

**Using the ACT-R Cognitive Architecture in combination with fMRI Data**

Jelmer P. Borst (jelmer@cmu.edu) & John R. Anderson

Carnegie Mellon University

Please cite as:

Borst, J.P, Taatgen & Anderson, J.R. (in press). Using the ACT-R Cognitive Architecture in combination with fMRI data. In B. U. Forstmann, & E.-J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience*. Springer: New York.

### **Abstract**

In this chapter we discuss how the ACT-R cognitive architecture can be used in combination with fMRI data. ACT-R is a cognitive architecture that can provide a description of the processes from perception through to action for a wide range of cognitive tasks. It has a computational implementation that can be used to create models of specific tasks, which yield exact predictions in the form of response times and accuracy measures. In the last decade, researchers have extended the predictive capabilities of ACT-R to fMRI data. Since ACT-R provides a model of all the components in task performance it can address brain-wide activation patterns. fMRI data can now be used to inform and constrain the architecture, and, on the other hand, the architecture can be used to interpret fMRI data in a principled manner. In the following sections we first introduce cognitive architectures, and ACT-R in particular. Then, on the basis of an example dataset, we explain how ACT-R can be used to create fMRI predictions. In the third and fourth section of this chapter we discuss two ways in which these predictions can be used: region-of-interest and model-based fMRI analysis, and how the results can be used to inform the architecture and to interpret fMRI data.

*Keywords:* ACT-R; Cognitive Architecture; fMRI; model-based fMRI; ROI analysis.

## Introduction

In 1973, Newell wrote a commentary in which he caricatured the current psychological practice as “playing a game of 20 questions [with nature]” [1]. While Newell considered the individual experiments and theories presented at the symposium to be “exceptionally fine” (p. 291), he was worried that the results would never be integrated into an overarching theory of the mind. As a solution, Newell proposed the idea of *cognitive architectures* (the actual term is not in his 1973 paper but was well in use at CMU when Anderson arrived in 1978; see for instance [2]).

A cognitive architecture is first and foremost a psychological theory: it explains for instance how our memory system works. Instead of being limited to a single psychological construct, however, architectures typically account for complete tasks, from perception to response execution. In addition – and unlike most classical psychological theories – a cognitive architecture is implemented as a computer simulation, which can be used to create cognitive models of specific tasks (e.g., the Stroop task, associative recognition, driving a car). This approach has multiple advantages. First, the models yield precise predictions, for instance reaction times and accuracy measures. Particularly when complete tasks are modeled – often models even interact with the same interface as human subjects – a direct comparison with human data is possible. Second, the underlying psychological components (e.g., memory, vision) are shared by the different tasks, and have to be truly general. If a simulated memory system only works for a single task it probably contains too many task-specific constructs. A cognitive architecture forces one to keep the components general enough to work for many different tasks. Third, because complete tasks are modeled, interactions between perception and central cognition (and between cognitive components themselves) arise naturally from the architecture, which can have a large impact on experimental results [3,4].

For decades, models developed in cognitive architectures were validated using response times, accuracy measures, and sometimes eye movements [e.g., 5]. However, behavioral data

does not always provide enough constraints to distinguish between different models [6; Chapter 13]. For example, the time leading up to a response typically consists of multiple cognitive steps, which can be arranged in different ways. Researchers turned to neuroimaging data for additional constraints and guidance in developing architectures [e.g., 6,7]. Cognitive architectures are well-matched to fMRI data: One cannot ignore any of the perceptual, cognitive, or motor components of a task when designing or interpreting fMRI experiments (because they all show up in brain activity) and a cognitive architecture requires that the modeler address all of these components (to get a running model).

In this chapter we describe how the cognitive architecture ACT-R can be used in combination with fMRI data. We will first explain ACT-R in some detail. Then, based on an example task, we will demonstrate the different steps of generating fMRI predictions from an ACT-R model. Subsequently, we discuss two different ways of using these predictions: a region-of-interest analysis and a model-based fMRI analysis. We conclude with a short section on how the two methods complement each other.

## **ACT-R**

Currently, several cognitive architectures are in use, for example SOAR [2], ACT-R [6], EPIC [8], and 4CAPS [7]. In this chapter we will focus on ACT-R<sup>1</sup>, because it has an explicit mapping between components of the architecture and brain regions. However, most ideas in this chapter are also applicable to other architectures.

---

<sup>1</sup> For the range of tasks (and associated publications) that have been modeled with ACT-R, see <http://act-r.psy.cmu.edu/>. ACT-R can also be downloaded from this website.

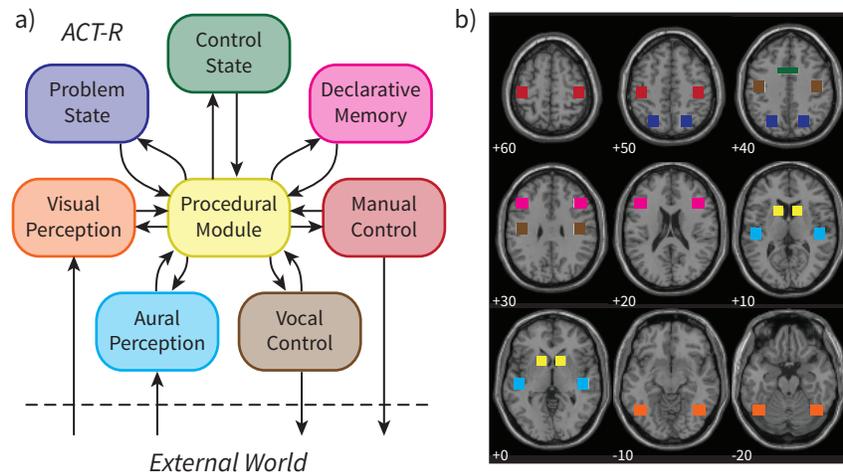


Figure 1. The main modules of ACT-R (a) and associated brain regions (b).

Numbers indicate the  $z$ -coordinate of each slice (MNI coordinates); the colors of the regions correspond to the colors in (a).

ACT-R consists of a set of independent modules that function around a central procedural module (Figure 1a). There are modules for perception (visual and aural) and action (manual and vocal), and several central cognitive modules (for details on the individual modules, see Anderson, 2007, or Anderson, 2005). The modules interact with the procedural module through buffers of limited size. The procedural module consists of rules that specify what cognitive action to take given the contents of the buffers. For instance, a rule might request the retrieval of the meaning of word encoded in the visual buffer. An ACT-R *model* consists of such rules and of knowledge in declarative memory (e.g., the meaning of the word ‘chair’). Thus, ACT-R itself can be seen as the fixed hardware – the architecture – of the mind, while the models function as software that runs on this hardware. The modules of ACT-R have been mapped onto small regions in the brain, which are shown in Figure 1b. These regions are assumed to be active when the corresponding module is active (see the section on region-of-interest analysis).

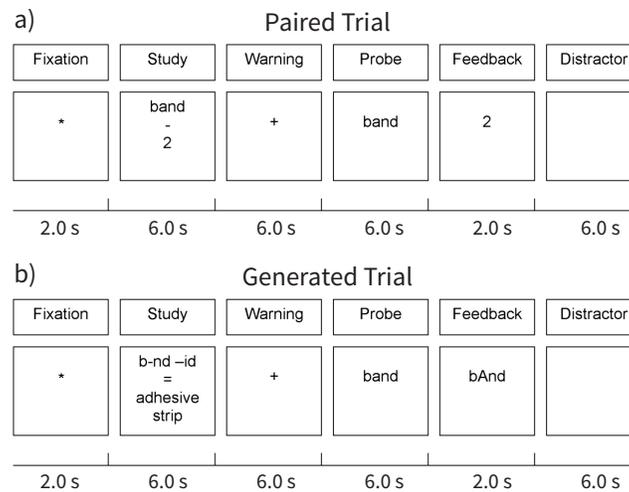


Figure 2. Experimental procedure. Adapted from Figure 1 in Anderson, Byrne, et al. (2008) by permission of the publisher. Copyright 2008 of the original by Oxford University Press.

### Using ACT-R to predict fMRI data

In this section we will describe how ACT-R can be used to predict fMRI data. First, we describe the task that we will use as an example throughout this chapter. We will then introduce the model, followed by how it can be used to generate fMRI predictions. The Lisp code for the model and Matlab code to generate the predictions can be downloaded from <http://act-r.psy.cmu.edu/>, under the title of this chapter.

### The Example Task: Associative Fan

To illustrate the analysis we will use a previously published experiment with an associated ACT-R model (Experiment 2, [9]). This experiment was designed to test the assumption that declarative memory activity is reflected by a region in the prefrontal cortex (see Figure 1b, the pink regions), while representational activity of the problem state module (roughly comparable to a capacity-limited working memory store, e.g., [10]) is reflected by a region in the

posterior parietal cortex (Figure 1b, dark blue regions). To this end, memory and representational requirements were independently manipulated in an associative recognition task.

Figure 2a shows the basic procedure. A trial started with a 2 second fixation screen, followed by a 6 second study presentation of a paired-associate. Subjects were asked to memorize the paired-associate that was presented, in this case ‘band – 2’. The study probe was followed by a 6-second fixation screen, after which a test probe was shown for a maximum of 6 seconds or until the response was given. The test probe consisted of a word (i.e., ‘band’); subjects had to respond with the associated number (i.e., ‘2’).

Memory requirements were manipulated within-subject by varying the delay between study and test items. The trial in Figure 2a is an example of having a study and test item in the same trial, but they could be as far as 7 trials apart. There were three levels: no delay, short delay (1-2 trials), and long delay (6-7 trials). Representational requirements were manipulated between-subject by contrasting a ‘paired’ with a ‘generated’ condition. Figure 2a shows the paired condition; Figure 2b the generated condition. Instead of showing the paired-associate directly, in the generated condition a word phrase was given: ‘b-nd -id = adhesive strip’. Subjects were asked to solve the phrase by finding a single letter to complete it. At test, ‘band’ was shown, and subjects had to respond at the recall test with the position of the letter they had filled in (i.e., ‘2’). Thus, responses were identical in the paired and the generated conditions. The assumption was that subjects would show greater representational activity in the generated case because they had to solve the phrase at study and extract the response at test.

Figure 3 shows the behavioral results, accuracy on the left and response times on the right [for details see 9]. The effect of the delay manipulation is clear: subjects made more errors and were slower to respond when the delay between study and test was longer (response times of correct responses are shown). The effect of the representational manipulation on behavior was more modest: no effects on accuracy, and only a marginal effect on response times ( $F(1,17) = 2.99, p = .10$ ), with the generated condition leading to slightly slower responses. This illustrates

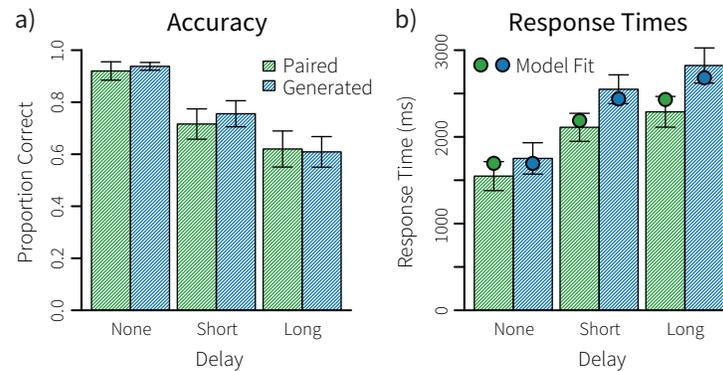


Figure 3. Behavioral results. Error bars indicate standard errors.

why behavioral data are often not detailed enough to constrain computational models: in these data there is almost no difference between the paired and the generated condition. However, the fMRI results will show that there are clear differences between these conditions that are captured in the ACT-R model of the task. Before we turn to the fMRI results we will discuss that ACT-R model, and how such a model can be used to generate fMRI predictions.

### The Model

As explained above, an ACT-R model consists of procedural rules and declarative knowledge that ‘runs’ on the cognitive architecture. Anderson, Byrne, et al. (2008) presented a model that performs the associative recognition task.<sup>2</sup> Figure 4 shows a schematic of model activity for four different trial types. In all conditions, the model starts with encoding the start fixation. When the pair or the phrase is presented two seconds later, it also encodes those, and represents the information in the problem state module (ACT-R’s capacity-limited working memory store). In the paired condition, the model then actively stores the pair in declarative memory. In the generated condition, it completes the phrased based on information retrieved from declarative memory, and stores the completed phrase in memory [see 9 for details]. As the figure indicates, the model assumes no difference in memory activity between the paired and the

<sup>2</sup> A version of the model that was adapted for this chapter can be downloaded from <http://act-r.psy.cmu.edu>, under the title of this chapter.

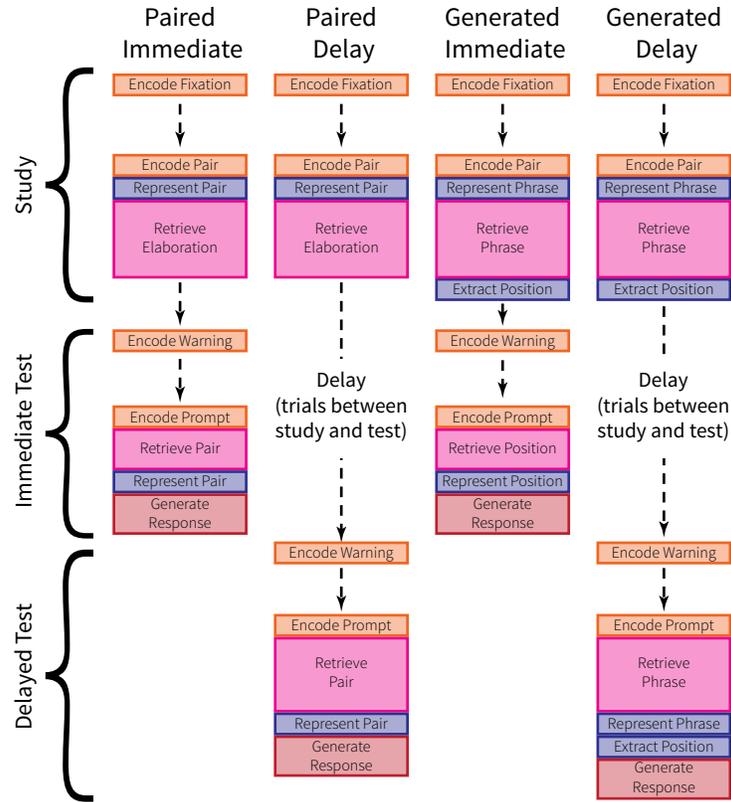


Figure 4. An overview of model activity in four representative conditions. Boxes are not drawn to scale, but indicate the general pattern of module activity. Colors correspond to Figure 1a; orange corresponds to visual activity, blue to problem state, pink to declarative memory, and red-brown to motor activity. Adapted from Figure 6 in Anderson, Byrne, et al. (2008) by permission of the publisher. Copyright 2008 of the original by Oxford University Press.

generated condition in the study phase. However, in the generated condition an extra problem state action is performed to extract the position of the letter that was filled in (i.e., bAnd => 2), resulting in more representational activity.

In the immediate test conditions (i.e., study and test are in the same trial, cf. Figure 2), the model retrieves the pair (in the paired condition) or the position (in the generated condition) from memory. It then represents this information in the problem state module and generates a response.

Thus, both memory and representational requirements are the same in the immediate test phase of the paired and generated conditions. If there is a delay between study and test, it is harder to retrieve the pair from memory, which results in longer declarative memory activity than in the immediate conditions. In addition, in the generated delay condition it is assumed that the model cannot directly remember the position that it filled in, but that has to retrieve the phrase from memory. As a consequence, an extra representational step has to be performed to extract the position again [see 9 for the rationale behind this]. Thus, in the generated delay condition additional representational activity is predicted in the test phase as compared to the other three conditions.

One of the advantages of using a cognitive architecture is that it provides us with latency information of the modules (e.g., visually encoding a stimulus, representing a pair, generating a response). For the current model all parameters were left at their default values, except for the time it takes to retrieve information from memory. This was estimated to fit the model to the behavioral data. The resulting fit is shown in Figure 3b (accuracy was not modeled). The correspondence between model and data is acceptable: the main effects in the data are reflected by the model (originally, the model was compared to data of two experiments, which yielded a better overall fit).

### **Predicting the BOLD response**

The next step in using ACT-R with fMRI data is generating BOLD (blood-oxygen-level-dependent) response predictions. As is known from the fMRI literature, the BOLD response is sluggish with respect to neural activity. This is illustrated in Figure 5a: if there is a spike of neural activity at time 0, the BOLD or hemodynamic response function (HRF) rises slowly to a peak around 5 seconds, declines again and dips under the baseline until it eventually comes back to baseline at around 30 seconds. The HRF is often described with a gamma function or a mix of

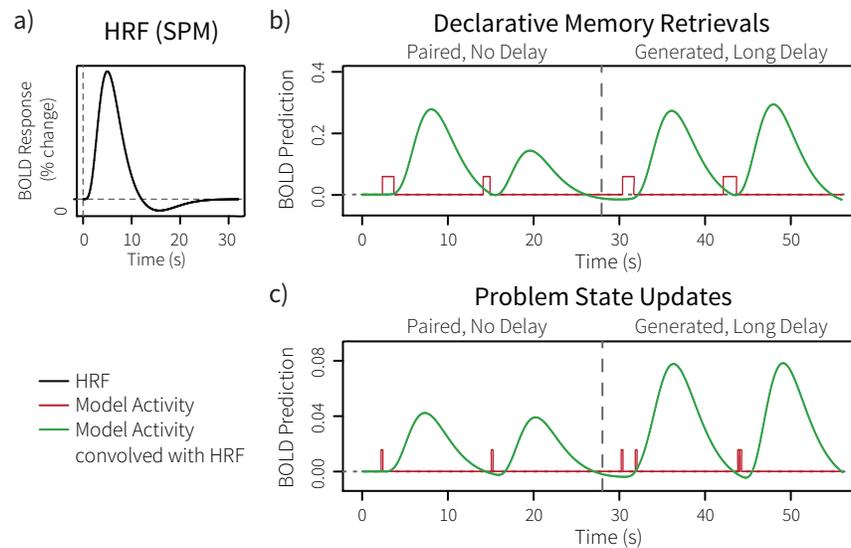


Figure 5. Predicting the BOLD response with ACT-R. Panel a) shows a typical hemodynamic response function; b) and c) show model activity in red and the model activity convolved with the HRF in green.

gamma functions [e.g., 11–13]. In this chapter we will use a difference of two gamma functions from the SPM software package [14].

To generate BOLD predictions, we convolve module activity – which resembles neural activity with respect to its direct timing – with the HRF. This is shown in Figure 5b (declarative memory retrievals) and Figure 5c (problem state updates). The red lines indicate the activity of the modules (cf. Figure 4). Two different trial types are shown: a paired trial without a delay between study and test on the left, and a generated trial with a long delay on the right. As explained in the model description above, a long delay leads to a slower second memory retrieval than no delay, and the generated condition shows more representational activity than the paired condition. The green lines depict the predicted BOLD response: longer activities lead to longer and higher BOLD predictions (declarative memory), and multiple short module activations can be

added up to a single large BOLD response (problem state).<sup>3</sup> In the next two sections we will describe how these predictions can be used for a region-of-interest analysis and for a model-based fMRI analysis.

### **Region-of-Interest Analysis**

Most ACT-R/fMRI papers to date have used so-called region-of-interest (ROI) analyses [e.g., 6,9,11,15,16]. For this analysis stream, all ACT-R modules have been mapped onto small regions of the brain (Figure 1b; see [6] for Talairach coordinates and details of the regions, or [15] for MNI coordinates). The assumption is that these regions are active when the corresponding module predicts activity. For example, declarative memory retrievals should lead to activity in the prefrontal cortex and problem state updates to activity in the posterior parietal cortex. Note that we do not assume that these are the only regions that are active in response to the modules, and neither that these regions exclusively indicate activity of ACT-R modules.

Using predefined ROIs has a number of advantages. By comparing model predictions and data one can validate and constrain models (i.e., if the predictions are off, the model should be improved), which would not be possible without a predefined mapping. In addition, because only a limited number of predefined regions are inspected the typical multiple-comparison problem of fMRI is avoided. This makes it possible to analyze much smaller differences than is possible with conventional fMRI analyses. The obvious disadvantage of an ROI analysis is that it is constrained to the predefined ROIs and ignores the rest of the brain.

---

<sup>3</sup> The amplitude of the predicted BOLD response for declarative memory is much larger than for the problem state module. However, the amplitude is typically fitted separately for each module in a region-of-interest analysis and is not important for the model-based fMRI analysis.

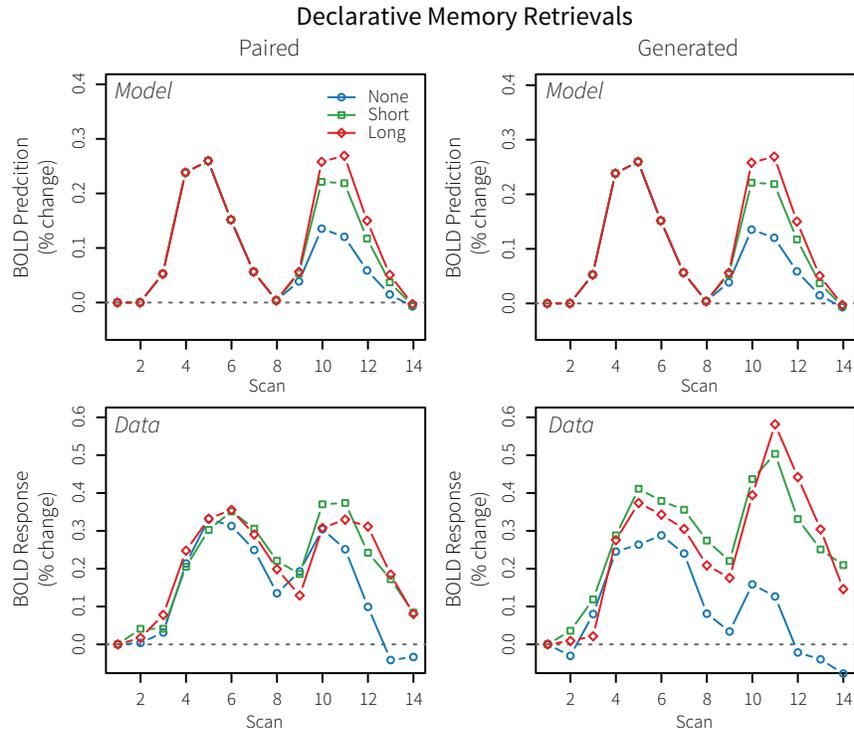


Figure 6. ROI results for declarative memory retrievals. Top panels show model predictions; bottom panels data. One scan = 2 seconds.

Figure 6 (declarative memory) and 7 (problem state updates) show the results for our example dataset. The top panels show the model predictions; the bottom panels show the data. For declarative memory, the model predicted no difference between the paired and the generated conditions, but a clear difference between the no, short, and long delay conditions at test. In the data, we see hardly any differences between these conditions at study – as predicted – and some differences – in the right order – at test, especially in the generated condition (if we average over the paired and the generated conditions the effect is clear, see also [9]). In addition, the peak seems to be larger in the generated condition, but this difference was not significant (see Anderson, Byrne, et al., 2008, for details). With respect to the problem state updates (Figure 7), the model predicted a larger response in the generated condition than in the paired condition, and only an effect of delay in the generated condition at test. These predictions were matched by the data.

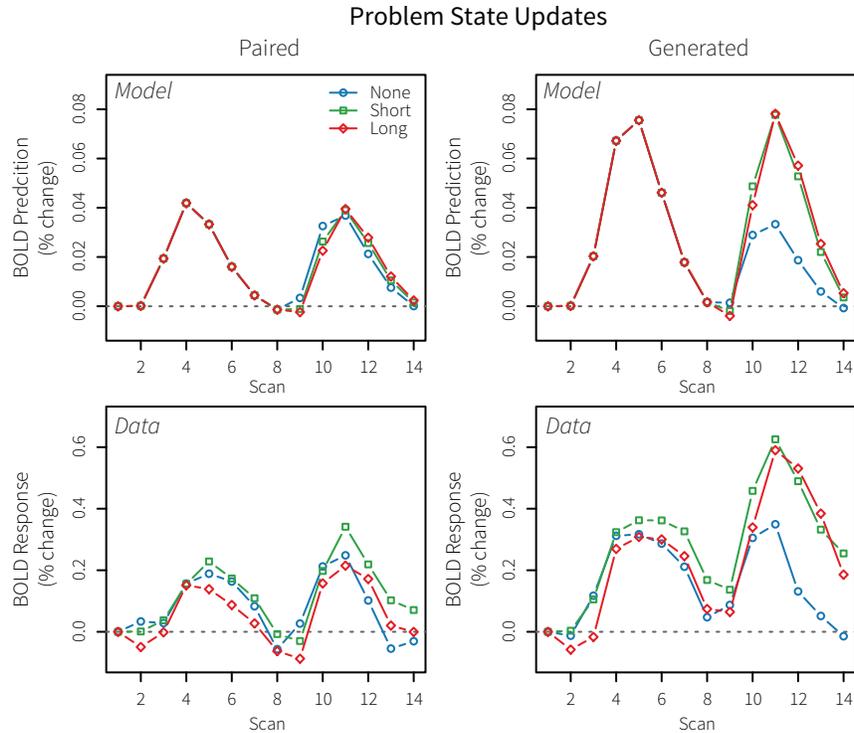


Figure 7. ROI results for problem state updates. Top panels show model predictions; bottom panels data. 1 scan = 2 seconds.

We can conclude that the prefrontal region indeed mostly reflects declarative memory retrievals, while the posterior parietal cortex reflects updating problem representations. The model's predictions matched the data reasonably well, although there were some discrepancies (e.g., for problem state updates the data shows a larger peak at test than at study in the generated condition, the model did not predict this). Such discrepancies can be due to several different reasons: the model might be inaccurate, the mapping of ACT-R on the brain might be incomplete, or – with respect to the noisy results for declarative memory at test – we might have to test more subjects. In addition, it is known that the shape of the BOLD response is different in different brain regions, as well as between different subjects. Here we presented a priori BOLD predictions based on SPM's HRF, but it would be reasonable to fit the shape and magnitude of the BOLD response separately for each region.

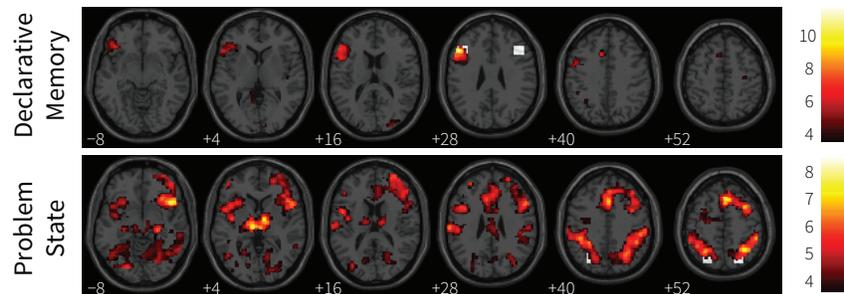


Figure 8. Model-based fMRI results. Statistical maps were thresholded at  $p < .001$  (uncorrected). White squares indicate predefined ACT-R regions.

### Model-Based fMRI Analysis

The ROI analysis has one clear disadvantage: it is dependent on the correctness of the predefined mapping. The current mapping was based on a reading of the literature on regional functions, and might therefore not be optimal. To find regions that map best on our module predictions we can use model-based fMRI [e.g., 17,18]. Whereas in conventional fMRI the experimental structure is typically used as the basis for the regressors in the general linear model (GLM, see Chapter 4), in model-based fMRI predictions stemming from a computational model are used. In the case of ACT-R this means that predictions such as the ones in Figure 5 are regressed against the BOLD response in all voxels in the brain. This shows which voxels correlate significantly with the predictions of a module, indicating that these voxels might implement the functionality of the module.

Using model-based fMRI with ACT-R involves generating model predictions for all trials for all subjects, because model activity is regressed against the BOLD response over the whole experiment. This has not typically been done in ROI analysis although it could be. Generating predictions for single trials requires representing any differences that might occur because of the specific stimuli on each trial, as these can lead to different behavior and different model predictions [see 19, for an example]. Second, because we are regressing the model predictions

directly against the brain data, it is important to have an exact time-mapping between data and model, to avoid, for instance, comparing a fixation in the data with a key-press in the model [17]. To this end, trial onset and key-presses were lined up between model and data. That is, the model predictions on each trial were subjected to a linear transformation to create a perfect response-time match to the data (i.e. all model activity was increased or decreased in length on each trial; see [19] for details). The resulting predictions were used for the model-based analysis.

Figure 8 shows the results of a model-based analysis for our example task (originally reported in [20]). The top panel shows that declarative memory updates were exclusively reflected by activity in a left prefrontal region, located directly on top of the predefined ACT-R region (indicated by white squares). The bottom panel shows that problem state activity was reflected by a large number of regions. This is not very surprising, given that many regions will be active in response to the task, and we search for correlating regions. A number of regions show a strong response: the largest and most significant regions included the inferior parietal lobule and the anterior cingulate. In addition, we see strong correlations in the thalamus (slice +4) and inferior frontal gyrus (-8). This illustrates a weakness of model-based fMRI: multiple regions might correlate with the predictions, yielding imprecise results. A meta-analysis combining these results with four additional studies indicated that the parietal and anterior cingulate activity was consistent over the five studies, while the other regions in Figure 8 are probably due to idiosyncrasies of the current task and model [20].

### **Concluding Remarks**

In this chapter we discussed how ACT-R can be used in combination with fMRI data. We described two different analyses: ROI analysis and model-based fMRI analysis. The remaining question is which analysis to use. This naturally depends on the situation: If one wants to test the predictions of a cognitive model it is more constraining to have pre-specified regions as in the ROI analysis. If one wants to understand how a cognitive function maps onto the brain, the

model-based approach allows one to see the full picture. With respect to model-based fMRI, the analysis is not limited to the current modules – any prediction from a model can be used (e.g., only numerical retrievals or representational activity in response to visual encoding). In effect, model-based fMRI yields a mapping that can be used for ROI analyses. However, the results of model-based fMRI are strongly dependent on the quality of the model and on how well the experiment dissociated different model processes. For this reason it either should be used over multiple different tasks as in [20] or in combination with conventional fMRI analyses.

### **Further Reading**

- ‘You can’t play 20 questions with nature and win’ [1] is Allen Newell’s commentary in which he argues for a cognitive architecture approach. Four decades later the paper is still a thought-provoking and entertaining must-read for every cognitive scientist. In a companion piece in the same volume he presents his initial production system approach to cognitive architectures [21].
- Chapter 1 of [6] gives a very clear introduction to cognitive architectures and ACT-R. In case you do not have the book available, [22] provides an introduction to ACT-R and its mapping on brain regions. For a more concise introduction to ACT-R’s mapping on brain regions, see [11].
- Gläscher and O’Doherty [17] give a general introduction to model-based fMRI analysis. Borst and Anderson [20] show how model-based fMRI can be applied in combination with an ACT-R model.

### References

1. Newell A. You can't play 20 questions with nature and win: projective comments on the papers of this symposium. In: Chase WG, editor. *Visual Information Processing*. New York: Academic Press; 1973. page 283–308.
2. Newell A. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press; 1990.
3. Kieras DE, Meyer DE. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*. 1997;12:391–438.
4. Van Maanen L, Van Rijn H, Borst JP. Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review*. 2009;16(6):987–99.
5. Salvucci DD. Modeling driver behavior in a cognitive architecture. *Human Factors*. 2006;48(2):362–80.
6. Anderson JR. *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press; 2007.
7. Just MA, Varma S. The organization of thinking: what functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*. 2007;7(3):153–91.
8. Meyer DE, Kieras DE. A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*. 1997;104(1):3–65.

9. Anderson JR, Byrne D, Fincham JM, Gunn P. Role of prefrontal and parietal cortices in associative learning. *Cereb Cortex*. 2008;18(4):904–14.
10. Borst JP, Taatgen NA, Van Rijn H. The Problem State: A Cognitive Bottleneck in Multitasking. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2010;36(2):363–82.
11. Anderson JR, Fincham JM, Qin Y, Stocco A. A central circuit of the mind. *Trends Cogn Sci*. 2008;12(4):136–43.
12. Cohen MS. Parametric analysis of fMRI data using linear systems methods. *Neuroimage*. 1997;6(2):93–103.
13. Friston KJ, Fletcher PC, Josephs O, Holmes A, Rugg MD, Turner R. Event-related fMRI: characterizing differential responses. *NeuroImage*. 1998;7(1):30–40.
14. Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD. *Statistical Parametric Mapping. The Analysis of Functional Brain Images*. Academic Press; 2007.
15. Borst JP, Taatgen NA, Stocco A, Van Rijn H. The Neural Correlates of Problem States: Testing fMRI Predictions of a Computational Model of Multitasking. *PLoS ONE*. 2010;5(9):e12966.
16. Sohn MH, Goode A, Stenger VA, Jung KJ, Carter CS, Anderson JR. An information-processing model of three cortical regions: evidence in episodic memory retrieval. *NeuroImage*. 2005;25(1):21–33.
17. Gläscher JP, O’Doherty JP. Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2010;1(4):501–10.

18. O'Doherty JP, Hampton A, Kim H. Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci.* 2007;1104:35–53.
19. Borst JP, Taatgen NA, Van Rijn H. Using a Symbolic Process Model as input for Model-Based fMRI Analysis: Locating the Neural Correlates of Problem State Replacements. *NeuroImage.* 2011;58(1):137–47.
20. Borst JP, Anderson JR. Using Model-Based functional MRI to locate Working Memory Updates and Declarative Memory Retrievals in the Fronto-Parietal Network. *PNAS USA.* 2013.
21. Newell A. Productions Systems: Models of Control Structures. In: Chase WG, editor. *Visual Information Processing.* New York: Academic Press; 1973. Page 527–46.
22. Anderson JR. Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science.* 2005;29:313–41.

### Exercises

1. Read Newell (1973a). Do you agree or disagree with Newell's diagnosis? Explain why.
2. Do you think cognitive architectures are a solution to Newell's problem?
3. Use the included Matlab code to generate model predictions for the manual and visual modules (like Figure 6 & 7). Do the predictions match your expectations?
4. Give three possible reasons for discrepancies between model predictions (such as in Figure 6 and 7) and fMRI data.
5. What do you think is more interesting: fitting the BOLD response of the model to the data by changing the parameters of the HRF, or only using a priori predictions? Explain why.
6. Discuss advantages of ROI analysis as compared to conventional exploratory fMRI analysis.
7. Discuss advantages of model-based fMRI analysis as compared to conventional exploratory fMRI analysis.
8. Discuss advantages of conventional fMRI analysis as compared to the methods described in this chapter.