Title
Processing Spatial Relations: A Meta-Analysis

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Kießner, Ann-Kathrin
Ragni, Marco

Publication Date
2017

Peer reviewed

Processing Spatial Relations: A Meta-Analysis

Ann-Kathrin Kießner (Ann-Kathrin.Kiessner@venus.uni-freiburg.de)
Center for Cognitive Science, University of Freiburg, 79098 Freiburg, Germany

Marco Ragni (ragni@informatik.uni-freiburg.de)
Cognitive Computation Lab, Technical Faculty, University of Freiburg, 79110 Freiburg, Germany

Abstract

The ability to reason about relations is relevant for many spatial cognitive processes. This can involve: (i) to represent spatial information mentally, (ii) to manipulate the spatial representation, and (iii) to infer new spatial information. Several cognitive theories make assumptions and predictions about the underlying processes. A detailed and systematic overview and analysis of ireliable effects across studies is missing. This article presents a meta-analysis of 35 studies about spatial relational reasoning. Studies were classified according to different factors including the *ambiguity of the spatial description*, i.e., if it the description allows for more than one representation, *the presentation of information*, i.e., if the information has been presented auditorily or in a written form, and the *task*, i.e., if a conclusion or model of the premises needs to be generated or verified. Implications of the findings for the mental model theory and working memory are discussed.

Keywords: reasoning; spatial relations; meta-analysis.

Introduction

Spatial cognition allows us to perform a variety of everyday actions, such as sharing spatial information, navigation, and even assembling diverse kinds of objects. A successful spatial interaction requires us to represent spatial relational information and to reason with and about this information. Human communication mainly uses *qualitative* descriptions to specify relationships between spatial objects[1] instead of a numerical or quantitative data description that is used in robot navigation. Relations which are expressed linguistically by the comparative, such as 'greater than', have been extensively investigated in the past century by using behavioral experiments (e.g., Hunter, 1957; Störring, 1908). Within an experiment, problems are often reduced to their essential characteristics limiting irrelevant information. Consider the following example:

Premise 1: The post office is left of the train station.
Premise 2: The train station is left of the main crossroad.
Conclusion: The post office is left of the main crossroad.

The premises contain spatial information ("left of") about the relationships among spatial objects (e.g.,"post office"). A deductive inference makes implicit given information, e.g., the relation between the post office and the main crossroad explicit. While this inference is easy, and most participants solve it correctly, such *transitive* inferences can be at the core of more difficult inference problems with more objects and more relations. In the following we will give a brief overview about reported factors of reasoning difficulty in the literature and present briefly implications of two cognitive theories relevant to our analysis. Cognitive psychologists have disagreed about the exact character of underlying mental processes (Goodwin & Johnson-Laird, 2005) and the main question is, what reliable findings need to be explained by cognitive theories?

Factors of Reasoning Difficulty

In general reasoning difficulty can appear on all levels: in comprehending the presented information (the language level), generating a mental representation (the representational level including working memory) or reasoning about the representation (reasoning level). The literature reports several factors that can be related to these levels.

Behavioral findings support that the *presentation format* affects spatial relational reasoning (Van der Henst & Schaeken, 2005): In comparison with simultaneous premise presentation, accuracy is significantly lower in sequential presentation of the premise information (Roberts & Sykes, 2003; Schaeken & Johnson-Laird, 2000; Van der Henst & Schaeken, 2005). This difference presumably reflects that during simultaneous premise presentation reasoners have all information available until they respond in contrast to sequential premise presentation posing more demand on working memory (Ormrod, 1979; Schaeken & Johnson-Laird, 2000). Models of working memory, e.g., Baddeley's Working Memory Model (WMM; Baddeley, 1986) support the assumption that human reasoning is restricted by the limited capacity of working memory (Klauer, 1997) and that there are different components with specific limitations and modalities. Based on Baddeley's WMM, factors such as presentation form, task type and number of terms may influence spatial relational reasoning. Similar to premise presentation, it can be assumed that the presentation form (written vs. spoken language) may affect the way a problem is processed (Ormrod, 1979).

If premises are presented auditorily the spatial information is presented sequentially and thus more load on working memory is placed (Ormrod, 1979). According to the WMM, the larger the number of terms within a problem, the higher is the amount of information that must be retained (Clevenger & Hummel, 2014). Lastly, the task type may have an influence as well. In conclusion generation tasks reasoners have to generate a conclusion. Whereas during verification tasks, reasoners have to check if a putative conclusions follows. This type of task captures only the ability to recognize a solution, but not producing it and thus requires smaller amounts of work-

[1] Although it would be more appropriate to speak about entities including humans, in the following we refer to spatial objects.

ing memory (Klauer, 1997; Kubinger & Wolfsbauer, 2010).

Another factor is the so-called *indeterminacy effect* (Byrne & Johnson-Laird, 1989). While Example 1 above allows only for one possible arrangement (that we call model) and is called a *determinate problem*, indeterminate problems are possible. Consider the following example:

Premise 1: The post office is left of the train station.
Premise 2: The train station is left of the police office.
Premise 3: The crossroad is right of the post office.
Conclusion: The police office is right of the crossroad.

An indeterminate problem is more difficult to solve than a determinate one (e.g., Boudreau & Pigeau, 2001; Johnson-Laird & Byrne, 1991). Determinate problems allow only one qualitative arrangement (in contrast to quantitative differences, e.g., metric distances), while indeterminate problems allow for multiple different arrangements – as explained by the (preferred) mental model theory (MMT: Byrne & Johnson-Laird, 1989; Ragni & Knauff, 2013):

(1) post office – crossroad – train station – police office

(2) post office – train station – crossroad – police office

Most people construct just one or two models at most and often neglect other models consistent with the premises. These models differ qualitatively, e.g., there is mentally a different arrangement of the train station and the crossroad possible from the indeterminate description.

So far no systematic review of the recent literature has been carried out and no uniform and unambiguous conclusions about differences in accuracy have been drawn. For this reason, a cross-study meta-analysis of behavioural data from spatial relational reasoning is conducted. The aim of this paper is to test whether the predictions of individual studies and the predictions of MMT and WMM hold generally and to give an detailed overview that can serve as a potential benchmark for theories about spatial reasoning. This analysis investigates differences in accuracy depending on indeterminacy, premise presentation, presentation form, and type of task as well as number of terms. Resulting from the theoretical background and empirical findings to spatial relational reasoning, predictions are:

1. In spatial relational reasoning, determinate problems are easier to solve than indeterminate problems (e.g., Boudreau & Pigeau, 2001; Johnson-Laird & Byrne, 1991).
2. Compared with problems presented in spoken language, problems in written language appear with higher accuracy (e.g., Ormrod, 1979; Van der Henst & Schaeken, 2005).
3. Reasoners solve more problems correctly when the task is to verify instead of generating conclusions/models.
4. Problems consisting of three terms are less difficult to solve than four-term problems.
5. In case of simultaneous premise presentation, accuracy is higher than in case of sequentially premise presentation.

In the following, these predictions are analyzed with the aim of gaining differences in accuracy specific to the various types. The results of the analysis are evaluated and interpreted with respect to predictions of the mental model theory and implications from working memory limitations.

The Meta-Analysis

Paper Acquisition In order to acquire sufficient and suitable data for the meta-analysis, we needed to find experiments in which the participants drew their own conclusions to all sorts of tasks in spatial relational reasoning. An initial set of eligible studies came from a meta-analysis database of coded studies about spatial relational reasoning from the Cognitive Computation Lab (University of Freiburg, Germany) the database incorporated a comprehensive search for studies reported until 2013. The database contained the literature using the online platforms PubMed and Google Scholar for entries by the following main query: '(relational) AND (reasoning) OR (reasoning) AND (about) AND (relations) OR (transitive reasoning)'. All the studies in that database were reviewed for eligibility and an independent search was conducted: Online literature searches were performed on the 29th of October, and 29th of November 2016 using PubMed and Google Scholar. For the first PubMed and Google Scholar search, the same term like in 2013. For the second PubMed search, conducted on the 29th of November, the query '(spatial) AND (reasoning) AND (relations) OR (spatial reasoning) AND (relations)' was used, since the initial query was too unspecific for this search engine.

Criteria for Inclusion of Studies Experiments were assessed and selected for this meta-analysis if they met not only the search terms but also the following criteria: Experiments containing spatial relations and in cases of visuo-spatial relations, experiments emphasizing on spatial representations were also considered. All the experiments had to involve healthy, adult participants and use a within-subjects design to keep the homogeneity among different conditions. Participants had to know beforehand that their task was to reach a conclusion and there had to be no secondary-tasks. These criteria were used to ensure that the reasoning process was actually taking place and to eliminate other cognitive processes as a biasing factor. Moreover, only peer-reviewed and published studies of both, behavioural and neurophysiological experiments conducted in any country were considered. Outcome results of accuracy must have been presented in a quantitative form that permitted computation or reasonable estimation of an effect size statistic representing the difference in accuracy. Finally, information on factors of interests had to be given in the study. The literature search identified 138 experiments of 84 articles reporting results from psychological studies. Of these, 32 experiments (23%) (e.g., Knauff & Johnson-Laird, 2002a) did not report behavioral data or did not present spatial realtions by means of language.

Twenty experiments (14%) were rejected because they did not report or measured accuracy (e.g., Brüssow et al.,

Table 1: Means and standard deviations of the overall accuracy for the studies (in %).

Dataset with indeterminacy condition	determinate problems	indeterminate problems
simultaneous-verbal presentation	70 (7)	52 (8)
sequential-verbal presentation	65 (4)	42 (13)
sequential-auditory presentation	63 (10)	34 (19)

Values are rounded to integers. *Indeterminacy (simultaneous-verbal)*: Experiments with simultaneous premise presentation, generation task, verbal presentation form and 5-term problems. *Indeterminacy (sequential-verbal)*: Experiments with sequential premise presentation, generation task, auditory presentation form and 5-term problems. *Indeterminacy (sequential-auditory)*: Experiments with sequential premise presentation, generation task, verbal presentation form and 5-term problems.

2013). Seventeen experiments (12%) did not report information about either indeterminacy, presentation form, type of task or number of terms nor premise presentation (e.g., Fangmeier et al., 2006). For 13 experiments (9%) the original study was not available (e.g., Hagert, 1984). Eight experiments (6%) used secondary-task methods (e.g., Knauff et al., 2004), six experiments (4%) used a between-subjects design (e.g., Boudreau & Pigeau, 2001) and three experiments (2%) included children or patients (e.g., Knauff & May, 2006). In two experiments (1%) a recognition task was performed (e.g., Mani & Johnson-Laird, 1982) and two other experiments (1%) used a visual presentation form (e.g., Knauff & May, 2006). In total, 206 raw differences in means of accuracy between different types of spatial relational reasoning problems and other types of (relational) reasoning met the inclusion criteria for this meta-analysis. An asterisk precedes each of these reports in the reference list.

Paper Classification After paper selection, experiment characteristics were coded by the authors for the following characteristics: indeterminacy, premise presentation, presentation form, number of terms, task and sample size. The sample size for each experiment was defined as the number of participants at the time of the final measure of logical correct answer. The first division of data was made between determinate and indeterminate problems. Furthermore, the data was subdivided into groups of simultaneous and sequential premise presentation. All premises were either displayed at the same time and remained available (simultaneous), or were presented one at a time and disappeared with the onset of a new premise (sequential). Moreover, it was coded whether the participants had to listen to the premises in form of audio recording using spoken language (auditory form) or whether the premises were presented by means of literacy language on screen or on paper (verbal form). Likewise, the data was grouped into experiments with either three, four, or five terms. Finally, experiments were assigned to the group of verification task when the reasoners had to verify a outative conclusion or to select the correct model from a given set of models. If the participants had to generate a model (an arrangement of objects) or to draw a conclusion, tasks were characterized as generation tasks. For each factor, the data was divided into subgroups finding the combination of variants that had the most values for comparison of two variants of a factor. For this purpose, cross-classifying factors were used to build a contingency table of the counts at each combination of factor levels resulting in eight combinations of data. In Table 2, the datasets for all types of factors and its characteristics are presented. The factor number of participants within an experiment was not included in matching since otherwise, the number of raw means within subgroups would have been too small for statistical analysis. For the results reported in this study, statistical analysis consisted of non-parametric tests using one-sided Wilcoxon Rank-Sum Tests and a significance level of $\alpha = .01$ was defined. All statistical analysis was performed using R (R version 3.2.2, 2015-08-14; R Core Team, 2015).

Results and Discussion

The data was corrected for outliers by excluding percentages of accuracy with values greater than two standard deviations from the mean value. All the statistical analysis reported is based on the data corrected by outliers. We analysed the percentage of correctness for all eight combinations of data.[2] Table 2 summarizes the descriptive statistics in all conditions.

The analysis shows that in the indeterminacy (simultaneous-verbal) sample, significantly more correct responses were given if the problem was determinate than indeterminate (Wilcoxon Rank-Sum Test $W = 168$, $p < .0001$, one-sided). The same trend is also visible for the sequential-verbal and sequential-auditory indeterminacy cases (Table 1). One-sided Wilcoxon Rank-Sum Tests showed that both trends were statistically significant (indeterminacy (sequential-verbal): $W = 30$, $p < .05$; indeterminacy (sequential-auditory): $W = 30$, $p = .01$). These results supports several empirical findings (e.g., Byrne & Johnson-Laird, 1989) and can be explained by MMT as well as Baddeley's WMM. For an indeterminate description, it is necessary to construct not one but several mental models in order to correctly represent its meaning. The larger the number of models that reasoners must consider, the higher is the load on working memory. The attempt to construct

[2] This was due to a better readability. For the purpose of calculating Odds Ratio and the further meta-analysis, absolute frequency of accuracy based on the number of participants and the percentage of the correctness of the given responses are used.

Table 2: Descriptive statistics of the overall accuracy for the studies (in %).

Dataset	Factor	$n_{rawmeans}$	$n_{participants}$	*Mean* (*SD*)	*Min*	*Q1*	*Median* (*MAD*)	*Q3*	*Max*
Indeterminacy***	det	12	2384	70 (7)	64	64	66 (3)	78	80
(simultaneous-verbal)	indet	14	2456	52 (8)	40	43	54 (7)	58	59
Presentation form 1 (n.s.)	verb	15	255	90 (10)	68	89	90 (11)	98	98
	audi	18	264	87 (7)	77	81	88 (9)	94	96
Presentation form 2**	verb	37	702	72 (15)	40	61	77 (17)	84	92
	audi	10	192	57 (14)	33	47	61 (12)	67	76
Task**	veri	15	255	90 (10)	68	89	90 (11)	98	98
	gen	30	856	76 (17)	54	58	75 (25)	91	99
Number of terms (n.s)	three	30	856	76 (17)	54	58	75 (25)	91	99
	four	37	702	72 (15)	40	61	77 (17)	84	92
Premise presentation (n.s.)	sim	24	3504	49 (12)	27	44	48 (10)	55	66
	seq	15	255	90 (10)	68	89	90 (11)	98	98

All values were rounded to integers. Factor labels refer to auditory vs. verbal presentation form (audi/verb), determinate vs. indeterminate problems (det/indet), model generation vs. model verification task (gen/veri) and sequential vs. simultaneous premise presentation (seq/sim). $n_{rawmeans}$: number of raw differences in means; $n_{participants}$: number of participants. *Indeterminacy (simultaneous-verbal)*: Experiments with simultaneous premise presentation, generation task, verbal presentation form and five-term problems. *Presentation form 1*: Experiments with sequential premise presentation, verification task, determinate and three-term problems. *Presentation form 2*: Experiments with sequential premise presentation, generation task, determinate and four-term problems. *Task*: Experiments with sequential premise presentation, verbal presentation form, determinate and three-term problems. *Number of terms*: Experiments with sequential premise presentation, verbal presentation form, verification task and determinate problems. *Premise presentation*: Experiments with verbal presentation form, verification task, determinate and three-term problems. *** significant $p < .001$, ** significant $p < .01$

several models may overload working memory capacities so that no models would be constructed. Likewise, the accuracy was significantly higher if the problems were presented verbally than during auditory presentation (Wilcoxon Rank-Sum Tests, presentation form 2: $W = 283$, $p < .01$, one-sided). A similar but not significant trend showed in the presentation form 1 condition (Wilcoxon Rank-Sum Tests, $W = 176$, $p = .07$, one-sided). Spoken language can only be presented in a serial way. Thus, the spatial description has to be stored in the phonological loop using a language-based form. At the same time, mental models are built in the visuo-spatial sketchpad. When working memory load increases it becomes harder to keep track of all the premises and inferences separately. Additionally, written language already implies information about spatial relations and is, therefore, more similar to the information contained in the problem description than in case of auditory presentation. Furthermore, a one-sided Wilcoxon Rank-Sum Test supports the prediction that reasoning difficulty is lower in verification tasks ($W = 332$, $p < .01$). Thereby, reasoners have to build a model based on the premises presented and verify a proposed model or conclusion. In addition to model construction, reasoners have also to draw a conclusion to solve the generation task correctly. This requires a larger amount of working memory (Klauer, 1997). The descriptive results for the dataset number of terms are consistent with the predictions. Three-term problems were higher in accuracy than four-term problems (Table 2). However, the Wilcoxon Rank-Sum Test showed no significant difference ($W = 641$, $p > .1$, one-sided). One explanation could be the limited capacity of the working memory that is roughly three to five objects or role bindings (Clevenger & Hummel, 2014). Both three-term and four-term problems do not exceed the capacities of working memory. However, descriptive results assume a tendency for an increasing memory load in four-term problems. Contrary to the predictions, there was no significant difference between accuracy in case of simultaneous and sequential premise presentation ($W = 0$, $p > .1$, one-sided Wilcoxon Rank-Sum Test). Furthermore, descriptive results indicate that problems with sequential premise presentation are less difficult to solve than simultaneously presented problems (Table 2). This result was unexpected and contradicts previous findings (Roberts & Sykes, 2003; Van der Henst & Schaeken, 2005). With regard to the data used in this study, the following differences can be observed which may have influenced the result: In case of simultaneous presentation, the total number of participants tested is higher than in sequential condition ($N = 3504$ vs. $N = 255$). The difference in sample size might have influenced the results. In addition, the factor premise order was not controlled in this study. The dataset of simultaneous premise presentation contained discontinuous problems. However, in the set of sequential premise presentation, the factor premise order was either unspecified or continuous. Studies have shown that accuracy is higher in case of continuous premise order (Knauff et al., 1998). Thus, an effect of premise order can not be excluded here. A further explanation may be that the amount of information must be processed is reduced as a result of sequential presentation. Hence, attention control may be facilitated and a model can be constructed incrementally from the premises (*principle of economicity,* Manktelow & Galbraith, 2012).

General Discussion

Despite a century of research on spatial relational reasoning and more than a hundred articles there is still no systematic analysis of which factors do contribute to the human difficulty in reasoning about spatial relations. The quality of the reported experimental data differs, sometimes all relevant information is reported, sometimes the standard deviation is missing. Hence, only a limited number of experiments could be included in our review. We identified reliable differences in accuracy between determinate and indeterminate problems, auditory and verbal presentation form and also between tasks of model generation and verification. This meta-analysis confirmed previous empirical findings that are predicted by the mental model theory and influenced by the limited-capacity working memory as predicted by WMM (see Klauer, 1997): The effect of indeterminacy is the strongest effect (see Table 2) that is directly related with the number of models that need to be generated in the reasoning process. The same holds for the task where the verification is easier (the constructed models need to be compared with a given one), while in the conclusion generation process all models need to be checked. And, finally a generation task with four terms (Presentation form 2 in Table 2) can lead to more demands on both the construction and storing the model in memory.

This systematic review has, however, a few limitations related to the results and their interpretation and leads to some new questions. First, the number of studies considered is a limiting factor to the expressive power of the analysis results. Furthermore, most of the studies included in this meta-analysis did not report any information about the premise order (e.g., if the premise information is continuosly), so it was not possible to control for this factor. Likewise, with regard to the small number of raw means, it was not possible to factor in the sample size of each experiment. The next step of analysis in this study will focus on this particular point and investigates questions, such as the variability in effects across studies and how this variability can be explored in terms of moderator variables. Identification of the moderator variables that describe the study characteristics associated with larger and smaller effects is another kind of contribution meta-analysis can make to understanding difficulties in spatial relational reasoning. Of particular importance is the role such moderator analysis can play in ascertaining which variants of spatial descriptions are most effective for reasoning.

Taken together this study illustrates a use of meta-analysis for data interpretation beyond conventional statistical analysis. Some cross-experimental results can be formulated: First, determinate problems are easier to solve than indeterminate problems. Second, compared with auditory presentation, problems in form of written language are less difficult. Further, the accuracy was better for tasks that require the verification of conclusions or models than in tasks that require to generate conclusions or models. This meta-analysis confirms some previous empirical findings, and supports predictions of the spatial mental model theory together with assumptions from a limited spatial working memory.

Acknowledgments

This work has been supported by a DFG Heisenberg fellowship RA 1934/3-1 and projects RA 1934/2-1 and 4-1.

References

Asterisks refer to studies included in the analysis.

Baddeley, A. (1986). *Working memory.* Clarendon Press, Oxford.

Boudreau, G., & Pigeau, R. (2001). The mental representation and processes of spatial deductive reasoning with diagrams and sentences. *International Journal of Psychology*, *36*(1), 42–52.

Brüssow, S., Ragni, M., Frorath, M., Konieczny, L., & Fangmeier, T. (2013). Premise annotation in mental model construction: An ACT-R approach to processing indeterminacy in spatial relational reasoning. *Cognitive Systems Research*, *24*, 52–61.

*Bucher, L., & Nejasmic, J. (2012). Relocating multiple objects during spatial belief revision. In U. D. Stachniss C. Schill K. (Ed.), *International conference on spatial cognition* (Vol. 7463, pp. 478–491). Springer.

*Byrne, R. M., & Johnson-Laird, P. (1989). Spatial reasoning. *Journal of memory and language*, *28*(5), 564–575.

Clevenger, P. E., & Hummel, J. E. (2014). Working memory for relations among objects. *Attention, Perception, & Psychophysics*, *76*(7), 1933–1953.

*De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, *2*(4), 513–521.

*Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of verbal learning and verbal behavior*, *21*(3), 296–306.

*Fangmeier, T., & Knauff, M. (2009). Neural correlates of acoustic reasoning. *Brain research*, *1249*, 181–190.

Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). fmri evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience*, *18*(3), 320–334.

Goodwin, G. P., & Johnson-Laird, P. (2005). Reasoning about relations. *Psychological Review*, *112*(2), 468–493.

Hagert, G. (1984). Modeling mental models: Experiments in cognitive modeling of spatial reasoning. In T. O'Shea (Ed.), *Proceedings of the 6th European Conference on Artificial Intelligence* (pp. 179–188).

Hunter, I. M. (1957). The solving of three-term series problems. *British Journal of Psychology*, *48*(4), 286–298.

Johnson-Laird, P. N., & Byrne, R. M. (1991). *Deduction.* Lawrence Erlbaum Associates, Inc.

Klauer, K. C. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking & Reasoning*, *3*(1), 9–47.

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

*Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience*, *15*(4), 559–573.

Knauff, M., & Johnson-Laird, P. (2002a). Reasoning and the visual-impedance hypothesis. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *International Conference on Spatial Cognition* (pp. 372–384). Springer.

*Knauff, M., & Johnson-Laird, P. (2002b). Visual imagery can impede reasoning. *Memory & Cognition*, *30*(3), 363–371.

*Knauff, M., & May, E. (2006). Mental imagery, reasoning, and blindness. *The Quarterly Journal of Experimental Psychology*, *59*(1), 161–177.

*Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., & Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: a functional mri study. *Cognitive Brain Research*, *13*(2), 203–212.

Knauff, M., Rauh, R., Schlieder, C., & Strube, G. (1998). Continuity effect and figural bias in spatial relational inference. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 573–578).

Knauff, M., Strube, G., Jola, C., Rauh, R., & Schlieder, C. (2004). The psychological validity of qualitative spatial reasoning in one dimension. *Spatial Cognition and Computation*, *4*(2), 167–188.

*Krumnack, A., Bucher, L., Nejasmic, J., Nebel, B., & Knauff, M. (2011). A model for relational reasoning as verbal reasoning. *Cognitive Systems Research*, *12*(3), 377–392.

Kubinger, K. D., & Wolfsbauer, C. (2010). On the risk of certain psychotechnological response options in multiple-choice tests. *European Journal of Psychological Assessment*, *26*, 302-308.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, *10*(2), 181–187.

Manktelow, K., & Galbraith, N. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgment and decision making.* Psychology Press.

*Morra, S. (1989). Developmental differences in the use of verbatim versus spatial representations in the recall of spatial descriptions: A probabilistic model and an experimental analysis. *Journal of Memory and Language*, *28*(1), 37–55.

*Morra, S. (2001). On the information-processing demands of spatial reasoning. *Thinking & reasoning*, *7*(4), 347–365.

*Nejasmic, J., Bucher, L., & Knauff, M. (2015). The construction of spatial mental models—a new view on the continuity effect. *The Quarterly Journal of Experimental Psychology*, *68*(9), 1794–1812.

*Nejasmic, J., Krumnack, A., Bucher, L., & Knauff, M. (2011). Cognitive processes underlying the continuity effect in spatial reasoning. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1127–1132).

Ormrod, J. E. (1979). Cognitive processes in the solution of three-term series problems. *The American Journal of Psychology*, 235–255.

*Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2010). Recomposing a fragmented literature: How conditional and relational arguments engage different neural systems for deductive reasoning. *Neuroimage*, *51*(3), 1213–1221.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

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

*Roberts, M. J. (2000). Strategies in relational inference. *Thinking & Reasoning*, *6*(1), 1–26.

Roberts, M. J., & Sykes, E. D. (2003). Belief bias and relational reasoning. *The Quarterly Journal of Experimental Psychology: Section A*, *56*(1), 131–154.

*Ruff, C. C., Knauff, M., Fangmeier, T., & Spreer, J. (2003). Reasoning and working memory: common and distinct neuronal processes. *Neuropsychologia*, *41*(9), 1241–1253.

*Schaeken, W., Girotto, V., & Johnson-Laird, P. N. (1998). The effect of an irrelevant premise on temporal and spatial reasoning. *Kognitionswissenschaft*, *7*(1), 27–32.

Schaeken, W., & Johnson-Laird, P. N. (2000). Strategies in temporal reasoning. *Thinking & Reasoning*, *6*(3), 193–219.

Störring, G. (1908). *Experimentelle Untersuchungen über einfache Schlussprozesse*. W. Engelmann.

Van Der Henst, J.-B. (2002). Mental model theory versus the inference rule approach in relational reasoning. *Thinking & Reasoning*, *8*(3), 193–203.

Van der Henst, J.-B., & Schaeken, W. (2005). The wording of conclusions in relational reasoning. *Cognition*, *97*(1), 1–22.

*Vandierendonck, A. (1996). Evidence for mental-model-based reasoning: A comparison of reasoning with time and space concepts. *Thinking & Reasoning*, *2*(4), 249–272.