Predictive Modeling of Individual Human Cognition: Upper Bounds and a New Perspective on Performance

Nicolas Riesterer* (riestern@cs.uni-freiburg.de)
Daniel Brand* (daniel.brand@cognition.uni-freiburg.de)
Marco Ragni (ragni@cs.uni-freiburg.de)
Cognitive Computation Lab, Georges-Köhler-Allee 79
Freiburg, 79110 Freiburg, Germany

Abstract

Model evaluation is commonly performed by relying on aggregated data as well as relative metrics for model comparison and selection. In light of recent criticism about the prevailing perspectives on cognitive modeling, we investigate models for human syllogistic reasoning in terms of predictive accuracy on individual responses. By contrasting cognitive models with statistical baselines such as random guessing or the most frequently selected response option as well as data-driven neural networks, we obtain information about the progress cognitive modeling could achieve for syllogistic reasoning up till now, its remaining potential, and upper bounds of performance future models should strive to exceed. The methods presented in this article are not restricted to the domains of reasoning but generalize to other fields of behavioral research and can serve as useful additions to the modern modeler's toolbox.

Keywords: syllogistic reasoning; neural networks; model evaluation; upper bounds

Introduction

"What I cannot create, I do not understand", the famous quote by Richard Feynman is one of the core maxims of model-driven research. Only if we are able to capture the fundamental mechanics of nature, effectively allowing us to simulate or re-create the associated behavior, we can speak of having gained true understanding. Translated to the domain of cognitive science, this quote is a reminder to constantly keep pushing cognitive models to their limits in order to improve not only their performance, but ultimately our understanding of the mental processes they reflect.

Recently, however, voices have surfaced questioning the merit of current modeling endeavors. For one, there is an ongoing debate about the role of individual data in modeling. Critics of the prevailing focus on data aggregation and corresponding population-based models have demonstrated a lack of group-to-individual generalizability both for experimental (Fisher, Medaglia, & Jeronimus, 2018) as well as for statistical research (Molenaar, 2004). They argue that while potentially useful for insight into typical human behavior, research on aggregates cannot be used to gain understanding about a single individual's cognitive system (Miller et al., 2002). On the other hand, though undoubtedly related, there is ongoing discussion about the methodologies used in cognitive modeling. For example, with the recent efforts to make Bayesian inference models applicable for the broader research community, probabilistic models and corresponding modeling paradigms (especially with respect to model evaluation and selection) have seen a surge in popularity (Vandekerckhove, Rouder, & Kruschke, 2018). However, critics argue that while ideal for discovering statistical relationships which can be tied to high-level theoretical assumptions, Bayesian models cannot be used as algorithmic or process-focused approximations of cognition (Stenning & Cox, 2006; Fugard & Stenning, 2013).

In this article we wish to add to the ongoing discussion about the explanatory power of current cognitive models. We adopt a bird's-eye view posing the fundamental question inspired by Richard Feynman's quote: To which degree are state-of-the-art models capable of reflecting what we are fundamentally interested in—the human mind? We investigate this for the exemplary domain of syllogistic reasoning, one of the core fields of human reasoning research.

With a long history of research stretching over 100 years and a state of the art encompassing at least twelve cognitive theories (Khemlani & Johnson-Laird, 2012), syllogistic reasoning lends itself as a demonstrative domain to investigate the levels of understanding research has achieved. In this domain, we define a prediction task querying models for responses to given syllogistic problems. The final model evaluation is performed by comparing the predictions with the actual human responses. To determine the absolute quality of models, we contrast cognitive accounts with data-driven methods from machine learning, namely a set of neural networks based on different features of the data. By comparing cognitive models with the data-driven results, we explore the potential that remains in the field and determine empirical upper bounds of performance to set goals of future modeling endeavors.

A syllogism is a form of categorical assertion consisting of two premises interrelating a set of three terms via quantifiers (All, Some, No, Some ... not). In experimental settings, participants are asked to relate the end terms of the premises (A and C in the example below), i.e., the terms occurring in only one of the premises:

All A are B
All B are C

What, if anything, follows?

Psychological research has shown that human syllogistic reasoning does not strictly follow formal logic principles (Wetherick & Gilhooly, 1995). Instead, past research has

*Both authors contributed equally to this manuscript.

produced various theories attempting to explain the cognitive principles underlying syllogistic inferences (Khemlani & Johnson-Laird, 2012). Since the domain is well-defined (taking the arrangement of terms into account, there are 64 distinct syllogistic problems and a total of nine possible responses including "No Valid Conclusion" indicating that the end terms cannot be related based on the premise information), syllogisms are an accessible domain for cognitive modeling to investigate what is assumed to be one of the fundamental concepts of human reasoning.

The remainder of this article is structured as follows. First, we introduce the state of the art in modeling human syllogistic reasoning. Second, we define the predictive modeling task as the foundation of our analysis and introduce the baseline models used to put cognitive model performances into perspective. Finally, we present the results of our analysis and discuss their implications for modeling syllogistic reasoning in particular and cognitive science in general.

Related Work

Traditionally, research on human syllogistic reasoning focuses on investigating deviations between human inferences and normative first order logic (Wetherick & Gilhooly, 1995). Over the course of time, the phenomena of syllogistic reasoning matured and were integrated into theories relating statistical effects such as the figural effect (Bara, Bucciarelli, & Johnson-Laird, 1995) with assumptions about mental representations (e.g., in the Mental Models Theory; Johnson-Laird, 1983) or fundamental principles of cognition (e.g., the Probability Heuristics Model by Chater & Oaksford, 1999).

A meta-analysis (Khemlani & Johnson-Laird, 2012) compiled a list of twelve contemporary theories along with the corresponding sets of derived conclusions for each syllogism. By comparison with a set of "liable pooled conclusions", i.e., a dichotomization based on which responses were selected by at least 16% of participants, they performed an analysis assessing how well individual theories were able to predict human responses. Employing classification metrics (hits, misses, correct predictions), the authors concluded that no single model clearly outperformed the others. Instead they found that depending on the metric of choice, all models exhibited distinct strengths and weaknesses rendering a conclusive ordering based on performance difficult.

More recent work leveraged the differences in predictive properties of heuristics for syllogistic reasoning by constructing portfolios exploiting the strengths while avoiding the weaknesses of individual models (Riesterer, Brand, & Ragni, 2018). We showed that the predictive accuracy of the resulting composite model (43%) clearly outperformed individual models (ranging between 37% and 18% for the best and worst cognitive model, respectively). In contrast to the meta-analysis discussed above, we directly based our analysis on individual responses instead of aggregates. The resulting accuracies demonstrated lacking capabilities of heuristic models when confronted with an individual prediction task.

This shift in perspective from modeling population data via pooled conclusions to modeling individual responses is motivated by the fact that the core objective of modeling human reasoning is the development of functionally equivalent computational formalisms capturing the essence of the processes driving human inferences. In today's research on syllogistic reasoning, process-driven performance analyses directly on the level of individuals are scarce. Especially in light of recent work in statistics showing that group-to-individual generalizability is limited if not impossible for parts of psychology and other empirical fields of science (Molenaar, 2004; Fisher et al., 2018), modeling individual data directly will become unavoidable.

In the following analyses, we investigate the potential remaining in the field by contrasting cognitive models with data-driven approaches in a prediction scenario focusing on individual human responses. It is important to note that the following work is not targeted towards model assessment in the traditional sense, but a comparison with methods that are expected to yield an upper bound for predictive performance.

Method

In this section we present the core modeling task of this article: predicting individual responses for given syllogistic reasoning problems. As the foundation for our evaluation we rely on a dataset supplied with the *Cognitive Computation for Behavioral Reasoning Analysis* (CCOBRA) Framework[1] consisting of 139 participants responding to the full set of 64 syllogisms by selecting which of the nine conclusion options could be followed from the premises. The model evaluation was performed in a leave-one-out crossvalidation setting where for each subject to be predicted, the models were fitted using the remaining 138 participants as training data. All code and data required for the analyses are made public on GitHub[2].

The Predictive Modeling Problem

The modeling problem is defined as the task to generate a conclusion for a given syllogism. More formally, the goal is to find a function $f : \mathcal{X} \rightarrow \mathcal{R}$ which transforms a problem input $x \in \mathcal{X}$ into a response $r \in \mathcal{R}$, where $\mathcal{X}$ and $\mathcal{R}$ correspond to the sets of 64 syllogistic problems and nine possible conclusions, respectively. Models are finally evaluated based on their predictive accuracy, i.e., the proportion of correct predictions on a given evaluation dataset. In sum, the modeling problem can be formulated in terms of an optimization problem for a prediction function $f(x)$ dependent on input x (syllogistic problem). The optimization procedure maximizes an accuracy score h, e.g., hits, dependent on the prediction $f(x_t)$ for problem x_t and target output y_t (human response) where t identifies the position in the experimental sequence for a dataset of size N:

$$\max_f \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T_i} \sum_{t=1}^{T_i} h(f(x_{i,t}|x_{i,1},...,x_{i,t-1};y_{i,1},...,y_{i,t-1}),y_{i,t})$$

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

[2] https://github.com/nriesterer/iccm-neural-bound

This problem definition has properties which are beneficial for cognitive modeling. First, it relies on a highly descriptive performance metric with a close connection to modern machine learning (error reduction). Consequently, good performance results (evaluated on unseen test data) are likely to translate to a sensible estimate of performance in application contexts. Second, the performance metric stretches over a clearly defined range of values between all misses (0%) and perfect prediction (100%) allowing for an assessment of absolute performance. The higher the score, the better a model is capable of approximating human reasoning behavior. The modeling task can be considered solved only if performance converges towards 100%. Finally, and arguably most importantly, it directly uses the data recorded in experiments without introducing the risk of misinterpretation due to making statements about populations or "average" reasoners which might not even exist (Miller et al., 2002).

Cognitive Models for Syllogistic Reasoning

As a starting point for our analysis, we relied on the prediction table reported in Khemlani and Johnson-Laird (2012, Table 7). To compile this list of predictions, Khemlani & Johnson-Laird went to great lengths collecting the most up-to-date versions of the respective approaches while maintaining close communication with the theories' inventors or current maintainers.

Unfortunately, however, the simplicity stemming from organizing model predictions in such a static tabular form fails to capture the intricacies of some methods (e.g., Baratgin et al., 2015). As a result, one should treat these representations as baselines for cognitive models' performances instead of comprehensive accounts reflecting their theoretical merit. Still, since prediction-oriented implementations of syllogistic models are rare, and custom implementation introduces the risk of integrating incorrect assumptions stemming from misconceptions about a theory's intent, we rely on the data from Khemlani and Johnson-Laird (2012) to obtain a conservative estimate of the general performance of cognitive models.

Baseline Models for Syllogistic Reasoning

In order to put the predictive performances of cognitive models into perspective, we introduce a set of baseline models. The *Random* model assumes a uniform distribution over the nine syllogistic responses. When queried for a response, one out of the nine options is randomly sampled from a uniform distribution with probabilities of $1/9$. This model serves as a random baseline all models are expected to exceed.

On the upper end of the performance spectrum, we provide the *Most-Frequent Answer* (MFA) model which computes the response distribution per syllogism from given training data. Predictions are generated by returning the response with highest probability mass (ties are resolved by uniform sampling). Since the predictive modeling scenario forces models to generate a single response to a given syllogism, the MFA is the optimal strategy when no information about the individual reasoner is provided.

Neural Models for Syllogistic Reasoning

To answer the question about remaining potential in the field of human syllogistic reasoning we need to provide upper bounds of performance. Since it is not trivially possible to quantify the numerous noise components in the data which stem from inconsistent responses or highly individual inference strategies, we focus on providing empirical upper bounds obtained from data-driven methods from machine learning. While not offering explanatory insight, the resulting accuracies give an indication about which proportion of the data can be successfully predicted by following the structural properties of the data. In particular, we introduce three neural networks focusing on three different perspectives of the predictive modeling problem. Even though neural networks are severely limited with respect to providing high-level explanation for cognitive processes, they have proven to be capable of achieving high levels of performance over the course of the last years and are suitable candidates for obtaining information about the potential remaining in the field.

The first neural network model is a *Multilayer Perceptron* (MLP), a standard feed-forward neural network featuring a topology of 12-256-256-9, i.e., a twelve-dimensional input consisting of three blocks of four bits each for the onehot-encoded quantifiers and figure[3], which is fed into two hidden layers of dimensionality 256 equipped with rectified linear activation units, and finally into the nine-dimensional output layer which indicates the generated response. The model is initially trained by providing syllogistic problems and corresponding human responses, and is optimized using the Adam optimizer (Kingma & Ba, 2014) with mean squared error as the loss function. After a prediction is obtained, the model is supplied with the true response in order to allow for an adaption to individual reasoning processes. This adaption step is realized by training the model for an additional epoch using the new datapoint.

Second, a *Recurrent Neural Network* (RNN) is employed, which explicitly integrates temporal dependencies into the conclusion generation process (for a conceptual introduction see Elman, 1990). The model features a 12-64-64-9 topology consisting of the twelve-dimensional inputs, two recurrent *Long Short-Term Memory* (LSTM) layers (Hochreiter & Schmidhuber, 1997), and the nine-dimensional outputs. Again, the model is trained using Adam, but uses categorical entropy as the error function (Deng, 2006). This model does not incorporate inter-individual differences. However, by actively modeling the task sequence, it is technically able to identify sequence effects which may be beneficial features for the prediction generation process.

Finally, a *Denoising Autoencoder* is applied which frames the predictive modeling problem as a reconstruction task. Similar to the domain of image restoration in which autoencoders have successfully been applied (Xie, Xu, & Chen, 2012), we supply the model with incomplete data about a reasoner.

[3] E.g., "All A are B; All B are C" is $(1,0,0,0,1,0,0,0,1,0,0,0)$, "Some B are A; Some B are not C" is $(0,1,0,0,0,0,0,1,0,0,1,0)$

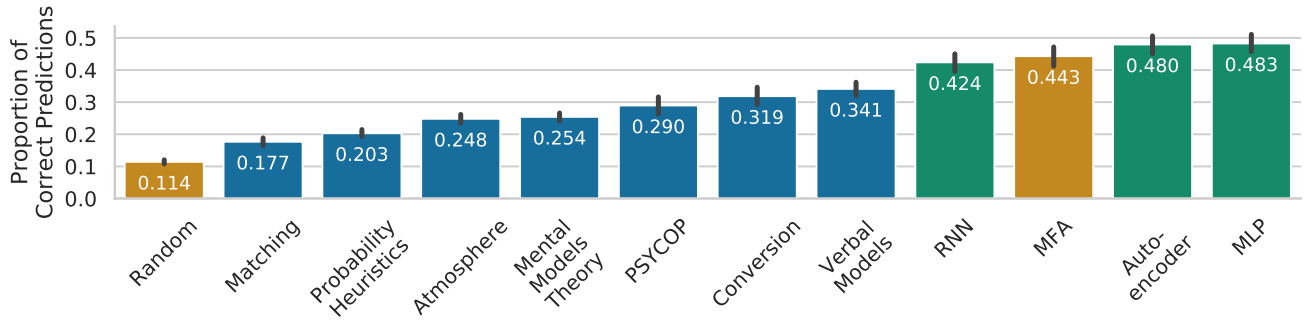

Figure 1: Predictive performance of the models for human syllogistic reasoning. Cognitive models are depicted in blue, baseline models in orange, and neural networks in green. Error bars denote 95% confidence intervals.

The goal of the model is to correctly fill in the blanks. This model is implemented as a 576-2000-576 network featuring a 576-dimensional input obtained by concatenating the onehot encoded responses of the 64 syllogistic problems. As such the inputs represent an individual reasoner's profile. In the hidden layer, this profile is expanded to a high-dimensional space in which relationships between the input dimensions become explicit. From this intermediate representation, the original input can be decoded again. During training, the model is presented with input vectors manipulated by setting values to zero. By training the model to approximate an identity function between noisy inputs and complete outputs by minimizing the mean squared error via Adam, it learns to associate the available information in a way enabling reconstruction of missing values. Over the course of the model evaluation, the autoencoder collects the individual's responses in the adaption step using completing the originally empty reasoner profile. Over time, it leverages the growing information about the individual continuously improving its predictive accuracy.

Results

Predictive Accuracies

The general evaluation results are depicted in Figure 1. The image shows that all models exceed the random model's predictive accuracy of 11% attesting the ability of models to capture the most basic properties of human syllogistic reasoning. The next block of models encompasses the entirety of the cognitive models spanning a range from 18% to 34%. Verbal Models, the best cognitive model, is followed by a substantial gap of performance to the RNN and more importantly MFA, the model always responding with the conclusion most frequently occurring in the training dataset. This constellation of model performances has a major implication for the state of the art in modeling syllogistic reasoning: There is considerable potential left to improve models even without taking inter-individual differences into consideration.

Going beyond MFA, the adaptive neural networks (autoencoder and MLP) demonstrate a basic capability to capture individual reasoning patterns and exploiting them to boost predictive accuracy. However, within this family of models, differences in performance emerge. Relying on temporal dependencies, the RNN model achieves the lowest accuracy scores falling even short of MFA. Reasons for this could be manifold ranging from the application of an unsuitable model topology to problems emerging from the limited amount of training data. However, a more data-centric argument could be that by increasing the data complexity due to the integration of a temporal axis, the models are presented with a problem that is much more difficult to learn than the basic syllogism-response transformation is. As a result, temporal dependencies, or more precisely sequence effects (Aczel & Palfi, 2016), cannot be recognized and leveraged to boost the predictor's accuracy.

The autoencoder which transforms the modeling problem into a reconstruction task achieves higher accuracies than the RNN exceeding the MFA strategy. It shows that the treatment of responses as some form of reasoning profile is a suitable representation to base predictors on that surpass the application of the MFA strategy.

Finally, the MLP achieves the highest accuracy overall (48%). It demonstrates that an integration of adaption to individual properties of cognition via continuous re-training with the newly obtained information can be successfully applied to boost model performance. This approach is not exclusively tied to neural network approaches but should generalize to arbitrary parameterized models which are fitted to training data.

Training Performance

Analyzing the reasons causing networks to perform poorly on data is a difficult task (Lee, Agarwal, & Kim, 2017). To rule out a network's inability to learn the fundamental properties of the syllogistic reasoning data, we investigate the training procedure illustrating accuracy progression on the training and test data per training epoch.

The accuracy progression of the network models during training is depicted in Figure 2. The blue and orange lines represent the mean accuracies (with the shaded band reflecting the 95% confidence interval) on the training and test datasets, respectively. For the RNN, the rise of the training dataset accuracy beyond 90% suggests that, in principle, the network is able to capture the properties of the training data. However,

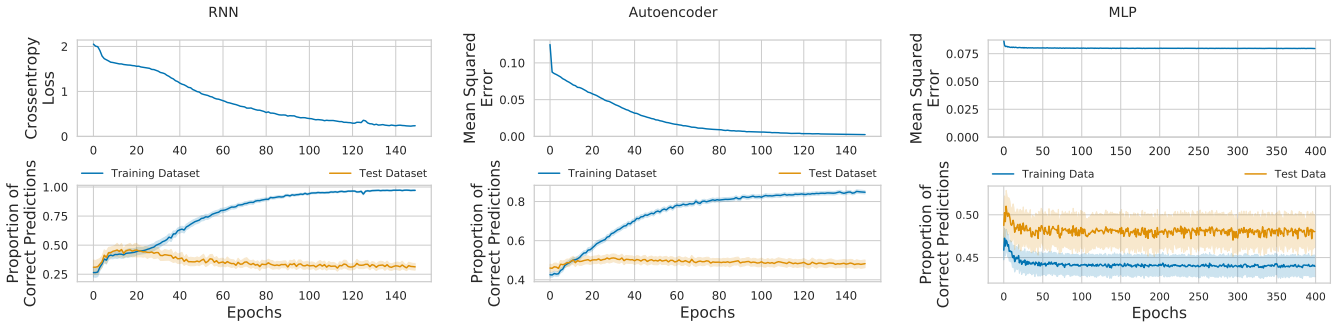

Figure 2: Training progression of the RNN and the autoencoder and MLP. The top plots depict the progression of the raw loss metric used for network optimization. Bottom plots represent the progression of prediction accuracy on training and test data.

the fact that the performance on the test data only rises for a short duration in the beginning of the training process indicates that the learned patterns cannot be generalized successfully to the test instances. The center plot for the autoencoder model paints a similar picture. Even though the effects of overfitting are not as dramatic as for the RNN, training accuracy is clearly improved while damaging the network's generalization capabilities to the test data. An alternative explanation for the superiority of the autoencoder could be that information about individual reasoners are more important for the prediction process or more directly related to specific responses. Finally, the MLP model, despite its predictive capabilities, shows the least amount of learning behavior. After a quick initial bump, the model drops in performance almost instantly and remains constant for the remainder of training. This is most likely due to the limited and inconsistent input and target data. Since in the case of the RNN and autoencoder each training example is high-dimensional and directly incorporates inter individual differences, it is unlikely to observe inconsistencies, i.e., different output for the same input. In contrast, the MLP is fed with 12-bit vectors representing syllogistic problems and produces response predictions for individuals. Since individuals respond differently to the same problems, this data is highly inconsistent and forces the model to adopt a strategy similar to MFA in which an average reasoner is approximated. Classical overfitting is not possible in this scenario.

The observed training performance leads to two conclusions. On the one hand, human syllogistic reasoning appears to follow systematic patterns, which, to some degree, can be leveraged by data-driven methods. The fact that both the RNN and autoencoder are able to learn to fit the training data up to nearly 100% additionally suggests that inconsistencies in the given sequence data (RNN) and reasoner profiles (autoencoder) are minimal. On the other hand, the raw training capabilities of the networks do not generalize well to unseen data. Even though the accuracy on the test data is substantially higher when compared to cognitive models, the training progression shows quick stagnation. Reasons for this could be numerous ranging from problems with respect to data complexity, informational content, or the small size of the dataset used (138 training instances).

In sum, the results show that a current upper bound in performance can be located at a predictive accuracy of roughly 50%. The fact that cognitive models fall significantly lower with a maximum of 35% highlights the potential remaining in the field. Even if the current focus on aggregate evaluation of models is continued, the models should be able to arrive at MFA's performance (44%). The network models demonstrate that by integrating assumptions about individuals even higher predictive accuracies can be achieved. However, even data-driven neural networks stagnate shortly after MFA. While this could be due to technicalities (e.g., network topologies or optimization methods), it could indicate that the purely response-focused data is approaching an upper bound of predictability.

General Discussion

We introduced a predictive modeling task to shift the focus of cognitive model evaluation from relative model selection to a form of model assessment based on absolute performance, i.e., predictive accuracy. In the demonstrative domain of syllogistic reasoning we illustrated that the current state of the art exhibits shortcomings with respect to the quality of model predictions. Without the intention of uncovering individual flaws of specific models, our analysis showed that at most 34% of our data could be successfully predicted by cognitive models. Especially when compared to baseline strategies such as responding with the most frequently chosen answer in the training dataset (MFA), which manages to achieve an accuracy of 44%, this performance is worrisome. For application in real-world scenarios such as in human-agent interaction, syllogistic models are far from being ready for deployment. Even if these theories are, in theory, able to account for core phenomena and statistical effects of syllogistic reasoning, they are of limited use if their assumptions cannot be generalized to useful predictions.

The lingering question is how much potential is left in the domain for future cognitive models to tap into. We introduced a set of neural network models focusing on different properties of the data. Since neural networks are known for being highly capable function approximators, we expected them to provide an upper bound of performance future generations of

cognitive models should be expected to achieve. Our results show that the networks were able to significantly outperform the cognitive models arriving at predictive accuracies of up to almost 50% for the adaptive MLP, the overall best predictor. Two of the networks, MLP and the autoencoder were able to leverage information about an individual's reasoning processes to a point that allowed them to surpass MFA. Finding optimal ways to integrate these inter-individual differences into models of cognition is key for achieving high accuracies. The discussion about which features allow for inter-individual differentiation has already begun (Bara et al., 1995; Stenning & Cox, 2006) and should become a central focus of future research in cognitive modeling.

In conclusion, our work illustrated that cognitive models for syllogistic reasoning have potential left for improvement. Currently, the state of the art is unable to reflect the processes underlying human syllogistic reasoning adequately. However, even if they manage to improve, without adjusting the modeling task to focus on individual responses, they will get stuck at the levels of MFA. The network models demonstrate that trivial individualization in the form of training continuation (MLP) is technically successful but does not lead to substantial improvements over MFA. Rather, future models and cognitive theories should integrate inter-individual differences into their core mechanics to give rise to the next level of cognitive models exhibiting properties useful for research (explainability) and application (predictive accuracy) alike.

We strongly feel that the discussed shortcomings originate from a prevailing focus on relative model evaluation and selection as well as statistical analyses and are not limited to the domain of syllogistic reasoning but could potentially generalize to other domains of cognitive modeling. As such, evaluations in terms of absolute performance scores such as predictive accuracies should be added to the toolbox of modelers in order to paint a more comprehensive picture about the capabilities of individual models.

Acknowledgements

This paper was supported by DFG grants RA 1934/3-1, RA 1934/2-1 and RA 1934/4-1 to MR.

References

Aczel, B., & Palfi, B. (2016). Studying the role of cognitive control in reasoning: Evidence for the congruency sequence effect in the ratio-bias task. *Thinking & Reasoning*, *23*(1), 81–97.

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). Development of syllogistic reasoning. *The American Journal of Psychology*, *108*(2), 157.

Baratgin, J., Douven, I., Evans, J. S. T., Oaksford, M., Over, D., & Politzer, G. (2015). The new paradigm and mental models. *Trends in Cognitive Sciences*, *19*(10), 547–548.

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

Deng, L.-Y. (2006). The cross-entropy method: A unified approach to combinatorial optimization, monte-carlo simulation, and machine learning. *Technometrics*, *48*(1), 147–148.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115.

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

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

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

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

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lee, H. S., Agarwal, A. A., & Kim, J. (2017). Why do deep neural networks still not recognize these images?: A qualitative analysis on failure cases of imagenet classification. *arXiv preprint arXiv:1709.03439*.

Miller, M. B., Horn, J. D. V., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., ... Gazzaniga, M. S. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience*, *14*(8), 1200–1214.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218.

Riesterer, N., Brand, D., & Ragni, M. (2018). The predictive power of heuristic portfolios in human syllogistic reasoning. In F. Trollmann & A.-Y. Turhan (Eds.), *Proceedings of the 41st German Conference on AI*. Berlin, Germany: Springer.

Stenning, K., & Cox, R. (2006). Reconnecting interpretation to reasoning through individual differences. *Quarterly Journal of Experimental Psychology*, *59*(8), 1454–1483.

Vandekerckhove, J., Rouder, J. N., & Kruschke, J. K. (2018). Editorial: Bayesian methods for advancing psychological science. *Psychonomic Bulletin & Review*, *25*(1), 1–4.

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

Xie, J., Xu, L., & Chen, E. (2012). Image denoising and inpainting with deep neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 25* (pp. 341–349). Curran Associates, Inc.