Deductive Spatial Reasoning: From Neurological Evidence to a Cognitive Model

Marco Ragni, Thomas Fangmeier, Sven Brüssow

{ragni, fangmeier, sven}@cognition.uni-freiburg.de Center for Cognitive Science, University of Freiburg Friedrichstr. 50, D-79098 Freiburg, Germany

Abstract

Cognitive modeling aims more and more to explain, predict and integrate behavioral data with brain activations found in fMRI studies. In this article we analyze transitive inferences (e.g. A is left of B and B is left of C then A is left of C) during the spatial reasoning processes. Behavioral findings suggest that reasoners tend to construct a mental model from the premises, which they in turn use to inspect to draw inferences. A reanalysis of our own previous fMRI-study investigating such examples provided us with brain activations pattern. A cognitive model using the (restricted) Bold-function in ACT-R 6.0 can partially predict and explain the results. The findings, limits and potentials of the current representation of the Bold-function in ACT-R are briefly discussed.

Keywords: Deductive reasoning; fMRI; ACT-R

Introduction

Assume you receive the following information:

The door is to the left of the garage.

The car is to the right of the garage.

Given this set of premises it is easy to draw an inference like "the car must be to the right of the door". But how do we reason about such so-called three-term problems? Which role plays working memory in such tasks? There are competing and different theories in cognitive science to explain the actual human reasoning process.

The *Theory of Mental Logic* introduced by (Rips, 1994) argues syntactically. This theory claims that humans apply transitivity rules to a given set of premises without constructing spatial representations, e.g. "If A is left of B and B is left of C then A is left of C". Standing in the tradition of AI-Approaches, there are, however, a number of problems involved, e.g. with regard to memory burden or the number of rules necessary to solve tasks (Ragni, 2008).

In contrast, the *Theory of Mental Models* (MMT) argues that humans construct mental models which are an internal representation of objects and relations in spatial working memory, matching the state of affairs given in the premises. The semantic theory of mental models is based on the mathematical definition of deduction, i.e. a propositional statement C is a consequence of a set of premises P, if in each model A of the premises P, the conclusion C is true. The mental model theory (MMT) assumes that the human reasoning process consists of three distinct phases: (1) the model generation phase, in which a first model is constructed out of the premises, (2) the inspection phase, in which the model is inspected to check if a putative conclusion is consistent with the current model. And (3) the validation phase, in which alternative models are generated from the premises that refute this putative conclusion (Johnson-Laird, 2001). A mental model is constructed incrementally from its premises (Ragni, Fangmeier, Webber, & Knauff, 2007) following the principle of economicity (Manktelow, 1999). Such a model construction process saves working memory capacities because new information is immediately processed and integrated into the model (Johnson-Laird & Byrne, 1991; Rauh, Knauff, Kuß, Schlieder, & Strube, 2005).

Both theories can explain a number of results but MMT is more widely accepted as the explaining theory in relational reasoning (e.g., Rauh et al., 2005; Jahn, Knauff, & Johnson-Laird, 2007; Goodwin & Johnson-Laird, 2005).

A *cognitive modeling* of this theory has several advantages: (i) this theory is more formally presented, (ii) it is fully specified in terms of necessary operations to process such problems as described above, and with the new Bold-functions in ACT-R 6.0 (iii) it allows for a prediction and model of the underlying brain activations. Especially, the last aspect has become more and more important in recent years. Foundational work has been done by Anderson, Qin, Stenger, and Carter who conducted and analyzed simple algebra tasks and developed a first model integrating fMRI-findings in ACT-R (Anderson, Qin, et al., 2004). More precise, based on ACT-R 6.0 they developed an information-processing model to predict the blood oxygenation level-dependent (BOLD) response of functional MRI in symbol manipulation tasks. Base-level activation learning in the ACT-R theory can predict the change of the BOLD response in practice in a left prefrontal region reflecting retrieval of information. In contrast, practice has relatively little effect on the form of BOLD response in the parietal region reflecting imagined transformations to the equation or the motor region reflecting manual programming.

In this article, we present a cognitive model for three-term series problems of spatial arrangements integrating a previous fMRI study. It is structured as follows: In the next section, we briefly introduce the experimental design, settings, and the fMRI-findings. Then, we proceed outlining our ACT-R model. Finally we compare the model results with the empirical results.

fMRI During Visual Relational Reasoning

We briefly report a study from our group (Fangmeier, Knauff, Ruff, & Sloutsky, 2006) in which different neural networks for three phases of the MMT during spatial relational reasoning were supported.

Participants. Twelve right-handed male students took part in the study. All were instructed and trained outside the scanner in order to minimize the learning process while scanning and to increase their accuracy.

Materials. The presented material in the original study consists of two conditions, 32 reasoning and 32 maintenance verification tasks for each subject. Since we just want to model the reasoning process in ACT-R we report only the reasoning task in detail. One reasoning task consists of two premises with three letters (V, X, Z in random order) in a spatial horizontal configuration as well as an offered conclusion. Each premise and the conclusion consists of two letters with a spatial relation. The spatial relation between the two letters of each premise or conclusion was coded by placing it right or left from the center of the screen. A sentential version of the given example in Fig. 1 would be: "X is to the left of V (premise 1) and "Z is to the right of V" (premise 2). For these premises, it follows "X is to the left of Z" (mental model which was constructed). Participants were asked to decide if an offered conclusion was correct. One of two alternative conclusions were offered: a valid one (as in Fig. 1) "X is to the left of Z" or an invalid one "Z is to the left of X".



Figure 1: Sequential presentation and timing of the premises and the conclusion (cp. section Materials).

Procedure and Data Acquisition. The participants were trained outside the scanner with 12 similar problems and had to reach at least 75% accuracy for participation. The trials were presented in an event-related design with four separate runs. Each run consist of eight reasoning and eight maintenance tasks in a random order. As noted before we report in this article only the procedure and results of the reasoning tasks.

The timeline of the complete task was as follows: Each task was introduced with the letter "S" in the center of the screen ("Schliessen" in German) for reasoning followed by a pause for 1 sec. Each premise and conclusion began with the presentation of the first letter for 1.5 sec, followed by

the second letter for 1.5 sec and a pause for 1 sec. Therefore each of the premises, and the conclusion lasted for about 4 sec. Overall the whole trial lasted for about 14 sec. In half of each premise or conclusion the first letter appeared on the left position, followed by the letter on the right position. In the other half of the tasks the first letter appeared on the right position. The term order variation prevented the participants from anticipating the next letter and from drawing the conclusion during the second premise. Further the variation of the term order is well established in the reasoning literature (Knauff, Rauh, Schlieder, & Strube, 1998). During presenting of the conclusion the accuracy was recorded via a two-button box. Scanning was performed on a 1.5 T Siemens Vision scanner. Functional images were collected with a gradient-recalled echo-planar imaging (EPI) sequence, allowing the sampling of 30 parallel slices covering the whole brain [TR repetition time): 4000 msec; TA (acquisition time): 3126 msec]. The exact scanning information can be seen in Fangmeier and colleagues (2006).

Design. Functional and anatomical images were reoriented so that the anterior commissure corresponded to the origin of the three-dimensional standard coordinate system used in the software SPM99 (1999). The four runs for each subject were separately realigned and corrected for motion, and underwent slice timing correction. Each subject's anatomical image was coregistered with a 40-slice EPI and the functional images of each run. The parameters for spatial normalization were determined from the anatomical images of each subject, and were applied to the corresponding functional images. Images were finally smoothed with a 8-mm full-width half-maximum Gaussian kernel.

fMRI Statistical Analyses. The hemodynamic response to the premises and conclusions was modeled with event-related delta functions, which were convolved with the canonical hemodynamic response function and its temporal derivative employed in SPM99. Low-frequency confounds were excluded from the model with a high-pass filter (192 sec cutoff), and an autoregression AR(1) model excluded the variance explained by the previous scan. The six realignment parameters for each run were included as covariates to avoid motion artifacts. First-level contrast images for every subject and contrast were then used for a random effects analysis to draw inferences on brain activation during the experimental problems. Only correctly answered problems were included in the analysis. All reported clusters within the conditions and the conjunction analysis are significant at the cluster level p.05, corrected for multiple comparisons (threshold t = 3.0). The contrasts were calculated as follows: premise processing phase (Premise 2 minus Premise 1), integration phase (Premise 2 minus Conclusion), validation phase (Conclusion minus Premise 2).

Further the beta values from the essential significant clusters were extracted. For each of the three different phases (premise 1, premise 2, conclusion) a cluster with ± 12 mm around the peak voxel was extracted from the beta images of the SPM statistic. The beta value for each phase represents the difference between brain activation during this phase and the overall mean derived from the whole brain, which is the actual value of the corresponding phase. The value is not a percent signal change but a difference to overall mean with an arbitrary unit. If the beta value is positive (or negative) the activation is higher (or lower, resp.) than the average activation as illustrated in the bar charts of the human data in Fig. 6.

Results. Our findings support the main assumptions of the MMT with respect to distinct phases.

During the initial *premise processing phase*¹ (see Fig. 2 A, B) for both presented premises occipito-temporal structures are activated with the following main Brodmann areas (BA 18, 19, and 37). These areas are active during tasks which are involved in visual working memory and imagery (Kosslyn, Ganis, & Thompson, 2001; Postle, Stern, Rosen, & Corkin, 2000) and with the ventral "what"-stream (Ungerleider, Courtney, & Haxby, 1998).

The following *integration phase* (see Fig. 2 B) shows an additional area in the anterior prefrontal cortex which covers the BA 32 and 10. Tasks in which multiple relations have to be hold simultaneously activated area 10 (Christoff et al., 2001; Prabhakaran, Rypma, & Gabrieli, 2001; Waltz et al., 1999) and a review of functions of the anterior prefrontal cortex assume that this area is responsible for the combination and coordination of multiple cognitive operations (Ramnani & Owen, 2004). Especially support for the premise integration comes from Kroger and colleagues (2002).

In the *validation phase* (Fig. 2 C) a putative conclusion has to be verified. The activation switched from the visual working memory (BAs 18, 19, and 37) to the posterior parietal cortex (BAs 7 and 40). This areas are frequently activated during spatial processing (Burgess, Maguire, Spiers, & O'Keefe, 2001) and the integration of sensory information from all modalities into an egocentric spatial representation (Xing & Andersen, 2000; Andersen, Snyder, Bradley, & Xing, 1997).

Cognitive Model

ACT-R is a cognitive architecture that consists of a number of modules each associated with certain cortical regions (Anderson, Bothell, et al., 2004; Anderson, Qin, et al., 2004; Anderson et al., 2008). When, for example, an ACT-R model that has been built is pressing some key on a keyboard, the manual module will be active and this predicts BOLD activity in the corresponding motor region in the brain. ACT-R's central executive—its procedural backbone—is a production system represented by the procedural module that is associated with the caudate region. Each time a production fires



Figure 2: Brain activation during reasoning. Activated regions are contrasts for the three phases calculated with SPM: premise processing (P1), integration (P2), and validation phase (C). The activations were significant at the cluster level calculated with SPM99 ($p \le .05$, corrected, threshold t = 3.0).

ACT-R predicts the BOLD rate in the caudate region is going up with a certain time lag as is known from real fMRI studies.

The procedural module controls ACT-R's strictly serial behavior; only one production in a time may fire. The modules, however, may operate in parallel and communicate over their buffers, each capable of holding one chunk of information. Hence, a production can require information from more than one module's buffer. Once a module is active, however, it only can become active again in a subsequent request, when it is free again.

The ACT-R model operates on three different kinds of chunks: (1) premise and conclusion chunks, (2) grid chunks, and (3) mental model chunks.

Premise and conclusion chunks are structurally equivalent and the corresponding chunk type defines two slots for the left and right term. Each time a term is presented the ACT-R model tries to integrate the term into the premise chunk or conclusion chunk respectively. After completion of P1 the corresponding chunk is integrated into the center of the second kind of chunk, a grid with four vacant positions (cp. Figure 3, P1). Once the mental model of P1 is complete it is cleared from the imaginal buffer by placing a new chunk of the type grid representing the position of the current model and adjacent free positions around it into the imaginal buffer. The first term of P2 is presented and if it has not been seen yet, the current grid is cleared from the the imaginal buffer. A new premise chunk with only one term is placed in the imaginal buffer instead.

This, however, is the first source of an possible error. Each

¹We denote the phases slightly different.



Figure 3: The ACT-R model processing the premises (P1, P2) and conclusion: the columns represent the different buffers each holding the respective chunks. While P1 is presented on the screen, both terms are successively placed into the slots of a two-model chunk that is generated in the imaginal buffer. In a subsequent step the information of this two-model chunk is integrated into a grid-chunk (indicated by 4 cells). After an analog processing of P2 the corresponding two-model is merged with the grid chunk. Then the conclusion chunk is iteratively compared with the conclusion chunk.

time a grid chunk is released to declarative memory and an identical one is detected both get merged to one and its activation is boosted. Hence, the more often two chunks have been merged in the past, the more dominant the result gets and interferes with the grid chunk that has most recently been created. Hence, recency is not necessarily a guarantee for successful retrieval.

The next term is presented and integrated into the premise chunk in the imaginal buffer. The grid chunk is retrieved from declarative memory and is now placed into the retrieval buffer. Now both chunks can be tested on the right hand side of a production and finally the open position in the grid can be filled according to the position in the premise chunk (cp. Figure 3, P2).

In a next step the imaginal buffer holding the current grid chunk, however, has to be cleared again in order to build the conclusion chunk. At the moment the first term of C is presented the model clears the grid chunk from the imaginal buffer in order to make it free for the creation of the conclusion chunk. The creation of the conclusion chunk is analogous to the creation of the premise chunks as described above.

Each cell of the built model of C is iteratively compared with the grid chunk in the retrieval buffer (cp. Figure 3, C). Here, however, a second source of a possible error can occur. In case of the release of a chunk to declarative memory and the retrieval from it in a subsequent step, the same problem occurs as described above. If the time between clearing a chunk from the imaginal buffer and being retrieved again is at a minimum, it can be retrieved again in order to be compared with the conclusion chunk. Otherwise the most general chunk that has repeatedly been merged in the past and that consequently is most dominant in terms of it strengthened activation may get retrieved erroneously. This chunk may cause an error at the comparison stage, because the cues in its slots may not match those from the conclusion model.

Empirical Evaluation and General Discussion

In the sense of Anderson and colleagues the presented model is not an attempt to cover all aspects of deductive reasoning and mental model theory but to add to a methodology that has recently attracted the attention of researchers: the evaluation of cognitive models with fMRI data and vice versa (Anderson et al., 2008, 1325). Nevertheless, the accuracy of the human and the model data fits quite well (human = 93%, model = 94%).

Table 1 shows the brain regions that are supposed to be linked to the buffers of ACT-R modules and Figure 6 illustrates the predicted BOLD responses of the model.

Table 1: Brain regions and corresponding Brodmann areas associated with ACT-R modules (Anderson et al., 2008, 1327).

Region	Brodmann	Module
Motor1	2, 4	Manual
ACC	24, 32	Goal
PPC	7, 39, 40	Imaginal
LIPFC	45, 46	Declarative
Caudate		Procedural
Fusiform	37	Visual

BOLD responses have been computed of a model run that simulates 32 trials of deductive reasoning tasks. Figure 4 shows the overall mean values for each phase P1, P2, and C. Figure 5 shows the continuous course of the overall mean BOLD response predictions for selected ACT-R modules. In all three phases there is almost no change in the rates for the manual module. The reason is that there is a lag of up to four seconds until the corresponding BOLD activity reaches its maximum. This, however, happens in the 12 seconds time window after an answer key has been pressed and consequently cannot be seen in the presented time frame. When the next trial starts the BOLD activity has decayed to its normal rate. This is analogous to the human data and therefore there is also no prominent BOLD activity for the three phases.

The declarative module, too, shows only low activity, slightly increasing towards the end. The reason is that model heavily relies on involving the imaginal module. Only when there are two chunks that have to be tested in a production concurrently there is the need to temporarily clear one chunk from the imaginal buffer because each buffer can only hold



Figure 4: The overall mean BOLD response predictions for six ACT-R modules for the first premise (P1), the second premise (P2), and the conclusion (C).



Figure 5: The course of the overall mean BOLD response predictions for selected ACT-R modules.

one chunk at a time. The model chunk gets retrieved immediately from declarative memory again via the retrieval buffer. There is, however, in most cases only one retrieval towards the end of P2 so that the predicted BOLD response is not comparable to that of the imaginal buffer. Only when the first term of the second premise presented on the screen has not been seen before a second retrieval is required. In addition, the maximum rate will, similar to that of the motor module, be in the lag of 12 seconds between two trials.

In the following, we concentrate the empirical evaluation of the correspondence between model and fMRI data on those three brain regions that have both been investigated in the study of Fangmeier et al. (2006) and that are also linked to the buffers of ACT-R modules. Figure 6 directly compares human data with model data. The scales for the human data, however, should be compared with caution, because typically in fMRI research the Δ -adjusted BOLD function with respect to mean activation is reported. For the present work this implied a transformation of the Δ -adjusted BOLD to absolute values in order to get comparable charts with the ACT-R BOLD response predictions. All predicted values of ACT-R were within an interval of [0.0-1.0], whereas in fMRI the beta means are not restricted to a fixed interval (i.e. values can also be negative or beyond 1.0). However, comparing the results at a qualitative level shows a similar pattern as is illustrated in Figure 6. An interesting difference between the predicted



¹ Calculated from the Δ -adjusted BOLD (Fangmeier et al., 2006).



Figure 6: The overall mean BOLD responses for three brain regions (top) and the corresponding predictions of three ACT-R modules (bottom) for the first premise (P1), the second premise (P2), and the conclusion (C): the occipito-temporal cortex (OTC) overlaps with Brodmann area (BA) 37 and is linked to the visual module; the anterior prefrontal cortex (APFC) overlaps with the anterior cingulate cortex, BA 32), that is linked with the goal module; the posterior parietal cortex (PPC) overlaps with BA 7, 40 and is linked to the imaginal module. Each phase (P1, P2, C) lasts 4 seconds resulting in a total presentation duration of 12 seconds (cf. Fig. 1 and 5).

BOLD function and the experimental results is within P2: the difference between the BOLD linked to the visual module and the corresponding brain region of occipito-temporal cortex (OTC). This remains still an open question.

Taken together, ACT-R 6.0 offers a powerful possibility to predict behavior and associated brain activations. This allows to model the different levels from neurological evidence to symbolic modeling. Integrating neurological findings have a main advantage for cognitive modeling: The goodness-to-fit can be extended far beyond the behavioral data, especially for the domain of complex cognition (Anderson et al., 2008, 1324). Differences in the setting can be traced back to different modules (which have different activation patterns). Certainly, a main problem is to compare results of the fMRI studies with predictions of the BOLD-function since additional work is necessary to identify the different scaling and intensity of the activations. So in some sense, the predicted BOLD function gives a good intuition, especially for qualitative comparison. Once a refinement of the modules in ACT-R is taken into account the fields of fMRI and cognitive modeling converge stronger.

Future work will integrate and compare the findings on the memory tasks to the deductive reasoning tasks.

Acknowledgements

This research was supported by the DFG (German National Research Foundation) in the Transregional Collaborative Research Center, SFB/TR 8 within project R8-[CSPACE] and the strategic project ActivationSpace. The authors are grateful to Matthias Frorath for assistance in the implementation of the ACT-R model.

References

- Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, 20, 303–30.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036-1060.
- 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.
- Anderson, J. R., Qin, Y., Stenger, A., & Carter, C. S. (2004). The relationship of three cortical regions to an informationprocessing model. *Journal of Cognitive Neuroscience*, 16(4), 637-653.
- Burgess, N., Maguire, E. A., Spiers, H. J., & O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage*, *14*(2), 439–53.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage*, 14(5), 1136–49.
- 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. N. (2005). Reasoning about relations. *Psychological Review*, 112(2), 468-493.
- Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, 35(8), 2075–87.
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, 5(10), 434-442.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Knauff, M., Rauh, R., Schlieder, C., & Strube, G. (1998). Continuity effect and figural bias in spatial relational inference. In *Proceedings of the twentieth annual conference of the cognitive science society* (p. 573-578). Mahwah, NJ: Erlbaum.

- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2(9), 635–42.
- Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cerebral Cortex*, 12(5), 477–85.
- Manktelow, K. I. (1999). *Reasoning and thinking*. Hove, UK: Psychology Press.
- Postle, B. R., Stern, C. E., Rosen, B. R., & Corkin, S. (2000). An fMRI investigation of cortical contributions to spatial and nonspatial visual working memory. *Neuroimage*, 11(5 Pt 1), 409–23.
- Prabhakaran, V., Rypma, B., & Gabrieli, J. D. (2001). Neural substrates of mathematical reasoning: a functional magnetic resonance imaging study of neocortical activation during performance of the necessary arithmetic operations test. *Neuropsychology*, 15(1), 115–27.
- Ragni, M. (2008). Human logic in spatial reasoning. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 933–939). Austin, TX: Cognitive Science Society.
- Ragni, M., Fangmeier, T., Webber, L., & Knauff, M. (2007).
 Preferred mental models: How and why they are so important in human reasoning with spatial relations. In C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, & T. Barkowsky (Eds.), *Spatial cognition v: Reasoning, action, interaction* (pp. 175–190). Berlin: Springer.
- Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, 5(3), 184–94.
- Rauh, R., Knauff, C. H. M., Kuß, T., Schlieder, C., & Strube, G. (2005). Preferred and alternative mental models in spatial reasoning. *Spatial Cognition and Computation*, 5, 239-269.
- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: The MIT Press.
- SPM. (1999). London, UK: Wellcome Department of Cognitive Neurology. (computer software)
- Ungerleider, L. G., Courtney, S. M., & Haxby, J. V. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences, U.S.A*, 95(3), 883–90.
- Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A System for Relational Reasoning in Human Prefrontal Cortex. *Psychological Science*, 10(2), 119–25.
- Xing, J., & Andersen, R. A. (2000). Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. *Journal of Cognitive Neuroscience*, 12(4), 601–14.