Preferred Inferences in Causal Relational Reasoning: Counting Model Operations

Marco Ragni (ragni@cs.uni-freiburg.de)
Foundations of AI, Technical Faculty, University of Freiburg and
Justus Liebig University, Experimental Psychology and Cognitive Science, Giessen
Stephanie Schwenke (schwenke@neptun.uni-freiburg.de)
Center for Cognitive Science, University of Freiburg
Christine Otieno (otieno@psychologie.uni-freiburg.de)
Department of Psychology, University of Freiburg

Abstract

Interpreting causal relations plays an important role in everyday life, for example in scientific inquiries and text comprehension. Errors in causal reasoning can be a recipe for disaster. Despite vast literature on the psychology of human causal reasoning, there are few investigations into preferred inferences in relational three-term problems. Based on a previous formal investigation about relevant causal relations we develop a cognitive modeling approach with mental models. The key principle for this approach proves to be the prediction of preferred inferences by model operations and the process of sub model integration. Subsequent experiments test preferred inferences, the number of model operations, and if concrete or generic problems make a difference in causal reasoning performance. Implications of the model are discussed.

Introduction

Causal reasoning has often been studied in both Artificial Intelligence and with human reasoners (e.g., Russell & Norvig, 2003). How we, as humans, reason and draw inferences is quite different from classical AI approaches. Nevertheless, being able to correctly interpret causal relations is an important everyday skill. For example, a production manager must be able to understand the complex interdependencies of supply chains in order to identify delivery bottlenecks in time. This way solutions can be found before serious supply problems occur in the production process. Consider another example: A clear stream flows through green countryside, its route is winding going this way and that, even splitting eventually into two creeks, as shown in Figure 1.

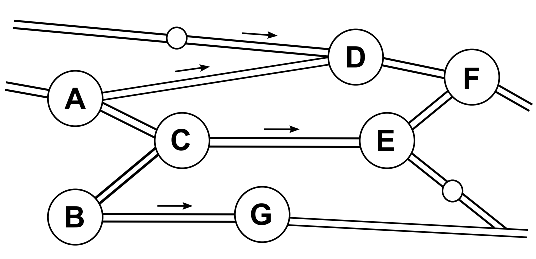

Figure 1: A river flow example.

One day people begin to notice dying cattails and reeds while picnicking at the park (point D). Downstream a farmer whose pasture (point F) is divided by the stream notices that several of his cows are sick. Something has polluted the water. There is much speculation about the culprit: A new factory (point B), the old tannery (point A), something even farther upstream? How can the townspeople determine what is polluting their stream? What areas should the townspeople investigate to glean the most information? If the new tannery (at A) caused the pollution at the park (point D), then it probably also caused the duck family to leave their home (at point C). However, the pollution is probably not to blame for the sheep missing from the green (point G, downstream from the factory). Why? If you walk upstream from the park or the ducks' home, you'll eventually get to the same point. Walking upstream from the green, will get you somewhere else.

When reasoning about causes and effects, elements can have various relationships to one another. One element may cause another (point A comes before point D in the river example; represented by $\prec$); conversely, one element may be caused by another (point D comes after point A; represented by $\succ$) – these are called dependent relations. Another type of dependency, like we saw with points D and C, is also thinkable; we call this relationship in which two elements share a common ancestor or cause fork (and represent it with $\curlywedge$). Two elements can also be independent of each other (like points B and D in the example, represented by $\asymp$). There are cases in which not all dependencies are known. For example we can imagine a fishing hole some ways distant from our stream and ask if the fishing hole is connected to the stream by water flowing underground? In these instances imprecise knowledge is described by unions of basic relations (in our example the fishing hole is $\succ$ or $\asymp$ of the tannery). Dependencies of probabilities (when observations depend on each other, as with our stream) are often described by a Bayesian network (Pearl, 2000). Looking at the structure of the network can reveal to us whether two random variables are independent. Orders (relationships between elements) may remain partly unspecified between elements when it does not matter. For example, we may not know whether the fishing hole is downstream, upstream or not at all connected to the park, but it doesn't really matter to us either.

In other cases the specific order is vital and is necessary to deduce the hidden truth. Causal networks are not just important for deducing sources of pollution, complex dependencies also play a role in identifying delivery bottlenecks in supply chain networks, minimizing delays in railways systems, and in inhibiting the spread of diseases. This last example pro-

Table 1: Relations of causal reasoning and their partial order definitions (cf. Ragni and Scivos, 2005).

Relation	Name	Partial Order
$X \prec V$	X causes V	$X \leq V$ and not $V \leq X$
$X \succ V$	X depends on V	$X \leq V$ and not $V \leq X$
$X \curlywedge V$	X and V have only a common cause	$\exists C\ C \leq V \wedge C \leq X$ and neither $X \leq V$ nor $V \leq X$
$X \asymp V$	X and V are independent	neither $\exists C\ C \leq V \wedge C \leq X$ nor $X \leq V$ nor $V \leq X$

Note. The relation "has a common cause" requires the introduction of an additional variable which we call C.

vides a situation in which specific orders are important: To stop a contagious disease it is desirable to identify a patient zero (the origin of the disease) and also know whether patients have had contact with one another. However, this is not always clear. Reasoning based on known dependencies can help detect formerly unknown causes or rule out possibilities, thus limiting their number.

While causal reasoning has often been examined with the assumption of base rates or probabilities, these are not always given; in the example above the townspeople do not know how probable it is that their stream will become poisoned, that cows will get sick because of pollution, etc. Syllogistic and relational reasoning have often been studied using three term series problems. This paradigm can be extended to causal reasoning. Consider the problem that emerges looking at some of the facts the townspeople have (the letters representing the facts are used here for the sake of simplicity):

G and C are independent of each other.
C and D have a common cause.

May we assume that G and D are independent of each other? The answer(s) to this problem, and how the average townsperson might solve it are explained in depth in the next section. Before moving on, let us consider an AI approach to the problem. In order for reasoning to be automated, a calculus that expresses the relation between a pair of nodes is needed to help distinguish between possibly affected areas and those that are unaffected. Do such efficient algorithms exist that can detect causes and implications, discover dependencies and find an order compliant with a given specification (i.e., was it actually necessary for the townspeople to invest so much time in investigation, or could a computer have solved their problem for them)? One such calculus, the dependency-calculus has been formally investigated. It can represent and deduce knowledge about dependencies and causal reasoning and can be formally grounded by extending the language of partial ordering. The calculus is NP-complete and all tractable subclasses have been identified (Ragni & Scivos, 2005). This calculus is useful in various applications dealing with reasoning about spatial, temporal, spatio-temporal, topological, competitive, and causal relations. In the next section we will review the theory of partial orders and define causal relations based on them. We will also extend the Theory of Mental Models to deal with preferred inferences. To more thoroughly investigate the theory and its extension, we present two experiments on human reasoning with causal relations and identify preferences in the next section. We will conclude by summarizing our results and raise questions left open.

Causal reasoning, partial orders, and theories

Causal relations can be formally described by an extension of a partial order: A *partial order* is a relation $\leq$ that satisfies the following three properties for any x, v, z: *reflexivity* ($x \leq x$), *antisymmetry* (if $x \leq v$ and $v \leq x$, then $x = v$), and *transitivity* (if $x \leq v$ and $v \leq z$, then $x \leq z$).[1]

We can base our interpretation of causal relations on the notion of partial order relations. We consider four basic relations, which we denote by $\prec, \succ, \curlywedge, \asymp$. If x, v are points in a partial order $\langle T, \leq \rangle$, then we define these relations in terms of the partial order as follows: Semantically, a constraint $v_1\ R\ v_2$ holds in a partial order $(T, \leq)$ that complies with this definition. All relations between nodes in Fig. 1 can be described by these basic relations. To describe trees, the relations $\{\prec, \succ, \curlywedge\}$ are sufficient. Therefore, whenever the relation $\asymp$ occurs, the graph cannot be a tree and requires separate representations.

It is possible to show that reasoning with such causal relations $\{\prec, \succ, \curlywedge, \asymp\}$ forms a relation algebra (Ladkin & Reinefeld, 1992), i.e., it contains all unions, intersections, and complements of a set of basic relations, and composition and inverse operations (Ragni & Scivos, 2005).

Transitive inferences are represented by applying the composition operator on two premises for the four causal relations $\prec, \succ, \curlywedge, \asymp$ and the associated composition as shown in Table 2. An interesting question is whether the relation "independent" ($\asymp$) provides additional complexity in comparison to classical partial orders.

A computational model

The graphical representation of causal relations (e.g., Figure 1) is not sufficient to explain preferred inferences in human reasoning. A systematic approach to representing causal reasoning by mental models was introduced by Goldvarg and Johnson-Laird (2001). They developed mental models for the four causal relations: causes, allows, prevents, and does not allow. To accommodate a few more relations (like those we saw above), we propose the following extension found in Table 3:

[1] Note that a total order is a partial order that satisfies a fourth property: *Comparability* (for any x, v, either $x \leq v$ or $v \leq x$).

Table 2: The formal composition table for transitive inferences with causal relations.

P1/P2	$\prec$	$\succ$	$\asymp$	$\curlywedge$
$\prec$	$\prec$	$\prec, \succ, \asymp, \curlywedge$	$\asymp$	$\prec, \curlywedge, \asymp$
$\succ$	$\prec, \succ, \curlywedge$	$\succ$	$\succ, \asymp, \curlywedge$	$\succ, \curlywedge$
$\asymp$	$\prec, \asymp, \curlywedge$	$\asymp$	$\prec, \succ, \asymp, \curlywedge$	$\prec, \asymp, \curlywedge$
$\curlywedge$	$\prec, \curlywedge$	$\succ, \asymp, \curlywedge$	$\succ, \asymp, \curlywedge$	$\prec, \succ, \asymp, \curlywedge$

Note. The leftmost column represents the relation of the first premise (abbreviated by P1, e.g., X $\prec$ V) and the upmost row the relation of the second premise (abbreviated by P2, e.g., V $\asymp$ Z). Each intersecting cell contains all possible relations for the given premises (e.g., X $\asymp$ Z). The relation are: $\prec$ (causes), $\succ$ (caused by), $\asymp$ (independent), $\curlywedge$ (have a common cause).

Table 3: Causal relations and possible mental models.

Relation	Name	Model	
X $\prec$ V	X causes V	X	V
X $\succ$ V	X is caused by V	V	X
X $\asymp$ V	X and V are independent		X
			V
X $\curlywedge$ V	X and V have a common cause	C	X
		C	V

Please note that for the relation "have a common cause" we assume[2] reasoners may use an additional place holder, a variable such as 'C' to represent this internally (recall that a common cause was something like the tannery which polluted water both at the pasture and at the park) .

(P)	*Premise 1:*	X and V are independent
	Premise 2:	V and Z have a common cause
	Conclusion:	X and Z are independent

The words in italics are only given here for representational purposes – they do not typically appear in an experiment. After reading the first premise, a reasoner might construct a mental representation of the form:

X
V

which represents that X and V are independent by having them in no horizontal order (cf. Table 3). According to the representation above, a model for the second premise might look like:

C V
C Z

In a second phase – the premise integration phase – the two representations are combined and form a greater model. The operations of the integration process are relevant and we hypothesize that reasoners try to match identical objects (in the sense of substituting them) in both sub models (in this case the "V"). The goal is to generate a minimal model representation, e.g., to reduce working memory load. We call this integration principle *sub model integration*. This operation leads to the following model (*), which we call the *preferred model*:

(*)
    X
C V
C Z

In this integrated representation the putative conclusion "X and Z are independent" holds (cf. Table 3). X and Z are not in a horizontal linear order (this would imply that Z, if it is to the right of X, is caused by X). Instead they are in vertical order; this implies that they are independent of one another.

From a logical perspective, however, there are other possible conclusions from the premises: "X is cause of Z" and "X and Z have a common cause" (cf. Table 2) are both as possible as "X and Z are independent." These conclusions do not hold in the preferred model and so should be less likely chosen by participants.

What precisely is a *preferred model*? Given a reasoning problem like the one above, a preferred model is reached by submodel integration. It is the easiest model to build, thus reasoners prefer it over models that take more effort to build (more on how other models may be built below). A preferred conclusion may be read from the model, in this case, that X and Z are independent. Practically, this means that we expect participants in experiments to most often choose this preferred conclusion.

As previously mentioned, reasoners can reach non-preferred conclusions. To this end a model revision process might take place, i.e., the participants may try to accommodate the new conclusion in the model by operations. Consider the first conclusion: "X is cause of Z". For this the reasoner would need to revise their preferred model representation (*) such that they can now "read" this conclusion from it. Two operations are necessary (delete the entry X in the first row and add the entry X to the third row):

C V
C X Z

If we consider the alternative conclusion "X and Z have a common cause," we again need two operations (delete the

[2] Note that *formally* it is a base relation, i.e., it excludes that additionally any other relation such as *causes* can hold at the same time. However, in Experiment 2 we found that when causes was the preferred conclusion, participants also considered "have a common cause" true significantly more often than not ($W = 22$, $p < .001$).

entry X in the first row; add the entry X to the row with V).

C V X
C Z

Model operations and manipulations cost time and can increase reasoning difficulty. In this case both alternatives should be rather neglected, and we can predict preferences. We hypothesize that there are two reasons for preferences: How a reasoner integrates the sub models and the number of operations necessary to construct a model for the conclusion. The preferred model is generated by the fewest model operations. Model representations give us an interesting advantage: We can derive reasoning difficulty based on the model operations necessary to construct them. If we construct models then, as outlined, the costs to integrate this information should reflect – at least to some extent – the reasoning difficulty which is a deterministic process (cp. Frosch & Johnson-Laird, 2011). In other words, we can derive from the model theory a principle of sub model integration. A hypothesis is that participants construct a linear ordering of the elements only when this is the only possibility. However, once the premises allow for separate entities (cp. Problem 1 above), they are kept separately and this property is inherited to connected sub models. In the next section we present a study designed to test this hypothesis.

Research Questions

- RQ 1: Do participants show a preference for linear over non-linear orders (i.e., $\prec, \succ$ over $\curlywedge, \asymp$) or vice versa?
- RQ 2: Can reasoning performance be explained by the construction of a preferred model and associated mental costs?
- RQ 3: Do generic or content problems have an effect on reasoning performance?

Empirical Study 1

Participants

We tested seventy-five participants (age: $M = 32.67$; $SD = 10.51$) on an online website (Amazon's Mechanical Turk). They received a nominal fee for their participation.

Material, Design, and Procedure

Participants were randomly assigned to four conditions in a 2×2 design: Type of text (with vs. without explicit causal relations) $\times$ order of tasks (generic vs. concrete tasks first). Participants read about the effects of the Asian ladybug in vineyards and the effects of the Horse-chestnut leafminer on chestnut trees. One version of the texts included explicit causal relations identical to the ones later given in the concrete tasks while the other included all of the content information without explicit causal relations. Each problem consisted of two premises and a putative conclusion. Each problem in the abstract version used the letters X, V, and Z and in each premise and conclusion we systematically varied the relations: cause of, caused by, is independent of, have a common cause based on a former formal analysis (cf. Ragni & Scivos, 2005). All possible combinations of these four causal relations (for two premises and one conclusion) allow for 64 different problems. Of the 64 problems 41 (64%) are correct conclusions (they can be found in the composition table), and for 23 problems the conclusion is incorrect (36 %, these are all relations left out in each associated cell in Table 2). Each participant was presented with 64 problems (whereby half of the problems were generic and half were concrete) consisting of two premises and one conclusion. The participants' task was to determine if, given the premises, the conclusion was possible.

Sentences were sequentially presented one by one: Participants first received premise 1; after pressing the spacebar the first premise disappeared and the second premise appeared. After pressing the spacebar again participants received the putative conclusion. The reason for this sequential order presentation was to require participants to store the information in working memory. This follows the *separate-stage-paradigm* (Potts & Scholz, 1975). Each problem reflected different possible causal relations and had the form of problem P above. For each problem participants were asked to decide if the putative conclusion was possible (correct) or impossible by pressing the keys "C" (correct) or "N" (not correct). Participants were asked to keep their forefingers on these keys.

Results

Research Questions 1 and 2 dealt with the overall preference effect. The derived conclusions and their acceptance rate for each three-term series problem can be found in the composition table (Table 4). Based on how often participants judged each conclusion to be correct, our results indicate that participants seem to prefer the independent relation ($\asymp$) and the fork relation ($\curlywedge$) over the linear relations $\prec$, $\succ$ (RQ1). Furthermore, the overall correctness was higher for a conclusion 'independent' ($\asymp$) vs. all other relations (65.3% vs. 52%, Wilcoxon, $W = 168$, $p < .05$). This is precisely what the Mental Model Theory predicts (the principle of sub model integration and more precisely the number of operations to transform the initial model of the first two premises into one which fits the conclusion). It yields a significant correlation with the accuracy data (Spearman's rho $\rho = .801$, $S = 20,000$, $p < .000001$).

Further support comes from a central prediction of the Mental Model Theory – namely that reasoners are better in single model cases (80.25%) than in multiple model cases (48.25%, Wilcoxon, $p < .01$). In addition to overall preference effects RQ 3 dealt with the influence of content and task order: There was a significant main effect of the order of tasks on reasoning difficulty assessed by accuracy on generic problems ($F(3,71) = 9.00$, $p = .004$; $\eta^2 = .114$) but not for concrete problems (F(3, 71) = 2.00, $p = .162$). That is, generic problems were perceived as significantly more difficult when generic problems were solved first.

Table 4: Composition table with human data (accuracy in percentage) for Exp 1. Relations predicted by our model are in bold.

*Premise*1 / 2	≺	≻	≍	⅄
≺	**≺: 73**	**≺: 60**, ≍: 34, ⅄ : 49, **≻: 58**	**≍: 79**	≺: 59, **⅄ : 75**, ≍: 56
≻	≺: 50 **≻: 50**, ⅄ : 30	**≻: 77**	≻: 40, **≍: 80**, ⅄ : 16	≻: 18, **⅄ : 50**
≍	≺: 29, **≍: 74**, ⅄ : 49	**≍: 92**	≺: 28, **≍: 71**, ⅄ : 23, ≻: 23	≺: 27, **≍: 74**, ⅄ : 31
⅄	≺: 50, **⅄ : 80**	≻: 53, ≍: 31, **⅄ : 55**	≻: 28, **≍: 69**, ⅄ : 43	≺: 49, **≍: 64**, ⅄ : 68, ≻: 39

Note. The leftmost column represents the relation of the first premise (abbreviated by P1, e.g., X ≺ V) and the upmost row the relation of the second premise (abbreviated by P2, e.g., V ≍ Z). Each intersecting cell contains all possible relations for the given premises (e.g., X ≍ Z) and accuracy in percentage.

Empirical Study 2

An additional important aspect in causal reasoning is that some events can prevent others (for example, the stream being polluted will prevent the annual River Festival from taking place). We conducted a second experiment to include this relation and assume the following mental model representation for X prevents V:

Relation	Model
X prevents V	X ¬ V

Participants

30 participants were recruited and tested on Amazon's Mechanical Turk. Before analysis, four datasets were eliminated because participants took an average of less than 2s per problem, so that data from 26 participants was analyzed (19 female; mean age 41 years, SD = 12.5 years).

Material, Design, and Procedure

This experiment worked with the four causal relations: causes, prevents, have a common cause, and independent. Problems were presented as questions of consistency. Three (abstract) assertions were given:

X causes V
V prevents Z
X causes Z

And followed by the question: Can all three assertions be true at the same time? As in Study 1, all 64 combinations of the four relations were presented to participants in a randomized order, in this study, however, only in abstract form (with X, V, and Z). Before beginning the experiment, participants were given two example problems. As previously, the three assertions and the question were given individually and self-paced such that participants had to press the space-bar to receive the next assertion. To answer the questions, participants clicked "Y" for "yes, all three assertions can be true at the same time" and "N" for "no, the three assertions cannot all be true at the same time." At the end of the experiment participants were asked several open-ended questions about the difficulty of the problems and what they interpreted the four relations to mean.

Results

Results from this second experiment offer some further support for the patterns identified in Study 1 with regard to preference effects (RQ 1). Problems in which X and Z are independent was the third assertion were given answers of "yes" significantly more often than any other sort of problem (in 82.2% of cases; Wilcoxon: W = 588, $p < .001$). The second most popular relation in this experiment, however, proved to be prevents which was accepted significantly more often than causes or have a common cause (in 64.9% of cases; Wilcoxon: W = 644, p = .046). Unlike above, there were no significant differences regarding preferences for common cause and causes (58.2% and 58.7% respectively). A second aspect of analysis dealt with preferences wrt. integration strategies (of the first two premises). Where the sub model integration strategy was possible (nine sets of the first two premises), results showed a clear preference for this strategy (84.6% correct) over others (63.1% correct; Wilcoxon: W = 178, $p < .001$). How participants react to the presence of both a term and its negation in one model is an interesting question that arises from the use of prevents. Two possibilities seemed particularly likely: 1) Reasoners would consider the term and its negation to be independent or 2) They would perform a full negation, that is, if X prevents V and V causes Z, then X prevents Z. Neither of these two scenarios could be confirmed or discarded – an almost equal number of problems supported each and this is an interesting matter to consider in future research. The third area of interest was the assessment of the difficulty of the problems. Acceptance of a set of assertions as true (RQ 3) did indeed prove to correlate significantly with the number of operations necessary (non-parametric bootstrapping with 2000 resamples [r = -.58, 95 % CI = (-.82,-.36)]). Closer analysis furthermore revealed that problems that required a modification of the initial model rather than the building of a completely new model, were answered correctly significantly more often (59% vs. 45%; Wilcoxon: W = 30, p = .036). No other significant differences between the model operations were found.

General Discussion

When reading and interpreting texts and when solving logical problems people form mental models by merging information at hand with prior knowledge (Graesser, Singer, & Trabasso, 1994). In text comprehension these mental models are often referred to as situation models (Kintsch, 1998). Humans draw inferences and reason about causes and effects on a daily basis; how we go about this is different from classical approaches in AI. Expanding on classical Mental Model Theory, we analyzed three-term series problems of five causal relations, namely causes, depends on, have a common cause, prevents, and independent. The computational complexity of the associated satisfiability problem is NP-complete (Ragni & Scivos, 2005). This means, in general, the problem of checking if there is a network that satisfies certain conditions is rather difficult. In many cases though, there are polynomial algorithms. Computational complexity is, however, an asymptotic measure, i.e., it makes no testable predictions for three-term series problems and especially does not make any predictions about differences in the preferred causal relations.

How do humans draw inferences for classical three-term-series problems if they use causal relations? We derived a prediction from the way model representations for the causal relations are built: The principle of sub model integration. Recall that sub model integration refers to a sort of "matching" of same elements to build a minimal representation. Whenever the relation "independent" appeared in a problem it seemed to trump the other relations. Participants then avoided constructing minimal representations, that is total orders, although they were possible. In all other cases they tended to perform a model integration leading to preferred answers. The introduced computational model based on the number of operations is a good predictor of reasoning difficulty. This computational model is an adaption of PRISM (Ragni & Knauff, 2013) that proved useful in explaining reasoning difficulty in spatial relational reasoning.

At a first glance, there is an alternative interpretation of the findings based on the surface or the form of the premise, namely the classical atmosphere hypothesis effect (Woodworth & Sells, 1935; Chapman & Chapman, 1959). This effect, however, does not explain the predominance of the two relations independent and have a common cause over linear relations in all but one case. From a modeling perspective models can be differentiated by their compactness, i.e., if an order is partial or total. Although a total order has a higher degree of informativity, as it allows a comparison of the relations between all nodes, it is rarely chosen. This is a consequence that requires further analysis. Performance on interpreting causal relations is also influenced by content: If the content supports the correct conclusion it facilitates performance, but it may also suggest an incorrect conclusion thereby impeding performance (Beller & Spada, 2003). Against the background of these potentially negative content effects and also considering cognitive load principles it seems reasonable to train learners on generic instantiations first before having them solve concrete tasks. An interesting consideration, perhaps especially in the case of content, is what cognitive processes underly model operations. When participants add an element to a model representation they may think about alternative causes for events derived from the context and content. Similarly they may derive further possible elements that prevent events. All these mental operations, however, may cause additional costs within our computational model and may explain reasoning difficulty and why some answers are strongly preferred over others.

Acknowledgments

The research was supported by the DFG with a Heisenberg grant (RA 1934/3-1) to the first author. The authors are grateful to Patrick Junker for developing experimental material, Franz Dietrich for experiment implementation, Sunny Khemlani and three anonymous reviewers for suggestions.

References

Beller, S., & Spada, H. (2003). The logic of content effects in propositional reasoning: The case of conditional reasoning with a point of view. *Thinking & Reasoning*, *9*(4), 335–378.

Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, *58*(3), 220–226.

Frosch, C. A., & Johnson-Laird, P. N. (2011). Is everyday causation deterministic or probabilistic? *Acta psychologica*, *137*(3), 280–291.

Goldvarg, E., & Johnson-Laird, P. (2001). Naive causality: a mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*(4), 565–610.

Graesser, A. C., Singer, M., & Trabasso, T. (1994, July). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395.

Kintsch, W. (1998). *Comprehension : a paradigm for cognition*. Cambridge, U.K. ; New York: Cambridge University Press.

Ladkin, P., & Reinefeld, A. (1992). Effective solution of qualitative constraint problems. *Artificial Intelligence*, *57*, 105–124.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge U.P.

Potts, G. R., & Scholz, K. W. (1975). The internal representation of a three-term series problem. *Journal of Verbal Learning and Verbal Behavior*, *14*, 439-452.

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

Ragni, M., & Scivos, A. (2005). Dependency Calculus: Reasoning in a General Point Algebra. In U. Furbach (Ed.), *KI 2005: Advances in Artificial Intelligence, Proceedings of the 28th Annual German Conference on AI* (p. 49-63). Berlin: Springer.

Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Prentice Hall.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, *18*(4), 451–460.