Combining Mental Models and Probabilities: A new Computational Cognitive Approach for Conditional Reasoning

Sara Todorovikj (sara.todorovik@gmail.com)
Paulina Friemann (friemanp@cs.uni-freiburg.de)
Marco Ragni (ragni@cs.uni-freiburg.de)
Cognitive Computation Lab, Technical Faculty, University of Freiburg, 79110 Freiburg, Germany

Abstract

Recent psychological experiments on conditional reasoning indicate the relevance of content, background knowledge and form on the sort of an individual's inference. Based on two of the most prominent theories, probabilistic and mental model based approaches, we develop a probabilistic mental model theory based on Pearl's ε-semantic. By modeling subjective belief in possible worlds, influenced by form and content of a conditional, our model is able to express numerically an individuals' degree of belief in a conditional, while providing an explainable semantics applicable to other domains.

Keywords: Probabilistic Cognitive Models; Reasoning; Conditionals; Predictive Modeling

Introduction

A core goal of cognitive science is to develop a unified theory of cognition. Johnson-Laird's Mental Model Theory (MMT, Johnson-Laird, 1983) is a key theory in the area of human reasoning, and a good candidate for offering a unified theory for a broad range of domains. Its intuitive, comprehensible form allows it to be applied to many domains within reasoning, and its algorithmic and predictive nature enables a qualitative evaluation on explored as well as novel tasks. The MMT assumes the creation and transformation of mental models to describe the scenario.

Another core unified theory for human reasoning are probabilistic approaches. The core assumption is that uncertain reasoning is the basis for rationality, rather than certain reasoning (e.g., Oaksford, Chater, & Larkin, 2000), which leads to the development of probabilistic reasoning models. These have been applied to a plenitude of reasoning domains (e.g., Elqayam & Over, 2013).

The advantages of probabilistic approaches to modeling human reasoning are, among others, the inherent ability to handle uncertain knowledge and the possible incorporation of a subjective degree of belief (e.g., Elqayam & Over, 2013).

In this paper, we propose a combination of the two approaches. We take the Mental Model Theory and extend it with a probabilistic account of the mental models using Pearl's ε-semantics (Pearl, 1991). We interpret the mental models in a conditional task, reported by Singmann, Klauer, and Beller (2016), as possible worlds in the sense of modal logic, and calculate their relative probabilities. One of the factors which define human reasoning is the type of content and the presentation form. More specifically, the individual's background knowledge, and the presentation form of the premises that participants are presented with, can heavily influence the drawn conclusions (e.g., Singmann et al., 2016). To account for this, we extend our model to give predictions for the different contents and presentation forms used in Singmann et al. (2016). The paper is structured as follows: First, we give a short introduction to conditionals, followed by two probabilistic approaches to conditional reasoning. Afterwards we introduce the ε-semantics and our cognitive model for the conditional reasoning task. Finally, we analyze the model's performance on the empirical data and compare it to the model in Singmann et al. (2016).

Reasoning with conditionals

Conditionals are statements of the form 'If p then q' (also written as $p \to q$), where p is called the antecedent, and q, the consequent. Given a conditional rule, i.e., 'if p then q', (also called a major premise) and a minor premise that describes the current situation, for example 'p is true' (given as p), individuals are asked to infer a conclusion. If an individual is given a conditional '$p \to q$' and a minor premise p, and they infer q, they followed the modus ponens inference form. If instead they were given the minor premise '$\neg q$' ('q is false'), and they conclude '$\neg p$', they followed the modus tollens inference form. There are four inference forms: *modus ponens* (MP), *modus tollens* (MT), *affirming the consequent* (AC), and *denying the antecedent* (DA), as shown in Table 1.

Table 1: The four inference forms.

MP	AC	DA	MT
$p \to q$	$p \to q$	$p \to q$	$p \to q$
p	q	$\neg p$	$\neg q$
q	p	$\neg q$	$\neg p$

When interpreting conditionals as causal relationships, in the real world we encounter so-called *disablers* and *alternatives*. Disablers are events that prevent q from happening, even if p has occurred, and alternatives describe events that enable q to happen, even if p has not, e.g.:

> If a balloon is pricked with a needle then it will pop.
> Disabler: The balloon was not inflated at all.
> Alternative: The balloon was pricked with a pen.

Data

The experiment we modeled in this paper, experiment 1 from Singmann et al. (2016), tested participant's endorsements for each of the four inference forms, depending on relative amount of disablers and alternatives, and the form of presentation. In the experiment, participants were asked to give an estimate of the probability, between 0% and 100%, for the different types of problems. The disablers and alternatives are expected to influence the estimates given by the participants. The second independent variable was the form: participants were given either no major premise (*reduced inference*), the major premise in form of a conditional, or the major premise as a *biconditional* (i.e., 'if and only if p, then q.'). In all three cases they were given a minor premise and a conclusion whose probability they were supposed to rate. The tasks used in the experiment are presented in Table 3[1].

Dual-Source model

Oaksford et al. (2000) proposed a probabilistic interpretation of conditional rules by using the probabilities of the antecedent ($a = P(p)$), consequent ($b = P(q)$), and exception ($\varepsilon = P(\neg q|p)$). Singmann et al. (2016) extended this model by disentangling the logical form and content of a conditional, by contrasting individuals' responses to regular conditional inferences, and, reduced inferences (which omit the conditional, and present only a minor premise). They are using three types of parameters: $\xi(C,x)$ (knowledge-based component, depending on the content C and inference x), $\tau(x)$ (form-based component reflecting the subjective degree of belief in the inference x), and λ (a weight given to the form-based knowledge). Endorsement of the reduced inference x with content C is expressed through the knowledge-based component, as shown in Eq. 1, and endorsement of the full inference x with content C is shown in Eq. 2.

$$E_r(C,x) = \xi(C,x) \tag{1}$$

$$E_f(C,x) = \lambda\{\tau(x) + (1-\tau(x)) \cdot \xi(C,x)\} + (1-\lambda)\xi(C,x) \tag{2}$$

ε-semantics

As described by Pearl (1991), ε-semantics is a 'formal framework for belief revision', where belief statements are statements of high probability, and belief revision conditions current beliefs based on new evidence. Simply put, we have a probability function P, which is defined over a set of possible world states, W. A probability $P(w)$ is assigned to each world state w as a polynomial function of some small, positive parameter ε. ε-semantics distinguishes between sentences that describe truths and general tendencies (e.g. 'Birds fly.'), and sentences that describe findings or observations in a specific situation (e.g. 'All blocks on this table are green.'). This is reflected in natural language when using the word 'If' (Pearl, 1991). A statement like 'If it's a bird, it flies' is reasonable, while 'If this block were on this table it would be green.' is not. In order to lay basis for our reasoning model, we will take into consideration the following definition according to Pearl (1991, p. 5):

> Let L be a language of propositional formulas, and let a truth-valuation for L be a function t, such that t maps the sentences in L to the set $\{0, 1\}$ (0 - 'false', 1 - 'true'). A probability assignment $P(w)$ is defined over the sentences in L, where each truth valuation t is regarded as a world w, and $\sum_w P(w) = 1$. This way a probability measure is assigned to each sentence l of L.

Model

In our model we aim to define worlds described by conditional rules following the definition given above. Given a conditional 'If p then q', we take into consideration all the possible worlds, i.e. all the combinations of truth-values for p and q, as shown in Table 2. As stated in the definition, we have a probability distribution P defined over all worlds, assigning a probability value p_i to each one of them.

Table 2: The possible worlds described by 'If p then q', the probability distribution P and probability values p_i, $1 \leq i \leq 4$.

p	q	P
0	0	p_1
0	1	p_2
1	0	p_3
1	1	p_4

For example, in the case of the conditional "If it is a bird, then it flies", the probability value assigned to the world where it is a bird and it is not flying ($p = 1, q = 0$) is p_3.

$$P(\beta|\alpha) = \frac{P(\alpha \wedge \beta)}{P(\alpha)} \tag{3}$$

As mentioned earlier, individuals are asked questions of the form 'Given p, how likely is it that q?', which is actually conditional probability, in this case noted as $P(q|p)$. Following the standard definition of conditional probability (Eq. 3), we obtain the four equations shown below, which describe the four inference forms using the probability distribution P of the conditional's worlds (Table 2):

$$\textbf{MP: } P(q|p) = \frac{p_4}{p_3 + p_4} \qquad \textbf{DA: } P(\neg q|\neg p) = \frac{p_1}{p_1 + p_2}$$

$$\textbf{AC: } P(p|q) = \frac{p_4}{p_4 + p_2} \qquad \textbf{MT: } P(\neg p|\neg q) = \frac{p_1}{p_1 + p_3}$$

Due to individual differences between reasoners, and a divergent background knowledge, it follows that every individual would have a different probability assignment for a certain world. Using these four equations, we can model each participant individually, and determine their personal probability

[1] We would like to note that the choice for the conditional content in the 'Girl' case can be thought of as slightly controversial, which unfortunately leads to some inconsistencies and difficulties when trying to model and/or analyze the data.

Table 3: Contents used in Singmann et al. (2016) experiments.

Keyword	Content	Disablers	Alternatives
Predator	If a predator is hungry then it will search for prey.	Few	Few
Balloon	If a balloon is pricked with a needle then it will pop.	Few	Many
Girl	If a girl has sexual intercourse then she will be pregnant	Many	Few
Coke	If a person drinks a lot of coke then the person will gain weight.	Many	Many

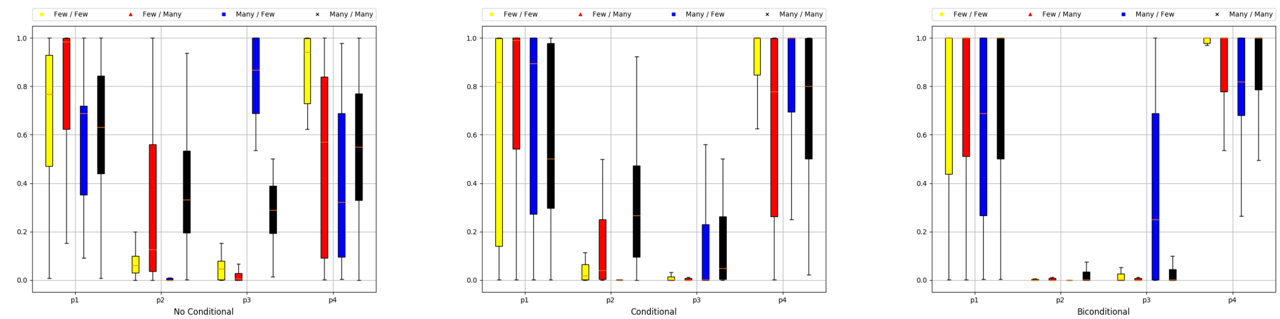

Figure 1: Box plots depicting individual world probability values p_i for every task. Probability values between 0 and 1 – no conditional (reduced inference) (left), conditional (center), and biconditional (right) case. Labels show the amount of disablers and alternatives for the task, e.g.: Few/Many $\rightarrow$ Few disablers / Many alternatives.

distribution, by taking the probability values p_1, p_2, p_3 and p_4 as our parameters and fitting them to their endorsements of the inference forms for every task. Since the parameters are bounded by the sum ($\sum_i p_i = 1$), we have only three free parameters in our model.

Parameters

The parameters in our model, p_1, p_2, p_3, p_4, describe the probability values an individual assigns to the possible worlds described by the conditional. In this section we examine our parameters in more depth and we aim to show that our model can account for the effect of disablers and alternatives on conditional reasoning, and, also, for the effect of individuals being presented with a reduced inference or a (bi)conditional.

Interpretation. Following Table 2, we will first focus on p_2 and p_3. Through these two parameters the effect of disablers and alternatives can be shown. p_2 is the probability that even though p happened, q did not, which is interpreted as the outcome of a disabler preventing q from happening. p_3, on the other hand, is the probability of the world when even though p did not happen, q did, which is interpreted as the outcome of an alternative enabling q to happen. The effect of the different presentation forms can be shown through p_4 and p_1. p_4 is the probability that both p and q happened, can be interpreted as an individual's degree of belief in the conditional rule. p_1 is the probability of the world where neither p, nor q happens, which can show an individual's belief in a rule as a biconditional (if and only if).

Hypotheses. We have two types of hypotheses about changes in the parameter values: First, tasks with few alternatives in contrast to tasks with many alternatives (and the same amount of disablers), may result in an increase in p_2. Similarly, we expect p_3 to be higher for tasks with many disablers in contrast to tasks with few disablers (for the same amount of alternatives). So, both p_2 and p_3 increase when comparing a task with few disablers and alternatives with a task with many disablers and alternatives. Second, the belief in the task's rule increases (and so does p_4) when individuals receive a (bi)conditional in contrast to a reduced inference. We expect an increase of p_1 in the biconditional case. In the reduced inference case, a belief that 'if p then q' is not present, which may lead to a high p_1 in some cases, as an expression of the lack of belief in the influence of disablers/alternatives. In the conditional case, the conditional still might be interpreted as a biconditional (e.g. Cummins, Lubart, Alksnis, & Rist, 1991). So, the change for p_1 depends on the type of conditional.

Fitting. The fitting of our model to the data was done using Python's `scipy.optimize.minimize` function, by minimizing the RMSE with the L-BFGS-B bound-constrained minimization method[2].

Figure 1 shows the individual parameter fit for every task in all three conditions. The first observation that we can make, when looking at the plots, is that the values of p_2 and p_3 are generally higher in the reduced inference case, compared to the conditional and biconditional case, where p_4 and p_1 are higher, as we assumed in our hypothesis. This also leads us to believe that individuals' reasoning is more 'logical' when they are presented with strict conditional rules, compared

[2] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html`

Table 4: Mean percentages of the individuals' values for p_1 and p_4 for every task in each condition (reduced inference, conditional, biconditional); Means of the differences between individuals' values for p_1 and p_4 for reduced inference - conditional, and conditional - biconditional, for every task. (D/A - Disablers/Alternatives; F - Few; M - Many; Red. Inf. - Reduced Inference; Cond. - Conditional; Bicond. - Biconditional)

Task	p_i	Red. Inf.	Cond.	Bicond.	Red. vs. Cond.		Cond. vs. Bicond.	
D/A		Mean	Mean	Mean	Mean	P-value	Mean	P-value
FF	p_1	64.60	60.60	75.20	-4.00	.869	14.60	**.039**
	p_4	79.20	87.40	87.20	8.20	.072	-.30	.715
FM	p_1	74.40	71.40	73.30	-2.90	.981	1.90	.903
	p_4	50.90	64.20	81.70	13.30	.159	17.40	.079
MF	p_1	58.60	66.50	63.70	7.90	.304	-2.80	.408
	p_4	39.20	81.50	71.30	42.30	$<$ **.001**	-10.20	.229
MM	p_1	61.80	58.20	73.50	-3.50	.688	15.30	.131
	p_4	54.50	71.50	81.70	17.00	**.004**	10.10	.082

to when they have to completely lean on their background knowledge, and be more creative when thinking about possible disablers and alternatives. Our next observation is about the change in values of p_2 and p_3 between tasks with varying amounts of disablers and alternatives. As we assumed in our hypothesis, it can be seen that the tasks with many alternatives have higher values for p_2, compared to the other tasks, and the tasks with many disablers have higher values for p_3.

Results and Discussion

Influence of disablers and alternatives. Our first hypothesis was about the influence of disablers and alternatives on p_2 and p_3. We first calculated the means of all individuals' values for p_2 and p_3, for each task, which are shown in Table 5, for the reduced inference case, since that is the case in which we can observe the application of individuals' background knowledge purely without having the influence of the (bi)conditional. We can immediately see that p_2 and p_3's values have a higher mean in the presence of many alternatives, or disablers, respectively.

Afterwards, we looked into pairs of tasks that differ in the amounts of disablers and alternatives, and how the p_2 and p_3 values change between them. Table 6 shows the means of the differences between the all p_i values of the pairs of tasks. In order to determine the statistical significance of the change of the probability values between tasks we performed the Wilcoxon signed-rank test on them, using Python's `scipy.stats.wilcoxon` method[3]. The analysis confirms our hypothesis that when increasing the amount of alternatives, the value of p_2 increases, and when increasing the amount of disablers, the value of p_3 increases.

[3] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html`

Table 5: The mean percentages of the values for p_2 and p_3 for every task in the reduced inference case.

Task		p_i	Reduced Inference
Disablers	Alternatives		Mean
Few	Few	p_2	8.70
		p_3	6.00
Few	Many	p_2	28.30
		p_3	5.10
Many	Few	p_2	9.00
		p_3	82.00
Many	Many	p_2	37.50
		p_3	32.10

Influence of a (bi)conditional. Our second hypothesis was about the influence of giving a (bi)conditional on p_1 and p_4. We first calculated the means of all individuals' values for p_1 and p_4, for each task, and every condition, which are shown in Table 4. The p_4 values in the (bi)conditional case are larger compared to the reduced inference case, as expected. However, the changes in the p_1 values, are not uniform.

Comparison

After fitting our model to the data, we compared its fit to the Dual-Source model (DSM) on the same data. We obtained the DSM parameter values for the participants from `https://osf.io/zcdfq/`, and used them accordingly in equations 1 and 2, as described above. The DSM uses 22 parameters to fit all four tasks for all conditions[4], and our model uses 3 parameters, but is fitted respectively to all problems. Hence, we can determine changes in the p_i across tasks. To determine the goodness of fit we calculated the RMSE and R^2. The mean RMSE for our model was .020, and R^2 was .963, compared to DSM's mean RMSE of .049, and R^2 of .815. It should be noted that the DSM has certain limitations – the $\xi(C,x)$ parameter values can only be obtained when fitting the reduced

[4] 16 values for $\xi(C,x)$, 2 values for λ, and 4 values for $\tau(x)$

inference case, so if the model was presented only with, e.g., the conditional case it would not be possible to fit it.

Table 6: Means of percentages of the differences Δp_i between individuals' values for all p_i for combinations of tasks in the reduced inference case, calculated as p_i(Task 1) - p_i(Task 2). (D/A - Disablers/Alternatives; F - Few; M - Many)

Task 1	Task 2	p_i	Reduced Inference	
D/A	D/A		Mean Δ	P-value
F/F	F/M	p_1	9.78	.178
		p_2	**19.60**	**.001**
		p_3	**-9.00**	**.016**
		p_4	**-28.35**	**.004**
F/F	M/F	p_1	-5.97	.347
		p_2	**-7.70**	$<$ **.001**
		p_3	**76.00**	$<$ **.001**
		p_4	**-40.04**	$<$**.001**
F/M	M/M	p_1	**-12.57**	**.036**
		p_2	9.20	.063
		p_3	**27.00**	$<$ **.001**
		p_4	3.65	.739
M/F	M/M	p_1	3.17	.769
		p_2	**36.60**	$<$ **.001**
		p_3	**-49.80**	$<$ **.001**
		p_4	**15.34**	**.011**
F/F	M/M	p_1	-2.80	.750
		p_2	**28.80**	$<$ **.001**
		p_3	**26.10**	$<$ **.001**
		p_4	**-24.70**	**.002**

Reducing the number of parameters

By fitting our 3 parameters to each task, we achieved a good fit, and our next goal will be to try to reduce the fitting to each task, while still obtaining satisfactory results, by challenging the predictive capabilities of our model. We tackle this by taking into consideration the earlier observations of how the probability distributions change between tasks and also between conditions, which leads to two different approaches. In the following, where necessary, a probability value will be denoted in the following way: $p_i^x(t)$, where t is the first letter of the task's keyword (*p, b, g, c*), x is the condition (r for reduced inference, c for conditional, and b for biconditional), and i is, as before, the probability value's index ($i \in [1,2,3,4]$). So, for example $p_3^r(g)$ denotes the p_3 value for the 'girl' task in the reduced inference case. To measure the goodness of prediction for every approach we calculated the RMSE[5].

Task probability distribution differences. The first approach focuses on how probability values change between tasks, especially based on the different number of disablers and alternatives. This can be done in two different ways which differ in the number of parameters used.

[5]In the case of prediction we do not take into consideration the R^2 measure, because, as shown in (Alexander, Tropsha, & Winkler, 2015), RMSE provides a better prediction quality measure.

Constant differences. When we fit the model initially we calculate the differences of the probability values between tasks, for the reduced inference, conditional, and biconditional case. By taking the means of those differences we obtain constant values which describe the general change of the probabilities among participants. For example, Table 7 shows the constants for probability value changes from the 'Predator' task to the other three tasks in the conditional case.

Table 7: Constants for probability value changes Δp_i between tasks for the conditional case. Values between 0 and 1. D/A - Disablers/Alternatives; F - Few; M - Many

Task 1 (D/A)	Task 2 (D/A)	Δp_1	Δp_2	Δp_3	Δp_4
F/F	F/M	.109	.113	.010	-.232
F/F	M/F	.06	-.035	.141	-.059
F/F	M/M	-.023	.237	.113	-.159

Now, we fit the participant's endorsements for one task, for which we need only three parameters, and predict the endorsements for the other tasks by using the constants, as shown in Eq. 4, where fit is the task we have already fitted, and *pred* is the task whose endorsements we predict.

$$p_i^x(pred) = p_i^x(fit) - const(fit, pred) \tag{4}$$

E.g., if we have fitted the probability values for the 'Predator' task in the conditional case, and we want to predict the p_2 value for the 'Balloon' task, we will calculate it by:

$$p_2^c(b) = p_2^c(p) - 0.113 \tag{5}$$

The RMSE values for this approach can be found in Table 8. Using the mean of the differences encourages an assumption that no matter how different individuals are, and how diverse their background knowledge is, there are still some similarities in their reasoning.

Individual differences. We are once again focusing on the differences in the probability values between tasks among participants, but now we are taking into consideration the individual differences. Here we are given the probability values for all tasks in the reduced inference/conditional case, and the probability values for one task in the conditional/biconditional case. Using that information, we aim to predict the endorsements for the other tasks in the conditional/biconditional case. In this case we need 2×3 (3 parameters for 2 tasks, reduced inference/conditional) + 3 (3 parameters for 1 task, conditional/biconditional) → 9 parameters. Eq. 6 shows how the calculations of the probability values for the conditional case are done, and Eq. 7 shows the same for the biconditional case.

$$p_i^c(t2) = p_i^c(t1) - (p_i^r(t1) - p_i^r(t2)) \tag{6}$$

$$p_i^b(t2) = p_i^b(t1) - (p_i^c(t1) - p_i^c(t2)) \tag{7}$$

Table 8: Prediction results when using constants to obtain probability values. 'Task' is the task to which we fit the model and use to predict the other three tasks. (D/A - Disablers/Alternatives; F - Few; M - Many; Red. Inf. - Reduced Inference)

	Red. Inf.	Conditional	Biconditional
Task (D/A)	RMSE	RMSE	RMSE
F/F	.245	.125	.231
F/M	.260	.118	.191
M/F	.226	.136	.200
M/M	.210	.099	.181

E.g., if we have fitted all the parameters for the reduced inference, and the 'Girl' task in the conditional case, and we want to predict the probabilities for the 'Coke' task in the conditional case, we would follow Eq. 8.

$$p_i^c(c) = p_i^c(g) - (p_i^r(g) - p_i^r(c)) \tag{8}$$

Table 9: Prediction results when using individual task differences to calculate probability values. 'Conditional' and 'Biconditional' denote predicting for that condition. (D/A - Disablers/Alternatives; F - Few; M - Many)

	Conditional	Biconditional
Fitted task (D/A)	RMSE	RMSE
F/F	.336	.255
F/M	.327	.238
M/F	.222	.336
M/M	.263	.203

The RMSE for this approach can be found in Table 9. This approach makes the assumption that there are similarities in the individual differences of probability values between tasks for all conditions.

Condition probability distribution differences

In this approach we focus on how the probability values change between the reduced inference and the conditional case and between the conditional and biconditional case. We are aiming to predict a task in the conditional/biconditional case, by fitting another task in both, the reduced inference and conditional/conditional and biconditional case, and the to-be predicted task in the reduced inference/conditional case, which totals to 9 parameters ($2\times3+3$). We will only take into consideration individual probability differences. Eq. 9 and 10 show how the calculations of the probability values are done.

$$p_i^c(t2) = p_i^r(t2) - (p_i^r(t1) - p_i^c(t1)) \tag{9}$$

$$p_i^b(t2) = p_i^c(t2) - (p_i^c(t1) - p_i^b(t1) \tag{10}$$

E.g., if we have fitted the probability values for the 'Balloon' task in the conditional and biconditional case, and the 'Coke' task in the conditional case, we can calculate the probability values for the 'Coke' task in the biconditional by 11.

$$p_i^b(c) = p_i^c(c) - (p_i^c(b) - p_i^b(b)) \tag{11}$$

Table 10: Prediction results when using individual condition differences to calculate probability values. 'Fitted task' is the task that is fitted in both conditions, whose parameter differences are used to predict other tasks. (D/A - Disablers/Alternatives; F - Few; M - Many; Red. - Reduced Inference; Cond. - Conditional; Bicond. - Biconditional)

	Red. to Cond.	Cond. to Bicond.
Fitted task (D/A)	RMSE	RMSE
F/F	.203	.152
F/M	.283	.261
M/F	.444	.167
M/M	.297	.322

The RMSE values for this approach can be found in Table 10. This approach makes the assumption that the individual differences of probability values when changing the type of rule are similar among different tasks.

Future work

In this paper we presented a combination of the Mental Model Theory and Pearl's ε-semantics. It is able to account for the influence of disablers and alternatives and the type of conditional. Using three parameters per task, we achieved a good fit. It is a starting point that will need more exploration to bring different cognitive computation theories closer together.

Acknowledgements

This work was supported by Heisenberg grants RA1934/3-1, RA1934/4-1, and RA1934/9-1 to MR.

References

Alexander, D., Tropsha, A., & Winkler, D. A. (2015). Beware of R^2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling*, *55*(7), 1316–1322.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19(3)*, 274–282.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning*, *19*(3-4), 249–265.

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

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: LMC*, *26(4)*, 883.

Pearl, J. (1991). *Epsilon-semantics* (Tech. Rep.). Computer Science Department, University of California.

Singmann, H., Klauer, K. C., & Beller, S. (2016). Probabilistic conditional reasoning: Disentangling form and content with the dual-source model. *Cognitive Psychology*, *88*, 61-87.