Predicting Spatial Belief Reasoning: Comparing Cognitive and AI Models

Johannes Mannhardt
University of Freiburg

Leandra Bucher
Universität Siegen

Daniel Brand
University of Freiburg

Predicting Spatial Belief Reasoning: Comparing Cognitive and AI Models

Johannes Mannhardt[1] **(johannes.mannhardt@students.uni-freiburg.de)**

Leandra Bucher[2] **(Leandra.Bucher@uni-siegen.de)**

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

Marco Ragni[1,3] **(ragni@sdu.dk)**

[1] Cognitive Computation Lab, University of Freiburg, Germany, [2] University Siegen, Germany,
[3] Danish Institute of Advanced Studies, Denmark

Abstract

Spatial relational descriptions in everyday life sometimes need to be revised in the light of new information. While there are cognitive models for reasoning about spatial descriptions, there are currently no models for belief revision for the spatial domain. This paper approaches this need by (i) revisiting existing models such as verbal model (Krumnack et al., 2010) and PRISM (Ragni and Knauff, 2013) and adapt them to deal with belief revision tasks, (ii) evaluate these models by testing the predictive accuracy for the individual reasoner on a previously conducted experiment by Bucher et al. (2013), (iii) provide baseline models and machine learning models, provide user-based collaborative filtering and content-based filtering methods, and provide an analysis on the individual level. Implications for predicting the individual and identifying strategies and shared similar reasoning patterns are discussed.

Introduction

Belief revision refers to the cognitive ability of reasoners to change existing beliefs to eliminate contradicting beliefs. Imagine you believe "London lies further north than Berlin. Berlin is north of Krakow". As you are in Krakow, you assume that it is no detour to first head towards Berlin. An expert in geography, however, tells you "Berlin actually lies north of London". How do you revise your beliefs, given that you trust the expert? In what way do you change your previous assumption? Do you try to preserve as much of your preexisting beliefs as possible? You can assume that there are differences between individuals in doing so. What are the underlying cognitive mechanisms? Is it possible to predict what different reasoners will do?

Human spatial relational reasoning has been well-studied from a neuro-cognitive, computational, and experimental perspective (Knauff, 2013; Ragni and Knauff, 2013). Since it is a field with a lot of practical relevance, it is well suited for the exploration of belief revision, and by modeling, it should be possible to identify the factors influencing it. While for spatial relational reasoning there exists a vast literature, there are only a handful of studies on belief revision (Knauff et al., 2013; Krumnack et al., 2010). Still, these studies have identified systematic and surprising effects on revision preferences, including aspects such as plausibility and the properties of relations such as visualizability (Bucher et al., 2013). This provides enough variation to possibly explain individual variation.

The paper is structured as follows: In the next section, we will introduce the experiment, its results, and the resulting data set. In section 3 we will introduce existing models for spatial relational reasoning and how they have been adapted for belief revision. This will be accompanied by models from machine learning such as recommender systems. We will present the evaluation in section 4 and finally discuss the results and their implications in Section 5.

The Data

The experimental data stems from Bucher et al. (2013). To determine the visualizability of the problems, a pilot study was carried out. 30 volunteers (14 male; aged from 19 to 55) evaluated 72 binary spatial and non-spatial statements related to their visualizability. Based on the results, a total of 192 problems were selected, 64 each for the categories visual, neutral, and spatial. Examples can be seen in Table 1.

Table 1: Examples for the three task categories.

Category	Examples
visual	The cucumber is thinner than the pumpkin.
	The asparagus is thinner than the cucumber.
neutral	The bird is weaker than the dog.
	The dog is weaker than the polar bear.
spatial	Russia is further east than Poland.
	Poland is further east than Germany.

The procedure of the experiment by Bucher et al. (2013) is now briefly described. Twenty volunteers (8 male; age 20-35; German as native language) were tested individually on 192 problems. Each participant was first presented with two statements (called premises), e.g.

Asparagus is thinner than cucumber
Pumpkin is thinner than asparagus

Each premise consisted of a reference object (RO) and a located object (LO). With the premise "asparagus is thinner than cucumber", asparagus serves as the LO and cucumber as the RO. The distinction between LO and RO is common in this field of research and was proposed by Landau and Jackendoff (1993), among others.

The subject was then presented with two arrangements of the three presented objects, so-called "models", e.g.,

Pumpkin-Asparagus-Cucumber
Cucumber-Asparagus-Pumpkin

one on each side of the monitor. Out of those two choices, only one was "correct", meaning that it was in accordance with the relation that the two premises established between the three objects. The task only continued if the correct model, Pumpkin-Asparagus-Cucumber, in this case, was recognized. Now each subject was confronted with a new premise, a counterfact, e.g.

Cucumber is thinner than pumpkin

Subjects were advised to treat this counterfact as indisputable. In half the cases, it was in accordance with the initial correct model, in the other half it wasn't. Only if the subject recognized an inconsistent fact as such, the experiment continued. Otherwise, the next problem started. In case the subject detected the inconsistency, the stage of belief revision followed. This stage was only reached if the subject identified the correct model in phase three and the counterfact from the last phase to be inconsistent with it. The subject was now presented with two new models, e.g.

Cucumber-Pumpkin-Asparagus
Asparagus-Cucumber-Pumpkin

one on each side of the monitor. Both of these models were created out of the correct model from phase three by the inclusion of the counterfact. One of these models was always plausible and the other one always implausible. In half of the cases, the plausible model was created by relocating the LO of the counterfact and the implausible model was created by relocating the RO. In the other half, it was reversed.

For plausibility, Bucher et al. (2013) relied on common knowledge. A premise such as "the tree is bigger than the flower" would generally be considered to be plausible, the premise "feather heavier than nail" to be implausible. The mental model "Feather-Nail-Hammer" is plausible with regard to the attribute weight. The mental model "Father-Son-Grandpa" is implausible with regard to the attribute age.

Now the participants had to choose the model that matched their expectation about the inclusion of the counterfact into the initial model. The experimental procedure included the stage of inconsistency detection because, as stated by Bucher et al. (2011), inconsistency detection is a prerequisite for belief revision. For revising one's assumptions, one, first of all, needs to recognize an inconsistency between initial assumptions and newly learned information. That's why the first phases of the experiment were conducted - to ensure that a participant was able to conclude from the two premises and then recognizes an inconsistency with that conclusion. This is when the process of belief revision happens. The approaches presented here aim to understand how preferences for a revised model are composed by comparing different modeling approaches in their accuracy of prediction. Different cognitive models were implemented/adapted and compared, including many simple models e.g. LO-preference, relocation of the object added last to the mental model, preference for the plausible model, etc., just to name a few. Also, the more advanced cognitive models PRISM (Ragni and Knauff, 2013) and the verbal model (Krumnack et al., 2010) were adapted. Following the cognitive models, four models from the machine learning area were implemented - content-based filtering (CBF), user-based collaborative filtering (UBCF), a multilayer perceptron (MLP), and an ensemble model.

In the data set used for modeling, the objects, e.g. cucumber, asparagus, etc., were replaced with A, B, and C. Also the premises have been reformulated, e.g. "A is to the left of C" was changed to "Left;A;C". A problem as it was presented in the data set is shown in Table 2.

Table 2: Structure of an experimental test problem.

Sequence	Task	Choices
1 Premises	Left;A;B/Right;C;B	CBA|ABC
2 Model	ABC	Left;C;A
3 Decision	ABC/Left;C;A	CAB|BCA

It consisted of three sequences, one for each time the participants had to make a decision. The first two sequences were largely ignored since we focused on modeling the sequence of belief revision. Predicting the first two sequences of constructing a model and detecting the inconsistency did not add any value since they only served as preliminary work for the last - the belief revision - sequence. Predicting the logically correct answer was in any case the most common one. All models did achieve the same accuracy for them, namely 0.927 for sequence one and 0.892 for sequence two.

Methods

Before presenting the cognitive and machine learning models, let's quickly compare both approaches. Cognitive models are trying to predict human behaviour by recreating the underlying cognitive process as best as possible. They are useful because they provide an accurate indication of the quality of a cognitive theory. Compared to machine learning models, however, which solely rely on statistical data, they are much less accurate in their prediction. While the big advantage of machine learning models is their great accuracy, for our purpose they have two big disadvantages. Firstly, machine learning models are often black boxes, meaning that it is not visible from the outside which patterns have been learned. This holds true especially for neural networks, but even with recommender approaches such as CBF, great effort is needed to find out exactly what has been learned. This is irrelevant for many applications, but since our use case is not only about high predictive power, but also about an understanding of the underlying cognitive processes, this is a core problem of the machine learning models. In addition, large amounts of data are required for training. Very large data sets with a large number of participants are rarely available in cognitive science since such extensive experiments are difficult to conduct.

To compare the performance of all models, we used CCOBRA[1], a python framework specifically created for behavioural reasoning analysis. This framework proposes few constraints to how the models need to be implemented, all that matters is the prediction. Accuracy is simply determined by dividing the number of correct predictions by the number of all predictions. This is done for each subject individually. For this purpose, CCOBRA provides different benchmark types. With "adaption", the model gets the know the correct answer after each prediction

[1] github.com/CognitiveComputationLab/ccobra

in order to gradually adapt to the current subject. With "coverage", a model gets to know all answers of the subject before prediction starts. This can be used to find out how well a model is generally able to represent an individual.

Simple Theories

To give an introduction to how a cognitive theory might be established and implemented, the following simple theories were instantiated and compared. These were also the models used for the ensemble model explained in the machine learning chapter.

LO-Preference: There exists a strong cognitive effect called LO-preference described by Bucher et al. (2011), where it was discovered that participants relocate the LO of the counterfact 87.78% of the time. It followed that the way the counterfact is formulated strongly influences the model revision process.

Relocation of the object added last to the model: The last inserted object was, in our case, the object introduced newly by the second premise (therefore always the right object C). According to Payne et al. (1993), the way the model is constructed plays a crucial role in the way the model is saved in the reasoner's mind and therefore has an influence on the way the model is revised when contradictory information is obtained. The hypothesis that the object added last to the model is the most likely one to be relocated is based on the assumption that it is the starting point for inspection of the mental model (Bucher et al., 2011; Knauff et al., 1998).

Preference for the plausible model: As explained in the experiment chapter, one of the revised models from sequence three was plausible while the other one was implausible. This theory states that there exists a preference for the plausible model. While Bucher et al. (2013) did find out that the preference for the plausible model was almost completely overwritten by LO-preference, it might still exist, especially without a strong contradicting effect.

Preference for the mental model presented on the left/right side of the monitor: In sequence three, the subjects were, as explained in the experiment chapter, presented with a choice for a model, one on the right side and one on the left side of the monitor. Perhaps some subjects, out of various reasons, did not construct the revised mental model in their mind before being presented with the choices but instead made a decision only after being shown both possibilities. This might possibly lead to either a left model or a right model preference.

First/Second premise rejection: Subjects had to reject one of the premises to include the counterfact. The question is if participants had a preference regarding the premise they wish to reject.

Verbal Model

Following Polk and Newell (1995) reasoning does not necessarily depend on domain-specific cognitive processes but on more general cognitive mechanisms. The authors introduced an approach which they called verbal reasoning that makes use of the cognitive mechanisms underlying verbal language (language comprehension) to draw conclusions from premises. The way in which an inference is made therefore depends to a large extent on the decoding of the verbal information. According to Polk and Newell (1995), this decoding of verbal information plays a crucial role in reasoning rather than domain-specific events in the brain. To instantiate the assumptions from the verbal model into a cognitive model that can be tested and evaluated, Krumnack et al. (2010) introduced a queue in which the objects of the premises are inserted. This queue displays an implicit direction that is, according to Maass and Russo (2003), determined by the cultural left/right difference, e.g. the direction in which scripture is read. In addition, there seems to be a natural tendency for a left-to-right direction when imagining spatial events since a right hemisphere dominance for attention often leads to slightly pronounced processing of objects in the – contralateral – left visual hemi-field (de Schotten et al., 2011). The verbal model implementation from Ragni et al. (2019) dealt with this personal preference by providing a compare-function, that tests the outcome of both possible implicit directions and matches the result with the actual subject's answer. For the purpose presented here, this implementation was adapted to fit belief revision in the following way: Inclusion of the counterfact happens by moving forward through the queue until the first object contained in the counterfact is found. It is then moved to the end of the queue. Following this process, the counterfact is included in any case. With the following example:

$$A^{*} \rightarrow B \rightarrow C$$

where the star marks the beginning of the queue and the arrows show the implicit direction, the counterfact "C is to the left of A" results in the model B-C-A, while with the queue

$$A \leftarrow B \leftarrow C^{*}$$

the same counterfact results in the model C-A-B. The preferred implicit direction of the current subject was determined before prediction started. This was possible with the CCOBRA benchmark type coverage.

PRISM

PRISM stands for "preferred inferences in reasoning with spatial mental models". It is a computational model that is based on the assumptions of the theory of preferred mental models, put forward by Ragni and Knauff (2013). The premises "Asparagus is thinner than cucumber" and "Pumpkin is thicker than cucumber" are ubiquitous, meaning that they only induce a single model, namely Asparagus-Cucumber-Pumpkin. When premises lead to multiple possible models like in the case of "Asparagus is thinner than Pumpkin" and "Pumpkin is thicker than cucumber", for which the two possible models are Asparagus-Cucumber-Pumpkin and Cucumber-Asparagus-Pumpkin, these problems are described as indeterminate. The theory of preferred mental models suggests that reasoners have a preference when deciding for one out of multiple possible models. The cognitive model that emerged from this theory was implemented in 2013 and then re-implemented in 2019 for Ragni et al. (2019). It is the later implementation the model presented here is based upon. The similarity between different models is determined by the number of swap operations needed to create one from the other. This approach was not applicable to our task, since both revised models required two swap-operations to get back to the initial model. The following, adaptive approach was used: in one of the variations, the relation of the two left objects to each other has remained the same, in the other the relation of the two right objects. Perhaps different participants had preferences with regard to which revised model "feels" more similar to the initial model.

$$\begin{tikzcd} & ABC+Left;C;A/Right;A;C \arrow[dl] \arrow[dr] & \\ CAB & & BCA \end{tikzcd}$$

For some people, keeping the relation of the two left objects as it is (B is to the right of A) might feel more similar to the initial model than keeping the relation of the two right objects (C is to the right of B). For other people, the opposite might hold. This theory was implemented through an adaptive approach, in which PRISM gradually learned which of the two similarity-functions fit best to the current subject.

Machine Learning Models

To achieve the best possible prediction, various machine learning models were tested on the data set, namely CBF, UBCF, an MLP, and an ensemble model. Additionally, another simple data-driven model that is used as a baseline is the most-frequent-answer model (MFA) which relies on predicting the most commonly selected response for each task. Generally, MFA can be considered to be an upper bound for models that are not able to adapt to individual participants.

Both CBF and UBCF are so-called recommender systems. While these approaches are usually used to suggest videos, products, images or other content to a user, they are used here to suggest/predict the most appropriate response based on the previous user's behaviour.

CBF (content-based filtering) is about making a decision in a situation in the same way a subject responded to a previously experienced, similar situation. For an online shop, CBF would suggest products similar to previously bought products. For a video portal, CBF would suggest videos similar to previously watched videos. To use CBF for our purpose, similarity between different tasks had to be determined. Since all objects of the premises/models were replaced with A, B, and C, the same tasks were repeated over and over again e.g. all tasks that looked like ABC/Left;C;A with the choices CAB/BCA (in that order) were treated as similar to one another, no matter the original objects behind A, B, and C (boxer, car, tree, etc.). Similarity in content between the different problems was therefore determined solely by the relations, by the side on the monitor on which both choices were located, and by the side on which the plausible model was located. CBF was tested with the benchmark types adaption and coverage.

UBCF (used-based collaborative filtering) is about finding similar users to the current user, and then take the behaviour of those users for prediction. For an online shop, UBCF would suggest those products that are bought by users who generally buy the same products as the current user. For a video portal, UBCF would suggest those videos that are watched by users who generally watch the same videos as the current user. For our purpose, the similarity of all subjects to the current subject was determined to form a subject neighborhood. Then, for prediction, the answers of the similar subjects were weighted more heavily than the answers of the not-so-similar subjects. This resulted in a so-called similarity matrix, in which pairwise similarity was determined between all participants. Simply put, UBCF worked similarly to MFA with the difference that the responses of participants who had previously behaved similarly as the current participant were weighted more heavily. This approach was interesting to find out to what extend different users shared similar reasoning patterns when confronted with the same tasks. UBCF was tested with the benchmark types adaption and coverage. In the first case, the subject-neighborhood gradually formed. In the second case, the subject-neighborhood was fully formed before the prediction started.

An MLP (multilayer perceptron) is a basic feed-forward neural network. Riesterer et al. (2020) compared multiple methods for predictive modeling. This comparison included, amongst various cognitive and statistical modeling approaches, a multilayer perceptron, which achieved the highest accuracy, outperforming MFA and an auto-encoder model. Although this comparison was done in the syllogistic domain, testing this MLP in our domain seemed promising. The MLP featured a topology of 10-256-256-4. The 10-dimensional input-layer one-hot-encoded the task presented in sequence three. The 4- dimensional output-layer encoded the response as a one-hot-encoded vector. Fig. 1 shows the topology of the neural network together with an example task.

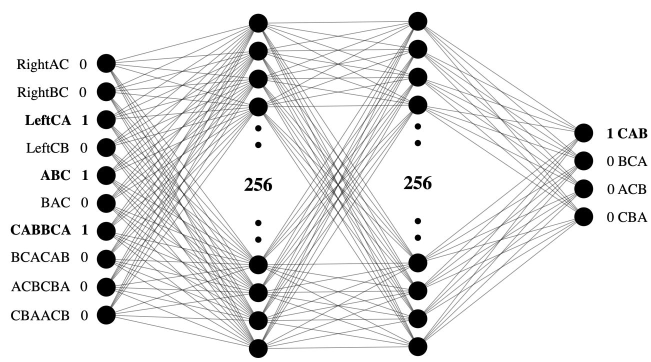

Figure 1: Neural network with the example task ABC/Left;C;A, encoded as a one-hot-encoded vector.

Before prediction, the MLP was trained on all 19 subjects for 30 epochs with a batch size of eight. Once prediction started, the MLP, in order to adapt to the current subject, trained for two epochs after every given answer.

Ensemble modeling is a machine learning technique about combining multiple models by aggregating all predictions into a single prediction. This can be done via multiple methods e.g. averaging/weighting the different predictions. For this purpose, all cognitive theories described in the simple theories section were adaptively combined by taking the model for prediction that so far performed best for the current subject. This approach was chosen to find out exactly which strategies underlie the reasoning process for every individual reasoner. The ensemble model was also tested with the benchmark type coverage, in which case the most fitting strategy was determined before prediction started.

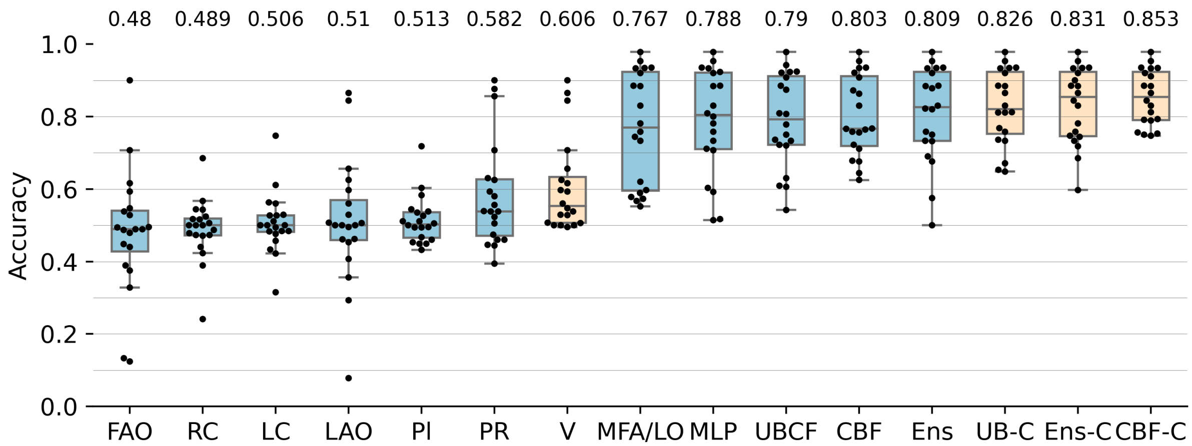

Figure 2: Accuracies of models for belief revision. Only sequence three is considered. Included are MFA (corresponds to LO-relocation), relocation of the object added first/last to the model (FAO, LAO), preference for the left/right choice (LC, RC), preference for the plausible model (Pl), PRISM (PR), the verbal model (V, coverage), MFA, CBF, UBCF, the MLP and the ensemble model (Ens).UB-C, Ens-C, and CBF-C are the same models as UBCF, Ens and CBF, but with the CCOBRA benchmark type coverage.

Evaluation

Fig. 2 shows the accuracies of all presented models. The orange boxes show the models with the benchmark type coverage, meaning that those models didn't gradually adapt to the current subject after every answer, but before the prediction started, which of course results in a higher accuracy since the model doesn't need to adapt over time. The machine learning models did surpass the cognitive models. LO-relocation performed just as well as MFA. Both models were, in any case, identical in their prediction, which is why only MFA is shown in Fig. 2. It stood out that while the simple cognitive theories did achieve accuracies not much higher than the random model, they had a high variance. Single few subjects were predicted very well, or very bad, by them, as can be seen by the few outliers. PRISM and the Verbal model lie between the simple theories and MFA. However, out of all cognitive models, they were the only adaptive ones.

All machine learning models were able to outperform MFA, but did also differ greatly from one another in their accuracies. With an accuracy of 0.8 (0.853 with coverage) , CBF did perform best out of all featured models. The ensemble model was a close second. With much distance to the ensemble model/CBF, but still significantly better than MFA, UBCF and the MLP did achieve similar accuracies of 0.79 (0.826 with coverage) and 0.788.

The small difference between the performance of the ensemble model and CBF leads to the assumption that it was possible to extract the main strategies CBF was able to leverage. Therefore, the dominant individual strategies for most subjects could be extracted, shown in Tab. 3.

It can be questioned whether the left model preference and the right model preference were actual strategies since they did solely depend on the way the task was presented on the screen. They did not add value to understanding the cognitive processes related to belief revision. Rather, they showed that two participants were probably not able to integrate the counterfact into the initial model according to their own preferences, perhaps due to an unwillingness towards expending cognitive effort.

Table 3: Dominant strategy for each subject.

Strategy	Test Subjects
LO-relocation	1,3,4,5,6,7,8,9,11,12,15,16,18,20
Last added object	14,17
First added object	2
Plausible model	13
Left model preference	10
Right model preference	19

Effect of Task Category on Dominant Strategy

While LO-relocation was the dominant strategy for subjects 16 and 20, CBF was able to predict them even better than MFA, which indicates that both subjects used different revision tactics for different tasks. To find out whether task characteristics had an effect on the dominating strategy, the effect of the task categories regarding LO-preference was investigated. As described in the experiment section, each task was assigned to one of three categories - visual, spatial, and neutral. We examined the effect of those categories on the dominating strategy, namely LO-preference. The results are shown in Fig. 3. Although there was wide variation among the subjects in terms of the dominant tactic with respect to the task category, no clear pattern could be discerned that would have applied to all subjects. Therefore, no clear statement could be made about the influence of the categories on the dominant strategy. However, it seemed that subjects 16 and 20 pursued different strategies for different tasks. Whether the category or other task characteristics played a greater role was not investigated.

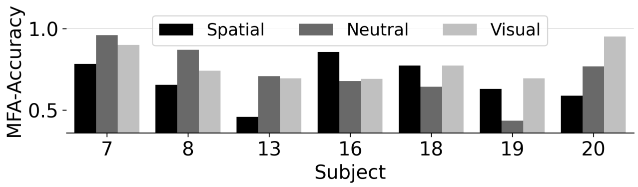

Figure 3: Effect of task category on accuracy of MFA.

Discussion

Various approaches from the fields of cognitive modelling and machine learning were adapted/implemented to test their applicability for belief revision. In addition, the behaviour of all test participants was analyzed individually. Comparing the accuracies of all models on the data set to an experiment conducted by Bucher et al. (2013) provided interesting results. PRISM and the verbal model were overshadowed not only by the statistical recommender approaches, but also by LO-relocation, which delivered the same accuracy as the upper bound MFA, making it the dominant strategy for counterfactual model variation. Adaptive approaches outperforming MFA shows that individual belief revision preferences follow individual reasoning strategies and that those strategies can be gradually learned and leveraged to enhance the quality of prediction.

Subject 12 did relocate the LO in 97.8%, subject 1 in 95.3% of the cases, which made them the most predictable participants. CBF-C reached an accuracy of 0.747 for subject 10, 0.75 for subject 11, and 0.753 for subject 19, which made them the most unpredictable participants. Fig. 4 shows the strengths and limitations of the machine learning models and provides interesting insights into their inner workings.

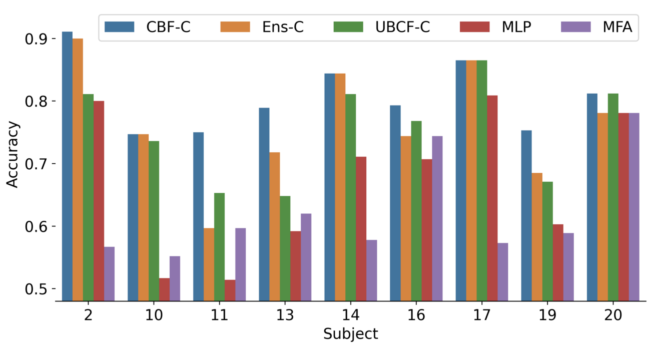

Figure 4: Accuracies of models for belief revision per subject. Only those subjects are presented for which there was a big difference in the accuracy of prediction.

CBF and the MLP performed better than UBCF, which shows that individual subjects were in fact very individual in their behaviour on the task - twenty test-subjects were too few to create a fitting user-neighborhood for each subject. The fact that some subjects were predicted very well by UBCF (2, 14 and 17) while other were not (8, 11, 13), reinforces this assumption.

As can be seen in Fig. 4, subject 11 was better predicted by MFA than by the ensemble model and UBCF. From this one can conclude that either the dominant strategy for subject 11 changed over the course of the experiment, meaning that the adaptive models learned something that was no longer true after a certain time, or that the main strategy used by subject 11 wasn't included in the ensemble model (and wasn't used by another subject). No clear pattern could be extracted according to which subject 11 acted. Subject 13 was better predicted by the ensemble model and CBF than by the other models. This was because those two models were the only ones that could, due to the way they were implemented, learn the preference for the plausible model.

The overall findings lead to the conclusion that it might be impossible to break down behavioural patterns to a single cognitive theory, at least on this domain, no matter how advanced and all-inclusive it might be. This was nicely shown with the ensemble model. Six different strategies were needed to achieve an accuracy close to CBF. Some test subjects deviated from the usual, dominant strategies and pursued their own tactics, which has led to four out of the six strategies representing only a single subject best.

A limitation lies in the underlying experiment and different approaches leading to the same result. The experiment was originally designed with the goal to investigate the effect of visualizability, of LO-preference and to see if it can be overwritten by plausibility - and not for differentiation between different cognitive models. Hence, left object relocation directly corresponded to first premise rejection and in some cases also to the verbal model and PRISM, while right object relocation directly corresponded to second premise rejection, relocation of the last added object, and, again, in some cases also to the verbal model and PRISM. This made it impossible to differentiate between the tactics the subjects used whose responses were better predicted by models other than LO-preference/MFA. Therefore little could be said about the applicability of PRISM and the verbal model to belief revision. While both models did perform well the way they were adapted to fit counterfactual model variation, they essentially did the same what much simpler and less polished models did and could not leverage their strengths. It would therefore be necessary to put the results of this work to the test in an experiment carried out differently e.g. with more than three different objects, three or more premises to form the initial model, more than two choices for the integration of the counterfact, more than 20 subjects, etc., to see whether the results still hold.

In conclusion, there exists no model complex enough that it's parameters are able to fully represent an individual, not least because of the inconsistencies inherent in human nature. Human spatial relational reasoning is sometimes illogical, contradictory, and subject to strong fluctuations. Nevertheless, the results of this work show that it's possible, at least within the presented domain, to extract many of the underlying behavioural patterns. Cognitive theories and the models derived from them provide interesting insights into what is normally taken for granted and provide a contribution to understanding human cognition, even if the findings, like it is the case with this work, only represent a constricted and clearly defined domain like spatial-relational reasoning.

References

Bucher, L., Krumnack, A., Nejasmic, J., and Knauff, M. (2011). Cognitive processes underlying spatial belief revision. *Proceedings of the 33rd CogSci*, pages 3477–3482.

Bucher, L., Nejasmic, J., Bertleff, S., and Knauff, M. (2013). Plausibility and visualizability in relational belief revision. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, pages 1946–1951.

de Schotten, M. T., Dell'Acqua, F., Forkel, S. J., Simmons, A., Vergani, F., Murphy, D. G. M., and Catani, M. (2011). A lateralized brain network for visuospatial attention. *Nature Neuroscience*, 14(10):1245–1246.

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. MIT Press.

Knauff, M., Bucher, L., Krumnack, A., and Nejasmic, J. (2013). Spatial belief revision. *Journal of Cognitive Psychology*, 25(2):147–156.

Knauff, M., Rauh, R., Schlieder, C., and Strube, G. (1998). Continuity effect and figural bias in spatial relational inference. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, pages 573–578, Mahwah, NJ. Mahwah, NJ: Erlbaum.

Krumnack, A., Bucher, L., Nejasmic, J., and Knauff, M. (2010). Spatial reasoning as verbal reasoning. *Proceedings of the 32th Cognitive Science Society*, pages 1002–1007.

Landau, B. and Jackendoff, R. (1993). "what" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2):217–238.

Maass, A. and Russo, A. (2003). Directional bias in the mental representation of spatial events nature or culture? *Psychological Science*, 14:296–301.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge University Press, Cambridge.

Polk, T. A. and Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, 102:533–566.

Ragni, M., Friemann, P., Bakija, E., Habibie, N., Leinhos, Y., Pohnke, D., Satyawan, Y., Schöchlin, M., and Turon, R. (2019). Predicting individual spatial reasoners: A comparison of five cognitive computational theories. In Stewart, T., editor, *Proc. 17th ICCM*, pages 157–162, Canada. University of Waterloo.

Ragni, M. and Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, 120(3):561–588.

Riesterer, N., Brand, D., and 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.