Covering strategy changes: From System 1 to System 2 in syllogistic reasoning

Evelyn Wiens (evelynwiens1@gmail.com)
Cognitive Computation Lab, University of Freiburg, Germany

Alice Ping Ping Tse (alice@hum.ku.dk)
Department of Nordic Studies and Linguistics, University of Copenhagen, Denmark

Marco Ragni (ragni@sdu.dk)
Danish Institute for Advanced Studies, South Denmark University, Odense, Denmark
Cognitive Computation Lab, Technical Faculty, University of Freiburg, Germany

Abstract

Most cognitive models for human syllogistic reasoning aim to explain an *average* reasoner, i.e., the responses given by aggregating the response of the majority of reasoners. Studies show that individuals can deviate a lot from this *average* reasoner. So far, there have been very few models to explain and predict the responses of individual reasoner. In empirical studies, it can be observed that participants often rely on heuristic strategies (System 1 processes) to solve syllogistic problems but participants switch to analytical strategies (System 2 processes) occasionally. The study by Tse et al. (2014) demonstrated that inhibition of the matching heuristic is necessary to switch to the analytical processes in conflict problems that the output from the heuristic does not agree with that from analytical processes. This paper presents four mechanisms to incorporate individual differences in reasoning strategies and effect induced by problem type of the syllogism in predictive computational models built according to the mental model theory, mReasoner, and verbal models theory. Among these models, the composite model, which takes the highest accuracy model for individual reasoner, can reach a median accuracy of 86% in predicting the conclusions given by individual reasoner in the study.

Keywords: predictive cognitive modeling; syllogistic reasoning; strategy changes; dual process theory; System 1 and 2

Introduction

Consider the following syllogistic reasoning example:

> All snakes are reptiles. [Premise 1]
> No rabbit is a snake. [Premise 2]
>
> ---
>
> Therefore, no rabbit is a reptile. [Conclusion]

For the example above, 93% of the 107 participants responded that the conclusion follows logically, which is termed "the conclusion is valid" (Tse, Ríos, García-Madruga, & Bajo-Molina, 2014). This example denotes a traditional syllogistic deduction consisting of two premises, featuring one of the four categorical quantifiers each (*All*, *Some*, *No*, or *Some ... not*, which are usually abbreviated as A, I, E, and O, respectively). Together, the two premises provide information about three terms (*reptile*, *snake*, *rabbit*), two of which only occur in one of the premises – the so-called end-terms, or the major terms (*reptiles* and *rabbit*). The goal of syllogistic reasoning is to connect the information conveyed by the premises, i.e. the categorical relationship between the two end-terms and the middle term (i.e. the term which occur in both premises), to deduce the relationship between the two end-terms. The middle term (also known as the minor term) does not appear in the conclusion. In the example, in addition to the premises, a conclusion candidate is presented below the horizontal line for the reasoner to verify. The term "mood" is used to describe the combination of the quantifiers in the premises and conclusion. The example above is of the mood AE-E for using the abbreviations above. A syllogism can be not only by using the four quantifiers for each premise and conclusion, the terms itself in the premises can be organised in four different ways, called figures:

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

By replacing *reptiles* with A (or a), *snakes* with B (or b) and *rabbits* with C (or c), the example above is a syllogism of figure 2, and can be denoted by AE-E2 and the premises and conclusion can be denoted by Aba, Ecb and Eca respectively. There are 256 possible syllogisms (64 different mood times 4 different figures) but only 27 of them have at least one valid conclusion.

Like many other common daily reasoning processes, humans tend to employ some heuristics when they want to solve a syllogistic problem, unless under certain circumstances. As proposed by Evans and his colleagues in the dual processing theory (Wason & Evans, 1974; Evans, 2006, 2008, 2011; Evans & Over, 2013), humans use the unconscious, intuitive, cognitive-resources-undemanding and rapid System 1 processes by default. The use of heuristics is among these processes. The output from these processes can be prone to biases arise from common sense, beliefs and previous experience. A classic example in human syllogistic reasoning is the belief bias effect that humans tend to accept more (invalid) conclusions which agree with their own beliefs and prior knowledge than otherwise (Morley, Evans, & Handley, 2004). However, humans can switch to the System 2 processes which are conscious, analytical, cognitive-demanding and rule-based under specific conditions, such as when they are told to solve the problems carefully (Evans, 2007).

Another example to illustrate dual processing processes in reasoning is the use of the matching heuristic to solve syllogistic problems. Humans choose the conclusion quantifier which matches the quantifier of the premise with *a lower*

number of entities, i.e. the more "conservative" premise, favouring the order $E > O = I >> A$. Therefore, the conclusion quantifier is the same as (matches) at least one of the premise' quantifiers. Therefore, for the AE example above, humans tend to accept or produce the E-conclusion (No A-C or No C-A conclusion). As mentioned, 93% of the participants accepted that the conclusion followed from the two premises but the conclusion is indeed invalid. That means, only 7% of the participants made the correct response – to reject the conclusion. The matching heuristic is syntactic in nature as it involves merely "matching" the conclusion quantifier with quantifiers of the two premises. Unlike the belief bias effect that the heuristic depends on the semantic of the conclusion, reasoners do not have to process the semantic information (i.e., reptiles, snakes, rabbits) in the premises and the conclusion when using the matching heuristic.

The matching heuristic is one out of 12 cognitive theories that aim to explain the aggregated response of participants in syllogistic reasoning and many of the 12 theories can cover a large number of responses given (Khemlani & Johnson-Laird, 2012). In this paper, we will extend a previous analysis conducted by Riesterer, Brand, and Ragni (2020) and Bischofberger and Ragni (2020) that focused on predicting individual reasoner, to test if incorporating individual reasoner's strategy change from system 1 to system 2 processes during reasoning can yield an adapted model of these theories with better predictive power.

The paper is structured as follows: In the next section we will introduce the data from the study by Tse et al. (2014) cognitive models on syllogistic reasoning. We will introduce how we adapted the models in Section 3 and report the results in Section 4, followed by a discussion (Section 5) which concludes the paper.

The Data

The data are from the study by Tse et al. (2014). 107 students from the University of Granada (mean age = 22.34 years, SD = 4.43; 89 females and 18 males) participated in the experiment. They were rewarded with course credits. The experiment was conducted in Spanish. The participants were all native speakers of Spanish and did not have any training in logic before.

Each participant had to judge the validity of 16 syllogisms, with each followed by a lexical decision task. The syllogisms were chosen to test the interplay between the use of matching heuristics (system 1 processes) and analytical strategy (system 2 processes). Therefore, conflict problems which are either match-invalid (the conclusion quantifier matches with the quantifier of the more conservative premise but it is logically invalid) or mismatch-valid (the conclusion does not agree with the matching heuristic but it is logically valid) and no-conflict problems which are either match-valid and mismatch-invalid were constructed. That means, participants can reach the same conclusion (accepting the conclusion) for no-conflict problems using both the matching heuristic or the analytical strategy; but for the conflict problems, participants have to inhibit the use of the matching heuristic (System 1 non-analytical default approach) and switch to the analytical strategy (System 2). Due to the aforementioned constraint, the only possible options are AE, EA and AA problems which allow the construction of conflict and no-conflict problems. AA problems were not chosen as the two premises have the same quantifier and the matching heuristic is based on matching the conclusion qualifier with the premise quantifiers. The syllogisms with the E conclusion were used as the matched syllogisms (AE-E and EA-E) while syllogisms with the O conclusion were used as the mismatched syllogisms (AE-O and EA-O). In order to prevent participants from guessing the experimental manipulation and as only one of the AE-O2 and EA-O1 syllogisms are invalid (it is not possible to have two mismatch-invalid AE-O and EA-O syllogisms), the AE-A1 and EA-A2 syllogisms were included as fillers to replace a AE-O and EA-O syllogism respectively. There were eight conflict problems, six no-conflict problems and two fillers, see Table 1. Half of the syllogisms (i.e. eight) had a valid conclusion while the other half were invalid.

Table 1: Types of problem used in the experiment Tse et al. (2014).

Problem Type	*Conclusion Type*	*Syllogism*
8 Conflict Problems (multiple-model)	4 Match-invalid	2 AE–E2
		2 EA–E1
	4 Mismatch-valid	2 AE–O1
		2 EA–O2
6 No-conflict Problems (single-model)	4 Match-valid	2 AE–E1
		2 EA–E2
	2 Mismatch-invalid	AE–O2
		EA–O1
2 fillers	invalid	AE–A1
		EA–A2

In the lexical decision task (LDT) after each syllogistic problem, participants were asked to judge whether 24 letter strings were real words in Spanish or not one by one. Half of them (i.e. twelve) were non-words while the other half were words in Spanish, with six of them related to the two terms in the conclusion while the other six were unrelated to the terms in the syllogisms.

The Predictive Model Task & Individualization

We use the CCOBRA-framework[1] to ensure a modeling evaluation standard as proposed by Riesterer et al. (2020). The model has then to predict the conclusion which should be drawn by the individual participant, before the he/she responds. In a predictive analysis, cognitive models need to be able to adapt to the individual they need to predict. This is in most cases done by a parameter optimization process.

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

CCOBRA uses a leave-one-out cross validation method and each test run generates automatically both the test- and training data. The output (the predicted response) from the model is then compared with participant's response.

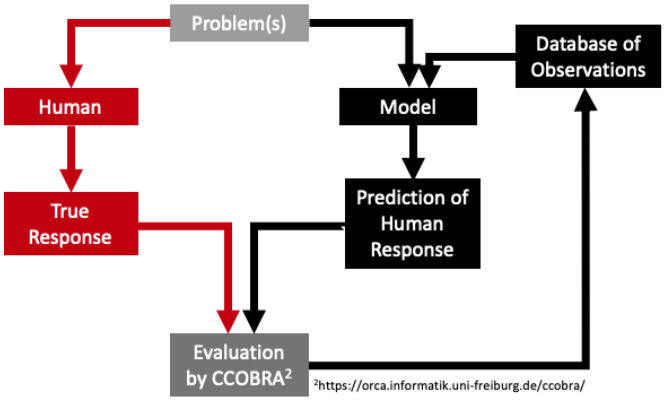

Figure 1: CCOBRA a system to evaluate the predictive accuracy of models.

To capture individual differences, the parameters of the implemented models have to be fitted to an individual. The best fit is determined by iterating over the prediction-response pairs and performing a grid search over the parameter space of the individual models. The parameter optimization for each individual is done before the actual prediction of the answers and therefore evaluate the overall ability of the model to account for individual data.

Cognitive Models

For an in-depth presentation of the existing cognitive models see (Khemlani & Johnson-Laird, 2012). We will present here briefly 3 core theories for System 2 (mental model theory, mReasoner, verbal models theory) that our models are based on.

Mental Model Theory

The mental model theory (MMT) (Johnson-Laird & Philip, 1983) postulates that people draw inferences with the help of mental models. A mental model consists of abstract tokens that reflect the situation asserted by the premises. For example, a (or an initial) mental model of the syllogism *All A are B. No B are C* can be:

a	b	
a	b	
	¬ b	c
	¬ b	c

A conclusion is either drawn based on the initial mental model or refuted with a search of alternative models. This implementation is based on a formalization of the classical theory of mental models (Sugimoto, Sato, & Nakayama, 2013) with some parameter adjustments (Bischofberger & Ragni, 2020). A parameter determines how likely certain individuals search for alternative models (i.e. search for counterexamples). Thus, the implementation can distinguish between people who consider alternative models and those who do not.

mReasoner

mReasoner (Khemlani & Johnson-Laird, 2013) is a more powerful model that follows the assumptions of the mental model theory. The construction of the initial mental model is individualized by two parameters. The size of the initial model is controlled through the parameter λ and the completeness of the encoded information is determined with the parameter ε. Conclusions are generated with heuristics and validated by the mental model. The parameter σ specifies the probability that individuals seek counterexamples after formulating a tentative conclusion. Furthermore, a parameter ω determines if falsified conclusions are weakened. Weakened conclusions are also validated with a search for counterexamples.

Verbal Models Theory

The verbal models theory (Polk & Newell, 1995) assumes that syllogistic reasoning is fundamentally verbal. The model implements multiple parameters that allow high adaptability to an individual. The relations of the premises are encoded into a mental model. Identifying tokens mark more easily accessible information that are derived from the subject of the premises. A conclusion is formulated from the marked tokens. If the program cannot formulate a conclusion, the mental model is reencoded. For this purpose, additional information is extracted from non-identifying tokens. Depending on the type of premise, internal parametrization, and reference token, a new premise is formulated to extend the mental model. The process is repeated until a conclusion can be formulated or reencoding fails. In the latter case, the program returns *no valid conclusion* (NVC).

Making Cognitive Models Adaptive

The cognitive models we just presented need to be made adaptive to predict individual reasoner in the aforementioned CCOBRA-framework. In the following, we describe four mechanisms on how the three models were made adaptive to the response of individual participant.

Probability for searching for counterexamples is adjusted individually

The first set of models are adjusted to model individual reasoning behaviour. To achieve this, parameter settings of the three theories (MMT, mReasoner, Verbal Models) were fitted to individual responses using CCOBRA. The flow structure is shown in Figure 2. In this adaption process, the possible effect of *matching bias* on reasoning, effect of problem type (conflict vs. no-conflict) and individual reasoning strategy (System 1 vs. System 2 processes) were not implemented. Therefore, these models adjust their internal parameterization

according to the response(s) of individual participant. The parameters are updated according to whether a participant just accept the initial conclusion or try to search for counterexamples. The probability for searching for counterexamples (i.e. the System 2 processing) of a participant would be higher if that particular participant gave more System 2 responses, i.e. a higher likelihood of System 2 responses.

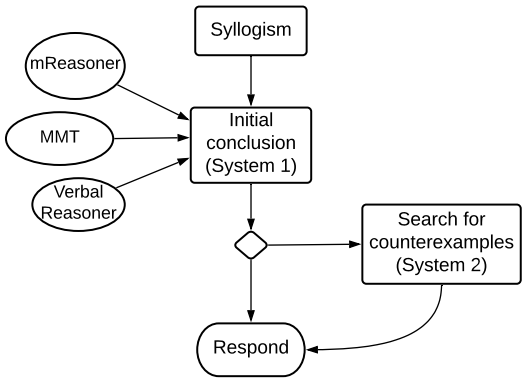

Figure 2: System 2 through individually adjusted probabilities

Probability for searching for counterexamples according to the selective processing model

The critical difference between conflict and no-conflict problems in the (Tse et al., 2014) study is that for no-conflict problems, participants can get the correct responses by either System 1 or System 2 processes (as the conclusions are congruent with the output from the matching heuristic); while for the conflict problems, participants must switch to System 2 processes in order to get the correct responses (as the conclusions are incongruent with the output from the matching heuristic). To resolve conflicts between the outputs from System 1 and System 2 processes in conflict-problems, the selective processing model (Evans, 2000) is generalized so that a conclusion is accepted as soon as there is at least one mental model (initial or alternative) that supports it. Originally selective processing is used in belief bias study to for the conflict between the belief bias and validity of the conclusion; while we have adapted it for the conflict resolution between the matching bias and validity of the conclusion. For match-invalid and match-valid syllogisms alternative models are not searched because the initial model (System 1) already support the given conclusion. In mismatch-problems alternative models (considered as the analytical System 2 processes) are searched. These set of models do not take into account individual differences, it focused on the property of the syllogisms.

Probability for searching for counterexamples is adjusted according to individual strategy and the selective processing model

These models take into account the response time of the participants to the syllogistic tasks and the results of the subsequent lexical decision task when predicting whether the participant would search for alternative models for MMT and mReasoner or reencode the mental models for verbal models. After an initial mental model is constructed by the respective cognitive model, the tentative conclusion is returned or refuted depending on the individual. If the response time of a particular trial is above a specified threshold (9000 ms in the implementation) and the semantic priming effect is diminished (the difference between the response times of the unrelated words and related words is smaller than 15 ms in the lexical decision task) the model tends to refute the initial conclusion (System 2 processes). For switching to System 2 processes, a longer response time is expected as System 2 processes are cognitive resources demanding and it also takes time to inhibit the output from System 1 processes. The missing of semantic priming effect indicates the inhibition of the heuristic (bias) processes (Tse et al., 2014; De Neys & Franssens, 2009). The threshold values were chosen based on evaluation results after a few tests.

The parameters of the respective theories are adjusted according to the behavioral information (response time and priming effect) to implement switches between System 1 and System 2. Additionally, the parameter settings are individualized (see "Probability for searching for counterexamples is adjusted individually" section) to ensure high predictive accuracy outside the threshold conditions.

The search for alternative models in the MMT and mReasoner models incorporated also the aforementioned selective processing model method. An overview over the reasoning process is shown in Figure 3.

Additive probability model

Another model that can describe the conflict between heuristic and analytic processes is the additive probability model (Evans, 2007). In this model, the underlying cognitive process is separated by an heuristic process and an analytical one. This implementation uses individual strategy (as mentioned in the previous paragraph) to determine from which process the predictions are computed. If the response time of the syllogistic task was low or the semantic priming effect occurred (in the LDT), a heuristic process computes the answer. In this case, a parameter determines if an answer consistent with the matching hypothesis is returned or the conclusion is blindly accepted. The analytical process is modeled using the mReasoner. Depending on the internal parameters of an individual, a conclusion can be either accepted or rejected. The selective processing model is not implemented here. For instance, selective processing always predicts acceptance of the conclusion in match-invalid syllogisms. That is because System 1 processes support the given conclusion. Without

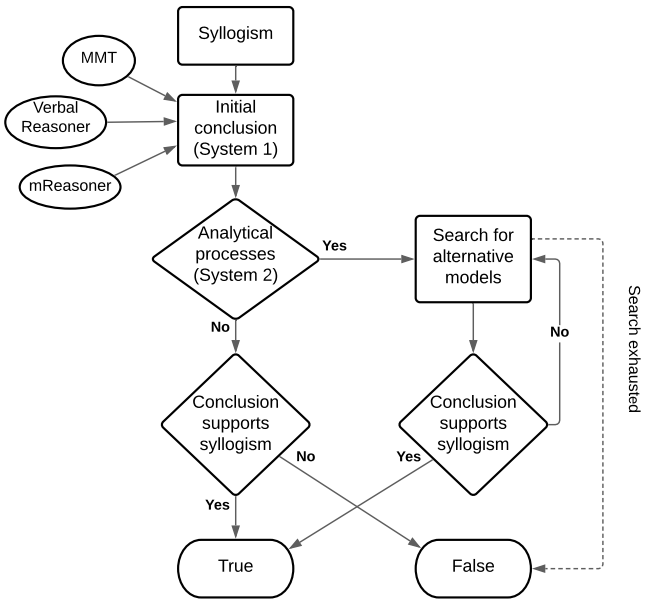

Figure 3: Alternative model search with individual strategy and selective processing.

selective processing there is an individual chance for System 2 processes for those syllogisms in this mechanism.

Composite Model

The previous models each describe individual strategies and theories of how people resolve conflicts between analytic and heuristic responses. While the previous models and extensions already implement individual differences, some models represent an individual's reasoning process better than others. To take advantage of the strengths of each model, a composite model was implemented. The composite model predicts the participants' responses based on the model that could achieve the highest accuracy for that person.

Evaluation and Discussion

Adaptive processes described in the last section were implemented in the three cognitive models built according to the mental model theory, mReasoner, and verbal models theory and were then evaluated in CCOBRA. To compare the overall performance, state-of-the-art models, incldluing PSYCOP, Matching, Conversion, and PHM; and two benchmark models, Uniform and Most Frequent Answer (MFA), were added. The Uniform Model assumes a uniform distribution over the set of responses. Any cognitive model should recognize basic patterns in the data set and outperform this model. Due to the noise in the data, it is unrealistic to assume that the models can perfectly predict participants' responses. Therefore, the MFA model serves as an empirical upper bound for models that do not implement inter-individual differences. The MFA returns the most frequent response given by the participants for each syllogism. A higher predictive power than MFA indicates that the specific models are able to capture different strategies of individuals (Riesterer et al., 2020).

Figure 4 shows the predictive power of the models. All models, except PSYCOP, make more accurate predictions than the uniform model and are therefore able to capture some properties of the data set. The low performance of PSYCOP is due to the high number of possible predictions for the syllogisms and the lack of adaptation to individuals in the implementation. The heuristic models PHM and Conversion achieve prediction accuracies of about 55%. The mental model theory and verbal models theory can predict 62% and 66% of the responses, respectively. Without individual strategies, the search for alternative models (counterexamples) is randomly determined and this leads to comparatively poor results. The matching heuristic, as well as atmosphere, achieve a prediction accuracy of 79%. Since many participants rarely considered alternative models, the accuracy of these heuristics is comparatively high. Moreover, these heuristics can perfectly replicate the answers for some participants as the experimental materials were designed to test the matching heuristic. Matching hypothesis is a modified version of the atmosphere hypothesis and thus the results of these two heuristics were expected to be similar.

The adaptive models all achieve higher accuracies than their static implementations. While the adaptive implementation of the verbal models is able to predict 68% of the data correctly, the adaptive version of mental model theory can achieve 77% accuracy. However, both of them cannot outperform the matching heuristic model. The adaptive implementation of mReasoner (80%) is able to outperform that of atmosphere and matching heuristics, and almost achieves MFA benchmark accuracy (81%).

The selective processing model achieves the same accuracy as the MFA benchmark. Although individual differences was not implemented in this model, it is able to predict most responses in the dataset, as well as the most frequent answers for a syllogism. The verbal models theory with individual strategy implemented (58%) performs worse than the static implementation. Although the re-encoding process of this theory relies on semantic processes, the participants' response times, as well as the results of the lexical decision task, cannot improve the predictive accuracy. The adaptive implementation of mReasoner and MMT with selective processing can outperform the results of the MFA. mReasoner with adaptive parameters and selective processing achieves 84% accuracy and MMT 81%. Thus, parameters of these models be adapted with individual strategies of the participants to obtain a better predictive power. They can capture individual responses and do not merely predict the *majority* response (average reasoner).

The additive probability model (80%) has about the same predictive power as the MFA model. Although the overall performance is worse compared to the other models, the performance of fewer participants can be improved. The composite model achieves a median accuracy of about 86%. The verbal models performs worse than MMT and mReasoner, in-

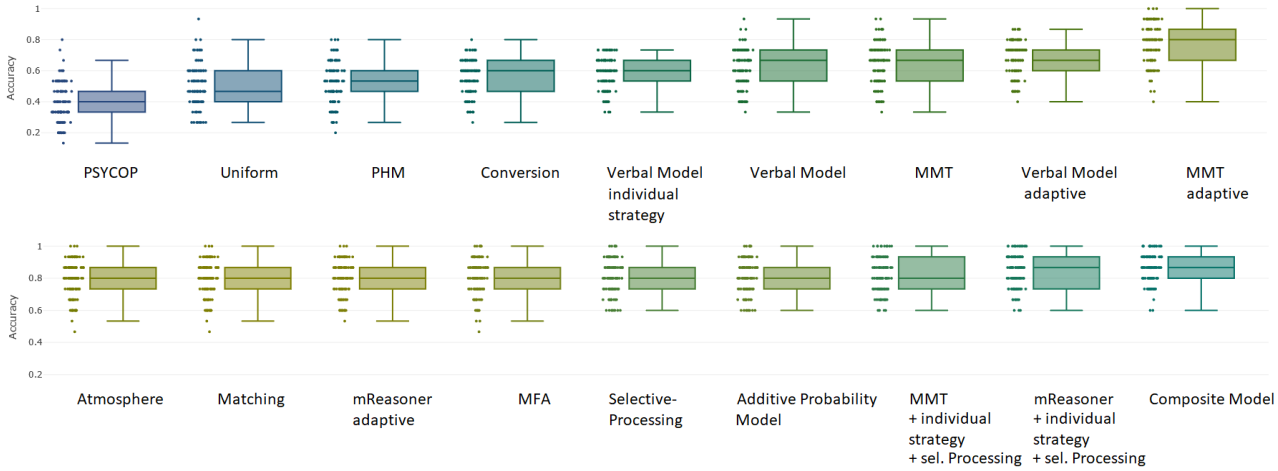

Figure 4: The predictive accuracies of the adjusted models and state-of-the-art models as given in Khemlani and Johnson-Laird (2012)

cluding the adaptive version. The mReasoner obtains the best performance in general. Comparing the four mechanisms, the cognitive models with individual strategy and selective processing implemented obtain the best results. Also, the models with individual strategy (based on the LDT and response times of the syllogisms) perform almost the same as those with the selective processing model.

In summary, the results demonstrate that incorporating individual strategies and effect of problem types can improve the performance of cognitive models to predict the responses of individual reasoners. This shows that the integration can improve state-of-the-art models to predict the responses of individual reasoner substantially and surpasses the MFA benchmark. Finding more optimized ways to integrate individual properties of a participant such as using results from the LDT and the response times to estimate individual strategy in reasoning, and to identify more individual differences can push the models to predict individual reasoners even further.

Acknowledgements

Support to MR by the Danish Institute of Advanced Studies and the DFG (RA 1934/4-1, RA1934/9-1) is gratefully acknowledged.

References

Bischofberger, J., & Ragni, M. (2020). Improving cognitive models for syllogistic reasoning. In *Proceedings of the 42th annual conference of the cognitive science society.*

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, *113*(1), 45–61.

Evans, J. S. B. T. (2000). Thinking and believing. *Mental models in reasoning*.

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395.

Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), 321–339. doi: 10.1080/13546780601008825

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, *59*, 255–278.

Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*(2-3), 86–102.

Evans, J. S. B. T., & Over, D. E. (2013). *Rationality and reasoning*. Psychology Press.

Johnson-Laird, P. N., & Philip, N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. *Harvard University Press*(6).

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

Khemlani, S., & Johnson-Laird, P. N. (2013). The processes of inference. *Argument & Computation*, *4*(1), 4–20. doi: 10.1080/19462166.2012.674060

Morley, N. J., Evans, J. S. B. T., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *57*(4), 666–692.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, *102*(3), 533–566. doi: 10.1037/0033-295x.102.3.533

Riesterer, N., Brand, D., & Ragni, M. (2020). Predictive modeling of individual human cognition: Upper bounds and a new perspective on performance. *Topics in Cognitive Science*, *12*(3), 960–974. doi: 10.1111/tops.12501

Sugimoto, Y., Sato, Y., & Nakayama, S. (2013). Towards

a formalization of mental model reasoning for syllogistic fragments. *Proceedings of the 1st International Workshop on Artificial Intelligence and Cognition (AIC 2013)*, *1100*, 140-145.

Tse, P. P., Ríos, S. M., García-Madruga, J. A., & Bajo-Molina, M. T. (2014). Inhibitory mechanism of the matching heuristic in syllogistic reasoning. *Acta Psychologica*, *153*, 95–106. doi: 10.1016/j.actpsy.2014.08.001

Wason, P. C., & Evans, J. S. B. T. (1974). Dual processes in reasoning? *Cognition*, *3*(2), 141–154.