Explaining Reasoning Effects: A Neural Cognitive Model of Spatial Reasoning

Julia Wertheim (wertheim@tf.uni-freiburg.de)
Cognitive Computation Lab, Department of Computer Science,
Albert-Ludwigs-University Freiburg, Georges-Köhler-Allee 79, 79110 Freiburg, Germany

Terrence C. Stewart (tcstewar@uwaterloo.ca)
Centre for Theoretical Neuroscience, University of Waterloo
200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

Abstract

According to mental model theory, spatial reasoning is based on the construction and variation of mental models representing spatial arrangements. Several effects in human spatial reasoning are known to support this theory, for example the ordering effect. Yet, reasoning effects have been observed for which the cognitive mechanisms are not entirely explained. To investigate how these effects can be attributed to neural computation, we modeled spatial reasoning in the Neural Engineering Framework.

We selected three experiments to simulate tasks in a cognitive model based on an internal display. In our model, performance declines with an increase of objects which is explained by the neural drift over time. We replicated effects from the studies which we have found to be due to continuous premise integration. By modeling and simulating spatial reasoning tasks, we showed that effects reported in psychological studies can be explained by the emergent properties of neural computation.

Keywords: Neural Engineering Framework; spatial reasoning; relational Reasoning; cognitive Modeling

Introduction

When interacting with the world, we are in need of navigating in the immediate moment or planning for the future. For that, we rarely have complete information about the relational dependencies so that we have to use the given information effectively. Deductive relational reasoning facilitates this process by decompressing implicit information from the given. By that, we mean inferring the connections between objects, in this case in space. For example, imagine that you are planning to visit a friend. Your friend lives behind the train station which is behind the old church. Standing in front of the church, you infer that your friend's home is behind the church because you know the relation between train station and church and your friend's home and church.

Although relational reasoning is so prevalent, its mechanisms are still relatively unclear (Hobeika, Diard-Detoeuf, Garcin, Levy & Volle, 2016). From a cognitive perspective, mental model theory (Johnson-Laird, 1983) provides a framework for conceptualizing relational reasoning as a mental action requiring the engagement of spatial cognition. It postulates that the solving of relational reasoning problems functions via the construction of mental models representing abstracted versions of the objects and their relations. According to this framework, first the mental model is constructed, and the abstracted version of the objects are generated. After that, the information is integrated into one mental model which is consequently validated by crosschecking with the given premises or the proposed conclusion (Knauff, 2013).

Computational models of higher cognition and reasoning add to the findings we gain from psychological testing and neuroimaging studies by giving us insights into the cognitive mechanisms underlying the solving of relational reasoning problems. Several attempts have been made to model relational reasoning in frameworks such as ACT-R for spatial reasoning (Boeddinghaus, Ragni, Knauff & Nebel, 2006) and Raven's progressive matrices (Rieman, Lewis, Young, & Polson, 1994, April) and LISA (Knowlton, Morrison, Hummel & Holyoak, 2012). Models in the Neural Engineering Framework (NEF), c.f. Rasmussen & Eliasmith (2011) for analogical reasoning, are particularly interesting because a biologically plausible modeling approach can be taken in which we can try to bridge subsymbolic and symbolic processing to gain insights into the effects of neural information processing. Our previous model of relational reasoning by Wertheim & Lohmeyer (2017) aimed at simulating relational reasoning but was intended as an exploration of modeling possibilities regarding relational reasoning in the NEF. Since mental model theory suggests that relational reasoning does not rely on specific rules but on spatial manipulation and verification, the modeling of relational reasoning in the NEF remains an interesting domain.

We aim at modeling relational spatial reasoning in the NEF for which we have constructed a model based on a spatial display representing the object's relations. We expect this to be representative of the mental model theory's account of reasoning. With subsequent simulations, we aim at observing that the model exhibits a sensitivity to cognitive load in terms of number of premises, the replication of continuity effects, meaning that continuous premises are easier to process than discontinuous ones (Nejasmic, Bucher and Knauff, 2015), and generally an enhanced understanding of these phenomena.

The Reasoning Tasks

We selected two articles investigating the principles on which human spatial reasoning is based (see Table 1). In the experiment by Bucher, Krumnack, Nejasmic and Knauff (2011), it was tested how reasoners, when confronted with inconsistent premises, adjust the previously built mental

model to fit to the last given premise. The reasoners were confronted with two initial premises:

P1: A is left of B.

P2: C is left of A.

The mental model based on this information is:

C – A – B

Afterwards, the last premise is presented to the reasoner which can either be consistent (C is left of B) or inconsistent (B is left of C) with the formerly build model. Afterwards, the reasoner is asked to adjust the previous mental model so that the objects' locations are consistent with the new premise. For this, two possible scenarios are hypothesized: either the reasoner relocates the last given object (A – B – C) or the to-be-located-object (B – C – A). Bucher et al. (2011) found that in 87.78% of the cases, participants relocated the to-be-located-object (LO).

The second experiment was conducted by Nejasmic, Bucher and Knauff (2015). In this experiment, it was examined how the continuity of the premises' content influences participants' reasoning capabilities. It was found that it is easier for participants (in terms of enhanced and faster performance) to construct models from premises that are continuous (A is left of B, B is left of C, C is left of D) than semicontinuous (B is left of C, C is left of D, A is left of B) or discontinuous (C is left of D, A is left of B, B is left of C) tasks. Reasoners seek to integrate all tokens as fast as possible and try to avoid relocating objects that are already represented in the mental model. Moreover, reasoners solve discontinuous tasks by first constructing one model and then modifying it if necessary, thereby saving cognitive resources by working as sparingly as possible (Goodwin & Johnson-Laird, 2005). In discontinuous tasks, partial models are connected by a temporary link which is weaker than the links between the explicitly related objects. If new premises confirm the link, it is strengthened; if not, the effort is made to relocate.

Table 1: Experimental tasks.

Exp.	Type	Premises	Concl.	Belief revision
Bucher et al., Exp. 1	Consistent	P1: A r B P2: C r A Model: C A B	C r B	
	Inconsis-tent	P1: A r B P2: C r A Model: C A B	B r C	Reloc. last object: A B C Reloc. LO: B C A
Nejasmic et al., Exp. 1	Contin-uous	P1: Ar B P2: B r C	C r D	
	Semi-continuous	P1: B r C P2: C r D	A r B	
Nejasmic et al., Exp. 1 & 4	Quasidis-continuous	P1: C r D P2: A r B	D r A	
	Discon-tinuous	P1: C r D P2: A r B	B r C	

Note: Exp.: Experiment, P1, P2: Premise 1, 2, concl.: conclusion, "r" denotes the relation 'is to the left of', reloc.: relocation, LO: to-be-located-object, "Model" indicates the mental model after the presentation of premise 2.

The Simulation

For modeling reasoning as explained by mental model theory, we decided to incorporate the idea of an internal display (Ragni & Knauff, 2013) for representing the object's locations, since the postulates of mental model theory are conceptually based on the objects being located in two-dimensional space. That is why we have initialized rules about the meaning of the four directions "to the left/right of" and "above/below of" in terms of the internal display (shift on the X/Y-axis).

To construct the model, the NEF (Eliasmith & Anderson, 2003) provides a method for connecting spiking neurons (here we use standard Leaky Integrate-and-Fire neurons) such that groups of neurons represent values, and connections between groups of neurons approximate computations on those values.

For this model, there are two basic things that need to be represented: the premise currently being considered, and the current state of the mental model. Importantly, we assume that only one premise is being considered at a time, and we do not model the process of deciding to switch from one premise to the next. As each premise is considered, the internal mental model should change, such that at the end we have a mental model that respects all the premises.

Representing Premises

To represent the symbolic nature of the premise, we use the idea of "semantic pointers" (Eliasmith, 2013). Here, the neurons are structured so as to represent a high-dimensional space (here, 256 dimensions), and every point in that space is a different possible premise. Specifically, we randomly choose unit vectors in the 256-dimensional space for each of the basic concepts (A, B, C, leftof, rightof, below, above, subject, etc.). We then form a premise such as "A leftof B" by computing A*subject + leftof*relation + B*object, where * is circular convolution. All of the steps in this process have been shown to be computable by spiking neurons, and this is the basis of many existing neural symbolic models (e.g. Rasmussen & Eliasmith, 2011), allowing for the compositionality, systematicity, and productivity seen in human symbol usage. Effectively, this approach means that each premise will have a unique spiking pattern, and that neurons can be organized to extract out this information and respond appropriately.

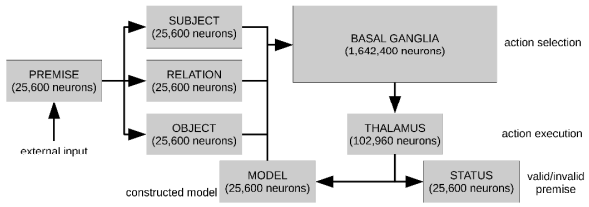

Figure 1: Visualization of our neural model of relational reasoning. Each box consists of the given number of spiking neurons and arrows show neural connectivity.

Figure 1 shows the overall structure of the neural model. The PREMISE neurons store the premise (e.g. "A leftof B") that is currently being input. The SUBJECT, RELATION, and OBJECT neurons extract out the individual parts of the premise (i.e. the connections from the PREMISE neurons to the SUBJECT neurons are set such that if "B leftof C" is in the PREMISE, then the SUBJECT neurons will be driven to fire with the randomly-chosen pattern for "B").

Representing a Mental Model

The MODEL neurons store the current mental model. This mental model is also represented as a 256-dimensional vector. Initially, the neural activity is just the default background activity which represents the vector 0 and the all objects are located at position X,Y (0, 0). As the mental model is built up, the objects are relocated according to the incoming spatial information from the premises. For example, if A is at an X,Y position of (0.3, 0.7) and B is at an X,Y position of (-0.2, -0.5), then we would have the vector 0.3A*X+0.7A*Y-0.2B*X-0.5B*Y, representing the object's positions on the display.

Importantly, this approach means that the MODEL neurons are highly generic. We can add more objects just by adding onto that vector. Indeed, we can also add more spatial relations (Z), or even abstract relations (height, age, cleanliness, etc.) without modifying the neurons representing the mental model, since all of those additions simply result in different 256-dimensional vectors, and the neurons are optimized to represent any vector in the 256-dimensional space. As with all semantic pointer approaches for representing symbol structures using vectors, the more information that is combined into the vector, the lower the accuracy of extracting information from the vector, and there will be a graceful degradation in this accuracy.

Action Selection and Execution

Given the above neural groups' capability of representing both the premise and the mental model, the remainder of the system involves deciding how to modify the mental model given the premises. We do this by implementing a Neural Production System using a model of the mammalian Basal Ganglia and Thalamus. This has been used to model tasks such as Tower of Hanoi (Stewart & Eliasmith, 2011) and the neuroanatomy, spiking activity, and temporal dynamics are matched to empirical data (e.g. Stewart, Choo, & Eliasmith, 2010).

The core idea is that we specify a set of possible actions, and the basal ganglia selects which one to perform, and then this action is executed by the output of the basal ganglia controlling routing in the model of the thalamus. In this case, there are two actions for each relation ("leftof", "rightof", "below", etc.). The first action for "leftof" is the action that should occur if the pattern in the RELATION neurons is "leftof" but those objects in the mental MODEL are the wrong way around (i.e. the X value for the SUBJECT is larger than the X value for the OBJECT). The second action is the one that should happen if the objects are the correct way around.

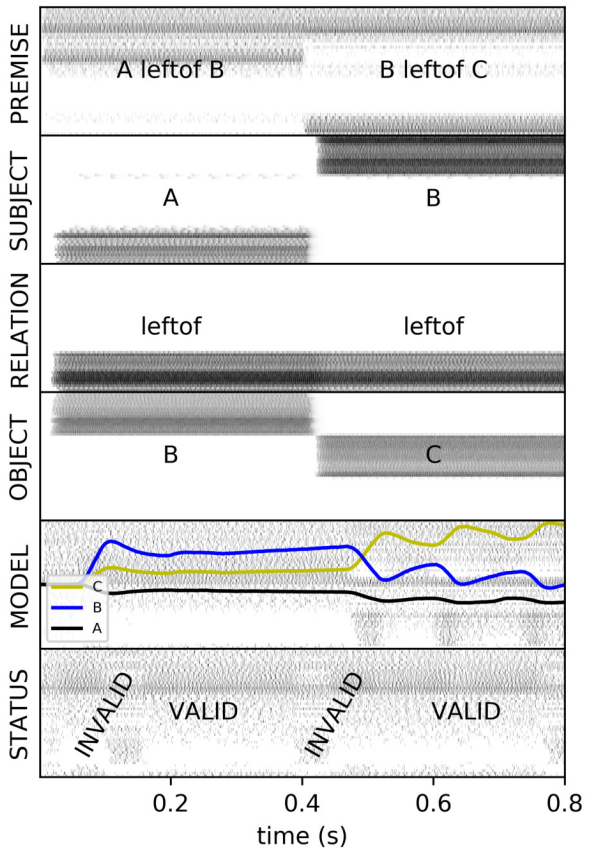

Figure 2: Output of the model presented with two premises ("A leftof B" and "B leftof C"). Each row shows spiking activity of some of the neurons in that region, plus text and lines indicating the meaning of those spikes.

To select among these actions, each action has a *utility function*: a mathematical calculation that takes the vectors currently stored in the SUBJECT, RELATION, OBJECT, and MODEL neurons and computes a single scalar value that represents how relevant this rule is to the current situation. This calculation is computed in the connections from those areas into the input to the basal ganglia. The rest

of the basal ganglia then determines which of those resulting *utility* values is largest and selects that action.

If the first action is selected, two things occur. First, the STATUS neurons are driven towards the pattern for INVALID (so that we know the neural system has decided there's something wrong with the mental model), and, more importantly, the mental model is adjusted. In particular, for "leftof", the neurons will have their activity adjusted towards the pattern for *object**X-*subject**X (where *subject* is the vector for the subject of the premise, such as A, and *object* is the object, such as B). This has the effect of gradually changing the mental model until such a time as the objects in the mental are the right way around, at which point the action selection system will stop choosing that first action, and instead select the second action. The second action does not change the mental model; instead it just adjusts the STATUS neurons towards the pattern for VALID.

Importantly, all of the model presented here is entirely implemented in spiking neurons with synaptic connections between them. The only input is the pattern of activity for the currently considered premise (into the PREMISE) neurons. This input is manually changed after 0.4 seconds to the next premise.

Figure 2 shows the neural activity of randomly selected neurons from different regions of the model as it is presented with the premises "A leftof B" and "B leftof C". Each row shows the spiking activity of individual neurons. The overlaid text indicates the interpretation of that activity (each possible value has a different ideal pattern of activity, and we show the value whose pattern is closest to the observed activity). For the MODEL neurons, we plot the current represented X value for the three objects in the model (A, B, and C).

For the first 0.1 seconds, the system recognizes that its current model (where A, B, and C are all at X=0) is INVALID for the given premise "A leftof B". The effect of this action is to decrease the X value for A and increase the X value for B. Eventually, these values are different enough that the basal ganglia changes to the VALID action, and it stops adjusting the mental model. At t=0.4s, the premise is changed to "B leftof C". Now the basal ganglia recognize that this premise is invalid, and starts changing the X values for B and C.

It should be noted that the odd oscillation pattern seen from t=0.5 to t=0.8 is because neurons do not perfectly store information. The mental model itself drifts slightly, just due to randomness in neural firing. This means that the system is finding the mental model slightly invalid, fixing it, recognizing it as valid, and then the model drifts back to being invalid. This sort of neural drift has been useful for modeling forgetting in working memory (Choo & Eliasmith 2010).

Concerning our tasks, we expect to observe degrading performance when the number of objects is increased, due to the inherent randomness of neural computation. Further, we expect to simulate the effects by Bucher et al. (2011) and Nejasmic et al. (2015). Considering the former, we expect to see 1) a correct ordering of the objects, 2) in inconsistency detection and a preferred relocation of the LO in contrast to the last object. Considering the latter, we expect a degradation of the model's performance from continuous, to semicontinuous, to quasicontinuous, to discontinuous tasks.

Evaluation and Comparison

After the development of the model, we evaluated it by comparing the outcomes of the simulation to behavioral data. For evaluating the model's performance, we used a score of either 0 for arrangements inconsistent with the premises or 1 for arrangements consistent with the premises. In the case of the inconsistent task by Bucher et al. (2011), score 0 denotes inconsistent arrangements, score 1 the relocation of the last object and a score of 2 for the relocation of the LOs. We have simulated each task 60 times to achieve robust results and evaluated how often the simulation's outcome was a correctly arranged set of objects. We have set the timepoint of evaluation for the tasks by Bucher et al. (2011) and Nejasmic et al. (2015) to 0.91 seconds, which is 0.11 seconds after the presentation of the last premise. For the tasks aiming at the cognitive load of reasoning, we have set the evaluation timepoint at 0.51 seconds for the task with three items and added 0.4 seconds per additional premise.

Table 2: Behavioral and simulation results.

		Correctness	
Task	Tested for	Exp.	Simulation [CI 95%]
Bucher et al., Exp. 1	Correct order	98.64%	93.34% [0.84, 0.98]
	Relocation LO	87.78%	94.00% [0.73, 0.99]
	Relocation last object	12.22%	5.56% [0.00, 0.27]
Nejasmic et al., Exp. 1, 4	Continuous	92%	78.33% [0.66, 0.88]
	Semicon-tinuous	79%	6.67% [0.02, 0.16]
	Quasidis-continuous	81%	80.00% [0.67, 0.89]
	Discon-tinuous	59% (Exp.1) / 67% (Exp.4)	76.67% [0.64, 0.87]
Cognitive load	2 premises		100% [0.95, 1]
	3 premises		80.00% [0.68, 0.89]
	4 premises		40.00% [0.28, 0.53]
	5 premises		15.00% [0.07, 0.27]
	6 premises		1.67% [0.00, 0.09]

Note: Exp.: Experiment.

Concerning the cognitive load, we have found that the more objects have to be considered by the model, the lower the performance (in terms of correct arrangement of the objects, see Table 2). This is because information is lost

after several seconds due to the instability accompanying neural computation, imitating the process of gradual forgetting. This means that information degenerates over time. Since each premise is only presented once, information from the first premises vanished over time (see Figure 3 for the arrangement of object 1 and object 3).

Concerning the tasks by Bucher et al. (2011), the correctness in consistent tasks was 98%, which is almost identical with the behavioral results (98.64%). In inconsistent tasks, the simulations revealed that the LO is relocated in 93.75% of the cases. This is because the information conveyed by the new (third) premise temporarily overrides the constellations from the former premises. This explains the findings of Bucher et al. (2011) why the LO, rather than the last object, is relocated during the processing of the inconsistent premise.

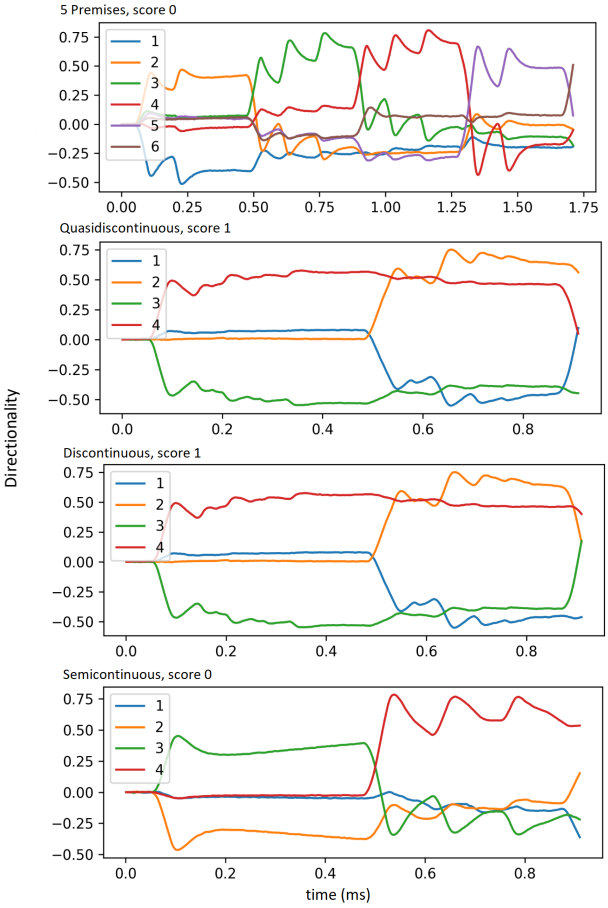

Figure 3: Internal displays during the simulations. Numbers 1-6 indicate the respective objects. Positive directionality (y-axis) indicates the right side and vice versa., the x-axis represents time in miliseconds.

Regarding the tasks by Nejasmic et al. (2015), the simulations showed that the model is most successful in predicting human performance (regarding correctness) in continuous tasks. Performance is second best in quasidiscontinuous and discontinuous tasks and lowest in semicontinuous tasks. This contrasts the behavioral results which exhibited the following order: continuous, quasidiscontinuous, semicontinuous and discontinuous tasks. The performance in continuous tasks has to be set in relation to the model's performance in tasks with three premises (four objects). The correctness levels (81.6% and 80.0%) are similar and performance in this range is comparable to the general model's performance when processing continuous relations with four objects. The performance in quasidiscontinuous tasks is second highest in the simulation, because these relations are easy to process for the model. As can be seen in Figure 3, since the first two objects are not considered in premise two, these relations can be held steadily during the processing of the second premise. Concerning discontinuous tasks, the model can hold four objects (and their relations) during the first two premises but when the third premise is presented, connecting two of the former objects, the former relations are in some cases confused. A distinctively low performance is exhibited in semicontinuous relations (6%). In the first premise, object three is on the (relative) right side (of object 2) and in the second premise, object three is on the (relative) left side (of object four). This poses a processing difficulty because information about object three is temporarily contradictory and the information about the relation between object two and object three is lost after the presentation of the second premise.

Conclusion

Our modeling idea of implementing a two-dimensional display to solve spatial relational reasoning tasks is based on findings of behavioral studies from cognitive psychology. Our intent was to gain insights into the effects of neural information processing in relational spatial reasoning. We found that model's performance deteriorating when more objects are to be considered in continuous premises. This is due to information from former premises vanishing over time due to the drift inherent in neural computation. Also, we replicated the findings by Bucher et al. (2011), thus our model relocates the LOs in indeterminate tasks since information from the premises is integrated continuously, while temporarily overriding information from former premises. This effect is also exhibited when considering the quasidiscontinuous tasks by Nejasmic et al. (2015) in which the model performs second best since spatial information about an object is stored most effectively when the next premise does not contain any spatial relational information about the former object.

Nonetheless, we faced some limitations concerning the model. The behavioral data were only reported as accumulated percentages, which limited our comparison between model and data. Concerning our decision when to stop the simulation and evaluate the results, we could have considered the participant's reaction times. We decided to not use this measure since we do not yet expect the model to reflect all cognitive processes contributing to the reaction time. Hence we chose preliminary timepoints corresponding

to the number of objects involved. Concerning the inconsistent tasks by Bucher et al. (2011), we diverged from the task presentation they have used. Bucher et al. (2011) asked participants whether the task would be consistent or inconsistent and then demanded them to adjust their mental model in case of inconsistent tasks. In the inconsistent tasks, we just presented the inconsistent premise and did not explicitly inquire whether it is inconsistent or not. Lastly, the model's performance was considerably lower in semicontinuous tasks compared to the behavioral results, due to the effects of continuous premise integration. This could be averted by magnifying the impact of the earlier premises and by that preventing the confusion of relational information about former objects.

Future Work

Since we tested our model on similar reasoning tasks, it would be interesting to test different task types in the model, such as tasks involving cardinal directions (Ligozat, 1998) or abstract relations (Knauff & Johnson-Laird, 2002). This would be feasible because of the vector space which does not only work for spatial tasks but also for abstract dichotomies (e.g. clean and dirty). Concerning the inference types, we envision improving the model in regard to the difficulties concerning processing semicontinuous tasks. For that, information from earlier premises could be valued more strongly than from latter premises so that later contradictory relations do not override them. Further, it would be interesting to conduct a thorough comparison between our model and the model built by Ragni & Knauff (2013). In this model, a two-dimensional display was implemented to simulate spatial relational reasoning. In contrast to our model, this one is based on different principles such as the solidity of the entities (they cannot easily cross each other) and rule-based, rather than dynamic processing.

Acknowledgments

This work was funded by the BrainLiksBrainTools Cluster of Excellence (DFG, grant #EXC1086), DFG Heisenberg RA 1934/4-1 and RA 1934/2-1.

References

Boeddinghaus, J., Ragni, M., Knauff, M., & Nebel, B. (2006). Simulating spatial reasoning using ACT-R. In *Proceedings of the seventh international conference on cognitive modeling* (pp. 62-67).

Bucher, L., Krumnack, A., Nejasmic, J., & Knauff, M. (2011, January). Cognitive processes underlying spatial belief revision. In *Proceedings of the Cognitive Science Society*, *33*(33), 3477-3482.

Choo, F. X., & Eliasmith, C. (2010, January). A spiking neuron model of serial-order recall. In *Proceedings of the Cognitive Science Society*, *32*(32).

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

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge: MIT Press.

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

Hobeika, L., Diard-Detoeuf, C., Garcin, B., Levy, R., & Volle, E. (2016). General and specialized brain correlates for analogical reasoning: A meta-analysis of functional imaging studies. *Human Brain Mapping*, *37*(5), 1953-1969.

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

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. MIT Press.

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

Knowlton, B. J., Morrison, R. G., Hummel, J. E., & Holyoak, K. J. (2012). A neurocomputational system for relational reasoning. *Trends in Cognitive Sciences*, *16*(7), 373-381.

Ligozat, G. É. (1998). Reasoning about cardinal directions. *Journal of Visual Languages & Computing*, *9*(1), 23-44.

Lohmeyer, M., & Wertheim, J. (2017, July). *Modeling Relational Reasoning in the Neural Engineering Framework*. Poster presented at MathPsych/ICCM 2017, Warwick, United Kingdom.

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.

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

Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generation in inductive reasoning. *Topics in Cognitive Science*, *3*(1), 140-153.

Stewart, T. C., Choo, X., & Eliasmith, C. (2010, August). Dynamic behaviour of a spiking model of action selection in the basal ganglia. In *Proceedings of the 10th international conference on cognitive modeling* (pp. 235-40).

Stewart, T., & Eliasmith, C. (2011, January). Neural cognitive modelling: A biologically constrained spiking neuron model of the Tower of Hanoi task. In *Proceedings of the Cognitive Science Society*, *33*(33).

Rieman, J., Lewis, C., Young, R. M., & Polson, P. G. (1994, April). Why is a raven like a writing desk?: lessons in interface consistency and analogical reasoning from two cognitive architectures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 438-444). ACM.