (especially Leslie Blaha) gave much needed logistical support.

Papers in this volume may be cited as:
Lastname, A., Lastname, B., & Lastname, C. (2023). Title of the paper. In Sibert, C. (Ed.). *Proceedings of the 21st International Conference on Cognitive Modelling* (pp. 6-12). University Park, PA: Applied Cognitive Science Lab, Penn State.

ISBN-13: 978-0-9985082-7-6, published by the Applied Cognitive Science Lab, Penn State.

Preferred Mental Models in Syllogistic Reasoning

Sara Todorovikj (sara.todorovikj@hsw.tu-chemnitz.de)
Daniel Brand (daniel.brand@hsw.tu-chemnitz.de)
Marco Ragni (marco.ragni@hsw.tu-chemnitz.de)
Department Behavioural and Social Sciences, Technische Universität Chemnitz,
Straße der Nationen 62, 09111 Chemnitz

Abstract

Inspired from previous research in the spatial reasoning domain, in this paper, we address the varying interpretations of premises of syllogistic problems among individuals and the differences in their resulting mental models. We conducted an experiment whose results show that model building is a relatively easy task for humans to do correctly and they do in fact have preferred models for most syllogisms, yet, without a relation to their responses. We report in-depth analysis of the models' canonicality in order to compare the model building behavior in humans to the processes implemented in mReasoner, a cognitive model that implements the Mental Model Theory.

Keywords: Syllogistic Reasoning; Preferred Mental Models; mReasoner

Introduction

With over a century of research history (Störring, 1908), syllogisms are one of the core domains of examining human reasoning abilities. A syllogism consists of two quantified premises describing the relationships between three terms through a common middle term. In a world of colourful shapes, consider the following syllogism:

All red shapes are circles.
Some red shapes are marked with a star.

What, if anything, follows?

The task at hand is to determine what kind of relation, if any, exists between the two end-terms, circles and (marked with a) star, also called subject and predicate, respectively.

There exist at least twelve theories that aim to explain and model the processes behind human syllogistic reasoning (for an overview, see Khemlani & Johnson-Laird, 2012). One of the most prominent theories among them is the Mental Model Theory (MMT; e.g., Johnson-Laird, 1975, 2010). MMT postulates that given some observations, individuals create iconic representations – *mental models* – of possibilities. They create their own subjective mental representation of the information presented in a reasoning task. Considering the example above, one possible representation would be:

circles	[red]	[star]
circles		

The square brackets around an instance denote that the set of entities described by it is exhaustively represented. Another possible mental model representation is:

circles		
circles	red	[star]
¬circles	¬red	

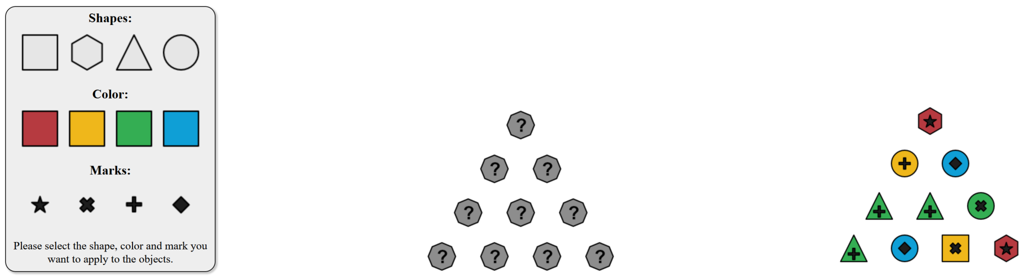
where ¬ denotes negation. Both mental representations support the conclusion "Some circles are marked with a star" - the logically valid conclusion to this syllogism. However, in order to confirm the validity, an individual should think of all possible premise interpretations and check if they hold. The expansion of the interpretation search space can make solving such problems difficult for humans (Johnson-Laird, 2006).

In the spatial relational reasoning domain researchers have repeatedly shown that individuals have *preferred mental models*, namely that they prefer creating some models while struggling with others (e.g., Ragni & Knauff, 2013; Jahn, Knauff, & Johnson-Laird, 2007; Rauh et al., 2005). Interestingly, experimental setups for the syllogistic domain do not typically address the model building process of reasoning. Namely, they do not involve examinations of which models the individuals create, if they are correct, or if they even have preferred models at all. To this end, we conducted an experiment where participants had to provide visual responses showing their representation of the given syllogistic premises, and with that we tackle our first research question:

[RQ1] Can we examine what kind of models do individuals create from the premises of syllogistic tasks and do they have preferred mental models?

In mathematical and computer sciences, the minimal, simplest representation of an expression is referred to as its canonical form. This concept is also discussed in the context of mental models in the syllogistic reasoning domain (Khemlani, Lotstein, Trafton, & Johnson-Laird, 2015). Namely, it denotes which entities form a canonical set for a given syllogism and which non-canonical instances do not have to be present in an individual's model but are still consistent with the premises. For example, in the representations above "circles red" is a canonical instance for the first syllogistic premise ("All red shapes are circles"), whereas "¬circles ¬red" is not. Thereby, canonicality can be interpreted as a mean to assess the "incompleteness" of the model in the sense of the coverage of all possible interpretations of the premises.

From the perspective of cognitive modeling, it is especially interesting if the canonicality of the models provided



(a) Selection box (b) Placeholder objects (c) Possible representation

Figure 1: Experimental design – Participants used the selection box to pick out their desired attributes that they can assign to placeholder objects in order to share their mental representation of a given syllogism.

by participants align with the assumptions made by MMT. The most prominent implementation of MMT for syllogistic reasoning is the LISP-based model *mReasoner*[1] (Khemlani & Johnson-Laird, 2013), which will therefore serve as a foundation of our analysis. Distinguishing between three systems, mReasoner creates intensional representations of the premises (System 0), builds and interprets an initial model (System 1) and performs a search for counterexamples (System 2) (Khemlani & Johnson-Laird, 2013). System 1 parameterizes the number of entities in a model and their canonicality - the likelihood whether they are drawn from a canonical set of typical entities or the full set of entities consistent with the premises (Khemlani et al., 2015). We analyze our participants' built models further and contrast them to the output of mReasoner's model building stage to address our bipartite second research question:

[RQ2.1] How influential is the canonicality of mental models that individuals build for syllogistic premises on the correctness of derived conclusions?

[RQ2.2] Is the model building behavior observed in humans in line with the model building processes of mReasoner?

Our paper is structured as follows – we first provide the necessary theoretical background regarding reasoning with syllogisms and mReasoner, followed by an in-depth explanation of our experiment. Afterwards, we analyze the experimental data and the participants' models (RQ1) and the correspondence of mReasoner's canonicality approach to the data (RQ2). We then conclude the article with a discussion of our findings.

Theoretical Background

Syllogisms

The two syllogistic premises and conclusion are characterized by their quantifiers and term order. We take into consideration the four first-order logic quantifiers *All*, *Some*, *No* and *Some not*, abbreviated by A, I, E and O, respectively. The order of the subject, predicate and middle terms in the premises determine the *figure* of the syllogism. We use the following notation (adopted from Khemlani & Johnson-Laird, 2012):

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

Using the abbreviations and figures, the example syllogism introduced above is denoted by AI4. Similarly, the conclusions are denoted by using the quantifier's abbreviation and take into consideration the direction of the end-terms – *ac* or *ca*, e.g. *Oca* indicates that Some C are not A. Finally, 'No valid conclusion' is abbreviated by NVC.

mReasoner

According to MMT, given syllogistic premises, individuals represent the entities described by the quantifiers using mental models and aim to derive a conclusion based on that. Before accepting a certain conclusion, they engage in a search for counterexamples, which, if successful, can lead to rejecting and correcting the original conclusion or concluding that there is no valid conclusion.

These processes are implemented within the cognitive model *mReasoner* (Khemlani & Johnson-Laird, 2013, 2016). Using the following four parameters it builds models and searches for counterexamples: λ determines the *size*, i.e. the number of entities as drawn from a Poisson distribution; ε determines the *canonicality*, i.e. how complete is the set of represented possibilities, given the premises; σ describes how likely is it to engage in a search for *counterexamples* and ω decides what happens when a counterexample is found – whether the conclusion is weakened or NVC is reported.

Experiment

The main objective of the experiment was to obtain a visual representation of the participants' (preferred) mental representation of given syllogisms. In order to achieve that, they were presented with a syllogism, whose terms are descriptions of objects and were asked to demonstrate what they imagined ten objects look like when taking into consideration the syllogistic premises.

An object is described using its *shape* (square, hexagon, triangle, circle), *color* (red, blue, green, yellow) and *mark*

[1]https://github.com/skhemlani/mReasoner

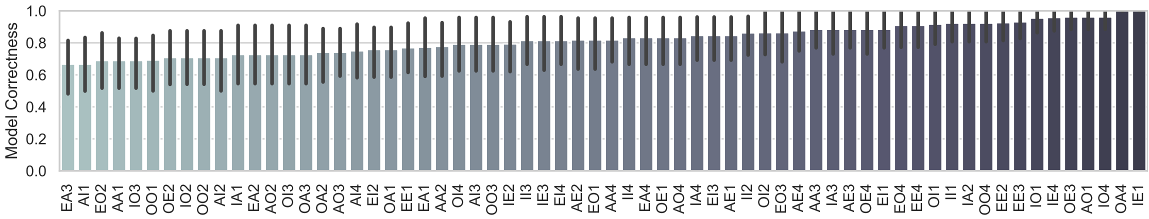


Figure 2: Mean correctness of the constructed models by syllogism.

(cross, plus, star, diamond). The presented syllogisms had the shapes as subjects, the colors as middle terms and the marks as predicates. For example, the syllogism AE3 with content 'triangles', 'green' and 'diamond' is as follows:

All triangles are green.
No shapes that are marked with a diamond are green.

The object attributes were randomized among syllogisms. Once presented with a syllogism, participants see a selection box and placeholder objects (Fig. 1a, 1b). By clicking on any shape, color and mark they select their desired object attributes that they can apply on a placeholder object by clicking on it. Once they are done defining the properties of each object, they end up with a visual representation of their mental model of the syllogism, i.e. what they imagine the 10 objects to look like based on information provided in the syllogism. For example, Fig. 1c depicts a possible mental representation of the syllogism AE3 presented above.

In a second part of the experiment, participants are once again presented with the same syllogisms and are prompted to select which of the 9 possible responses follow (e.g. Brand, Riesterer, & Ragni, 2022).

Participants are divided in eight groups based on the presented syllogisms. In order to maintain a similar experience among participants, we used the `Ragni-2016` dataset obtained from the Cognitive Computation for Behavioral Reasoning Analysis (CCOBRA) Framework[2], to determine the difficulty of syllogisms based on the amount of correct responses. The `Ragni-2016` experimental data provides responses from 139 participants for all 64 syllogisms. We divided all syllogisms in eight difficulty groups and created the final sets of presented syllogisms by selecting one from each group.

Participants

We obtained data from 200 participants (age 19-76, 42% female) recruited on Prolific[3] and the experiment was performed online as a web-experiment. After completing the experiment, the participants received compensation of 3 GBP. All of them were native English speakers.

Procedure

Participants were first introduced to all possible attribute options in terms of shapes, colors and marks that an object can have. Following is an explanation on how to select and apply the desired object attributes on the placeholder objects. It was emphasized that they must select an option for each attribute and the appearance of all ten objects has to be specified. Then the experiment started and they had to show what they imagine the objects look like, using the introduced selection-box and placeholder objects, for 8 syllogisms. Once these tasks were completed, participants started the second portion of the experiment – the single choice tasks for the same 8 syllogisms.

Analysis

General Experimental Data Analysis

Since the object attribute descriptions were randomized among tasks, throughout our analysis and model comparisons we focus on whether the attributes in the responses correspond to the attributes presented in premises or not. That means, for example the model instance "square red" for the syllogistic premise "All squares are red" is treated equally with the model instance "circle green" for the premise "All circles are green".

We analyzed the correctness of the provided representations. Given a syllogistic premise with terms X and Y, we distinguish the following scenarios in which the representations are correct, based on the quantifier. For *All*, there must be no $X\neg Y$ instances. In the case of *No*, there must be no XY instances. Finally, for *Some* and *Some not*, there should be at least one XY or $X\neg Y$ instance, respectively. Out of 1600 observations, participants provided a correct representation in 1314 of them (82.12%). The mean correctness of the models by syllogism is visually represented in Figure 2. In 497 cases (31.06%) the participants gave a logically correct response and for only 408 (25.50%) they provided a correct representation *and* a logically correct response. Despite substantial differences in the model correctness between different syllogisms, it does not appear to be related to the difficulty of the syllogism: The correctness of the representations and responses do not have a significant correlation (Spearman's $r = -.0005$, $p = .9819$). Besides no apparent connection to task difficulty, a comparison between the best and worst 32 tasks also indicates that the model correctness seem to not be affected by negativity of quantifiers (24/24 in best/worst, respectively), particularity (25/23) or validity (17/20 invalid syllogisms). The only peculiarity is related to the figure of the syllogism, with figure 2 leading to more incorrect models

[2] https://orca.informatik.uni-freiburg.de/ccobra/
[3] https://www.prolific.co/

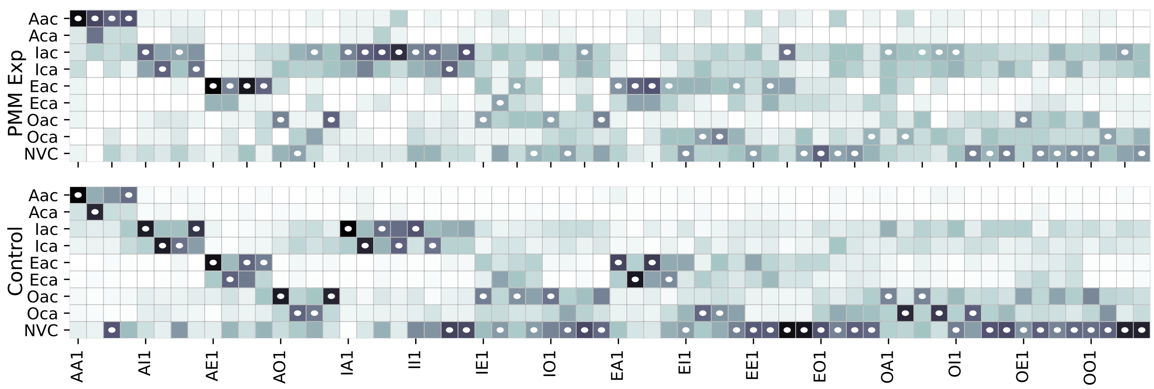


Figure 3: Response distributions for all syllogisms in the conducted experiment and the `Ragni-2016` control dataset. Higher percentage of given response is depicted with a darker color and the most frequently selected response for each syllogism is denoted with a white dot.

(12/4) and figure 4 seemingly being easier for model construction (3/13). Overall, the correctness of the models remains arbitrary with respect to typical structural properties of syllogisms commonly known to affect syllogistic reasoning performance. The response distribution among all participants is illustrated in Figure 3. We contrast the obtained responses to neutral, control data – the `Ragni-2016` data from CCOBRA, as introduced above. It can be immediately observed there is a tendency for participants to not choose NVC answers as often as in the control data and to avoid the *ca* direction in their responses. This implies that there was a belief bias effect among participants, namely that some superficial beliefs and background knowledge were induced by introducing a world with a discrete amount of possible object attributes.

Preferred Mental Models

Similarly to above, in the following analysis we focus solely on the presence and absence of the attributes in the participants' responses, without considering the specific contents. That narrows down the instance space to 8 different entities, for a syllogism with terms X, Y and Z:

X Y Z	X Y ¬Z	X ¬Y Z	X ¬Y ¬Z
¬X Y Z	¬X Y ¬Z	¬X ¬Y Z	¬X ¬Y ¬Z

For each representation, we created a binary vector of size 8 that indicates whether an instance was present in the model (= 1) or not (0) denoting an individual's preferred model pattern. We then counted among all participants, how many times each pattern occurred for each syllogism. The one pattern with the most occurrences is then the preferred mental model for a given syllogism. Figure 4 shows a visual representation of the participants' preferred mental models. Note that not all 64 syllogisms are represented, only those that have only one preferred model and more than 2 individuals have given them as responses (binomial test with likelihood 2^8).

Honorable Mentions Here, we briefly report on interesting findings among the other syllogisms that did not have a clearly preferred mental model. Starting with AA1, which had a tie for a preferred model - 24% of participants created an "XYZ" representation, and another 24% added the entity "¬X¬Y¬Z" to it. In other words, one part of the population represented only the terms they were presented with, whereas the other part made a point to include terms not mentioned at all, as an offset. For EA4, 21% of the participants created an "¬XYZ" representation and other 21% added the instances "X¬Y¬Z" and "¬X¬Y¬Z" to it. Namely, the second group explicitly represented that "No Y are X", but both X and Y can exist without the other one. For the rest of the syllogisms, no particularly interesting patterns were found – there are no preferred mental models for them.

mReasoner

When building a model, two of mReasoner's parameters are relevant: λ - which controls the number of instances in the model and ε - which determines the likelihood that the model representation is constructed with instances from the full set in contrast to only canonical ones (Khemlani et al., 2015). In Table 1 we show the canonical and noncanonical instances that can be drawn from the sets, according to the LISP implementation of mReasoner.

First, we looked into the instances of the participants' representations in terms of canonicality. Based on the amount of noncanonical models, we derived which ε value would be used according to mReasoner's postulates. For example, for the premise "All circles are red", if we have 8 instances of "circles red" and 2 instances of "diamonds red", following Table 1, we have 8 canonical and 2 noncanonical entities, out of 10. That means that the assigned[4] ε value would be 0.2. The distribution of obtained ε values is shown in the left-most barplot in Figure 5. We did not find any correlation between the assigned ε values and the correctness of responses (Spearman's $r = -.0343, p = .1702$).

[4]Please note that mReasoner's ε value is a *likelihood* - what we assign is a value based on the proportion of noncanonical instances we observe in one specific individual outcome, not in terms of probabilities.

Afterwards, we fit mReasoner to all task response pairs using a grid-search to determine the parameters. For ε, we used values in the range of 0 to 0.9 with steps of 0.1. While the maximum value for ε is 1.0, it was omitted since mReasoner frequently fails the model creation phase. The parameters associated with the search for counterexamples (σ and ω) had a less fine-grained stepsize of 0.25. With respect to λ, which controls the size of the constructed model, we used two different approaches:

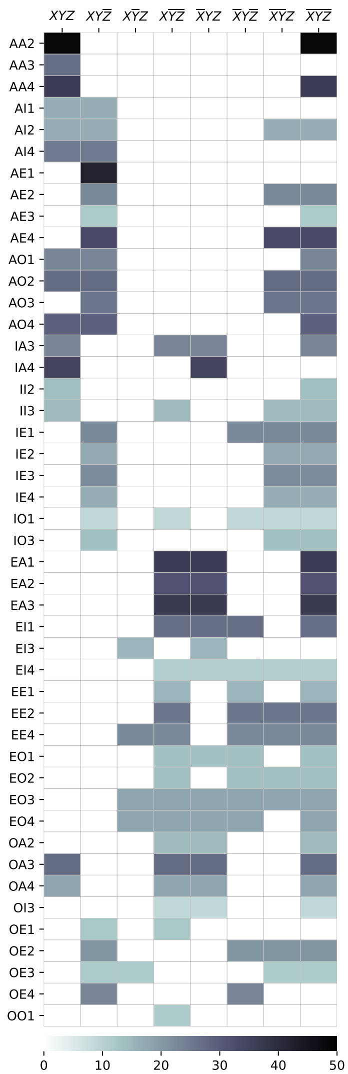


Figure 4: Preferred models provided by the participants for different syllogisms. Only syllogisms with a unique preferred model are shown. The columns denote possible instances present in the provided model. Shading illustrates the proportion of participants creating a model containing the respective instances.

First, we bypassed the λ-parameter and instead “forced” mReasoner to create exactly 10 instances to reflect the experimental setup. Furthermore, we ensured that the constructed instances precisely reflect the value of ε-parameter (i.e., instead of using ε as the probability to draw from the full set including non-canonical interpretations, it now defines the proportion of non-canonical instances). This provides an opportunity for a direct comparison between the participants' representations and mReasoner's created models, especially in terms of parameter values. The distribution of the proportion of non-canonical instances in the models created by participants as well as the distributions of ε is shown in the first and second barplot of Figure 5, respectively. Visually, the distributions seem to differ substantially with little common trends observable. This is supported statistically, since no significant correlation between the distributions was found (Spearman's $r = .0058, p = .0694$).

Quantifier	Canonical	Noncanonical
All	X Y	¬X Y ¬X ¬Y
Some	X Y X ¬Y	¬X Y ¬X ¬Y
No	¬X Y X ¬Y	¬X ¬Y
Some not	X Y X ¬Y ¬X Y	¬X ¬Y

Table 1: Canonical and noncanonical instances for a syllogistic premise with terms X and Y according to mReasoner (Khemlani et al., 2015)

Second, we fitted mReasoner with an active λ parameter fitting, obtaining results with the intended configuration and thereby eliminates potentially introduced problems due to our manipulation. Additionally, participants might not use all 10 instances to reason about the conclusion, even if the scenario suggests it. Here we report the distribution of the best-fitting ε values (for any λ) in the third plot of Figure 5, followed by the distribution with “estimated” ε values based on the actual proportion of noncanonical instances (ε_{est}). Thereby, ε_{est} resembles the same interpretation as the ε-values in the previous scenario with a fixed size. Note that in same cases, ε_{est} can still have the value 1.0, since it reflects the actually created model and not the likelihood. We did not find correlation neither between our assigned ε values and mReasoner's ε (Spearman's $r = .0330, p = .2980$) nor with ε_{est} (Spearman's $r = .0576, p = .0720$). We ensure that forcing mReasoner to work with exactly 10 instances does not influence the ε distributions (Spearman's $r = .7564, p < .0001$). It is important to note however that there are multiple potential ε values that could be used for fitting to a task response pair. In the case of a fixed model size, there were on average 6.1 values leading to the same response, while there are 6.6. for the regular approach (out of 10 possible values for ε in the grid-search).

Discussion

In this paper we investigate two research questions regarding the mental model building process in syllogistic reasoning. For **RQ1** we examined what kind of mental models individuals create when presented with syllogistic premises. Towards that we designed and conducted an experiment centered around an imaginary world of colourful shapes with marks, where participants had to provide their visual representation of syllogisms first, and afterwards gave their conclusions. We noted a tendency for a belief bias effect in their conclusions. Namely, this suggests that an individual might be hesitant to conclude NVC, when some background knowledge regarding the existence other shapes might go against it. This is of interest for potential investigations of belief bias effect in a controlled content environment. Regarding the mental models, 82% of them were correctly representing what is stated in the syllogistic premises, indicating a general ability to correctly interpret them, and no particular syllogistic property was found to affect the correctness. We found preferred mental models for 46 out of 64 syllogisms, some occurring within a larger proportion of participants than others. There is a noticeable tendency among syllogisms with an A-premise to include noncanonical instances with terms that were not presented at all, likely due to them being an easy addition without introducing errors. We note a weakness in the PMMs for syllogisms with particular quantifiers (I, O) – though a preferred model was found, it was a smaller proportion of participants, i.e. their interpretation is rather varying. This could be associated with the quantifier's low informativeness allowing for more possible models without a clear preference. This in turn might be a reason for a lower confidence in an individual's interpretation, which is a proposition by another prominent syllogistic reasoning model - the Probability Heuristics Model (PHM; Chater & Oaksford, 1999; Oaksford & Chater, 2001).

For **RQ2** we looked into the canonicality of the individuals' mental models, whether that ties into their responses and ultimately whether the observed behavior is in line with the model building process of mReasoner. In order to quantitatively analyze the canonicality of the models, we leaned on mReasoner's canonicality parameter, ε. We contrast correlation analyses between response correctness and a) ε values assigned based on observed noncanonicality proportions; b) fitted mReasoner ε values on task responses, with "forced" 10 instances and with the regular intended configuration. We did not find any significant correlation in any scenario, pointing to a potential lack of relevance of the models for the responses. On the other hand, another reason might be that we cannot confirm with confidence that the built models in the experiment were indeed used for the reasoning portion of it. In MMT, the model building process is rather important, however in the mReasoner implementation (and our grid-search when fitting), we have more than 6 values out of 10 that can be used on average. This leads to the question if having two parameters for the model building process is really necessary, from a complexity perspective. However, one of mReasoner's assumptions is that humans build correct representations, which is mostly in line with our observations. In case of errors, a potential source can be incomplete representation, which is also in line with our observations. As an example, we take the syllogism AA4 and its preferred mental model that consists solely of "XYZ" and "¬X¬Y¬Z" meaning that no instance supports the logically correct conclusions, *Iac* and *Ica*. In order for an individual to conclude NVC, only one single model is not sufficient in order to deduce that there is a contradiction. There are two possibilities, we either use some additional processes (e.g. heuristics, search for counterexamples) or we create and test multiple models. By definition, mReasoner does not build two models for NVC. In the initial phase of model building, it assumes a correct construction and then uses epsilon to draw the exact instances. Later on, the initial representation is manipulated to e.g. add counterexamples and enable the conclusion of other responses.

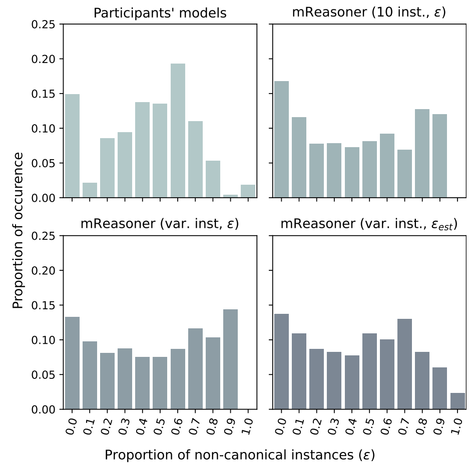


Figure 5: Distribution of ε (proportion of noncanonical instances) of the instances directly provided by participants and fits of mReasoner to their responses. For mReasoner, distinctions between a fit with 10 instances and a variable number of instances are made. In the case of a variable number of instances, the ε-parameters used by mReasoner and the estimated ε based on the resulting models is shown.

To summarize, we can conclude that while individuals do have preferred mental models for a large portion of syllogisms, the initially built mental models are not substantial for finding conclusions. It is very likely that this is due to syllogisms being generally imbalanced in terms of validity, meaning that the majority of them can not be solved straightforwardly with an initial model anyway. It is, however, important to know that even when a final representation might be incomplete, its instances are still appropriately chosen in line with the premises. In contrast to manipulation of an existing model, as proposed and implemented by mReasoner, the model building phase seems to be a rather easy task for humans, so certainly, a plausible way to solve the tasks would in fact be a repeated construction of models.

References

Brand, D., Riesterer, N., & Ragni, M. (2022). Model-based explanation of feedback effects in syllogistic reasoning. *Topics in Cognitive Science*, *14*(4), 828-844. doi: https://doi.org/10.1111/tops.12624

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

Jahn, G., Knauff, M., & Johnson-Laird, P. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*(8), 2075–2087. doi: 10.3758/BF03192939

Johnson-Laird, P. (1975). Models of deduction. In R. J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults* (pp. 7–54). New York, US: Psychology Press.

Johnson-Laird, P. (2006). *How We Reason*. Oxford, New York: Oxford University Press.

Johnson-Laird, P. (2010). Mental models and human reasoning. In *National academy of sciences* (Vol. 107, pp. 18243–18250).

Khemlani, S., & Johnson-Laird, P. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*(3), 427–457.

Khemlani, S., & Johnson-Laird, P. (2013). The processes of inference. *Argument & Computation*, *4*(1), 4–20.

Khemlani, S., & Johnson-Laird, P. (2016). How people differ in syllogistic reasoning. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *38th annual conference of the cognitive science society* (p. 2165–2170). Austin, TX: Cognitive Science Society.

Khemlani, S., Lotstein, M., Trafton, J. G., & Johnson-Laird, P. N. (2015). Immediate inferences from quantified assertions. *The Quarterly Journal of Experimental Psychology*, *68*(10), 2073-2096. doi: 10.1080/17470218.2015.1007151

Oaksford, M., & Chater, N. (2001, aug). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, *5*(8), 349–357.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*(3), 561–588. doi: 10.1037/a0032460

Rauh, R., Hagen, C., Knauff, M., Kuss, T., Schlieder, C., & Strube, G. (2005). Preferred and Alternative Mental Models in Spatial Reasoning. *Spatial Cognition and Computation*, *5*, 239–269. doi: 10.1080/13875868.2005.9683805

Störring, G. (1908). *Experimentelle untersuchungen über einfache schlussprozesse*. W. Engelmann.