# Unifying Models for Belief and Syllogistic Reasoning

**Daniel Brand**[1,2] **(daniel.brand@cognition.uni-freiburg.de)**

**Nicolas Riesterer**[1] **(riestern@cs.uni-freiburg.de)**

**Marco Ragni**[1,2] **(ragni@sdu.dk)**
[1]Cognitive Computation Lab, University of Freiburg, 79110 Freiburg, Germany
[2]South Denmark University, Denmark

**Abstract**

Judging if a conclusion follows logically from a given set of premises can depend much more on the believability than on the logical validity of the conclusion. This so-called belief bias effect has been replicated repeatedly for many decades now. An interesting observation is, however, that process models for deductive reasoning and models for the belief bias have not much of an overlap—they have largely been developed independently. Models for the belief bias often just implement first order logic for the reasoning part, thereby neglecting a whole research field. This paper aims to change that by presenting a first attempt at substituting the first order logic components of two models for belief, selective scrutiny and misinterpreted necessity, with two state of the art approaches for modeling human syllogistic reasoning, mReasoner and PHM. In addition, we propose an approach for extending the traditionally dichotomous predictions to numerical rating scales thereby enabling more detailed analysis. Evaluating the models on a dataset published with a recent meta-analysis on the belief bias effect, we demonstrate the general success of the augmented models and discuss the implication of our extensions in terms of the limitations of the current focus of research as well as the potential for future investigation of human reasoning.

**Keywords:** syllogistic reasoning; belief bias; cognitive process models; individual prediction

## Introduction

How do humans reason? This question has been approached from a variety of directions ranging from formal logical explanations (Inhelder & Piaget, 1958; Rips, 1994) to analogical model-based explanations (Johnson-Laird, 1983). What they have in common is that they focus on interpreting and assessing the logical structure of problems to either derive representations in terms of logics or analogies. However, there is strong evidence that human reasoning behavior is not only dependent on logical properties of the problems but also on interpretations of the content provided by the premises (Morgan & Morton, 1944). Consider the following problem:

No addictive things are inexpensive. [Premise]
Some cigarettes are inexpensive. [Premise]

---

Therefore, some addictive things are not cigarettes.

Would you agree with 92% of participants (Evans et al., 1983) that the conclusion (the statement under the line) follows from the two premises (the statements above the line)? This example denotes a traditional syllogistic deduction consisting of two premises featuring one of four categorical quantifiers each (*All*, *Some*, *No*, or *Some ... not*, which are usually abbreviated as A, I, E, and O, respectively). Together, both premises provide information about three terms (*addictive things*, *inexpensive*, *cigarettes*), two of which only occur in a single premise—the so-called end-terms (*addictive things* and *cigarettes*).

Judged on the basis of formal logics, the inferred relationship between the end terms, i.e., the conclusion, is invalid given the premises. But the conclusion is believable, and a majority of reasoners accepted the logically invalid conclusion as valid. Maintaining the logical form of the problem but losing the believability of the content produced the complementary endorsement probability (8%; Evans et al., 1983).

Recent research has established the belief bias effect as a robust and systematic interaction between logic and belief (Evans et al., 1983; Klauer et al., 2000; Dube et al., 2010a; Trippas et al., 2018). Consequently, belief is an important influential factor for human syllogistic inferences. Despite this fact, it is usually neglected by theories for human reasoning and their model implementations. Similarly, research on the belief bias has traditionally neglected the insight gained on the basis of theories and models by focusing on the interaction with formal logics alone (Klauer et al., 2000; Trippas et al., 2018). The only exception to this independence is the *Mental Models Theory* for syllogistic reasoning, which provided an explanation for the effects of belief (Klauer et al., 2000; Oakhill & Garnham, 1993). mReasoner, the implementation of the *Mental Models Theory* (MMT; Johnson-Laird, 1983), does not offer mechanisms for handling different beliefs so far (Khemlani & Johnson-Laird, 2013).

In this paper we attempt to bridge the current independence of theories for syllogistic reasoning and the belief bias. Inspecting the composition of the traditional theory-agnostic accounts for the belief bias, *selective scrutiny* and *misinterpreted necessity* (Evans et al., 1983), we argue that they offer the potential for natural extension with the mechanisms postulated by most process models for syllogistic reasoning. Based on this extension, the resulting models are provided with the opportunity to leverage a novel kind of feature, which should lead to an increase in predictive accuracy.

The following text is structured into four parts. Firstly, general background into cognitive modeling of syllogistic reasoning and the theoretical foundation of the belief bias is presented. Secondly, our method for combining models for syllogistic reasoning and belief is presented. Thirdly, this approach is evaluated based on predictive performance by relying on the model benchmarking framework CCOBRA[1]. Our results are discussed and put into context with respect to the current state of the art and its implications for the field.

## Theoretical Background

The question to which degree the human ability to *reason* is influenced by personal beliefs is fundamental to reasoning research (Morgan & Morton, 1944). However, throughout the investigation of human reasoning, questions about the processes underlying the human ability to reasoning and the interaction between logic and belief have largely been pursued independently for methodological reasons despite their obvious relationship and importance in our daily lives. To illustrate the state of the art, we will briefly present the most prominent explanatory approaches for syllogistic reasoning and the belief bias in the following.

### Models for Syllogistic Reasoning

Syllogistic reasoning is a traditional domain of human reasoning research that has brought up a multitude of explanatory theories and corresponding models throughout its investigation. A meta-analysis recently provided an overview over twelve of the most prominent theories of syllogistic reasoning (Khemlani & Johnson-Laird, 2012). In the following, two of the most influential models for syllogistic reasoning are introduced: *mReasoner*, the official implementation of the *Mental Models Theory* (MMT; Johnson-Laird, 1983), and *PHM*, a model based on the probabilistic approach to cognitive science.

**mReasoner.** The LISP-based model mReasoner (Khemlani & Johnson-Laird, 2013) follows the four inferential stages proposed by MMT (e.g. Copeland, 2006). First, a mental representation, i.e., the mental model, for the first premise is constructed. Second, the information for the second premise is integrated into the mental model. Third, a candidate conclusion is inferred from the mental model. Fourth, MMT attempts to falsify this candidate conclusion via a search for counterexamples that constructs alternative models for the premises. If the conclusion is falsified, MMT proceeds with a different conclusion candidate. If all conclusion candidates can be falsified, "No Valid Conclusion" (NVC) is returned to indicate that no quantified conclusion follows from the premises.

mReasoner realizes the principle ideas of MMT by relying on four parameters to optimize its inferential behavior to datasets (Khemlani & Johnson-Laird, 2016) and even individuals (Riesterer et al., 2020): $\lambda$ specifies a Poisson-distribution from which the number of entities to represent in the mental model is drawn. For each entity to be created, $\varepsilon$ specifies the probability for it to be sampled from a fully specified logically consistent reference model instead of a canonical one only representing limited and potentially incomplete information about the premises. Based on the constructed mental model, a putative conclusion can be checked via a search for counterexamples. Parameter $\sigma$ denotes the probability of mReasoner engaging in this search for counterexamples. If a counterexample is found, $\omega$ is the probability to continue the search for a consistent conclusion based on a weakened version of the conclusion candidate. Otherwise, NVC is returned. If the search for counterexamples turns out unsuccessful, the candidate is accepted as the conclusion.

**Probability Heuristics Model.** PHM (Chater & Oaksford, 1999) is a model for human syllogistic reasoning that is fundamentally based on the concept of *probabilistic validity*. At its core, PHM approximates the computationally expensive probabilistic inferences by following a set of three generation heuristics (G1-G3) and two test heuristics (T1, T2). First the *min-heuristic* (G1) identifies the least informative premise, the *min-premise*, based on the quantifier informativeness ranking A>I>E>O. Based on this information it defines the conclusion quantifier to be the quantifier of the min-premise. *p-entailment* (G2) proposes the quantifier probabilistically following (Chater & Oaksford, 1999) from the min-heuristic conclusion candidate as an alternative candidate. As the final generative heuristic, *attachment* (G3) specifies the order of terms in the conclusion by postulating that whenever the min-premise begins with an end-term, this end-term is used as the subject of the conclusion. Otherwise, the end-term of the most informative premise, the *max-premise*, is used.

After the conclusion candidate is generated, it is subjected to the *max-heuristic* (T1) that assesses a reasoner's confidence in it. PHM assumes this confidence to be proportional to the informativeness of the max-premise. Consequently, if the max-premise features an uninformative quantifier, the likelihood of the reasoner to reject it in favor of NVC is high (Copeland, 2006). In similar spirit, the *O-heuristic* (T2) postulates that "Some ... not" conclusions should generally be avoided because of their extreme uninformativeness (Chater & Oaksford, 1999).

For this work, we use a recently published implementation of PHM, which makes the model applicable for the evaluation of individual predictions by utilizing five parameters to fit the heuristics to individual reasoners (Riesterer et al., 2020).

### Models for Belief

First experiments that indicated the effect of content on reasoning are almost a hundred years old (Wilkins, 1928). Since then, a large variety of studies have provided evidence for that finding (for overviews, see Evans et al., 1983; Trippas et al., 2018). Some research, such as the work by Gorden (1953), has shown that problems that allow for the application of fast-and-frugal heuristics can outweigh the impact of the belief bias. To exclude the potentially confounding effects of heuristics, the seminal work by Evans et al. (1983), which has framed the experimental investigation on the interaction between logic and belief, excluded the affected syllogisms from its investigation. Resulting from this field of research are theories that attempt to explain the influence of

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

belief based on its interaction with formal first order logic (Evans, 2007; Klauer et al., 2000).

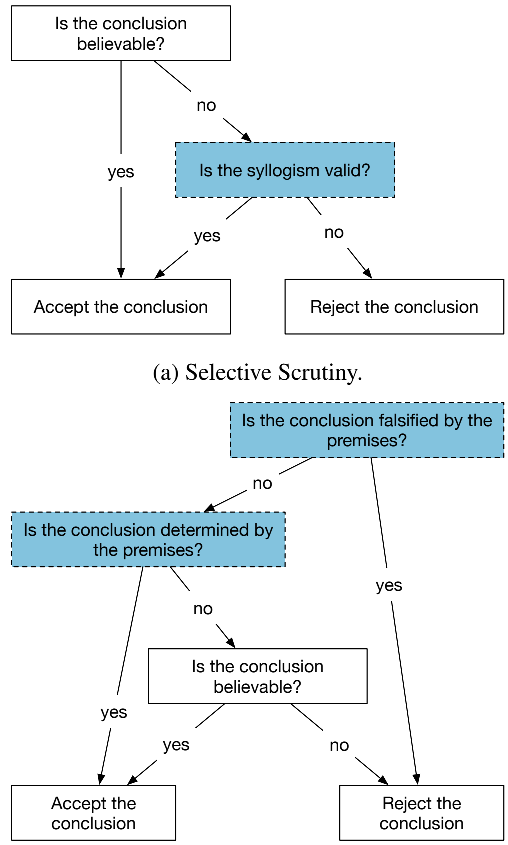
The question about the influence of belief on reasoning is traditionally approached on the basis of two theories: *selective scrutiny* and *misinterpreted necessity* (Evans et al., 1983):

**Selective Scrutiny.** The selective scrutiny account (Evans et al., 1983) assumes that the effects of belief precede the actual processes of reasoning in the human mind. When confronted with syllogistic problems, reasoners first assess the believability of a putative conclusion in terms of the premises. In the case of a believable problem, reasoners dismiss the necessity to perform reasoning and conclude the problem's validity directly. In the case of an unbelievable problem, reasoners perform the actual task of reasoning and generate a conclusion based on their understanding of logic.

**Misinterpreted Necessity.** While selective scrutiny puts the effects of belief before the reasoning process, misinterpreted necessity (Evans et al., 1983) assumes that belief affects its results. If a conclusion is falsified by the premises, it can be rejected directly. Similarly, if the conclusion necessarily follows from the premises, it can be accepted directly. Belief comes into play when a conclusion is possible but not necessary. In this case of inherent uncertainty, belief is used to determine whether the conclusion is accepted or rejected.

Figure 1 visualizes the processes assumed by selective scrutiny (left) and misinterpreted necessity (models) in form of a graph (adapted from Klauer et al., 2000). It is important to note that both accounts do not incorporate assumptions based on existing theories of syllogistic reasoning but rely on logic alone (blue boxes). Both, selective scrutiny and misinterpreted necessity, assume that the believe effects influence the validation process.

Recent research on the belief bias effect has brought up additional theories and models. Some of them are based on theoretical assumptions about the processes underlying human reasoning (Klauer et al., 2000; Oakhill & Garnham, 1993). Others, especially the most recent instances, investigate the belief bias effect on the basis of elaborate probabilistic models such as *Multinomial Processing Trees* (Evans, 2007; Klauer et al., 2000; Trippas et al., 2018). A prominent account based on the Signal detection theory (SDT) explains the believe effect as a response bias (Dube et al., 2010b), which would imply that the believe effect influences the decision stage rather than the determination of the validity. While there is another prominent interpretations using the SDT framework based on the argument strength (Klauer & Kellen, 2011), the response bias account was supported by a recent study (Stephens et al., 2019). This is problematic for selective scrutiny and misinterpreted necessity, as their traditional interpretation focus on influences of the validation process. However, while these newer accounts of the belief bias effect might provide a better grasp of the data, they inherently operate on a group-level by being fitted to large corpora of data. This also hold for the accounts based on the SDT framework, as they are defined on distributions rather than distinct processes. For our purposes, however, we require accounts that are theory-agnostic and allow for a direct extraction of predictions for individual human behavior, which were provided by the traditional approaches.




(a) Selective Scrutiny.

(b) Misinterpreted Necessity.

Figure 1: Illustrations of the selective scrutiny and misinterpreted necessity accounts (adapted from Klauer et al., 2000). The blue boxes with dotted edges reflect the locations where inferential models can be integrated.

## Method

While both, selective scrutiny and misinterpreted necessity, assume access to an assessment of a problem's logical validity, humans often fail to give logical correct responses. Therefore, models for human syllogistic reasoning are aimed at accounting for the systematically illogical behavior of human reasoners. In doing so, they can be regarded as formalizing accounts for *human logic* instead of *formal logic*. This observation allows for a natural way to integrate syllogistic models into the models for belief by replacing the parts relying on logic (represented by the blue boxes in Figure 1) with the respective mechanisms from syllogistic models. Note that in the individual paradigm, the conflict between the response bias and accounts assuming an influence on the determination of validity are resolved. The prominent model of syllogistic reasoning do not incorporate a distinct decision phase, which implies that the integration of a response bias would have to take place after the general reasoning process provided by the model. As distribution-based explanations have to be made deterministic for application in an individual prediction scenario, the account based on a response bias would align with the misinterpreted necessity. Put differently, when combining accounts for believability and reasoning without substantially changing the underlying models, the most general approaches would be to include belief before (selective scrutiny) or after the reasoning process (misinterpreted necessity). This makes selective scrutiny and misinterpreted necessity good promising for the extension of reasoning models.

There are two requirements that syllogistic models have to meet in order to be usable in this approach: First, they have to provide a mechanism for determining the validity of a conclusion. This requirement is usually met by models capable of solving syllogistic problems. All conclusions that a model is able to derive from a given syllogism can be considered *valid*, as the model would predict humans to consider these conclusions valid. Second, the model needs to provide a mechanism to determine if a conclusion is *possible*. This requirement is not generally met by models for syllogistic reasoning. For example, heuristics such as *Atmosphere* or *Matching* (Wetherick & Gilhooly, 1995) derive conclusions directly from the quantifiers in the premises, which allows for checking validity but does not provide a distinct concept of possibility. However, it is important to note that only misinterpreted necessity relies on the concept of possibility, which means that the selective scrutiny model can be applied to a wider range of models.

For this work, we use *mReasoner* and *PHM*, which satisfy both requirements. While mReasoner directly offers support for determining if a conclusion is possible, PHM was extended to allow for determining the possibility of a conclusion. To this end, we relied on its generative heuristics. For a given syllogism, PHM can generate a set of conclusion candidates that can be interpreted as *possible*. This set is then tested, which leads to the selection of the final *valid* conclusion.

## Dataset

Our analysis relies on the dataset that was published along with the meta-analysis on the belief bias effect in Trippas et al. (2018). The dataset contains pairs of syllogistic problems (consisting of premises and a putative conclusion with believable or unbelievable content to verify) and corresponding human responses from a set of 22 studies. In total, the dataset contains responses from 993 individuals who responded to up to 16 syllogistic problems, each. In some of the underlying studies, participants were presented with structurally equivalent problems twice. Responses were originally collected as ratings on scale ranging between 1 and 6 in the studies underlying the dataset. In addition to this, Trippas et al. (2018) also provided a dichotomization of these ratings where validity was determined from the rating values (invalidity for values $\leq 3$, validity for values $> 3$).

Table 1: Rankings for the different belief models for different combinations of attributes. The validity, possibility and believability of conclusion for a given syllogism is used to derive the rankings for selective scrutiny, misinterpreted necessity and a baseline model representing the absence of belief effects. For not uniquely specified ratings the median was used. Other possible values are shown in parentheses.

| Valid | Poss. | Believ. | Misinterp. Necess. | Selective Scrutiny | No Belief |
|---|---|---|---|---|---|
| ✓ | - | ✓ | 6 | 6 | 5 (6, 4) |
| ✓ | - | ✗ | 5 | 4 | 5 (6, 4) |
| ✗ | ✓ | ✓ | 4 | 5 | 2 (3, 1) |
| ✗ | ✓ | ✗ | 3 | 2 (3) | 2 (3, 1) |
| ✗ | ✗ | ✓ | 2 | 5 | 2 (3, 1) |
| ✗ | ✗ | ✗ | 1 | 1 (2) | 2 (3, 1) |

## Extension to Ratings

While the dataset does not only contain dichotomized decisions but also ratings for each conclusion, models for belief are generally only able to predict the acceptance or the rejection of a conclusion without gradations (see Figure 1). In our approach, syllogistic models are used as replacements for the first order logic parts inside of a superordinate belief model, which then derives the final decision. Therefore, we also needed to develop an approach for incorporating ratings into the belief models.

Since the structure of the belief models proposes hierarchies based on the attributes of a conclusion (i.e., validity, possibility, and believability), they allow for the derivation of ratings depending on the combinations of these attributes. Analogously to the dataset used for the evaluation, we used ratings ranging between 1 and 6 and constrained the ratings so that $\leq 3$ is considered as a rejection, while ratings $> 3$ accept the conclusion. This constraint guarantees that the extended belief models are equivalent to the original ones when used for dichotomous verification tasks. In case of ties the median value for the respective attribute combination was used. Table 1 shows the ratings for all combinations of the conclusion attributes.

## Evaluation Scenario

To evaluate the models, we applied them to a prediction task where they had to predict the human decisions to accept or reject a given conclusion for a syllogism. Our analysis allows the syllogistic models to fit their parameters to each individual participant before querying for predictions. Thus, we evaluate each model's performance to accommodate individuals behavior in their parameter space (*coverage analysis*; Riesterer et al., 2020).

Besides PHM, mReasoner, a model using first-order logic, and the respective augmented models based on selective scrutiny and misinterpreted necessity, an individually fitted variant was also included. Here, the best belief model (selective scrutiny, misinterpreted necessity or no influence of belief) for each individual reasoner was determined post-hoc.

As the influence of belief is likely to exhibit inter-individual differences, this model can serve as a theoretical upper bound of the performance achievable with the current models for belief.

In addition to the cognitive models, several baselines were included: First, a *Random* model was included to denote the lower bound of acceptable performance by randomly accepting and rejecting conclusions. Second, as 59% of the problems were responded to twice in the dataset, the *PersonMean* model was included. It uses the mean response of a person for a given problem. If participants are consistent with their responses, *PersonMean* is able to always correctly predict the decision. Therefore, the model can serve as an upper bound and an assessment of the consistency of the participants. The last baseline model is a *Portfolio* which selects the optimal combination of syllogistic models and belief models. As for the individually fitted models, this selection was performed post-hoc and should therefore only serve as a theoretical bound of the performance that could be achieved by the present models.

## Results

First, we discuss the results[2] obtained from evaluations the predictive ability of conclusion acceptances and rejections without taking rating into account. Since the models for syllogistic reasoning and belief were designed for this task paradigm, the results are well suited to assess the potential that lies in their combination. Second, in order to evaluate the proposed extension to ratings, we analyze the model performances when applied to the rating task.

### Verification Results

Figure 2 shows the predictive accuracy of the original syllogistic reasoning models and the augmented belief models that were combined with the respective reasoning models. Overall, the belief variants ($Mdn = 59\%$) achieved significantly higher predictive accuracies than the variants ignoring beliefs ($Mdn = 55\%$, Mann-Whitney test, $U = 359165$, $p < .0001$). Inspected separately, the predictive accuracies achieved by selective scrutiny ($Mdn = 61\%$) outperformed the variants relying on misinterpreted necessity ($Mdn = 56\%$, Mann-Whitney test, $U = 308196$, $p < .0001$) due to the PHM-based misinterpreted necessity model performing worse than its non-belief variant. As misinterpreted necessity relies on a mechanism to determine if a conclusion is possible, which is not directly provided by PHM, the poor performance indicates that our proposed way of determining the possibility based on the generative heuristics of PHM is insufficient.

The *individualized belief* models, which select the best belief model for each individual reasoner post-hoc, slightly outperform the overall best belief model (misinterpreted necessity for mReasoner and first-order logic and selective scrutiny

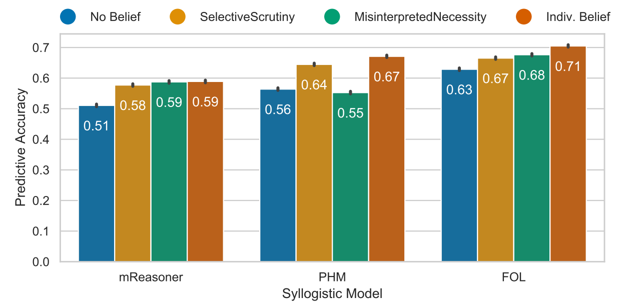



Figure 2: Comparison of the predictive accuracy for the belief × syllogistic model combinations when predicting whether an individual human reasoner accepts or rejects a conclusion.

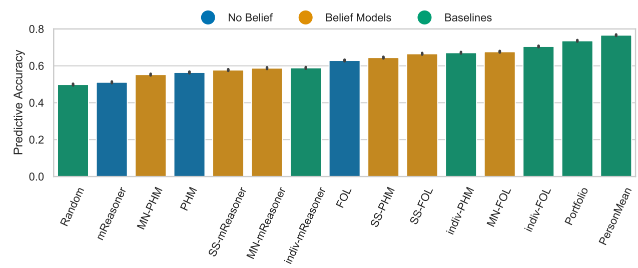



Figure 3: Overall performance of models when predicting the acceptance of the conclusion for a syllogistic task with (yellow) and without (blue) augmentation by a belief model. Additionally, baseline models for estimating the lower and the upper bounds are included (green).

for PHM), which indicates that there are in fact, although rather weak, inter-individual differences with respect to belief.

Figure 3 includes the baseline models to allow for a better understanding of the absolute performances. When compared to the upper bound given by *PersonMean*, which directly uses the individual reasoner as its own predictor, the performance of the first-order logic-based models (*FOL*) are not far behind.

### Rating Results

The results for the rating task are shown in Figure 4. Instead of predictive accuracy, we scored model performances in terms of the absolute difference between the prediction and the rating given by a reasoner. Overall, the results are in line with the results from the verification task. As the order of the models is exactly the same as for the verification task, it indicates that our extension to ratings does indeed preserve the underlying principles of the models. Besides the combination of misinterpreted necessity and PHM, which suffers from the aforementioned insufficient determination of possibility, the error of all models stays below 1.89, the standard deviation of the ratings in the dataset. Using the mean rating of a per-

[2]The scripts and data underlying the analyses of this article are openly available on GitHub: github.com/Shadownox/cogsci-beliefmodeling.

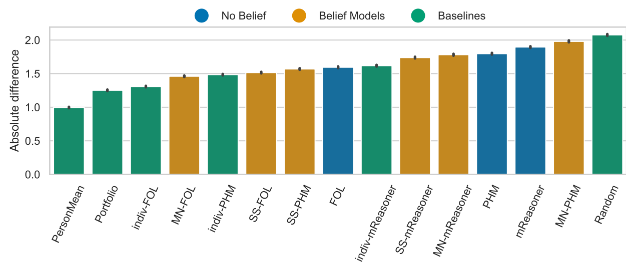



Figure 4: Overall performance of models when predicting the rating of a conclusion for a syllogistic task with (yellow) and without (blue) augmentation by a belief model. Additionally, baseline models for estimating the lower and the upper bounds are included (green).

son for each unique task as a predictor (*PersonMean*) leads to an error of 1.00, which quantifies the inconsistency of individuals when rating the structurally same task (with respect to syllogism, conclusion, and believability). Therefore, the FOL-based models and selective scrutiny with PHM, which achieve an absolute difference of about 1.5, perform reasonably well. This indicates the general success of our approach for deriving ratings from the hierarchical assumptions of the models if no direct possibility to derive ratings or confidences is provided by theories and models.

## General Discussion

In this article, we introduced an approach for combining models for belief and reasoning in syllogistic reasoning—two lines of research that have operated mostly independently. Our analyses showed that information about belief is generally beneficial for optimizing the prediction generation process of models. In our model evaluations that were based on the extensive meta-analysis data from Trippas et al. (2018), the integrative approaches based on selective scrutiny and misinterpreted necessity (Evans et al., 1983) were able to significantly outperform their counterparts, which were restricted to the usual problem description data.

One requirement of misinterpreted necessity for the combination's success that could be identified from the evaluation results is the availability of *possibility* in a model's underlying conceptual foundation. One of the applied models, mReasoner, readily provides a notion of possibility from the interpretation of conclusion consistency with the internal representation of the premise information (e.g., Schaeken et al., 1996). In PHM, the other applied model for syllogistic reasoning, possibility is not directly accessible from within the model's internal processes. As such, the model applied in this article featured an experimental approach for extracting possibilities that turned out poorly (no improvement over the variant not relying on belief). Selective scrutiny does not require information about the possibility of a conclusion candidate, which makes it well-suited for a use in combination with a wide range of models. Its variants were able to significantly outperform the traditional model variants neglecting information about belief.

Predictions for the verification data could be extracted naturally from most of the predictive accounts. Ratings, however, turned out to be more challenging due to rarely being in the focus of process modeling of syllogistic reasoning. Ideally, future models for syllogistic reasoning and belief should aim at providing explanations for human behavior on a deeper level such as by providing weightings or confidences for their outcomes. More complex models for belief, such as the dual-process parallel-competitive model by (Evans, 2007), lend themselves as a starting point for this type of application. While not originally being developed for it, the inherent perspective of conflict resolution given the individual outcomes of the reasoning and belief processes can still be used as a foundation for the extraction of rating information.

In our evaluation, first-order logic appears to be superior to traditional syllogistic reasoning models when combined with a belief model. However, this is likely an artifact of the dataset. While the dataset provided by Trippas et al. (2018) is extensive in the sense that it includes a large number of individuals, only a small set of syllogisms and conclusions was selected in order to "minimize figure, atmosphere, and conversion effects". Therefore, a bias towards FOL is likely to exist, as problems relating to effects that are usually at the core of syllogistic reasoning research were excluded. As such the results should not be interpreted in terms of performance measures for the syllogistic models. The bias towards FOL also has an impact on the *Portfolio*, which selects the best individual combination for each reasoner. As FOL is disproportionately often the best account, the potential of the portfolio is severely reduced. However, it still manages to outperform the best belief model (FOL with misinterpreted necessity), showing that inter-individual differences should not be neglected.

This also reveals a general problem of current research in syllogistic reasoning and the belief bias. Each field actively considers the effects of the other as confounders. For instance, datasets collected for the evaluation of syllogistic theories and models carefully use terms that do not trigger individually differing beliefs (Copeland & Radvansky, 2004). Similarly, as discussed above, studies investigating belief bias try to exclude syllogisms known to trigger robust response biases. Ultimately, however, theories and models need to integrate belief into their respective inferential mechanisms to eventually provide a unified approach to cognition. The results of this work demonstrate the success resulting from a generic integration of belief into computational models of syllogistic reasoning. Knowledge, belief, and reasoning have now been investigated in isolated research for too long. The methods are available, and the time has come to bring the fields closer again and to consider more unified cognitive approaches at the *individual human level*.

## Acknowledgements

Support to MR by the Danish Institute of Advanced Studies and the DFG (RA 1934/4-1, RA1934/8-1) is gratefully acknowledged.

## References

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

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

Copeland, D. E., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *57*(8), 1437–1457.

Dube, C., Rotello, C. M., & Heit, E. (2010a). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review*, *117*(3), 831–863.

Dube, C., Rotello, C. M., & Heit, E. (2010b). Assessing the belief bias effect with rocs: It's a response bias effect. *Psychological review*, *117*(3), 831.

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), 321–339.

Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306.

Gorden, R. L. (1953). The effect of attitude toward russia on logical reasoning. *The Journal of Social Psychology*, *37*(1), 103-111.

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. Basic Books.

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

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.

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

Klauer, K. C., & Kellen, D. (2011). Assessing the belief bias effect with ROCs: Reply to dube, rotello, and heit (2010). *Psychological Review*, *118*(1), 164–173.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852–884.

Morgan, J. J. B., & Morton, J. T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, *20*(1), 39–59.

Oakhill, J. V., & Garnham, A. (1993). On theories of belief bias in syllogistic reasoning. *Cognition*, *46*(1), 87–92.

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.

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. MIT Press.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition*, *60*(3), 205–234.

Stephens, R. G., Dunn, J. C., & Hayes, B. K. (2019). Belief bias is response bias: Evidence from a two-step signal detection model. *Journal of experimental psychology: learning, memory, and cognition*, *45*(2), 320.

Trippas, D., Kellen, D., Singmann, H., Pennycook, G., Koehler, D. J., Fugelsang, J. A., & Dubé, C. (2018). Characterizing belief bias in syllogistic reasoning: A hierarchical Bayesian meta-analysis of ROC data. *Psychonomic Bulletin & Review*, *25*(6), 2141–2174.

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

Wilkins, M. C. (1928). *The effect of changed material on ability to do formal syllogistic reasoning*. New York.