Title

Data-Driven Construction of Individualized Process Models for Human Reasoning

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Brand, Daniel
Bombach, Clemens
Ragni, Marco

Publication Date

2026

Data-Driven Construction of Individualized Process Models for Human Reasoning

Daniel Brand (daniel.brand@psychologie.tu-chemnitz.de)[1],
Clemens Bombach (clemens.bombach@hsw.tu-chemnitz.de)[1] &
Marco Ragni (marco.ragni@hsw.tu-chemnitz.de)[1]
[1]Predictive Analytics, Chemnitz University of Technology, Germany

Abstract

Over the past decades, human reasoning research has identified a variety of effects and processes, of which several have been compiled into comprehensive theories. Based on such theories, cognitive models were developed that made the theoretical findings applicable and testable. However, the models often consist of a variety on sub-processes and effects internally, but are not built in a modular way, hindering the transfer of findings between different models and their comparability. We approach this problem by proposing a different perspective: By treating the generation of cognitive process models as a search problem, process models can be derived from cognitive operations automatically in an objective way. Our method is illustrated on the domain of syllogistic reasoning, where we show that it generates a process model that outperforms state-of-the-art models while preserving their explanatory meaning. Finally, we discuss our approach as a framework for streamlining and facilitating cognitive modeling endeavors.

Keywords: Cognitive Modeling; Human Reasoning; Syllogistic Reasoning; Process Models

Introduction

Undoubtedly, the ability to reason and draw inferences is one of the most important abilities for the advanced life we have developed. Therefore, it is not surprising that research of human reasoning is a wide discipline that covers several aspects, encompassing reasoning over spatial relations, conditional statements, syllogisms, decision making and more. Across all those domains, research identified strategies, mechanisms and effects (e.g., see Manktelow, 1999). Findings were then refined into theories of (the respective type of) reasoning, which then in turn were used to create cognitive models from. In this paper, we use syllogistic reasoning as an exemplary domain to tackle a problem that is inherent to the other domains of reasoning as well: When theories and models are created, they only consider a subset of the processes and effects identified —and are often specifically tailored to those, making it hard to align different theories and transfer between them.

Dating back to the ancient Greeks, syllogisms are concerned with reasoning over two quantified statements as illustrated in the following example:

All samurai are warriors.
Some warriors are ninjas.

What, if anything, follows?

The two statements (*premises*) thereby connect two end-terms (*samurai* and *ninjas*) via a common middle term (*warriors*). The task is usually to derive the relationship between the two end-terms —or conclude that there is *no valid conclusion* (NVC). In traditional syllogisms, the statements use first-order logic quantifiers, which can be abbreviated by a single letter: *All* (A), *Some* (I), *No* (E) and *Some not* (O). Quantifiers may be classified according to universality or negativity: The quantifiers A and E are universal, I and O are particular. The quantifiers A and I are positive, whereas E and O are negative. Additionally, the arrangement of the terms is referred to as the *figure* of the syllogism, with the following notation used in this paper:

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

Combining figure and the abbreviations of the quantifiers, the syllogism in the example would be abbreviated by *AI1*. Likewise, conclusions can be abbreviated by combining a quantifier and *ac* and *ca* to denote the direction: "Some samurai are ninjas" would therefore be abbreviated by *Iac*. All combinations of figures and quantifiers finally build the full set of the 64 traditional syllogisms.

As one of the oldest reasoning domains to be researched (e.g., Störring, 1908), syllogistic reasoning has no shortage of theories and models, ranging from heuristic over model-based to logic-based ones (e.g., see Khemlani & Johnson-Laird, 2012). However, each of those covers only aspects of the whole spectrum, and not all parts being rigorously formalized or algorithmically specified. This leads to the proposed processes, effects and biases often not being transferrable across different theories. Thus, while the models can be evaluated in terms of their ability to account for the observed behavior (e.g., Khemlani & Johnson-Laird, 2012; Riesterer, Brand, & Ragni, 2020), the results can only make a statement about which model is the current best account, taken as a whole. In reality, the performance may hinge on small details though, of which some could potentially fit to other theories as well, but simply were not considered yet. Therefore, such evaluations give limited insight into what *would* be possible with the knowledge present in the state-of-the-art so far. While portfolios can approximate this (Ragni et al., 2021; Riesterer et al., 2018), they combine complete models and not their components.

In this work, we tackle this problem by going one step further from the idea of portfolios: We propose a unified pipeline

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

that serves as a generalized process model across the different theories. Our method then automatically finds a process model configuration accounting for given observations, essentially treating the modeling problem as a search problem. Thereby, it uses an objective, data-driven way while still preserving most of the explainability offered by cognitive models. In the following, we will introduce the dataset and describe the method in detail. Subsequently, we will evaluate it with respect to its ability to account for individual reasoning behavior, and to assess how much of the observed behavior can be covered by the current state-of-the-art. Finally, the results are discussed with respect to cognitive modeling as a whole.

Dataset

We used a dataset from previous work that contains the responses of 95 participants to all 64 syllogisms (Brand & Ragni, 2025). The response format was multiple-choice, so that participants had to select all conclusions that they believed to follow from the premises, which has been shown to be a response format that provides more information (and does avoid forcing participants to potentially make an arbitrary choice) while preserving the commonly observed effects (see Brand & Ragni, 2023). The syllogisms were presented using hobbies and professions as contents to avoid biases. The dataset is openly available to support modeling endeavors and contains behavioral data for other reasoning domains as well. All materials, scripts and the model implementations of this work are also publicly available on GitHub[1].

Method

At the heart of our approach lies the idea of a general structure, in which the majority of findings from the current state-of-the-art can be accommodated. Thereby, the general idea is to see modeling of reasoning processes in a more abstract sense: Every process model assumes a certain sequence of steps (e.g., Brand et al., 2020; Johnson-Laird, 1983; Oaksford & Chater, 2001) and assumes specific operations or sub-processes for each step. By standardizing these steps, we allow for recombination of those steps, and can reformulate cognitive modeling itself as a search problem: In experiments, participants' response patterns are observed, essentially providing a list of task-response pairs. A respective cognitive process model now consists of a combination of sub-processes, effects and biases that replicate those patterns as closely as possible (i.e., trying to derive the same response for a given task as the participants). Cognitive modeling, thereby, is then the search of a suitable selection of those components, with the goal of maximizing the ability to replicate the observations. Based on the inherent similarity to a planning problem, as common in AI, we will refer to such sub-processes, effects and biases as *actions*, indicating that they are the steps necessary to derive a response prediction from the task itself.

[1] https://github.com/brand-d/2026-indiv-process-models

Types of Actions

We identified 3 types of actions that are sufficient to formulate a majority of the processes, effects and biases that are prevalent in the state-of-the-art.

1. **Task Processing**: The first type of actions is referring to any kind of processes that deal with interpreting the task itself and its meaning. In the case of syllogistic reasoning, this entails the understanding of the quantifiers, and potential transformations to make reasoning easier (while potentially introducing errors).
2. **Response Candidate Generation** actions build the initial mental representations and solution candidates. Those candidates can be complete, but can also be partial solutions instead (i.e., define certain features of the solution). In those cases, other actions of this kind have to complete the response candidates before subsequent actions can happen. In the case of syllogistic reasoning, such actions may conclude a quantifier, but not decide on a direction yet.
3. **Response Candidate Alteration** is the last type of actions, which derive the final responses (if not directly done before). Here, the conclusion candidates are tested, filtered, altered or even extended (e.g., by utilizing logical entailment). An example in the domain of syllogistic reasoning would be the deliberate search for counterexamples as proposed by MMT (e.g., Khemlani & Johnson-Laird, 2013).

Those three types of actions are suitable to capture most of the models in several domains of reasoning and correspond to general phases of a reasoning model. Naturally, especially heuristic models and models already consisting of modular sub-processes are easy to translate, while self-contained models with rather specific internal representations require more effort. It is important to note that, at this stage, it is expected that the correct formalization of the sub-processes is ensured by cognitive modelers. The three types of actions also correspond to phases, since actions of a specific type are only applicable at this stage. Once the state progressed to a subsequent phase, they will no longer be applicable. Thereby, the phases are meant to serve as a framework to guide cognitive modelers to implement sub-processes in. We argue that using intermediate (potentially partial) response candidates is a format that can be utilized by the vast majority of process models, acting as a common denominator.

Implementation for Syllogistic Reasoning

For syllogistic reasoning, we implemented a variety of actions originating from state-of-the-art theories and models, as well as several important effects and biases. We identified additional distinctions between actions within the same phase that are specific for syllogistic reasoning and the respective models. We therefore divided the three phases into additional *stages*. The resulting pipeline showing the stages within the phases as well as the specific actions within them is shown in Figure 1. In the following, the stages as well as the actions and their underlying cognitive theories are briefly introduced.

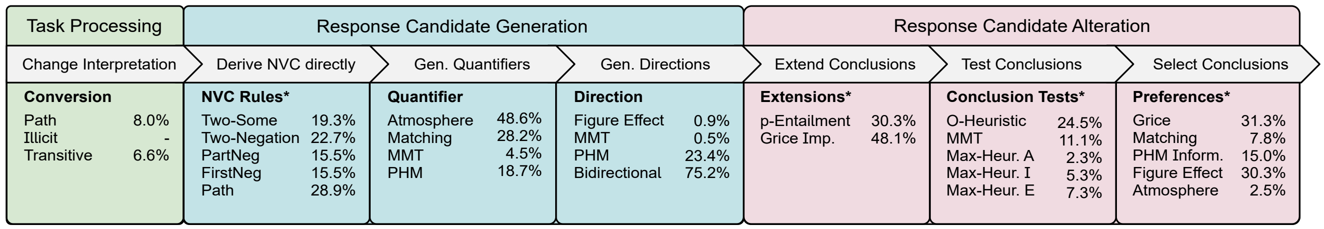

Figure 1: Schematic overview of the pipeline used to find the process model for an individual reasoner in the syllogistic reasoning domain. On top, the general phase is shown, below the stages that are specific for syllogistic reasoning with the respective actions/sub-processes. Stages denoted with an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

Note that our selection of actions is by definition not complete, but reflects, to our knowledge, the most relevant mechanisms of the current state-of-the-art.

Task Processing All of the findings we identified that fit into the first step of the pipeline are *Conversions*, i.e., different interpretations of the task itself. Thereby, a syllogism is simply mapped to another syllogism, which is then used in the following steps. We assume that only one conversion can be present, since each conversion represents a different interpretation and not an iterative reasoning process.

Note that, in this work, we only included Conversions as means to process the task, making the task processing phase rather limited. However, this can be different in other domains. For example, when considering syllogism with biased contents (i.e., to introduce belief biases), interactions between the formal task structure and the contents could be included in this phase.

All of the conversions we identified are based on the idea of transitivity: In line with the idea of TransSet (Brand et al., 2020), it is assumed that reasoners try to re-arrange the task to reflect a direct transitive form (i.e., reformulate the syllogism so that it is figure 1 or 2). The first, *Path* is directly adopted from *TransSet*, and 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). This can potentially lead to logical errors, for example, when a premise with the quantifier *All* is inverted.

The second conversion also aims to establish a transitive path, but instead reverts the direction of premises to preferably obtain a syllogism with figure one (instead of preferring a certain quantifier). Thereby, a logically valid option is used (i.e., only the quantifiers *E* and *I* are reverted, called *Transitive* in the pipeline overview) and a version based on the idea of *Illicit Conversion* (Revlis, 1975). Here, the quantifiers *A* and *O* can also be interpreted bidirectionally, even though it is not logically warranted. Note that *Illicit Conversion* only assumes the presence of the bidirectional interpretation in human reasoners' representation, but does not specify when it is used (Khemlani & Johnson-Laird, 2012). In our implementation, we thereby included it with the purpose of achieving a transitive connection, but other interpretations are possible.

Response Candidate Generation Actions In this phase, the first set of conclusion candidates is generated. For syllogistic reasoning, a conclusion can be described by three features: the quantifier, the direction, and if it is a NVC response instead. Mechanisms present in cognitive models thereby treat them independently, which we reflected by separating them into three stages. Note that the majority of syllogisms are invalid, thus making NVC an integral response to capture correctly (Ragni et al., 2019). Commonly, it is assumed that NVC is either concluded at the end, as a result of ruling the respective conclusion candidates out (e.g., due to low confidence in PHM Copeland, 2006; Oaksford & Chater, 2001), or as the result of fast-and-frugal strategies participants might apply to test whether the task even warrants a (cognitively more demanding) deliberate reasoning process (e.g., TransSet Brand et al., 2020).

The majority of the NVC Rules are part of TransSet (Brand et al., 2020), which accounts for 5 of the 6 rules. First, two well-established rules for NVC —the *Two-Some-Rule* (Dickstein, 1978) and the *Two-Negations-Rule*, which test if both quantifiers of the premises are particular or negative, respectively, are included. An extension of it is *PartNeg*, which responds for every syllogism not containing *All* in the premises with a NVC response. All three rules were tested as successful NVC strategies that could improve model performance (Riesterer, Brand, Dames, & Ragni, 2020). Furthermore, *FirstNeg* tests only the quantifier of the first premise for negativity, as long as the second quantifier is not *All*. Finally, the *Transitive Path* rule relates to the respective conversion described before: It tests if TransSet's conversion had failed and a transitive path is not possible.

For the quantifier generation, we implemented the *Atmosphere Effect*, which derives the quantifier based on the *mood* of the syllogism (i.e., combining particularity and negativity of both quantifiers; Woodworth & Sells, 1935). The atmosphere effect is also equivalent with the way TransSet derives quantifiers (Brand et al., 2020). Second, we implemented the *Matching Hypothesis* (Wetherick & Gilhooly, 1995), which selects the quantifier selecting the least amount of elements that is present in the premises (with order $E > I = O > A$). Third, we added the *Min-Heuristic* of the Probability Heuristics Model (PHM) (Oaksford & Chater, 2001), which is based on the idea of probabilistic validity. Similarly, we implemented PHM's *Attachment heuristic* for determining the direction of the conclusions, as well as using MMT again based on the provided table. Additionally, we included the *Figure Effect* (Dickstein, 1978; Johnson-Laird & Bara, 1984), which predicts the A-C direction for figure 1 syllogisms and the C-A direction for figure 2. Finally, as before, we added the MMT based on the table provided by Khemlani and Johnson-Laird (Khemlani & Johnson-Laird, 2012).

Response Candidate Alteration Actions Within this phase, we identified three sub-types of actions: Extensions, Conclusion Tests and Preferences. Based on the initial set of conclusion candidates, the first stage, *extension* sub-processes can take place. It contains actions that derive additional conclusions based on the already selected ones. Here, we included PHM's p-entailment (Chater & Oaksford, 1999; Oaksford & Chater, 2001), which allows to derive conclusions that are probabilistically entailed (e.g., *All A are C* warrants *Some A are C*). Additionally, two Gricean implicatures were included: "Some A are C" implicates "Some A are not C" and vice versa (although not logically warranted).

The second stage, *Conclusion Tests*, features filters that reduce the set of conclusion candidates (which may lead to NVC if no conclusion remains). Those represent critical tests, where participants are convinced that something is opposing the conclusion candidate.

From PHM, we added the test-heuristics, namely the O-Heuristic, which states that conclusions with "Some not" should be avoided (and can therefore lead to NVC; see Copeland, 2006) and the Max-Heuristic, which states that reasoners respond based on their confidence in the quantifier of the max-premise (i.e., the premise with the maximum informativeness following the order $A > I > E > O$; see Oaksford & Chater, 2001). Since the threshold for the confidence needs to be fitted to individuals (e.g., Riesterer, Brand, & Ragni, 2020), we added versions with different confidence thresholds. Note that the Max-Heuristic could be also treated as a rule to directly derive NVC, since it does technically not consider the conclusion and only is based on the premises. However, the position within the *Conclusion Tests* is arguably closer to the intended function as a test-heuristic within PHM's theoretical foundation.

Furthermore, a rule derived from MMT is added. As before, w utilized the response table provided by Khemlani and Johnson-Laird (2012). Thereby, if NVC is a possible answer in the table, NVC is assumed to be concluded. Additionally, all conclusions not in line with those of MMT are excluded. The rule thereby works as both, a preference/filter and test for NVC.

Finally, the last stage are *Preferences*, which behave similar to conclusion tests, but only remove candidates if conclusions remain (i.e., NVC cannot be concluded by them). Preferences thereby account for conclusions that participants have little trust in, but prefer them over NVC nonetheless. Here, one action implements the *Gricean maxime of quantity* (Grice, 1975), so that A is preferred over I, and E over O (e.g., if both, "All" and "Some", are possible, it would be expected to only state "All" in a conversation, since it is the more specific case). Besides that, we included several of the quantifier and direction generation actions as preferences. For many, it is ambiguous in the state-of-the-art if they are actual generative processes or preferences aside from actual reasoning processes, which is reflected here.

Evaluation

In order to utilize the process model pipeline, it needs to be fitted to the respective experimental data. For this, a configuration (i.e., a specific selection of actions within the pipeline that is active) that optimally accounts for the data at hand is searched (potentially within the space of all possible combinations of actions). Thereby, a general problem in modeling human reasoning needed to be handled: In cases where participants solve the majority of tasks correctly, results can reflect knowledge rather than actual reasoning. To account for them, we added first-order logic as an alternative to the process model that bypasses all the steps except the preferences. However, this was only relevant for 6 out of the 95 participants.

We performed two types of evaluations, by creating two different types of process models using the pipeline: *Local model* and *Global models*. Both types and the respective evaluations will be described in the following.

Local Models

First, one can seek an optimal configuration for each observation on its own. For example, if for the task *AE1* the conclusion *Eac* and *Oac* was observed, a possible explanation would be that the quantifier was determined using the *Atmosphere theory*, the direction followed the *Figure Effect*, and finally the conclusions were extended using *p-Entailment*. This kind of evaluation allows to gain insights into which observations are currently not possible at all to be accounted for (i.e., there is no explanations based on the considered theories and models). Respectively, it also allows to identify actions that are essential for specific observations. We refer to this as a *local model*, since it only has a single goal, and corresponds to an isolated search for a sequence of actions.

We used a breadth-first search to find sequences of actions that are able to generate a participant's behavior for a given task. The search was performed twice: once with first-order logic as a valid strategy and once without it. Furthermore, only exact matches with the observations were considered.

Global Models

In contrast, a different goal is the more common task to account for an individual participant as a whole. In these cases, the difficulty arises that actions may interfere with each other. For example, actions that predict *NVC* early on essentially block any other conclusion. Therefore, an action may be required for certain observations, but detrimental to account for others. In consequence, a *global model* needs to be found, meaning that an overarching optimization criterion needs to be considered instead of an isolated search for an optimal sequence for each observation individually.

Since the search for a global model is computationally expensive, we applied several domain-specific adjustments to reduce the search space. We first find a partial solution only for the stages that build up conclusion candidate sets: conversion, quantifier and direction generators, as well as extensions. This sums up all actions that are not concerned with generating *NVC* responses. Then, in a second step, the remaining stages were fitted based on the previously identified configuration, adding all actions that remove conclusion candidates or conclude NVC. This separation is a good approximation, since NVC overwrites any other conclusion: Due to that, it can be fitted independently from the conclusion set generation. The main drawback is that the interaction between conversions and NVC rules, conclusion tests and preferences is not considered. However, this can be argued to make sense from a cognitive standpoint: since the conversion aims to make the task easier to reason about, it is thematically closer to processes that actually aim at deriving conclusions, compared to those that prematurely abort the task.

For evaluating the *Global Model*, which is supposed to represent an individual participant's reasoning process, we used a leave-one-out cross-validation on the basis of tasks: For each participant, a process model was fitted for each of the 64 tasks based on the other 63 tasks. This was done to provide the model with as much information about the individual participant's reasoning behavior as possible. For the respective task, a prediction was made and compared using the Jaccard coefficient as a metric to compare sets of responses (e.g., see Brand & Ragni, 2023). We used the CCOBRA framework for our evaluation, which is implemented in Python and facilitates the benchmarking of cognitive reasoning models [2].

Results

Local Model

Overall, when including the option of using first-order logic, 78.8% of the observations could be accounted for by sequences of actions. Without first-order logic, the percentage dropped only by one percent to 77.8%. This is not surprising, since many of the cognitive models also can account for at least some of the logically correct responses, and hence are able to compensate the lack of a dedicated first-order logic strategy. At first glance, the results seem to show that there is still substantial potential for improvements, since over 20% of the observations cannot be explained by the current processes and mechanisms. However, it is important to note that human reasoning behavior can also be noisy, being influenced by factors outside of experimental control, so that not every observation can be seen as an accurate reflection of an associated reasoning process.

[2]https://github.com/CognitiveComputationLab/ccobra

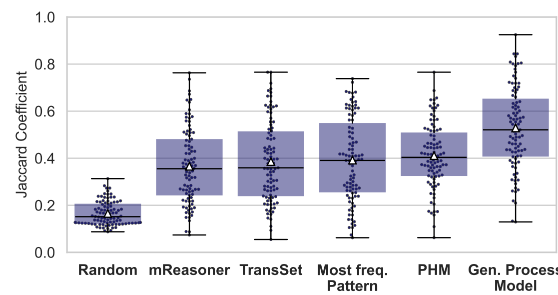

Figure 2: Results of a leave-one-out cross-validation of state-on-the-art cognitive models and the models created by our approach when constructing a global model. Boxplots show inter-quartile ranges as well as medians, points represent individual participants, and the respective mean values are denoted by triangles.

Note that this evaluation also allows for a deeper analysis of the weaknesses and strengths of the state-of-the-art (e.g., for which tasks and situations the actions account well and which processes are most important), which are, however, out of the scope of this work.

Global Model

The results of the global model are shown in Figure 2. For comparison, we included three state-of-the-art cognitive models that most of the actions are based on and have publicly available implementations. Thereby, *mReasoner* as an implementation for MMT (Khemlani & Johnson-Laird, 2013) and *PHM* have been applied to multiple-choice data before (Brand & Ragni, 2023), while *TransSet* performs well on single-choice tasks, but can only predict a single conclusion (Brand et al., 2020). Furthermore, a *random* baseline was added, which selects combinations of conclusions randomly, as well as a baseline in the form of the *most frequent pattern* (MFP). The latter, when making a prediction, uses the most frequently selected conclusion combination for the task at hand based on the other participants' responses. Thereby, it denotes the best strategy when no adaption to an individual reasoner is performed, making it an important bar to beat.

The results indicate that our approach can indeed account for the observed behavior substantially better than the cognitive models on their own, achieving an average Jaccard coefficient of 0.528. Thereby, PHM, the best performing cognitive model, achieves a score of 0.411, barely surpassing the MFP (0.392). TransSet (0.385) and MMT (0.366) do not surpass that mark.

An advantage of our approach is that it allows to interpret the respective process models, gaining insight into which processes were integral for the performance. Percentages of occurrences in the global modal fits are shown in the pipeline illustration (see Figure 1). Thereby, it becomes apparent that direction is mostly determined as a preference (via the *Figure Effect*). Unsurprisingly, processes of PHM, as the best cognitive model in this setting, are commonly selected. However, despite not being adjusted for multiple-choice tasks and thereby performing suboptimally, the components of TransSet individually contribute even more to the result: Not only are the various NVC rules utilized well, also the well-established *Atmosphere effect*, which is also incorporated in TransSet, is the most commonly selected actions for the quantifier generation. This shows that our approach is able to help integrating existing findings into a new variation: a theory like TransSet could be incrementally extended to multiple choice by integrating the respective extensions and preferences. On the other hand, it shows also that many specific mechanisms are indeed transferable and perform well when integrated within other theories.

Discussion

In the present work, we presented a method to generate process models that are adaptable to the individual reasoner from existing cognitive models and reasoning data. While comparable to portfolio approaches (e.g., Riesterer et al., 2018), our approach goes one step further and is applied to the level of sub-processes and effects, essentially combining identified components from the state-of-the-art to an individual process models for each reasoner. Thereby, it offers a well-ordered pipeline, which can serve as a framework for modeling itself: New sub-processes can be easily added to the respective steps in the pipeline to accommodate new findings. Since the pipeline is defined in a neutral way, only using the task and the conclusions themselves as intermediate data structures, it imposes a minimum of restrictions to the type of sub-processes that can be included. Furthermore, the well-ordered pipeline allows to trace which action was responsible for specific conclusions, allowing to pinpoint weaknesses and strengths of specific sub-processes, effects and biases for the respective predictions.

When applied to the domain of syllogistic reasoning, the resulting model achieves a Jaccard coefficient of .528 when predicting the responses of individual participants in a multiple-choice setting. The generated model thereby substantially outperforms state-of-the-art models, which achieve at most a score of .411 (PHM). When used for each observation on its own instead of individual participants, we found that a ratio of 0.788 of the observations can be accounted for using the identified processes. This thereby denotes the amount of responses that can be explained at all by any of the included findings from the state-of-the-art. Thereby, it is important to consider that the gap is not solely due to shortcomings of the state-of-the-art itself, but also due to the noise and inconsistencies that are to be expected in human reasoning data.

However, the presented approach is not meant as a cognitive model itself since it still requires a careful interpretation if the combination is cognitively plausible. The main advantages lie in the insights that can be gained from it: Not only does it show what the current state-of-the-art can account for in total, the respective configurations can also be interpreted as hints for future modeling endeavors: By showing which combinations of sub-processes and effects would account for the behavior best, cognitive modelers can consider incorporating the respective sub-processes into their models in a way that is in line with their theoretical foundation. In a similar fashion, comparisons between different datasets are possible (e.g., for comparing different manipulations and what differentiates them on a process-level). For this work, a natural extension would be the comparison between single-choice and multiple-choice.

The process models generated by our approach still offer the explanatory power of the underlying theories, since the respective components are directly created in line with those. For example, the mechanisms of PHM and TransSet can already be represented as a configuration of the pipeline. It thereby offers an easy way to extend upon those models simply by implementing additional actions and quickly evaluating the potential improvements.

While heuristic models like PHM and TransSet fit easily due to being modular in nature, theories like the Mental Model Theory (MMT) that rely heavily on their own internal representations can prove to be more challenging. However, the proposed structure at least allows to integrate them directly as a "black box" (e.g., using prediction tables). Nevertheless, working on (re)-implementing them within the structure of the pipeline is an ongoing process of this work.

At its core, our approach hinges on the idea to consider cognitive modeling of human reasoning itself as a search and optimization problem. When adopting that perspective, a framework for formulating cognitive processes and effects in a modular and unified way can be created, that not only allows to easier transfer processes, mechanisms and effects between competing cognitive theories and facilitates efforts to unify them, but also offers the possibility to automatically find a configuration for an individual reasoner. In doing so, it can provide an explanation about *why* a reasoner might have come to the respective conclusions by finding a corresponding sequence of actions that suggests which processes may underlie the observed behavior.

Furthermore, adopting the proposed perspective and adhering to the suggested framework allows to streamline cognitive modeling in order to better keep track with the rapid progress in the field: Integrating findings requires simply formulating them as actions in the respective phases. The approach itself thereby is not restricted to a single domain: Since the pipeline is mostly dependent on the structure of the task, similar pipelines can be created for many tasks in human reasoning research that are commonly modeled by step-by-step process models. As such, we present our approach not as a high-performance model, but as a framework, adding another tool to the toolbox of cognitive modelers.

Acknowledgments

This project has been funded by grants to MR in the DFG projects 529624975 and 283135041.

References

Brand, D., & Ragni, M. (2023). Effect of response format on syllogistic reasoning. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society* (pp. 2408–2414).

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.

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

Copeland, D. E. (2006). Theories of categorical reasoning and extended syllogisms. *Thinking & Reasoning*, *12*(4), 379–412.

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

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *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., Brand, D., & Riesterer, N. (2021). The predictive power of spatial relational reasoning models: A new evaluation approach. *Frontiers in Psychology*, *12*, 626292.

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., Dames, H., & Ragni, M. (2020). Modeling human syllogistic reasoning: The role of "No Valid Conclusion". *Topics in Cognitive Science*, *12*(1), 446–459.

Riesterer, N., Brand, D., & Ragni, M. (2018). The predictive power of heuristic portfolios in human syllogistic reasoning. *KI 2018: Advances in Artificial Intelligence: 41st German Conference on AI, Berlin, Germany, September 24–28, 2018, Proceedings 41*, 415–421.

Riesterer, N., Brand, D., & Ragni, M. (2020). 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.

Störring, G. (1908). *Experimentelle Untersuchungen über einfache Schlussprozesse*. W. Engelmann.

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.