Modeling Transfer and Learning Effects in Syllogistic Reasoning Using Automatically Generated Process Models

Daniel Brand (daniel.brand@metech.tu-chemnitz.de)
Clemens Bombach (clemens.bombach@hsw.tu-chemnitz.de)
Marco Ragni (marco.ragni@hsw.tu-chemnitz.de)
Predictive Analytics, Chemnitz University of Technology, Germany

Abstract

Building cognitive process models for reasoning traditionally involves hand-crafting them from underlying theories and assumptions, thereby limiting comparability and transfer between competing theories. To address this, we present a novel approach that automatically constructs process models: Process models are treated as a sequence of cognitive actions/operations that are extracted from cognitive theories and represented in a unified framework while preserving the explanatory merits of the underlying theories. Our method then searches for an optimal sequence to fit the observed reasoning behavior, providing insights into inter-individual differences on the level of reasoning processes. We apply our approach to the domain of syllogistic reasoning and use it to obtain insights into which processes and parts of state-of-the-art models account best for individual reasoning behavior. We use a dataset including other reasoning domains as well as the progression of syllogistic reasoning performance over time. This allows us to investigate how processes and components of state-of-the-art models correspond to an individual's performance in related domains and how they can account for learning effects. Finally, we discuss the potential of our approach for cognitive modeling: First, by utilizing a unified framework for cognitive actions, we can obtain process models combining the best components of state-of-the-art models to account for observed phenomena. Second, it allows us to relate specific processes better to other domains of reasoning, advancing a unified understanding of reasoning. Finally, in-depth insights into the respective theories and models are gained in an objective manner that facilitates the development of cognitive process models.

Keywords: Reasoning; Syllogisms; Process Models; Individual Differences

Introduction

At the core of cognitive modeling lies the motivation to encapsulate facets of the processes driving our minds into a simple, understandable model that still maintains a grasp on the relevant principles and mechanisms. Out of all abilities of the human mind, the ability to reason about information is undoubtedly one of the most fundamental abilities of our intelligence. Consequently, it was researched extensively, leading to a variety of mechanisms, processes and strategies being identified (e.g., see Manktelow, 1999).

In the spirit of Richard Feynman's famous quote "What I cannot create, I do not understand", cognitive models have the important role of concretizing and compiling the findings into testable accounts for the observed behavior. Thereby, cognitive modeling of reasoning was traditionally revolving around process models, which make assumptions about the underlying mechanisms on an algorithmic or process-oriented level. With the uprising of Bayesian models, however, research moved towards probabilistic models as powerful accounts for the observed behavior. While successful in discovering statistical relationships, this has been criticized due to their lack of explanatory power on the level of mental processes (e.g., Fugard & Stenning, 2013).

Conversely, cognitive process models come with the problem of incompatibility among each other. Due to their more complex and algorithmic structure, different process models rarely share comparable structures, making it difficult to unify findings from the different underlying theories. Especially in reasoning, where higher level processes are involved, specific accounts for strategies are also unlikely to be generalizable across individuals: inter-individual differences in strategies may differentiate so substantially that an *average* reasoner with a respective reasoning process may not exist (Riesterer et al., 2020b).

In recent work (Brand et al., 2025), we approached this problem by re-thinking cognitive process modeling as an optimization problem: When treating sub-processes and effects as modular actions, creating a process model explaining an observed behavior to a task corresponds to finding a sequence of those actions that generate the observed behavior adequately when applied to the task at hand. That way, process models that fit individual participants best can be automatically generated, while preserving most of the explanatory value of the actions (i.e., the respective sub-processes).

Our work uses syllogistic reasoning as an exemplary domain, which is one of the oldest and best-researched domain of reasoning (for an overview, see Khemlani & Johnson-Laird, 2012). Syllogisms are tasks requiring reasoning over two quantified statements as illustrated in the following example:

Some A are B.
All B are C.

What, if anything, follows?

The two statements (*premises*) connect two end-terms via a middle term present in both premises. The task is traditionally to derive a conclusion about the relationship between the two end-terms or conclude that *no valid conclusion* (NVC) exists. The arrangement of the terms is referred to as the *figure* of the syllogism, and is one of the following four figures (in the paper, we use the same notation as Khemlani and Johnson-Laird, 2012):

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

The statements use first-order logic quantifiers and can be abbreviated by a single letter: *All* (A), *Some* (I), *No* (E) and *Some not* (O). By combining the quantifier and the figure, the syllogism in the example would thereby be abbreviated by IA1.

In this work, using data from work by Dames et al. (2022), who investigated the relation to other reasoning tasks and re-test effects in syllogistic reasoning, we use our method to investigate those on process level. Thereby, the structure is as follows: First, we introduce the dataset our work is based on as well as our method and the respective actions derived from the state-of-the-art. Second, we assess the performance of the generated models compared to cognitive models and investigate how the changes in performance over time can be explained by the identified differences in (sub-)processes. Third, we relate the differences in the individual process models to the general performance of those participants in other reasoning tasks. Finally, the results are critically discussed and put into a general perspective for cognitive process modeling of reasoning.

Method

Dataset

For our analysis, we use a publicly available dataset by Dames et al. (2022) that tested all 64 syllogisms twice to investigate retest effects. Additionally, other tasks were administered. Besides questionnaires assessing the personality traits, reasoning-related tasks were also included, which we will consider in this work. These consisted of the *Cognitive Reflection Test* (CRT; Frederick, 2005) with additional questions by Toplak et al. (2014), conditional reasoning tasks (e.g., see Singmann et al., 2016), a short version of the Advanced Progressive Matrices (referred to as Raven in the remainder; Bors & Stokes, 1998), the CORSI block tapping task (Corsi, 1972) and a operation span test (e.g., see Turner & Engle, 1989). Overall, the task contains complete responses of 106 participants.

Process Model Generation

We utilize a recently proposed method (Brand et al., 2025) to automatically generate process models. At its core, the method revolves around the idea of reinterpreting cognitive modeling itself as an optimization problem: In experiments, a set of tasks X is tested, and for each $x \in X$ a respective response behavior y_p of the participant is observed. A cognitive process model predicting the observed behavior consists of actions $(a_1, \ldots, a_n)$ that generate a prediction y_a from x, such that each action describes a potential cognitive (sub)process. Although it is surely up to debate which actions can be considered as such, we consider commonly observed effects and assumed processes from the state of the art as valid actions.

Whereas for a single task $x \in X$ with a corresponding response y the problem can be solved by an exhaustive search for an action sequence leading from x to y, this is no longer the case when considering a full response pattern across all tasks. In this case, the problem becomes an optimization problem, where some actions/processes may solve certain tasks well but introduce errors for others. Therefore, an optimal selection of actions has to be found by maximizing a suitable objective function.

To structure the problem and create a unified framework that can incorporate as many findings, theories and models as possible, we identified three types of actions that are prevalent in the state-of-the-art for modeling human reasoning. They correspond to three phases:

1. *Task Processing* contains actions that represent cognitive processes dealing with interpreting the task itself and its meaning. They essentially realize a function $a(x) = x'$, mapping a task x to another task x' (e.g., by interpreting "Some B are A" as "Some A are B" in the case of syllogistic reasoning) and can be considered as preprocessing steps.

2. *Response Candidate Generation* contains actions that build the initial mental representations and respective solution candidates (i.e., conclusions in the case of syllogistic reasoning). Those candidates can be partial or complete solutions. In the case of partial solutions, other actions have to be used within this phase, until complete solution candidates are derived (e.g., in syllogistic reasoning, the quantifier may be decided by one action, while subsequent actions derive a direction). Actions in this phase realize a function $a(x) = (x, y_a)$, where y_a is a, potentially incomplete, representation of response candidates.

3. *Response Candidate Alteration* contains actions that work on already existing response candidates and reason over them again in the context of the task. They realize a function $a(x, y) = (x, y')$, where y' is an altered version of y. In the case of syllogistic reasoning, such actions can, for example, be entailments (e.g., p-entailment; see Chater & Oaksford, 1999; Oaksford & Chater, 2001) or tests of the conclusions (e.g., the

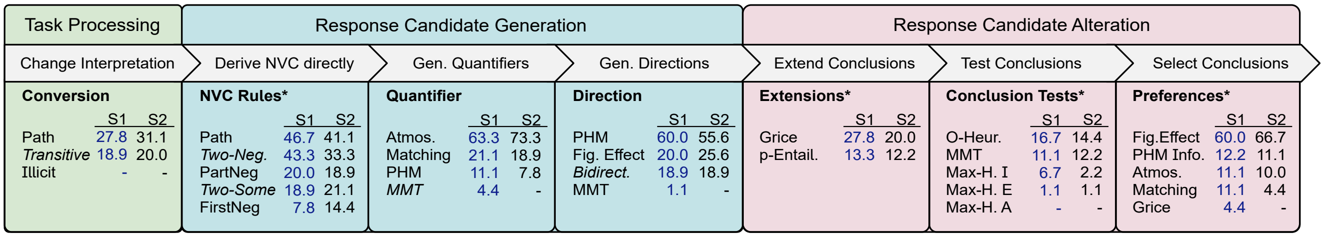

Figure 1: Schematic overview of the pipeline used to find the process model for an individual reasoner in the syllogistic reasoning domain. On top, the general phase is shown, below the stages that are specific for syllogistic reasoning with the respective actions/sub-processes. Stages denoted with an asterisk can have multiple actions/sub-processes active. Percentage values denote the ratio of occurrences in the global model fits in both sessions. Processes in italic reconstruct the logically correct solutions the best.

search for counterexamples proposed by Khemlani & Johnson-Laird, 2013).

In the case of syllogistic reasoning, we identified additional stages within each phase, which simplify the optimization process by constraining the search space. Figure 1 illustrates the pipeline with the respective stages and actions that we isolated and modularized from the current state-of-the-art. In the following, we briefly introduce the respective actions.

Task Processing Actions

The task processing actions we implemented are based on the idea of *conversion*, which means that reasoners re-interpret the task in a fashion that simplifies the subsequent reasoning process. The first conversion, the *Path* heuristic, is directly adopted from *TransSet*, which proposes that reasoners convert syllogisms without a transitive path (i.e., figure 3 & 4) into ones with a transitive path by changing the direction of one of the two premises that possess a universal quantifier (with priority $A > E$; Brand et al., 2020). The two remaining, *Transitive* and *Illicit* conversion also aim to establish a transitive path, but with a preference towards figure 1. In the case of the illicit conversion (see Revlis, 1975), it may introduce reasoning errors when the direction of a universal and positive or a particular and negative premise is altered. For *Transitive*, only logically valid conversions are carried out.

Response Candidate Generation

Responses in syllogistic reasoning are of the form "All A are C", "Some C are not A", described by one of the four quantifiers and a direction (A-C or C-A). Furthermore participants may respond that no valid conclusion may be drawn (NVC). Therefore, the response candidate generation can be separated into a direct test for NVC, a quantifier generation step and a direction generation step. Since 37 of the 64 syllogisms only have NVC as a valid response, it is essential to include processes that lead to NVC responses. In the response generation phase, this is reflected by five distinct *NVC rules* that attempt to derive NVC directly from the (possibly converted) task description. All of these rules are derived from TransSet. For instance, the *Two-Neg* and *Two-Some* rules conclude NVC if both quantifiers are negative or particular respectively.

Quantifier selection rules were derived from the Atmosphere theory (Woodworth & Sells, 1935), Matching heuristic (Wetherick & Gilhooly, 1995), Probability Heuristics Model (Chater & Oaksford, 1999) and Mental Models Theory (Johnson-Laird, 1983; Johnson-Laird & Bara, 1984). For instance, the Atmosphere heuristic can be described as follows: The conclusion quantifier is universal if and only if both premise quantifiers are universal. The analogous rule holds for positivity. In a similar fashion, we obtained direction generation rules from the PHM and MMT theory. Moreover, we implemented the *Figure Effect* (Dickstein, 1978) which selects an A-C direction for figure 1 syllogisms and C-A direction for figure 2 syllogisms. The *Bidirectional* process simply denotes that both directions are selected during response candidate generation.

We note that conclusion generation in MMT is somewhat more involved since it relies on mental models as an intermediate representation. Therefore, we relied on the prediction table given in Khemlani and Johnson-Laird, 2012 to determine the set of quantifiers and directions that can follow from a given premise.

Response Candidate Alteration

Processes that alter generated responses can again be divided into three stages: *Extensions* act by adding conclusions to the responses generated in the previous step. An example is the *Grice* implicature which consists of inferring "Some A are not C" from "Some A are C" and vice versa. The *p-Entailment* heuristic is derived from PHM. The second stage consists of *conclusion tests* that filter conclusion candidates to model, for instance, a search for counterexamples as in MMT. This can possibly lead to a NVC response. The other conclusion tests are derived from PHM: For instance, the *O-Heuristic* of PHM states that reasoners prefer not to conclude responses involving the "Some...not" quantifier due to the low informativity of the resulting statement. Finally, we included *preferences*. These also act as filters, but do not lead to a NVC conclusion. One example is the *Gricean maxime of quantity* (Grice, 1975) which states that reasoners prefer universal to particular responses. Thus, for instance, if "All A are C" and "Some A are C" were both included the responses previously, the latter one is deleted. The other preferences are derived from the Figure Effect, PHM, the Atmosphere theory and the Matching heuristic.

Evaluation

First, we fitted process models for every participant on all 64 tasks optimizing accuracy of the reconstruction of the respective participant's responses. We included PHM, MMT and TransSet using openly available implementations (Brand et al., 2020; Riesterer et al., 2020a) for comparison, since most actions were adopted from these models. Additionally, a random model as the lower baseline as well as the most frequent answer (MFA) is included as the best strategy when no adaption to an individual reasoner is performed. Since the behavior of participants already seeming to know (formal) strategies for syllogistic tasks can not be attributed to any specific reasoning process (i.e., it may also be explicitly learned rules, knowledge about the tasks, etc), those that are best described by pure first-order logic in the first session were excluded for the evaluation of second session ($n = 15$).

Second, we analyze the resulting sequences of actions of the respective generated process models. For those, we assess if they capture the findings by Dames et al. (2022): An improved performance was reported in the retest, but mostly for valid syllogisms. Dames et al. (2022) propose two explanations for this: first, it can be due an increased usage of simple heuristics like the Atmosphere theory (Woodworth & Sells, 1935), which do benefit valid but not invalid syllogisms. This especially showed in increased O-responses ("some... not") in the second session. Second, aversions to NVC (e.g., by interpreting it as "giving up"; see also Ragni et al., 2019) may have caused participants to find efficient strategies for non-NVC responses, but not for the invalid syllogisms that require NVC. Put together, we expect to find an increase in heuristics like the Atmosphere theory, and a decrease in mechanisms to derive NVC. Especially the O-Heuristic as well as Max-Heuristics of PHM would be in line with the proposed explanation.

Finally, we investigate the relation between the identified processes and the performance in the other reasoning tasks. Here, especially CRT and Raven performance were found to be related to syllogistic reasoning performance. Conditional reasoning, on the other hand, was found to correlate with syllogistic reasoning performance by Brand and Ragni, 2025. For this, we use the k-means clustering ($k = 2$) on the configurations of the process model pipeline (see Figure 1) identified for each individual participant to find two centroids (i.e., two prototypical process models). The assignments of participants to the respective centroids are then related to the respective scores in the other reasoning tasks. For this analysis, we use the data of the first syllogistic task session in order to assess the transferability of other reasoning domains to the initial reasoning behavior in syllogistic reasoning (rather than learning effects). All materials and analysis scripts are available publicly[1].

In order to investigate whether the relation of processes to performance can be disentangled further, we used a linear regression model, applied to the data of the first session.. Our initial set of predictors was comprised of all processes and performance on the syllogistic reasoning tasks, with an aim to predict CRT, Raven and conditional reasoning performance. We included syllogistic reasoning performance as a control variable to see whether the processes add predictive performance beyond the correlation between raw performance scores. Since the design matrix we obtained exhibited multicollinearity between the features, we used a linear regression with both ℓ^1 and ℓ^2 penalty (ElasticNet) for regularization and feature selection. For the selected coefficients, we calculated variance inflation factors (VIF) to measure the remaining multicollinearity and removed process variables with high VIF. After this feature selection step, we ran (unregularized) ordinary least squares regression with the selected features onto the target features. The results were compared to a baseline model that only used syllogistic correctness as a predictor. The regularization parameters were selected via leave-one-out cross-validation.

Results

When fitting our method to the participants of both sessions, we obtained individual process models for each individual participant. The models' accuracies are shown in Figure 2, where it becomes apparent that our method was able to find configurations of the pipeline that are better able to account for the observed reasoning behavior than state-of-the-art cognitive models on their own. This indicates that none of the individual cognitive models currently is an optimal account, despite TransSet being relatively close to the performance of the generated process models (62.4 vs 59.4 for the initial syllogistic task session and 63.3 vs 61.9 for the retest session, respectively). Both models thereby surpass the most frequent answer (MFA; 55.2 and 56.7, respectively) in both sessions, indicating that they successfully capture at least some individual traits.

[1]https://github.com/brand-d/iccm-2026-autoprocessmodels

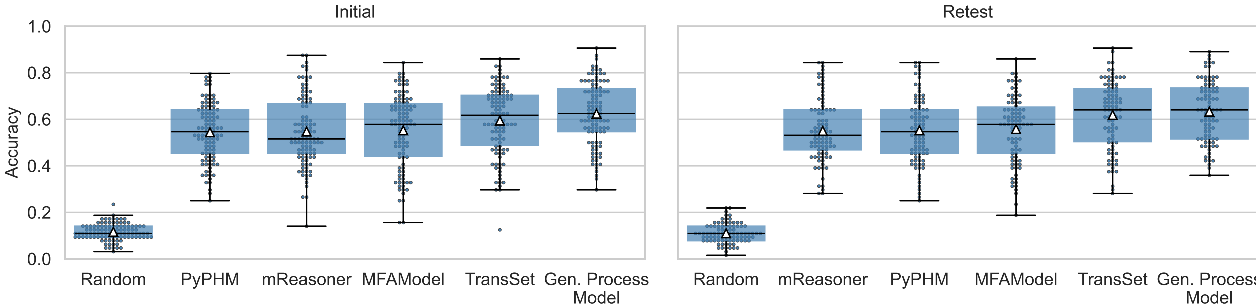

Figure 2: Accuracy of cognitive models and the automatically generated process model when replicating the response patterns in both sessions. Boxplots show inter-quartile ranges and medians and the respective mean values are denoted by triangles. Points denote accuracies for individual participant.

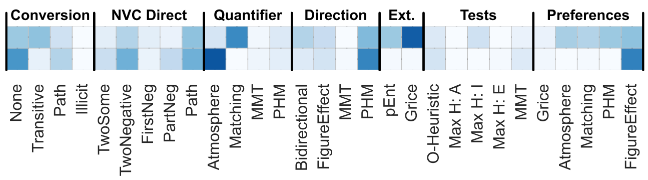

Figure 3: The two clusters found by the k-means algorithm, represented as rows in the heatmap. Both centroids correspond to configurations for the pipeline (see Figure 1). Darker shades denote higher values.

Process Model Analysis

Figure 1 shows the proportion of the actions being selected for both sessions (S1/S2). Note that an action for both quantifier and direction always needs to be selected, while a conversion action is not mandatory. Furthermore, the remaining stages can have none, one or multiple actions selected. An interesting finding to note for both sessions is that participants adhered strongly to the Figure Effect as their final decision step (i.e., by having to select a single response), but not as the initial direction method.

Overall, the changes between the two sessions are subtle, with some of the more notable ones being an overall decrease of NVC rules and conclusion tests, a further shift towards the Atmosphere theory for quantifiers and a reduction of extensions overall. Those results are confirming the findings of Dames et al. (2022) on a process level: There seems to be a refinement of the NVC strategies, reducing errors caused by responding NVC to a valid syllogism. Thereby, the proposed increase in heuristics (e.g., Atmosphere) can be supported by our method. However, given the accuracy of around $60-65\%$ accuracy of our method, it is important to note that such subtle differences have to be treated with caution. The overall change in performance between sessions in the data set is about 2.6%, which is hard to grasp reliably on the level of response behavior.

Table 1: Rank biserial correlation of the participants' performance in reasoning tasks to their assignment to one of the two clusters in Figure 3. Positive effect sizes denote a correlation with the second (lower) centroid. All p-values are corrected using the Bonferroni-Holm-correction and significance is marked by an asterisk.

Task	U	r_{rb}	p
Conditionals*	466.5	.473	$< .001$
CORSI*	528.5	.402	.005
CRT*	617.0	.302	.040
Operation Span	784.0	.114	.388
Raven*	447.5	.494	$< .001$
Syllog Correct*	308.5	.651	$< .001$

Relation to other tasks

We clustered the individual process models (i.e., the selected actions in the pipeline) into two clusters using k-means clustering of the first session. The resulting centroids are shown in Figure 3. It becomes apparent that, indeed, little differences between participants exist with respect to their NVC processes. This is surprising, since NVC is reported to be one of the main inter-individual differences (e.g., Brand & Ragni, 2025; Ragni et al., 2019) and also was identified as one of the major factors, as well as an effect of feedback, leading to improved performance (e.g., Brand et al., 2022) with participants usually giving too few NVC responses in the beginning. It is, however, consistent with the reported results by Dames et al. (2022), where NVC responses were already consistently given in the first session, thereby indicating that NVC processes seem to be present across participants. On the other hand, main differences show in the quantifier selection (Matching vs. Atmosphere), and the utilization of Extensions like Grice and p-Entailment. Finally, more mixed preference/filter processes are present in the first of the clusters, whereas the second cluster distinctively uses the Figure Effect only. Given the more distinct processes of the second cluster, and the presence of the logically not warranted Grice extension in the first cluster, we assume that the second cluster corresponds to participants with higher overall reasoning performance.

Table 1 shows the correlation between the cluster assignments and performance in the other reasoning tasks. Indeed, it shows that there was a significant correlation between an assignment to the second cluster and performance in conditional reasoning ($r_{rb} = .473$, $p < .001$), the CORSI block tapping task ($r_{rb} = .402$, $p = .005$), the cognitive reflection test (CRT; $r_{rb} = .302$, $p = .04$) and the Raven test ($r_{rb} = .494$, $p < .001$). However, when taking the syllogistic correctness itself into account ($r_{rb} = .651$, $p < .001$), it becomes apparent that the findings most likely simply reflect the respective correctness, which has the strongest relation to the respective process model configuration. This in line with findings by Brand and Ragni (2025), in which results between reasoning domains were transferable, but barely beyond the level of performance/correctness.

Regression Analysis

Considering the regression analysis, it is important to note that p-values and performance measures such as R^2 may be too optimistic due to the effect of the preliminary feature selection procedure. Thus, the results should be interpreted in a purely descriptive manner. The regression analysis was carried out to investigate the relationship of syllogistic reasoning performance and the identified processes on performance on other reasoning tasks such as conditional reasoning, CRT and Raven intelligence test. For the conditional score ($R^2 = 0.374$) and CRT score ($R^2 = 0.161$), we found no significant effects beyond correctness. However, regarding the Raven score, we observed that the R^2 almost doubled compared to the baseline model ($R^2_{\text{full}} = 0.412$ versus $R^2_{\text{base}} = 0.211$). Moreover, of the 10 selected processes, we found significant effects for the Grice implicature ($\beta = -0.13, p = 0.012$) and Matching heuristic ($\beta = 0.19, p = 0.004$) in addition to a significant effect of syllogistic correctness ($\beta = 0.33, p = 0.012$). In connection with the clustering analysis, this suggests that use of Grice implicature both distinguishes reasoners and may have some relation to performance.

Discussion

Our method was able to account well for the observed behavior of individuals, as apparent by surpassing the accuracy of relying on the most frequent answer as well as state-of-the-art cognitive models for syllogistic reasoning. We believe that the utility of our approach lies in its ability to map the large space of possible reasoning patterns onto a more manageable space of theory-derived cognitive processes. This allows cognitive scientists to study individual differences in reasoning in terms of processes, instead of relying on statistical relationships only. To this end, we could support the central findings by Dames et al. (2022), finding the subtle differences between both experimental sessions (initial test and retest) reflecting in the processes. As hypothesized by Dames et al. (2022), we indeed found an increase in the Atmosphere theory as well as the decline of the Gricean implicature extension, which can be understood as a refinement of strategies and explain the reported improved performance on valid syllogism.

For our clustering analysis on the selected processes we split participants into two groups. The resulting two groups coincided with lower and higher reasoning performance well, indicating that it was the main factor behind the inter-individual differences. This is supported by the significant correlations between the clusters with other reasoning tasks. With this, the analysis supports findings by Brand and Ragni (2025), who reported that a general reasoning ability that is transferable across domains indeed appears to be present, but that the information obtained from related domains is often limited to a split in "correct" and "incorrect".

Of note is the role played by the Gricean implicature in the lower performing cluster, suggesting that its use could be avoided by participants with higher reasoning ability who identify the potential fallacy. This is also reflected in our regression analysis of the Raven score onto the underlying processes, where it also was one of the only related processes.

We agree with Fugard and Stenning (2013) that process modeling should remain integral for cognitive modeling of reasoning and argue that our method has shown to be a useful tool to gain additional insights into observed reasoning patterns on the level of processes. As such, it can be an addition to a modelers toolkit, streamlining analyses with process models, extending and improving them, easing the transition to modeling individual reasoning behavior, and facilitating the transfer of findings between different theories and models. Thereby, it can be one step towards bringing process modeling back to the heart of modeling human reasoning.

Acknowledgements

The authors thank the German Research Foundation (DFG) for financial support via grants awarded to MR within the projects 283135041 and 529624975. DB was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 529624975. CB was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) -– Project number 283135041.

References

Bors, D. A., & Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58*(3), 382–398. https://doi.org/10.1177/0013164498058003002

Brand, D., Bombach, C., & Ragni, M. (2025). Data-driven construction of individualized process models for human reasoning. https://doi.org/10.31234/osf.io/d2smc_v1

Brand, D., & Ragni, M. (2025). Using cross-domain data to predict syllogistic reasoning behavior. In D. Barner, N. R. Bramley, A. Ruggeri, & C. M. Walker (Eds.), *Proceedings of the 47th Annual Meeting of the Cognitive Science Society* (pp. 4370–4377).

Brand, D., Riesterer, N., & Ragni, M. (2020). Extending TransSet: An individualized model for human syllogistic reasoning. In T. C. Stewart (Ed.), *Proceedings of the 18th International Conference on Cognitive Modeling* (pp. 17–22). Applied Cognitive Science Lab, Penn State.

Brand, D., Riesterer, N., & Ragni, M. (2022). Model-based explanation of feedback effects in syllogistic reasoning. *Topics in Cognitive Science*, *14*(4), 828–844. https://doi.org/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. https://doi.org/https://doi.org/10.1006/cogp.1998.0696

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain.

Dames, H., Klauer, K. C., & Ragni, M. (2022). The stability of syllogistic reasoning performance over time. *Thinking & Reasoning*, *28*(4), 529–568.

Dickstein, L. S. (1978). The effect of figure on syllogistic reasoning. *Memory & Cognition*, *6*, 76–83.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25–42.

Fugard, A. J., & Stenning, K. (2013). Statistical models as cognitive models of individual differences in reasoning. *Argument & Computation*, *4*(1), 89–102. https://doi.org/10.1080/19462166.2012.674061

Grice, H. P. (1975). Logic and conversation. In *Syntax and Semantics. Vol. 3 : Speech Acts* (pp. 41–58). Academic Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, *16*(1), 1–61.

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

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

Manktelow, K. I. (1999). *Reasoning and thinking*. Psychology Press.

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

Ragni, M., Dames, H., Brand, D., & Riesterer, N. (2019). When does a reasoner respond: Nothing follows? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2640–2646). Cognitive Science Society.

Revlis, R. (1975). Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, *14*(2), 180–195.

Riesterer, N., Brand, D., & Ragni, M. (2020a). Do models capture individuals? Evaluating parameterized models for syllogistic reasoning. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 3377–3383). Cognitive Science Society.

Riesterer, N., Brand, D., & Ragni, M. (2020b). Predictive modeling of individual human cognition: Upper bounds and a new perspective on performance. *Topics in Cognitive Science*, *12*(3), 960–974.

Singmann, H., Klauer, K. C., & Beller, S. (2016). Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology*, *88*, 61–87. https://doi.org/https://doi.org/10.1016/j.cogpsych.2016.06.005

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, *20*(2), 147–168.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Wetherick, N. E., & Gilhooly, K. J. (1995). 'Atmosphere', matching, and logic in syllogistic reasoning. *Current Psychology*, *14*(3), 169–178.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, *18*(4), 451–460.