LNAI 11669

Kouzou Ohara
Quan Bai (Eds.)

Knowledge Management and Acquisition for Intelligent Systems

16th Pacific Rim Knowledge Acquisition Workshop, PKAW 2019
Cuvu, Fiji, August 26–27, 2019
Proceedings

Springer

Check for updates

A Cross-Domain Theory of Mental Models

Sara Todorovikj(✉), Paulina Friemann, and Marco Ragni

Cognitive Computation Lab, University of Freiburg, 79110 Freiburg, Germany
{todorovs,friemanp,ragni}@cs.uni-freiburg.de

Abstract. Cognitive models for human reasoning are often specialized and domain-specific. So the question whether human reasoning across domains shares the same (or at least a similar) mental representation and inference mechanism is still an unexplored territory, as is the endeavor to create cognitive computational models for multiple domains of human reasoning. In this paper, we consider the theory of mental models for conditionals as a test-case and aim to extend it towards syllogistic reasoning using a formal translation. The performance of this new cross-domain theory is comparable to the performance of state-of-the-art domain-specific theories. Potentials and limitations are discussed.

Keywords: Predictive modeling · Human reasoning · Conditionals · Syllogisms · Mental models

1 Introduction

How do people produce conclusions from prior information? Humans do not always follow the steps proposed by formal logic, and often make logical errors [14]. One of the goals of cognitive science is to have a better comprehension of the way that humans reason. One mean of doing so is by developing cognitive models that would account for the errors and ultimately predict the way an individual would reason. That is very important for predicting human behavior, which, in turn, helps with successful interaction and collaboration with intelligent systems.

Highly specialized cognitive models of human reasoning are developed for various domains, e.g.: conditional [8,12], syllogistic [9], spatial [3]. These cognitive models tackle the specific reasoning domain for which they are designed using different approaches, such as heuristics or probabilistic updating mechanisms. A generalizability to explain human reasoning in a different domain is often not given. However, the question whether human reasoning processes share specifics across domains is still an open question.

© Springer Nature Switzerland AG 2019
K. Ohara and Q. Bai (Eds.): PKAW 2019, LNAI 11669, pp. 184–194, 2019.
https://doi.org/10.1007/978-3-030-30639-7_16

Example 1. Conditional and syllogistic reasoning problems:

Conditional	Syllogism
If the number on the card is 3, then the card is colored red. The number on the card is 3.	Some artists are bakers. All bakers are chemists.
Therefore, the card is colored red.	Therefore, some artists are chemists.

A widely acknowledged account of human reasoning is the Mental Model Theory (MMT) [5–7]. The MMT suggests that humans construct mental models of the given information, and inspect and possibly manipulate them mentally to reach a conclusion. We consider the MMT-based cognitive model for conditional reasoning proposed by [8], and develop a generalization for syllogistic reasoning based on a conditional reformulation of the quantifiers. Similar work has been done by Bara et al. [1], where they created a computational model, also based on mental models, which is used to make predictions in the syllogistic, relational and propositional domain. They make an assumption that individuals always construct the correct mental representation of the premises that they are given. Since that assumption is rather unlikely to be true, we have a different approach – we try to reverse engineer the construction of the mental representation, by adapting to the individuals' responses. That way we also account for logically erroneous representations individuals might construct, and even different interpretations of the same premise by the individuals, e.g., the case when some individuals interpret a conditional as a bi-conditional.

The remainder of this article is structured as follows: First, we give a brief introduction to conditionals and syllogisms, followed by an introduction to the relevant points of the Mental Model Theory and its application in both domains, conditional and syllogistic. A proposal for a generalized cross-domain model follows. Finally, we present the results in a prediction task for both domains.

1.1 Conditionals

Conditionals are statements, usually of the form "If X then Y" (also written as $X \rightarrow Y$, where X is called the antecedent, and Y, the consequent), often used to describe a causal relationship between any two propositions. In this paper,

Table 1. The four inference forms for "If X then Y" (short: $X \rightarrow Y$)

Inference form	Conditional	Minor premise	Conclusion
Modus Ponens (MP)	$X \rightarrow Y$	X	Y
Modus Tollens (MT)	$X \rightarrow Y$	$\neg Y$	$\neg X$
Affirmation of the Consequent (AC)	$X \rightarrow Y$	Y	X
Denial of the Antecedent (DA)	$X \rightarrow Y$	$\neg X$	$\neg Y$

we also consider the form "If X then *possibly* Y". Research on conditional reasoning often relies on acceptance rates for the four inference forms (see Table 1) that can follow when given a conditional along with a minor premise. The conditional in Example 1 is an example of the logically valid *Modus Ponens* (MP). The other logically valid inference form is the *Modus Tollens* (MT). Logically invalid inference forms are *Affirmation of the Consequent* (AC), and *Denial of the Antecedent* (DA).

1.2 Syllogisms

Syllogisms are quantified assertions consisting of two premises and a conclusion. They are used to reason about properties of entities by using quantifiers. Here we take into consideration only the standard quantifiers: "All", "Some", "Some not" and "No". The syllogisms have been a popular psychological research target for over 100 years [9], and their general analysis goes all the way back to Aristotelian times. Example 1. gives an example of a syllogism.

Table 2. Moods of premises and figures of a syllogism.

(a) Moods of a premise/conclusion.

Mood	Premise
Universally affirmative (A)	All X are Y
Particular affirmative (I)	Some X are Y
Universal negative (E)	No X are Y
Particular negative (O)	Some X are not Y

(b) Figures of a syllogism.

Figure	Premise 1	Premise 2
1	X-Y	Y-Z
2	Y-X	Z-Y
3	X-Y	Z-Y
4	Y-X	Y-Z

Each premise in a syllogism (and the conclusion) can be in one of the four moods shown in Table 2a. The research done on syllogistic reasoning is focused on acceptance rates for the possible conclusions. Conclusions contain two terms, a subject and a predicate ('artists' and 'chemists' in Example 1). The premises contain the conclusion's subject and predicate and relate them to a middle term, that appears in both premises ('bakers'). A syllogism can have four different figures, based on the order of the terms, as shown in Table 2b. In the syllogism example in the introduction, the moods of the two premises are I, and A, respectively. The order of the terms in the premises is: X-Y, and Y-Z, which is Fig. 1. This syllogism is of type IA1. The conclusion is in the mood I, and the order of terms X-Z, which we denote as IXZ.

1.3 Mental Model Theory

The theory of mental models [5–7] is a cognitive theory that assumes that an individual reasoner constructs an analogous mental representation of the state of affairs. A reader more familiar with formal logic may think of a truth-table like representation that is iconic [5]. Let us consider first a mental model for the conditional "If X then Y". According to the mental model theory (MMT), a reasoner who processes this information constructs an initial mental model consisting of the antecedent X and the consequent Y, i.e., the reasoner represents first the case where both are true (and nothing is false): the model X Y. Other possible interpretations, e.g., the case where X is false (written as $\neg X$) are abbreviated by an ellipsis.

Premise	Mental model	Fleshed-out models
If X then Y	X Y	X Y
	...	¬X ¬Y
		¬X Y

The mental model represents what is true according to the information, but not what is false. This can be fleshed-out in a second process, leading to all possible interpretations (see right column). This can explain why the inference processes MP and AC can be immediately inferred, while for MT and DA the flesh-out process is necessary. The processes of the Mental Model Theory for Conditional Reasoning have been formalized as a multinomial process tree by Oberauer [12] (see Fig. 1)[1]. The MPT can be interpreted as a binary decision diagram with parameters on its edges (see Fig. 1). Specifically, in the case of MMT,

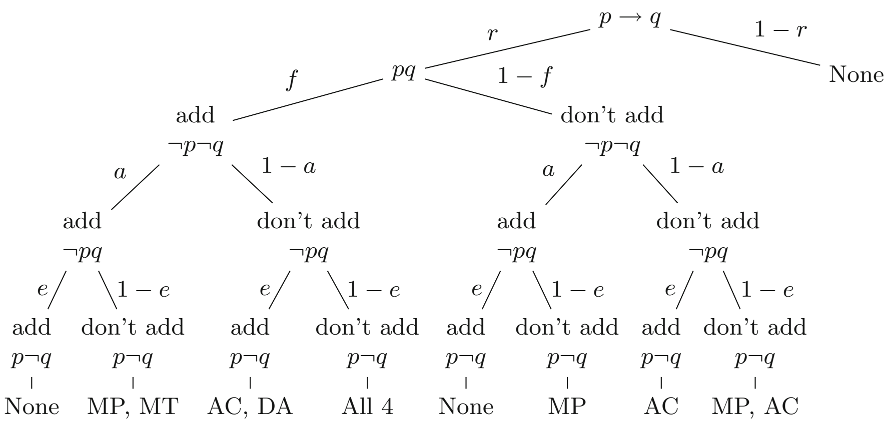


Fig. 1. Oberauer's formalization of the MMT [12] for the conditional "If p then q". The parameters r, f, a, e take on values in the interval $[0, 1]$, indicating the probability of taking the respective decision path in the model. The leafs represent the responses.

[1] In the original model by Oberauer [12], the parameter $1-r$ describes the probability that an individual will not reason, but guess. In our implementation of the model we do not use guessing.

the decisions correspond to whether a human reasoner will add a certain model to their mental representation of the conditional or not. Given a conditional "If p then q", there are four possible mental models that an individual can add to their mental model representation of the conditional (pq, $\neg p \neg q$, $\neg pq$, $p \neg q$). For each one of those mental models, there is a certain parameter that describes the probability of that model being added to the mental representation.

Individuals aim to maintain the information that is provided to them with the conditional or syllogism and try to reach a conclusion based on that. Often, individuals would engage in a search for counterexamples. If their search is successful, the conclusion is no longer accepted by the individual [9]. Quantifiers are interpreted by the representation of single entities, representing the respective set. Consider the premise "All X are Y".

Premise	Mental Models
All X are Y	X Y X Y X Y Y ...

Again, in a mental model, X and Y means that both properties are true, and each line represents an entity and contains the properties which are true for this entity. The first three rows represent a set of entities which are described by the properties X and Y, whereas the fourth row represents an entity described only by Y, but not by X. In order to represent a set of entities that have the same properties, three entities are used[2]. For a second premise, another mental model is constructed and integrated into the original. Consider the example in Table 3 "All X are Y" and "Some Y are Z". The final mental model yields the conclusion *"Some X are Z"* (IXZ), or *"Some X are not Z"* (OXZ). There is no counterexample to either of these conclusions.

Table 3. Representation of the syllogism AI1 using mental models

Premise 1	Mental Models	Premise 2	Mental Models	Combined Models
All X are Y	X Y Y	Some Y are Z	Y Z Y Y	X Y Z X Y Y Y

[2] As discussed by Johnson-Laird, no iconic model can show that it represents an entire set, we have no way of knowing whether a model describes the whole set, or just a small number of entities that belong to it [5].

A Common Ground. There is a close relation between conditional premises and the premises of syllogisms [2]. Consider the conditional premise "If it is a dog, then it is a mammal", and the syllogism's premise "All dogs are mammals", their representation using mental models is equal[3].

$$\begin{array}{c}\text{dog mammal}\\ \dots\end{array}$$

We will build upon this in the following cross-domain modeling.

2 Cross-Domain Reasoning

Our goal is to demonstrate how a cross-domain cognitive theory can be built for conditional and syllogistic reasoning. We take an already existing formalized approach for reasoning with conditionals using mental models (see Fig. 1) and extend its application to syllogisms.

Note that p and q in Fig. 1 stand for the antecedent and the consequent respectively, which can take different forms in the concrete application. For example, for the syllogism AI1 (cf. Table 3), in the first premise, "All X are Y", p corresponds to X and q corresponds to Y, whereas for the second premise, "Some Y are Z", p corresponds to Y and q corresponds to Z.

Example 2. Adding models:

"If X then Y" $\xrightarrow{\text{add } pq}$ $X\ Y$ $\xrightarrow{\text{add } \neg pq}$ $\begin{array}{c}\text{X Y}\\ \neg\text{X Y}\end{array}$ (p being X, and, q being Y)

"If Y then possibly Z" $\xrightarrow{\text{add } pq}$ Y Z $\xrightarrow{\text{add } p\neg q}$ $\begin{array}{cc}\text{Y} & \text{Z}\\ \text{Y} & \neg\text{Z}\end{array}$ (p being Y, and, q being Z)

In the following we describe how we apply the principle of adding models, as described above, to syllogisms.

Step 1: Translation of syllogistic premises to a pair of conditionals. The translations of all the possible moods of a syllogism's premise to a conditional premise, based on their equal representation using mental models, are shown below:

Mood	Premise	Conditional
A	All X are Y	If X then Y
I	Some X are Y	If X then possibly Y
E	No X are Y	If X then not Y
O	Some X are not Y	If X then possibly not Y

This process yields two conditional statements which can be treated as conditional reasoning problems, and can therefore be modeled using the formalization in Fig. 1.

[3] From now on we use a compressed version of the models representing sets of the same entity, i.e., one unique entity only, as we do not consider quantifiers like "Most".

Step 2: Obtaining mental model representations. Below, we show a representation of the conditionals that describe syllogism AI1. This corresponds to adding the *pq* model for the first premise in the syllogism, and adding the *pq* and $p\neg q$ models for the second one.

Syllogistic Premise	Conditional	Mental Models
All X are Y	If X then Y	X Y
Some Y are Z	If Y then possibly Z	Y Z Y

Step 3: Merging. After obtaining the two mental models representing the two premises, the next step is to merge them based on the middle term Y which appears in both premises, in order to construct the final mental model representation. In the table above we have the model X Y for the first premise. We will merge that model with all models of the second premise that also contain Y. The first such model is Y Z. Merging based on Y, we obtain X Y Z. The second model is Y, leading to a merged representation X Y. We obtain the following mental model representation of the syllogism: $\begin{array}{l}\text{X Y Z}\\ \text{X Y}\end{array}$

Step 4: Answer prediction. Since what we are interested in is the relation between X and Z, once the full representation is obtained, we do not take Y into consideration anymore (e.g. we would only consider $\begin{array}{l}\text{X Z}\\ \text{X}\end{array}$ from the representation shown above). Also, model duplicates are eliminated, and models with both elements being negative are not considered (¬X ¬Z). Based on this representation, a prediction of the individual's answer about what follows from the given syllogism can be made. Possible final representations and the corresponding answers can be found in Table 4. In the cases where there is more than one choice for a possible answer, one is chosen randomly.

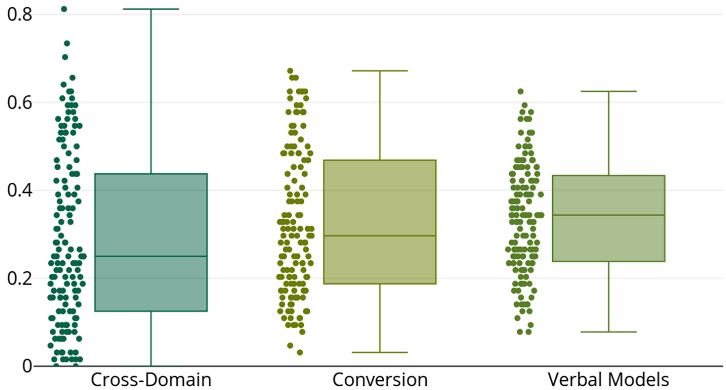
Table 4. Final representations and corresponding possible answers.

	Number of unique models in the representation							
	0	1			2			3
Models	-	X Z	X (¬Z)	(¬X) Z	X Z (¬X) Z	X Z X (¬Z)	(¬X) Z X (¬Z)	X Z (¬X) Z X (¬Z)
Answers	NVC	AXZ AZX	EXZ	EZX	IZX OZX	IXZ OXZ	NVC	NVC

Note. NVC – No valid conclusion can be drawn from the premises.

3 Results

To test, how good our cross-domain model is able to perform, we used the CCOBRA-framework benchmarking tool[4] which provides empirical benchmark data from psychological experiments for cognitive models in different domains. A cognitive model can be trained on specific training data, and needs to predict for each individual reasoner her putative conclusion for the respective inference problem. In a first step, we trained our model on the conditional data set which yielded the probabilities $P(pq) = 1.0$, $P(\neg p \neg q) = .65$, $P(\neg pq) = .5$, $P(p\neg q) = .05$. In the conditional domain, using these parameters yields an accuracy of 63%, which is only one percent lower than the best performing model (see Table 5b). Applying the same parameter distribution achieved an accuracy of about 21% on the syllogistic data. Therefore, four further parameter optimizations were performed in the syllogistic domain, based on different criteria, in order to examine the relation between mental models and type of syllogism. The criteria taken into consideration are mood and figure of the syllogism, and whether the currently analyzed premise is the first or second one. The last optimization was using no specific criterion, i.e. optimized the four parameters $(P(pq), P(\neg p\neg q), P(\neg pq), P(p\neg q))$ for all criteria at once. For the parameter optimization based on mood and figure, four separate parameters were fit for each mood or figure, respectively, totaling to 16 parameters. In the case of premise number, there is a total of 8 parameters, four parameters for each premise. Optimization was done with a randomized search with 10^6 iterations on values in the interval $[0, 1]$ with an increment of 0.1. All model specific optimizations had only an average accuracy of at most 27% which is lower than the general optimization. For the general optimization, i.e., without differentiation of mood, figure or premise number, we conducted a grid search with the same specifics. The best parameter values for the general optimization were $P(pq) = 1.0$, $P(\neg p\neg q) = 1.0$, $P(\neg pq) = .2$, $P(p\neg q) = 1.0$



[4] orca.informatik.uni-freiburg.de/ccobra/.

The model was tested once again on the conditional data set using these parameters, and its performance dropped down to an accuracy of 38% (compared to the initial 63%). In Table 5, the results of running the Cross-Domain model on a syllogistic and conditional benchmark are shown.

Table 5. Prediction results, given as percentage of correct predictions, and best and worst predictive accuracy for individual reasoners.

(a) Selection of Syllogistic Models

Model[a]	Accuracy	Best	Worst
Verbal [13]	34%	61%	7%
Conversion [15]	32%	67%	3%
Cross-Domain	**29%**	**81%**	**0%**
PSYCOP [16]	29%	70%	6%
Atmosphere [17]	24%	44%	4%

(b) Selection of Conditional Models

Model	Accuracy
Probabilistic [11]	64%
Dependence [10]	64%
Cross-Domain	**63%**
Independence [10]	63%
Suppositional [4]	62%

[a] The models are taken from the meta-analysis by Khemlani and Johnson-Laird [9].

4 Discussion and Conclusion

Our motivation for our paper was the question, if human reasoners employ a similar or a different model representation for conditional and syllogistic reasoning. To analyze this, we used a computational formalization of the cognitive theory of mental models for human conditional reasoning and extended it towards reasoning with syllogisms, by translating quantified assertions into conditionals. While the fitting of our cross-domain model for conditionals demonstrated a performance comparable to other domain-specific models, the initial transfer of the respective parameters yielded just 21% prediction performance of the cross-domain model for the syllogistic data. By fitting it to the syllogisms, the model reached 29%, with several participants fitted better than by the two best-predictive cognitive models – the Verbal and the Conversion model.

What have we learned? It is possible to bridge the gap between domain-specific cognitive models by using a theory-preserving translation. It is even possible to outperform domain-specific models for some individuals. The cognitive processes and interpretations related for each quantifiers are, however, different. This may mean that the internal representations (for our mental model) are not the same, hence, a general unified theory of human deductive reasoning is still an open question. For the general optimization process, three out of four parameters have the value 1.0, which means that the three corresponding models are always added and only the model $\neg pq$ is added only with a probability of 0.2. This needs to be further investigated empirically. Our results indicate, however, that the outlined idea needs to be applied to other cognitive theories as well, such that the cross-domain power (or generalizability) of approaches can be better estimated. For this a general theory of how to generalize a cognitive model across domains is necessary and identifying general principles across formal and cognitive theories might be a first step. Still, this work is only a first try towards successful cognitive modeling of cross-domain human reasoning.

To summarize: Our model has a predictive accuracy comparable to state-of-the-art cognitive models in both domains. However, our model is capable of modeling human reasoning in two different domains, whereas the rest of the models in the benchmark are highly specialized, domain-specific cognitive models. On the basis of our model lies a reduction of the two tasks to a common interpretation. This made it possible to compare the mental representations.

While cross-domain data of individual reasoners is rare, in the future, this type of modeling should be performed on a data set where the same individual gives responses to both, conditional and syllogistic tasks. This way, we can learn more about individual differences in reasoning, which would aid in a more successful simulation of the human mind.

Acknowledgments. This paper was supported by DFG grants RA 1934/3-1, RA 1934/2-1 and RA 1934/4-1 to MR. We are also grateful to Lukas Elflein for helpful comments.

References

1. Bara, B.G., Bucciarelli, M., Lombardo, V.: Model theory of deduction: a unified computational approach. Cogn. Sci. **25**(6), 839–901 (2001)
2. Bucciarelli, M., Johnson-Laird, P.N.: Strategies in syllogistic reasoning. Cogn. Sci. **23**(3), 247–303 (1999)
3. Byrne, R.M., Johnson-Laird, P.N.: Spatial reasoning. J. Mem. Lang. **28**(5), 564–575 (1989)
4. Evans, J.S.B., Over, D.E.: If: Supposition, Pragmatics, and Dual Processes. Oxford Cognitive Science, Oxford (2004)
5. Johnson-Laird, P.N.: How We Reason, 1st edn. Oxford University Press, Oxford (2006)
6. Johnson-Laird, P.N.: Mental models in cognitive science. Cogn. Sci. **4**(1), 71–115 (1980)
7. Johnson-Laird, P.N.: Mental models and human reasoning. Proc. Nat. Acad. Sci. **107**(43), 18243–18250 (2010)
8. Johnson-Laird, P.N., Byrne, R.M.: Conditionals: a theory of meaning, pragmatics, and inference. Psychol. Rev. **109**(4), 646–678 (2002)
9. Khemlani, S., Johnson-Laird, P.N.: Theories of the syllogism: a meta-analysis. Psychol. Bull. **138**(3), 427–458 (2012)
10. Oaksford, M., Chater, N.: A rational analysis of the selection task as optimal data selection. Psychol. Rev. **101**(4), 608 (1994)
11. Oaksford, M., Chater, N., Larkin, J.: Probabilities and polarity biases in conditional inference. J. Exp. Psychol. Learn. Mem. Cogn. **26**(4), 883 (2000)
12. Oberauer, K.: Reasoning with conditionals: a test of formal models of four theories. Cogn. Psychol. **53**(3), 238–283 (2006)
13. Polk, T.A., Newell, A.: Deduction as verbal reasoning. Psychol. Rev. **102**(3), 533 (1995)

14. Ragni, M., Kola, I., Johnson-Laird, P.N.: On selecting evidence to test hypotheses: a theory of selection tasks. Psychol. Bull. **144**(8), 779 (2018)
15. Revlis, R.: Two models of syllogistic reasoning: feature selection and conversion. J. Verbal Learn. Verbal Behav. **14**(2), 180–195 (1975)
16. Rips, L.J.: The Psychology of Proof: Deductive Reasoning in Human Thinking. MIT Press, Cambridge (1994)
17. Sells, S.B.: The Atmosphere Effect: an Experimental Study of Reasoning. Archives of Psychology. (Columbia University), New York (1936)