Neural Principles for Modeling Relational Reasoning: Lesson learned from Cognitive Neuroscience

Julia Wertheim (wertheim@cs.uni-freiburg.de)
Cognitive Computation Lab, Technical Faculty, Albert-Ludwigs-Universität Freiburg,
Georges-Köhler-Allee 79, 79110 Freiburg, Germany

Marco Ragni (ragni@cs.uni-freiburg.de)
Cognitive Computation Lab, Technical Faculty, Albert-Ludwigs-Universität Freiburg,
Georges-Köhler-Allee 52, 79110 Freiburg, Germany

Abstract

Cognitive models serve the purpose of implementing theories of human cognition and give the opportunity to simulate reasoning processes for comparing them to participant data. Relational reasoning is particularly relevant, because it is closely connected to spatial navigation and planning. In modeling relational reasoning, findings from neuroscience have been largely neglected. As we are showing, the connection between neuroimaging and cognitive modeling has been elementary so far. We aim at bridging the gap between the neurocognitive correlates of relational reasoning and cognitive models thereof. Computational models and architectures, as well as the recent neuroimaging literature investigating relational reasoning are reviewed. By identifying functional modules, we postulate the neuroscientific loci which a modeler aiming at simulating reasoning should consider before conceptualizing a neurocognitive model of relational reasoning.

Keywords: Cognitive modeling, Relational reasoning, Neuroimaging, Cognitive architecture

Introduction

Cognitive models enable the testing of cognitive theories and the comparison to psychological findings. In the last decade, the interest in biologically plausible cognitive modeling has been rising, not least because of transnational research projects such as the Blue Brain Project (Markram, 2006). Especially interesting is the modeling of higher cognitive processes such as relational reasoning, usually taking the form of premises like "Tom is to the right of Sally. Sally is to the right of George.", from which the inference "Tom is to the right of George." can be derived. For successfully solving this task, participants are asked to infer explicit knowledge about the objects' relations to each other, which is implicitly given in the premises. Relational reasoning is closely linked to spatial navigation and hence motor function, as well as to analogy and language processing. Thereby, it is more versatile than other reasoning types and particularly promising for cognitive modeling since multiple cognitive abilities are recruited. The expected findings of modeling relational reasoning could greatly contribute to fields such as Brain-Computer-Interfaces or medical diagnostics. For example, if a patient suffers from a brain lesion, the only information available so far is a potential function loss in cognitive abilities associated with the respective region. Detailed information about the region's function in terms of more complex cognitive abilities such as reasoning is not yet available. Biologically plausible cognitive model could provide details about the wide-ranging cognitive deficits resulting from the loss when informed by the lesion site. Hence, investigating the connection between models of cognition and neuroscience are beneficial for developing medically relevant models of neuropathology and diagnostic purposes.

But how *do* we go about investigating human cognition, specifically relational reasoning? According to Marr's analysis, there are three levels to be considered (Marr, 1982, see Figure 1). On the computational level, the strategic aim of the cognitive effort is evaluated. This involves a formal model or theoretical framework of relational reasoning, such as mental model theory, as well as cognitive architectures in which these can be implemented. On the algorithmic level, human performance is assessed. This involves reasoning effects and models describing and explaining the processes. On an implementation level, the 'hardware' in which cognitive process are implemented is considered, namely neuroanatomy. Regarding relational reasoning, this results in finding the neural correlates of these processes and assessing their neuroanatomical feasibility.

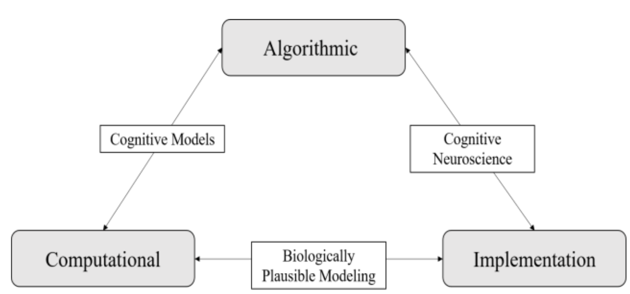

Figure 1: Representation of the analysis levels.

By modeling cognitive processes, we predominantly work on the computational level, but they are inevitable intertwined. In the case of relational reasoning, cognitive models are described to solve reasoning tasks which have previously been tested on human participants, hence the algorithmic and computational level are connected. In cognitive neuroscience, the algorithmic and implementation levels are conjoined by identifying the relevant brain areas. As we are going to show in the following, the link between the implementation and computational level is currently underdeveloped, although both levels bear important insights for the other. Our article serves as a resource for modelers aiming at the development of a cognitive model for relational reasoning which is based on neuroscientific insights.

Mutually informing each other: cognitive neuroscience and cognitive modeling

Cognitive neuroscience can greatly benefit from the insights of cognitive models, for modeling provides more sophisticated analyses of cognitive processes. On this basis, neuroscientific hypotheses can be formulated more accurately and tested on more precise levels of analysis by revealing hidden cognitive processes and fostering more accurate study designs (Forstmann et al., 2011). Conversely, cognitive modeling likewise benefits from neuroscientific findings in using them to restrain and inform the set-up and conceptualization of cognitive models, thereby making them more viable (Forstmann et al., 2011). This includes systematic reviews of cognitive models, some of which can be favored based on neuroscientific plausibility (Forstmann et al., 2011).

In this article, we aim at bridging the gap between cognitive models and insights from cognitive neuroscience about the neural underpinnings of relational reasoning. For this, we systematically review cognitive models and neuroimaging studies of relational reasoning and identify the most central brain regions. We characterize the regions in terms of their functionality to the task and summarize by establishing a neuroscience-based standard functions and brain regions required for modeling relational reasoning.

Cognitive Models of Relational Reasoning

Regarding the algorithmic level, Friemann and Ragni (2018) have recently published a review of spatial relational reasoning models. Their collection of models was adapted to review the current state of the art, Further, we conducted an online researched via Google Scholar and Pubmed. As Table 1 displays, the models vary greatly along different parameter such as the number of dimensions of processable relations and their inclusion of findings from cognitive neuroscience. The listed models are going to be categorized in terms of the aforementioned dimensionality, working memory capacity and whether findings from cognitive neuroscience are incorporated in the model.

Dimensionality

The lowest dimensionality of one, enabling the processing of relational dichotomies is offered by the models of Schlieder and Berendt (1998), Bara, Bucciarelli and Lombardo (2001), Hummel and Holyoak (2001), Morrison et al., (2004), Krumnack, Bucher, Nejasmic, Nebel and Knauff (2011) and Dietz, Hölldobler and Höps (2015). An additional dimension, allowing for, e.g., the processing of cardinal directions, is featured by the models by Ioerger (1994), Schultheis and Barkowsky (2011), Wertheim and Stewart (2018) and Kounatidou, Richter and Schöner (2018). Only the models by Johnson-Laird and Byrne (1991) and Ragni and Knauff (2013) provide a three-dimensional space in which relational reasoning operations are represented.

Working memory

Concerning working memory, most models exhibit a limited capacity inspired by human processing. In the models by Schlieder and Berendt (1998), Dietz et al. (2015) and Kounatidou et al. (2018), the capacity of the working memory is unspecified and therefore not explicitly adapted to human performance.

Table 1: Overview of cognitive models of spatial relational reasoning.

Authors	Dim.	WM	ND
Johnson-Laird & Byrne, 1991	3	Lim	✗
Ioerger, 1994	2	Lim	✗
Schlieder & Berendt, 1998	1	n/a	✗
Bara et al., 2001	1	Lim	✗
Hummel & Holyoak, 2001	1	Lim	✗
Morrison et al., 2004	1	Lim	✓
Krumnack et al., 2011	1	Lim	✗
Schultheis & Barkowsky, 2011	2	Lim	✓
Ragni & Knauff, 2013	3	Lim	✗
Dietz et al., 2015	1	n/a	✗
Wertheim & Stewart, 2018	2	Lim	✓
Kounatidou et al., 2018	2	n/a	✓

Note. Dim.: Number of dimensions; WM: Working memory; Lim: Limited capacity; ND: Inclusion of neuroscience data.

Inclusion of neuroscience data

Concerning the implementation of neuroscience data, only few models appear to be relevant. The model by Morrison et al. (2004) includes lesion patient data, whereas the model by Schultheis and Barkowsky (2011) is explicitly based on the modularity hypothesis. Apart from that, only the models by Wertheim and Stewart (2018) and Kounatidou et al. (2018) run on artificial neural networks which are (partially) based on the mechanisms of actual neurons.

We conclude that the inclusion of neuroscience data has not yet been widely used in the development of cognitive models. Nonetheless, cognitive models can theoretically be implemented in current cognitive architectures such as the Turing-complete ACT-R. Also, this has already been done by, e.g., Wertheim and Stewart (2018) in the Neural Engineering Framework (NEF). Hence, an online research was conducted via Google Scholar and Pubmed to review cognitive architecture providing a programming framework and to investigate to what extend insights from neuroscience have been used to restrain frameworks or provide predictions. A notable review has been on cognitive architectures has been published by Samsonovich (2010), but not all aim at biological plausibility.

Cognitive Architectures

Some frameworks consider brain function on either a level of restraining implementation possibilities or in modeling neurocognitive processes. These include the architectures 4CAPS (Just, Carpenter & Varma, 1999) and conceptually also the precursor 3CAPS (Just & Carpenter, 1992) by approximating the BOLD response. ACT-R (Anderson, 2007) and EPIC (Meyer & Kieras, 1999) incorporate anatomically and functionally plausible correspondents to brain regions. The NEF (Eliasmith, 2013) simulates neuronal activity and connectivity, whereas SOAR (Newell, 1992) restrains working memory to a neurobiologically plausible time span. CLARION (Sun, 2002) is based upon the modularity hypothesis and Sigma (Rosenbloom, 2013) features neural networks. In Table 2, we evaluate the cognitive architectures towards features essential to biologically plausible computation of relational reasoning.

Table 2: Overview of cognitive architectures suitable for processing relational reasoning

Model	Module	BOLD
3CAPS	✓	✗
SOAR	✓	✗
4CAPS	✓	✓
EPIC	✓	✓
CLARION	✓	✗
ACT-R	✓	✓
NEF	✓	✗
Sigma	✗	✗

Note. Module: Does it feature separate interacting modules?; BOLD: BOLD function predicted?

Modularization

Modularization of functional components has been a common practice when designing cognitive architectures. This is based upon the modularity hypothesis stemming from evolutionary psychology which claims that cognition is facilitated by function-specific brain regions serving as modules (Fodor, 1985). We have found that almost all architectures share this basic trait, except for Sigma (Rosenbloom, 2013).

BOLD prediction

A common approach in biologically plausible cognitive modeling is the prediction and approximation of the BOLD response derived from fMRI studies. So far, his has only been accomplished in the framework 4CAPS (Just et al., 1999) and ACT-R (Anderson, 2007).

Functional Brain Regions for Relational Reasoning

The meta-analysis by Wertheim and Ragni (2018) examines the neural correlates of relational reasoning and was used to identify the brain regions active during task solving. In the following, the regions are examined based on their involvement in cognitive processes with regard to their relevance to cognitive modeling.

Frontal lobe

The most wide-spread activation was found in the bilateral, but mostly left prefrontal cortex. According to O'Reilly and Munakata (2000), this region is functionally responsible for active and flexible maintenance of complex mental representations, as well as goal-directed executive control, especially regarding the monitoring of overall processing (Eriksson et al., 2015). Further, a left-sided activation in reasoning has been previously assumed and supported by lesion studies. For example, Goel et al. (2006) showed that left-, in contrast to right-sided lesions hinder participants to correctly decide whether determinate tasks are correct.

Particularly relevant to relational reasoning is the dorsolateral prefrontal cortex and middle frontal gyrus (DLPFC/MFG, BA 9, 8). It is active during executive functioning and cognitive monitoring (Prabhakaran, Smith, Desmond, Glover & Gabrieli, 1997) and in maintaining multiple relations (Waltz et al., 1999), which is proposed to translate to the entertainment of a mental model and integration of several relations (Wertheim & Ragni, 2018). For example, the MFG is used in the architecture ACT-R serving as a declarative memory module (Anderson, 2007). BA 6 (Supplementary motor area, SMA) is involved in task planning (Hanakawa et al., 2002), whereas the precentral gyrus (PreCG, BA 9, 8) facilitates attention management (Acuna, Eliassen, Donoghue & Sanes, 2002).

Parietal lobe

The posterior parietal cortex (PPC) is typically associated with the (repetitive) processing of spatial information and scenarios, such as mental rotation (O'Reilly & Munakata, 2000). Specifically, activation was found in the bilateral superior parietal lobule (SPL), inferior parietal lobule (IPL), precuneus (PCUN, BA 7, 40). It is involved in executive working memory and sustained attention (Koenigs, Barbey, Postle & Grafman, 2009) and linked to the selection of the attention focus (Awh, Vogel & Oh, 2006). From a modeling perspective, this region is specifically involved in the construction and manipulation of mental models (Ragni, Franzmeier, Maier, & Knauff, 2016). Concerning the precuneus, neuroimaging studies have found its specific involvement in abstract tasks, as well as episodic memory retrieval (Cavanna & Trimble, 2006). Henceforth, the PPC facilitates a mental space in which model representation and manipulation takes place.

Basal ganglia

The right claustrum shares extensive structural connections to the prefrontal cortex (Ullman, 2006). In a computational sense, the selection of actions is assigned to this region (O'Reilly & Munakata, 2000). This assumption is supported by further neuroimaging studies of reasoning, e.g., Jia et al. (2011) assign rule induction to the basal ganglia. From a computational perspective, O'Reilly (2006) specified its role of gating of mental representations coming from the PFC.

Table 3: Overview of brain regions central to relational reasoning found by Wertheim and Ragni (2018).

Regions	BA	Function
SPL	7	Processing spatial information, construction and manipulation of mental models
IPL	40	
SMA	6	Task planning
DLPFC	9, 8, 46	Executive functioning, cognitive monitoring, maintaining of complex information, attention management Declarative memory
Claustrum	-	Action selection, representation gating

Note. BA: Brodmann area; SPL: Superior parietal lobule, IPL: Inferior parietal lobule, SMA: Supplementary motor area, DLPFC: dorsolateral prefrontal cortex.

Occipital lobe

Although Wertheim and Ragni (2018) did not find any significant clusters in the occipital lobe, processing-wise it is interesting because it shares connections with the PPC (Culham & Kanwisher, 2001). As it is active during the processing and abstraction of visual input (O'Reilly & Munakata, 2000), it should be considered for processing visual inputs and redirecting information to the PPC.

Discussion

In this article, we reviewed the current state of the art of cognitive models, architectures and the neuroscience of relational reasoning and hence provide a guideline for programmers aiming at building biologically plausible models of relational reasoning. Although there exists a considerable selection of models explaining the cognitive mechanisms underlying reasoning on the algorithmic level, only few have been implemented. Hence, we continued by reviewing cognitive architectures and found that there have been various approaches to including neuroscience results by either restraining programming environments or deriving predictions relevant for neuroscience. Nonetheless, the interface between cognitive architectures and cognitive neuroscience is sparse and only a synthesis of different approaches could foster the current state of the art. Concerning cognitive models, this would be by assuming three dimensions on which relations can be represented, a psychologically plausible constraint on working memory, was well as the potential implementation with neural networks. Regarding cognitive architectures, starting points are anatomically and functionally specified modules, as well as the prediction of the BOLD response (for examples see, Anderson et al., 2008 and Borst & Anderson, 2015).

Cognitive modeling would benefit from conjoining preexisting approaches to integrating neuroscience. Similarly, cognitive neuroscience can benefit by informing experimental set ups from cognitive simulations. For example, O'Doherty, Hampton and Kim (2007) developed model-based neuroimaging for correlating assumed cognitive processes with actual scanning data. Concerning our review of neuroimaging studies, the most relevant regions which need to be considered in cognitive models have been identified and examined for their involvement in functions relevant to relational reasoning (see Figure 2). These include the PPC for abstracting and processing spatial information and working memory and the basal ganglia for action selection and information gating. The basal ganglia have already been implemented as an action selection system in the NEF (Senft, Stwart, Bekolay, Eliasmith & Kröger, 2016), whereas the imaginal buffer in ACT-R serves as a correspondent to the PPC (Anderson, 2007). Further identified regions are the SMA for task planning, and DLPFC for meta-cognitive functions such as cognitive monitoring and attention management. This region-function matching can inform cognitive models by a precise selection of actions and associated brain regions. Since we did not only identify the relevant regions but also their function specifically in relational reasoning, modelers can either only include the regions' respective function or decide to consider neuroanatomical details as well.

We initially claimed that there exists a gap between the implementation and computational level of investigating cognition. We have contributed to closing this gap by identifying biologically relevant features of architectures which should be merged and extended, as well the most functionally relevant brain regions from neuroimaging. By this, we have established a first example of a necessary precondition to neurocognitive modeling and proposed guidelines from which both domains can benefit. Cognitive models could be improved by this localization of activation foci by constraining the models based on the cognitive succession of cognitive demands needed to fulfill a task which is examined by data derived from neuroimaging and corresponding cognitive theories.

From a practical perspective, biologically plausible cognitive models could be used for diagnostic purposes in medical environments. From these models, we could infer more detailed cognitive impairments in higher cognitive functions. So far, it is only possible to identify basic impairments following the damage of brain tissue, such as impairments in processing language and working memory. By developing a more detailed account of the neurological, as well as cognitively functional subunits of the mind-brain, diagnoses and decisions can be improved and more elaborate restorative and preventive therapies can be developed. Potentially, this program might develop into a resource for the structure-function mapping between brain regions and their involvement in specific tasks which would foster the mutual exchange between these two vibrant fields of research, as well as increase the practical usage of neuroscientific data for cognitive modeling.

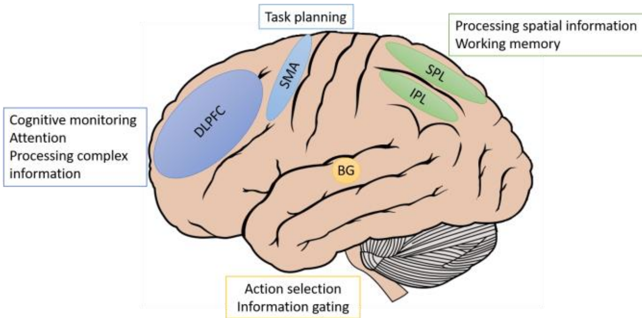

Figure 2: Representation of the brain regions active during relational reasoning and associated functions thereof.

The merging of the computational and implementation domains can be deepened and fostered by analyzing further meta-analyses on the neurocognitive correlates of reasoning tasks which can be theoretically or have been practically implemented in cognitive architectures. This would improve the specificity of determining which brain regions subserve cognitive functions, thus merging the approaches by brain mapping via neuroimaging and bypassing the difficulties of small sample sizes in neuroimaging studies (by meta-analyses) and cognitive modeling (by task specificity). Another domain of future work could be the more in-depth connection between implementation and computation by investigating the structural properties of the respective regions (e.g., arrangement of layer cells, interaction between inhibitory and excitatory mechanisms) for examining potential structure-function dependencies relevant to cognitive architectures per se and for spatial relational reasoning in particular.

Acknowledgments

This research was supported by the BrainLinks-BrainTools Cluster of Excellence funded by the German Research Foundation (DFG, grant number EXC 1086) to JW, a Heisenberg scholarship (RA 1934/3-1 and RA 1934/4-1) to MR and the DFG priority program New Frameworks of Rationality (SPP 1516).

References

Acuna, B. D., Eliassen, J. C., Donoghue, J. P., & Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cerebral Cortex*, *12*(12), 1312-1321.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?*. New York: Oxford University Press.

Anderson, J. R., Carter, C. S., Fincham, J. M., Qin, Y., Ravizza, S. M., & Rosenberg-Lee, M. (2008). Using fMRI to test models of complex cognition. *Cognitive Science*, *32*(8), 1323-1348.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201-208.

Bara, B. G., Bucciarelli, M., & Lombardo, V. (2001). Model theory of deduction: A unified computational approach. *Cognitive Science*, *25*(6), 839-901.

Borst, J. P., & Anderson, J. R. (2015). Using the ACT-R Cognitive Architecture in combination with fMRI data. In *An introduction to model-based cognitive neuroscience* (pp. 339-352). Springer, New York, NY.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564-583.

Charlton, R. A., Barrick, T. R., Lawes, I. N. C., Markus, H. S., & Morris, R. G. (2010). White matter pathways associated with working memory in normal aging. *Cortex*, *46*(4), 474-489.

Dietz, E.-A., Hölldobler, S., & Höps, R. (2015). A computational logic approach to human spatial reasoning. In *IEEE symposium series on computational intelligence* (pp. 1627–1634). IEEE.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford, UK: Oxford University Press.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron*, *88*(1), 33-46.

Fodor, J. A. (1985). Precis of the modularity of mind. *Behavioral and Brain Sciences*, *8*(1), 1-5.

Forstmann, B. U., Wagenmakers, E. J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract?. *Trends in Cognitive Sciences*, *15*(6), 272-279.

Friemann, P., & Ragni, M. (2018). Cognitive Computational Models of Spatial Relational Reasoning: A Review. *In* Thrash, Kelleher, & Dobnik (Eds.). *The 3rd Workshop on Models and Representations in Spatial Cognition (MRSC-3)*.

Goel, V., Tierney, M., Sheesley, L., Bartolo, A., Vartanian, O., & Grafman, J. (2006). Hemispheric specialization in human prefrontal cortex for resolving certain and uncertain inferences. *Cerebral Cortex*, *17*(10), 2245-2250.

Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H., & Shibasaki, H. (2002). The role of rostral Brodmann area 6 in mental-operation tasks: an integrative neuroimaging approach. *Cerebral Cortex*, *12*(11), 1157-1170.

Hummel, J. E., & Holyoak, K. J. (1996). LISA: A computational model of analogical inference and schema induction. In G. W. Cottrell (Ed.), In *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 352-357). Hillsdale, NJ: Erlbaum

Ioerger, T. R. (1994). The manipulation of images to handle indeterminacy in spatial reasoning. *Cognitive Science*, *18*(4), 551-593.

Johnson-Laird, P. N. & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Jia, X., Liang, P., Lu, J., Yang, Y., Zhong, N., & Li, K. (2011). Common and dissociable neural correlates associated with component processes of inductive reasoning. *NeuroImage*, *56*(4), 2292-2299.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122-149.

Just, M. A., Carpenter, P. A., & Varma, S. (1999). Computational modeling of high-level cognition and brain function. *Human Brain Mapping*, *8*(2-3), 128-136.

Koenigs, M., Barbey, A. K., Postle, B. R., & Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *Journal of Neuroscience*, *29*(47), 14980-14986.

Kounatidou, P., Richter, M. & Schöner, G. A Neural Dynamic Architecture That Autonomously Builds Mental Models in (2018). *In* T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 641-646). Austin, TX: Cognitive Science Socitey.

Krumnack, A., Bucher, L., Nejasmic, J., Nebel, B., & Knauff, M. (2011). A model for relational reasoning as verbal reasoning. *Cognitive Systems Research*, *12*(3-4), 377-392.

Marr, D. (1982). A computational investigation into the human representation and processing of visual information. *WH Freeman and Company, San Francisco, CA*.

Markram, H. (2006). The blue brain project. *Nature Reviews Neuroscience*, *7*(2), 153-160.

Meyer, D. E., & Kieras, D. E. (1999). Précis to a practical unified theory of cognition and action: Some lessons from EPIC computational models of human multiple-task performance. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance* (pp. 17-88). Cambridge, MA: MIT Press.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, *16*(2), 260-271.

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., & Jonides, J. (2012). A meta-analysis of executive components of working memory. *Cerebral Cortex*, *23*(2), 264-282.

Newell, A. (1992). SOAR as a unified theory of cognition: Issues and explanations. *Behavioral and Brain Sciences*, *15*(3), 464-492.

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of sciences*, *1104*(1), 35-53.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. MIT press.

Prabhakaran, V., Smith, J. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1997). Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. *Cognitive Psychology*, *33*(1), 43-63.

Ragni, M., Franzmeier, I., Maier, S., & Knauff, M. (2016). Uncertain relational reasoning in the parietal cortex. *Brain and Cognition*, *104*, 72-81.

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

Rosenbloom, P. S. (2013). The Sigma cognitive architecture and system. *AISB Quarterly*, *136*, 4-13.

Samsonovich, A. V. (2010). Toward a unified catalog of implemented cognitive architectures. *BICA*, *221*(2010), 195-244.

Schlieder, C., & Berendt, B. (1998). Mental model construction in spatial reasoning: A comparison of two computational theories. In U. Schmid, Krems, J. F., & F. Wysotzki (Eds.), *Mind modelling: A cognitive science approach to reasoning, learning and discovery* (pp. 133-162). Lengerich (Germany): Pabst Science Publishers

Schultheis, H., & Barkowsky, T. (2011). Casimir: an architecture for mental spatial knowledge processing. *Topics in Cognitive Science*, *3*(4), 778-795.

Senft, V., Stewart, T. C., Bekolay, T., Eliasmith, C., & Kröger, B. J. (2016). Reduction of dopamine in basal ganglia and its effects on syllable sequencing in speech: A computer simulation study. *Basal Ganglia*, *6*(1), 7-17.

Sun, R. (2001). *Duality of the mind: A bottom-up approach toward cognition*. Mahwah, NJ: Erlbaum.

Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit?. *Cortex*, *42*(4), 480-485.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., ... & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*(2), 119-125.

Wertheim, J., & Ragni, M. (2018). The neural correlates of relational reasoning: A meta-analysis of 47 functional magnetic resonance studies. *Journal of Cognitive Neuroscience*, *30*(11), 1734-1748.

Wertheim, J. & Stewart, T. (2018). Explaining Reasoning Effects: A Neural Cognitive Model of Spatial Reasoning. *In* T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2675-2680). Austin, TX: Cognitive Science Society.