NeuroImage 58 (2011) 137–147

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Using a symbolic process model as input for model-based fMRI analysis: Locating the neural correlates of problem state replacements

Jelmer P. Borst [a,*], Niels A. Taatgen [b], Hedderik van Rijn [c]

[a] Dept. of Artificial Intelligence, University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands
[b] Dept. of Artificial Intelligence, University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands
[c] Dept. of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 20 December 2010
Revised 19 April 2011
Accepted 23 May 2011
Available online 16 June 2011

Keywords:
Model-based fMRI
Cognitive model
Problem state

ABSTRACT

In this paper, a model-based analysis method for fMRI is used with a high-level symbolic process model. Participants performed a triple-task in which intermediate task information needs to be updated frequently. Previous work has shown that the associated resource – the problem state resource – acts as a bottleneck in multitasking. The model-based method was used to locate the neural correlates of 'problem state replacements'. To analyze the fMRI data, we fit the computational process model to the behavioral data and regressed the model's activity against the fMRI data. The brain region responsible for the temporary representation of problem states, the inferior parietal lobule, and the brain region responsible for long-term storage of problem states, the inferior frontal gyrus were thus identified. These results show that model-based fMRI analyses can be performed using high-level symbolic cognitive models, enabling fine-grained exploratory fMRI research.

Introduction

If one wants to find the neural correlates of a theorized cognitive process using the classical fMRI analysis method of *cognitive subtraction* (e.g., Cabeza and Nyberg, 2000; Logothetis, 2008), the first step is to translate the theory into suitable experimental conditions. Then, an experimental condition placing demands on the process of interest is compared to a control condition. The control condition is the same as the experimental condition except for the absence of the process under investigation. Brain areas that are more active in the experimental condition than in the control condition are assumed to be involved in the cognitive process of interest (e.g., Friston et al., 2007). However, it would be better to localize cognitive functions in a more direct way. Especially for more complex tasks, the translation of theory into experimental conditions is non-trivial. In complex tasks, central cognitive processes are often used in all experimental conditions (although with a different frequency or temporal pattern), which makes it difficult to find a good control condition that does not include the process of interest. A way to address this problem and to localize brain functions in a more direct way is to use model-based fMRI analysis (e.g., Gläscher and O'Doherty, 2010; O'Doherty et al., 2007).

In model-based fMRI analysis, information coming from a computational model that simulates the process of interest is correlated against fMRI data, showing which brain areas show activation patterns that are consistent with the process of interest. This method has proven to be very successful in locating brain areas involved in reinforcement learning (e.g., Daw et al., 2006; Hampton et al., 2006; Haruno and Kawato, 2006; Kim et al., 2006; Wunderlich et al., 2009). Parameters of mathematical reinforcement models were correlated against brain data, showing which brain areas are involved in the reinforcement learning process. In this paper we will use the model-based method with a higher-level symbolic cognitive model. Such a higher-level model not only simulates a particular process, but the whole task including, for example, visual and motor processes. Instead of correlating model *parameters*, we will correlate the presence and absence of activity of cognitive resources against brain data, showing where the cognitive resources are best represented in the brain. This way, we will investigate whether predictions derived from a high-level process model can be used for model-based fMRI, and whether this combination allows for more direct exploratory fMRI analyses.

The problem state resource

We will use model-based fMRI to analyze data of a relatively complex experimental paradigm, which was developed to investigate the neural correlates of the 'problem state resource' (Borst et al., 2010a). The problem state resource is defined as the part of working

Abbreviations: ACT-R, Adaptive Control of Thought-Rational; FWE, Family-Wise Error correction; GLM, General Linear Model; HRF, Hemodynamic Response Function; LME, Linear Mixed Effects model.

* Corresponding author.
E-mail addresses: jpborst@ai.rug.nl (J.P. Borst), niels@ai.rug.nl (N.A. Taatgen), hedderik@van-rijn.org (H. van Rijn).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.05.084

memory that is available at no time cost (Anderson, 2005), as opposed to other elements in working memory that take time to retrieve and use (e.g., McElree, 2001). It is normally used to represent intermediate information in a task, and can at most contain one chunk of information (Borst et al., 2010b). Thus, the concept of a problem state resource is comparable to the focus of attention in working memory theories that pose an extremely limited focus of attention (e.g., Garavan, 1998; McElree, 2001). The concept of a central problem state resource originates from a series of neuroimaging experiments by Anderson and colleagues, who found that activity in the posterior parietal cortex correlates with the number of transformations of mental representations (Anderson, 2005; Anderson et al., 2003, 2005; Sohn et al., 2005).

Although Anderson and colleagues assumed on functional grounds that the problem state resource contains at most one chunk of information, we have recently provided empirical evidence for this assumption. In a series of experiments, we showed that the problem state resource is a source of interference when required by multiple tasks at the same time (Borst et al., 2010b; Borst and Taatgen, 2007). A computational cognitive model was developed to account for the observed multitasking interference. The basic assumption of the model is that when multiple tasks require a problem state, the contents of the problem state resource have to be replaced on each switch between tasks. That is, on every alternation the problem state of the previous task is stored in declarative memory, while the problem state of the current task is recalled from declarative memory and restored to the problem state resource. The model incorporating these time-consuming and error-prone problem state replacements provided a good match with the interference effects in the data. The current experiment was performed to find the neural correlates of the resources that are used by the model, which are, apart from the problem state resource, associated with vision, manual action, and declarative memory.

Materials and methods

Behavioral experiment

To locate the neural correlates of the model's resources, we used a triple-task design in which participants alternated between solving subtraction problems and entering text, while performing a listening comprehension task simultaneously (Fig. 1 shows a screenshot of the experiment). Both the subtraction task and the text-entry task had two versions: an easy version that did not require maintenance of a problem state and a hard version that did.

In the subtraction task participants had to solve 10-column subtraction problems. Although participants were shown only one column at a time to minimize eye and head movements, participants were trained to perceive these columns as part of a full 10-column subtraction problem. In the easy version the upper term was always larger or equal to the lower term: no carrying was required. However, in the hard version participants had to carry in 6 out of the 10 columns; thus, participants had to remember whether a carry was in progress while performing the text-entry task.

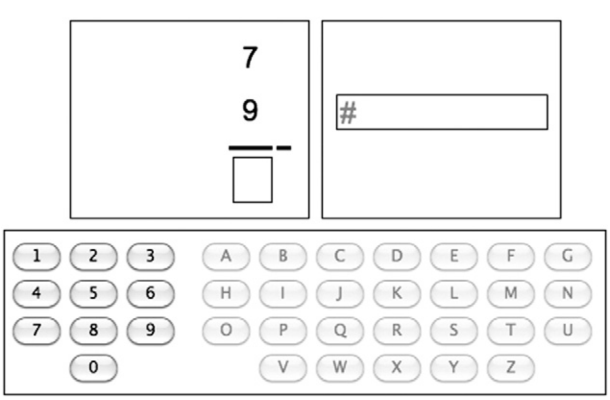

Fig. 1. Screenshot of the experiment.

In the text-entry task participants had to enter 10-letter strings. In the easy version a single letter was shown, which the participants had to enter. After solving one column of the subtraction problem, a new letter was shown, etc. In the hard version, a complete 10-letter word was shown once at the start of a trial, but as soon as the participant entered the first letter, the word disappeared and had to be entered letter by letter without feedback. Thus, in the hard version of the text-entry task participants had to remember what word they were entering.

Because participants had to alternate between the tasks after every number and letter, they had to keep track of whether a carry was in progress and what word they were entering (and the position within the word) in the hard versions of the tasks while giving a response on the other task. Supported by results from previous experiments (Borst et al., 2010b), we assumed that participants used their problem state resource to keep track of the absence or presence of a carry and the words, but did not use this resource in the easy conditions. In half of the trials participants also had to perform a listening comprehension task. As the listening task is not the focus of the current paper, we collapsed over this task if not mentioned otherwise. A detailed description of the Methods and discussion of the setup can be found in Borst et al. (2010a).

fMRI procedures and analysis

The fMRI data were collected with a Siemens 3T Allegra Scanner using a standard radio frequency head coil. Each functional volume existed of 34 axial slices (3.2 mm thickness, 64×64 matrix, 3.125×3.125 mm per voxel), acquired using echo-planar imaging (2000 ms TR, 30 ms TE, 79° flip angle, 200 mm field of view, 0 slice gap, with AC-PC on the 11th slice from the bottom). Functional acquisition was event-related; scanning onset was synchronized with stimulus onset. Anatomical images were acquired using a T1-weighted spin-echo pulse sequence at the same location as the functional images but with a finer resolution (3.2 mm thickness, 200 mm field of view, 256×256 matrix, 0.78125×0.78125 mm in-plane resolution).

The data were analyzed using SPM5 (Wellcome Trust Centre for Neuroimaging, London). This included realigning the functional images, coregistering them with the structural images, normalizing the images to the MNI (Montreal Neurological Institute) ICBM 152 template, and smoothing them with an 8 mm FWHM Gaussian kernel.

Participants

Thirteen students of Carnegie Mellon University participated in the experiment. The data of three participants were excluded (one participant fell asleep in the MRI scanner, one ignored the listening task, and with one fMRI recording problems were encountered) leaving 10 complete datasets (3 women, average age 21.9, range 19–28, right-handed). All participants had normal or corrected-to-normal vision and normal hearing. Informed consent as approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh was obtained before the experiment. Participants received US$ 100 compensation.

Results

Behavioral results

All reported *F*- and *p*-values are from repeated measure analyses of variance (ANOVAs), effects were judged significant when a .05 significance level was reached, and accuracy data were transformed

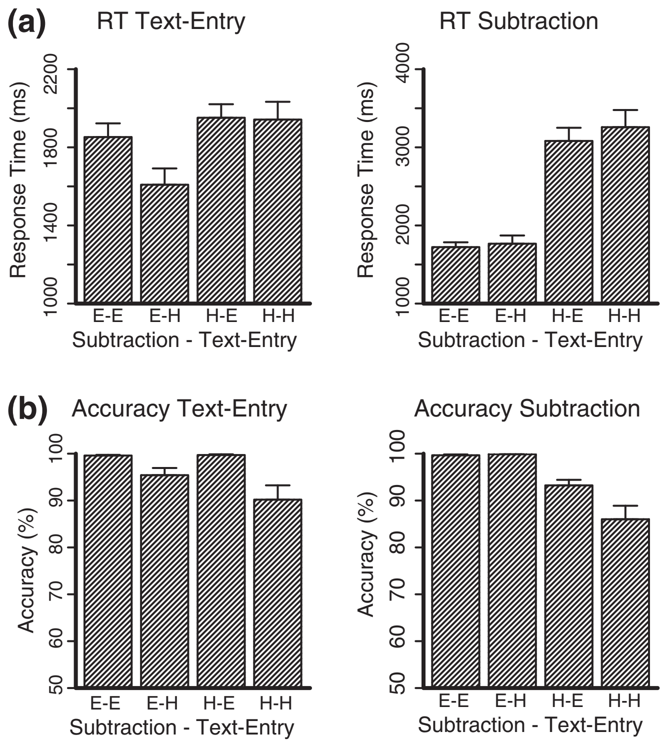

Fig. 2. Behavioral results. RT = response time, E/E = *easy-easy*, E/H = *easy-hard*, etc., error bars indicate standard error.

using an arcsine transformation before performing ANOVAs. Outliers in response times were eliminated by means of a two-step procedure. First, response times faster than 250 ms and slower than 10,000 ms were removed. Then, data exceeding 3 standard deviations from the mean per condition per participant were excluded. Overall, 2.4% of the data was discarded. First responses on both tasks were removed per trial.

As the listening task is not the focus of the current paper, we collapsed over levels of difficulty for this task if not mentioned otherwise.[1] Response time on the text-entry task was defined as the time between entering a digit in the subtraction task and entering a letter in the text-entry task, while a response time on the subtraction task was defined as the time between entering a letter in the text-entry task and entering a digit in the subtraction task. Fig. 2(a) shows the response times on the text-entry task (left) and the subtraction task (right); Table 1 contains the results of the ANOVAs. The interaction between Subtraction Difficulty and Text-Entry Difficulty was significant in the text-entry task, as were both main effects. Interestingly, response times decreased when the text-entry task was hard but increased when the subtraction task was hard. We have come across this effect before (e.g., Experiment 2 in Borst et al., 2010b). It can be explained by assuming that in the hard conditions participants know what word they are entering and thus do not need any additional visual input, but in the easy condition participants first have to look at the screen to see which letter they have to enter next. Our computational model (Borst et al., 2010b) fitted these results. In the subtraction task the interaction effect failed to reach significance, but both main effects did: a small increase of response times when text entry was hard, and a large increase when the subtraction task was hard.

Fig. 2(b) shows the accuracies, in which the interaction effects between Subtraction Difficulty and Text-Entry Difficulty reached significance for both tasks. The main effects of the tasks were also significant: Subtraction Difficulty for the subtraction task and Text-Entry Difficulty for the text entry task.

[1] Please note that the listening task has hardly any influence on the subtraction and text-entry results, see Borst et al. (2010a, 2010b).

Table 1
ANOVA results of the behavioral data. Interaction is the interaction between Subtraction Difficulty and Text-Entry Difficulty.

Source	Response times			Accuracy		
	F(1,9)	*p*	η_p^2	F(1,9)	*p*	η_p^2
Text-entry task						
Subtraction	32.13	<.001	.78	5.20	.049	.37
Text-entry	5.62	.042	.38	44.27	<.001	.83
Interaction	12.10	.007	.57	6.38	.03	.41
Subtraction task						
Subtraction	83.94	<.001	.90	96.33	<.001	.91
Text-entry	5.36	.046	.37	4.04	.075	.31
Interaction	2.66	.137	.23	21.16	.001	.70

Based on similar effects on response times and accuracy we previously argued that the results of this type of experiment support the idea of a problem state bottleneck (Borst et al., 2010b). In the easy tasks, no intermediate results need to be stored in the problem state resource. If a task is hard, accurate performance on a task requires storing intermediate results. Although participants had to alternate between tasks, the combination of one hard and one easy task is not problematic, since the problem state resource is not overwritten during the easy task. If, however, both tasks are hard, both tasks require the use of the problem state resource. Therefore, on each step in the task in the *hard-hard* condition the problem state resource has to be swapped out: on each step an old problem state is retrieved from declarative memory and restored to the problem state resource, overwriting the problem state of the other task. This results in the typically observed over-additive interaction effects.[2] To test whether this effect can indeed explain the observed data, we developed a computational cognitive model, which we will discuss in the next section.

Cognitive model

To account for the data, we used a high-level symbolic cognitive model (Borst et al., 2010b), developed in the cognitive architecture ACT-R (e.g., Anderson, 2007). Most important for the task at hand are the problem state resource and resources associated with vision, manual actions, and declarative memory. The model uses the visual resource to perceive the stimuli; this resource is assumed to do focused processing of attended stimuli. The manual resource was used to make responses; it operates the 'hands' of the model. The declarative memory resource was used to retrieve facts (e.g., '5 − 2 = 3'). Facts in ACT-R have a certain activation level, which represents frequency and recency of use (e.g., Anderson and Schooler, 1991). This activation level determines the probability of retrieving a fact, and the speed with which a fact is retrieved. For example, a simple subtraction fact such as '5 − 2' is probably used very often, and has therefore a high activation level. In contrast, '17 − 8' will have been used less often in the past, and will therefore have a lower activation level and take a little more time to retrieve. The problem state resource is used to maintain intermediate information and is therefore of particular interest for the current paper. Information in the problem state resource can be accessed at no time cost, but it takes 200 ms to replace it (e.g., Anderson, 2005). Problem states that are discarded from the problem state resource are still available in declarative memory, and can be retrieved and restored later.

[2] That the interaction for the response times of the subtraction task in the current experiment failed to reach significance is probably a power issue, as previously reported experiments showed significant effects (Borst et al., 2010b). Moreover, an extensive behavioral pilot experiment with the exact same experimental setup as the current experiment also showed both interactions. We report the results of this pilot experiment in the Appendix.

Using a cognitive architecture makes it meaningful to model all components relevant for a task such as visual, manual, and declarative memory processes (e.g., Cooper, 2007; Newell, 1990): the architecture provides the time it takes to move the mouse or retrieve a fact from memory, and as such the time courses of when the different resources are used. These elements of the architecture have received extensive experimental support (e.g., Anderson, 2007 and see http://act-r.psy.cmu.edu/).

To account for multitasking aspects the model uses threaded cognition theory (Salvucci et al., 2009; Salvucci and Taatgen, 2008, 2011). According to threaded cognition theory, multiple tasks can be active concurrently, but a cognitive resource can only be used by one task at a time. Thus, the problem state resource can only maintain information for a single task. However, in the *hard-hard* condition of the present experiment, a problem state is required for both tasks. The problem state then has to be replaced on each step of a trial (participants had to alternate between the subtraction and text-entry tasks), which takes time (a declarative retrieval and 200 ms problem state restoration, see Borst et al., 2010b for details). In contrast, in the *easy-easy* condition the problem state resource is not used at all, while in the *easy-hard* and *hard-easy* conditions it is only used for one task. Because problem states have to be restored and replaced at each step in the *hard-hard* condition, this leads to an over-additive interaction effect on response times. Furthermore, because the model sometimes retrieves an incorrect problem state from declarative memory, it also gives an explanation for the interaction effect on accuracy.

Previously, the model was fitted to data of Experiment 1 in Borst et al. (2010b, p. 370), and shown to give a good account of the data (R^2 for response times and accuracy approached 1). Subsequently, the model was used to predict the data of Experiments 2 and 3 in the same paper, showing that the model was capable of generalizing to different datasets. Experiment 3 in that paper also included the listening task; the model fit well to the data of all three tasks in that dataset. We use the exact same model in the current paper to analyze the fMRI data. Thus, the model takes the small influence of the listening task on the timing of the other tasks into account. The listening task itself, at least for the purposes of the current model, only requires use of the declarative memory resource, and does not use any of the other resources. However, we analyzed declarative memory only in the non-listening condition. An extensive description of the (lack of) influence of listening can be found in Borst et al. (2010b).

Model-based fMRI analysis — method

We will now turn to the model-based fMRI analysis data to locate the model's resources. For the classical fMRI analysis method of cognitive subtraction, one typically defines stimulus functions that correspond to the experimental conditions (e.g., Friston et al., 2007). These stimulus functions are entered into a general linear model (GLM), which shows brain areas in which activity correlates with the conditions of the experiment. Stimulus functions used for classical fMRI analyses are coarse in the sense that they assume stable activation during a complete trial — an assumption that often does not hold. During a trial in the current experiment, for example, participants solve a 10-column subtraction problem and enter a 10-letter word. These processes involve multiple fixations, manual actions, memory retrievals, and continuous maintenance and updating of problem states, and can thus not be characterized by constant activation throughout a trial. To construct more detailed stimulus functions, we fitted the cognitive model to the data, and entered the activity of the model's resources as stimulus functions into the general linear model. This method takes into account when and how often a resource is used during a trial, instead of assuming constant activation.

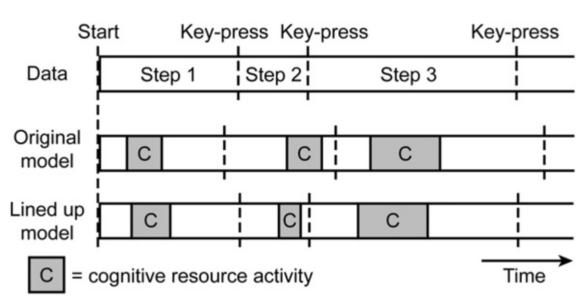

Fig. 3. Demonstration of the linear transformation that was used to line up the model data with the participants' data.

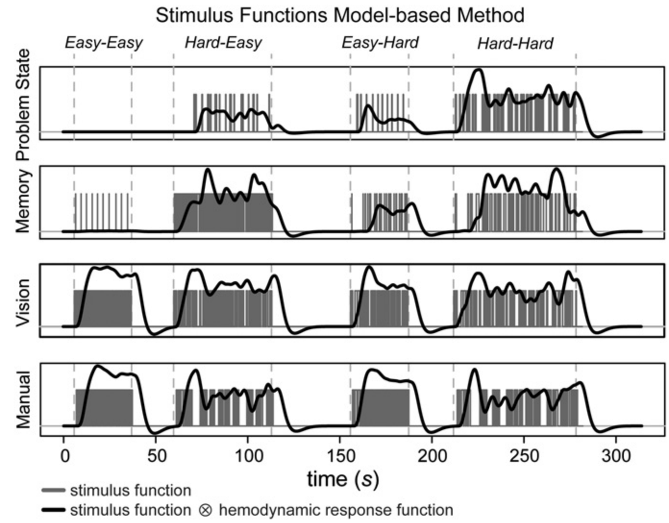

Fig. 4. Stimulus functions and convolved stimulus functions for the model-based analysis method, for four cognitive resources: the problem state resource, declarative memory, vision, and the manual motor resource. *Easy-Hard* etc. = *Easy Subtraction – Hard Text-Entry.*

To approximate the cognitive processes at trial level, we ran the model for each participant on the same stimuli as the participant received, in the same order, including all non-experimental components, such as fixation and feedback screens. To further improve the timing of the model, we lined up the model's responses with the participant's responses. Fig. 3 gives a schematic overview of the procedure. The first line represents data, with key-presses as dashed lines. The second line shows a model simulation of these steps. As the model is regressed directly against brain data it is important to have a correct time mapping between model and data (Gläscher and O'Doherty, 2010): it does not make sense to compare a fixation in the model to a key-press in the data. Therefore, we used a linear transformation to line up the key-presses of the model to the key-presses of the participants. Line 3 in Fig. 3 shows the result: the transformation causes Step 1 of the model to increase a little in length and Step 2 to decrease in length. Not only the key-presses of the model are shifted, but also cognitive resource activity within a step is shifted and in- or decreased in length (represented by the gray boxes in Fig. 3). The resulting activity for four cognitive resources during four different trials in the experiment is shown as gray lines in Fig. 4.

As the next step, the stimulus functions were convolved with a hemodynamic response function. The convolved stimulus functions are displayed in black in Fig. 4. For model-based fMRI analysis it is crucial that the different resources of the model make different predictions, because otherwise different resources cannot be distinguished. Fig. 4 shows clearly different patterns between the problem state resource and declarative memory[3] on the one hand, and the

[3] The fact that there is hardly any BOLD response predicted in the *easy–easy* condition of the declarative memory resource is caused by the very short (~10 ms) retrievals in that condition (only subtraction facts under 10 have to be retrieved, e.g., 5 − 2 = 3).

Table 2
Correlations between the different regions and predictions over all data points. Please note that correlations with declarative memory were calculated only on the non-listening data, because that was the data that was used for the model-based analysis.

Correlation between:	Prediction	Data
Problem state — declarative memory	.69	.65
Problem state — vision	.51	.23
Problem state — manual	.43	.44
Declarative memory — vision	.22	.35
Declarative memory — manual	.21	.62
Vision — manual	.93	.38

visual and manual resources on the other hand. It is important to note that problem state activity of the model relates to changing the contents of the problem state resource, not to passive maintenance of information. The problem state resource and declarative memory are used most in the *hard-hard* condition because of repeated problem states replacements, while they are used less in the easier conditions. The visual and manual resources show a different pattern: they are used for roughly the same amount of time in all conditions, but their activity is spread over a larger period of time in the harder conditions, resulting in a lower BOLD prediction (the same amount of visual information has to be processed and the same amount of key-presses have to be made, but as response times are higher more time is available in the more difficult conditions). The problem state resource and declarative memory can also be distinguished from each other: declarative memory is used often in the *hard subtraction – easy text-entry* condition (many subtraction facts have to be retrieved from memory to process the carries), resulting in high BOLD predictions. The problem state resource only shows intermediate BOLD levels in this condition (see Fig. 4). The visual and manual stimulus functions, on the other hand, are very similar. Table 2 shows the inter-correlations between the stimulus functions (the prediction column shows the correlations between the predictions). The correlation between the visual and the manual resource is .93, which makes it difficult to identify separate areas for these resources.

These stimulus functions were then entered one by one into a GLM to see which brain areas correlated with the predicted activity of the resources. Each stimulus function was accompanied by its 'opposite': a function showing when the resource was not active.[4] Feedback screens and the screens indicating conditions were also entered into the GLM; fixation screens formed the baseline. Contrast images were made for the individual participants, and entered into second level random-effect group analyses.

Model-based fMRI analysis — results

Fig. 5 shows the results for (a) the problem state resource, (b) declarative memory, (c) vision, and (d) the manual resource. The column on the left shows the regions that were identified by the model-based fMRI analysis. A threshold of $p<.01$ (FWE-corrected) and 100 contiguous voxels was applied to the results, with the exception of declarative memory, for which a threshold of .05 was used (FWE-corrected; see below for an explanation). Crosshairs in Fig. 5 are located at the most significant voxel, except for the manual resource and the visual resource (see below). The *xyz*-coordinates indicate the most significant voxel in MNI-coordinates in the located region. The white squares show the existing mapping between ACT-R's resources and brain regions (e.g., Anderson, 2007), which can be used for confirmatory analyses (e.g., Anderson et al., 2008; Borst et al., 2010a).

The area that corresponded best to problem state activity was located in the inferior parietal lobule, around the intraparietal sulcus. Declarative memory also showed activation in that area, but the best fitting area was located around the inferior frontal gyrus. The threshold for declarative memory was increased to $p<.05$.[5] This was necessary due to a more limited dataset, as we only used the trials in which the listening task was not present. If we included the trials with the listening task, the best fitting area for declarative memory coincided with the aural regions, because the model is not able to separate auditory processing of the incoming speech and the subsequent updating of declarative memory.

Fig. 5(c) and (d) show the results of the visual and manual resources. As discussed above, the stimulus functions were very similar as moving the mouse in the model is almost always accompanied by moving the eyes, and this was reinforced by the convolution with the HRF. It was therefore not surprising that the analyses yielded very similar results. The most significant area for both was the occipital visual area, but both also showed a fitting area in the motor cortex. This area was a little larger for the manual resource than for the visual resource, and as we know that manual actions are represented in the motor cortex, we centered the results of the manual resource on that area. The crosshairs of the visual area were moved down 13 mm, to enable comparison with ACT-R's predefined region (however, the coordinates indicate the most significant voxel).

The middle column of Fig. 5 shows the stimulus functions that were entered into the GLM, averaged per condition per trial. Thus, what is shown here is the activity of the model convolved with the BOLD response over the course of entering 10 numbers in the subtraction task and entering 10 letters in the text-entry task. The x-axis represents time in scans (1 scan = 2 s). The y-axis shows % BOLD change (the height of the curves is not important for the GLM, only the relative magnitude of the curves). The right column of Fig. 5 shows the measured BOLD response in the 100 most significant voxels for the located regions. ANOVA results of the area under the curve (reflecting total activation in a trial, see e.g., Anderson, 2005; Stocco and Anderson, 2008) are reported in Table 3. In general, the graphs show that the model-based fMRI method is able to identify patterns of activation in the brain that are very similar to the predictions of the model. Where for the problem state and declarative memory resources the *hard-hard* condition shows most activation, this is not the case for the manual and visual resources,[6] as predicted by the model. On the other hand, the fit is not perfect. For instance, while the model predicted no activity at all in the *easy-easy* condition for the problem state resource, this was not found in the located region. This indicates that the model-based analysis is not limited to identifying regions with a perfect fit, but locates regions that correlate significantly with the model predictions.

Discussion

The model-based method was able to find neural correlates corresponding to the cognitive resources of the model. Instead of using the experimental conditions as a basis for the analysis, the model-based method allows for assessing directly which parts of the brain correlate significantly with model predictions. In this paper we have shown that this is possible with a detailed high-level symbolic

[4] While the 'non-activity' of a resource is implicitly modeled by its stimulus function (1 = resource active, 0 = not active), we added a non-activity regressor to distinguish it from the fixation screens. Because non-activity often takes place in between activity of a resource, due to the convolution with the HRF there is usually some activity present on these scans, unlike on the fixation scans. To account for this, we added 'non-activity' as a separate regressor.

[5] If we decrease the p-values to .001 the same areas are found for all cognitive resources (except that the number of consecutive voxels had to be lowered to 25 for declarative memory).

[6] The graph for declarative memory ends earlier than the graphs of the other resources because only trials without the listening task are taken into account; these trials are shorter than the trials with the listening task.

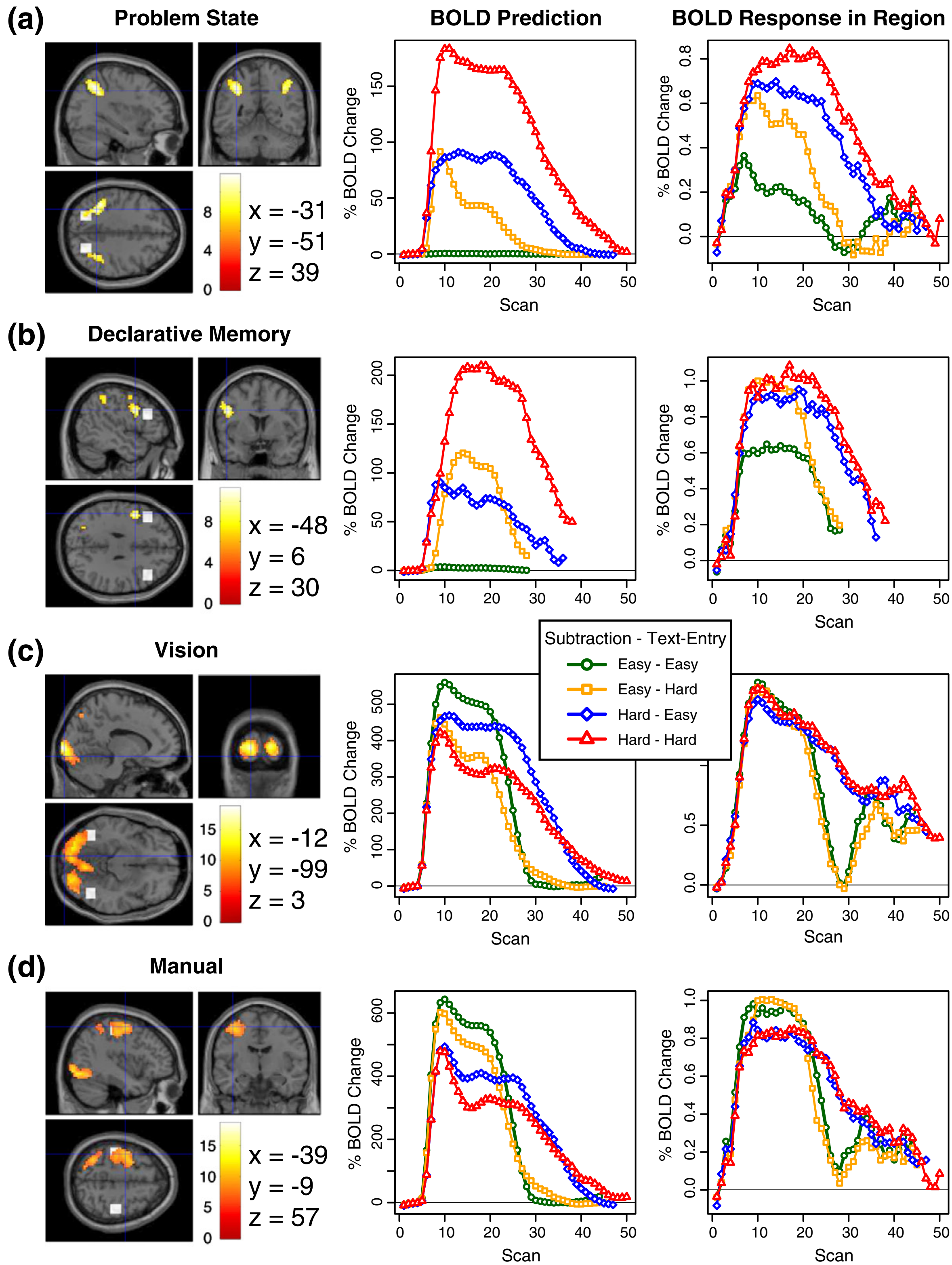

Fig. 5. Results of the model-based analysis method, for (a) the problem state resource, (b) declarative memory, (c) vision, and (d) the manual motor resource. (a), (c), and (d) with a Family Wise Error threshold of $p<.01$ and 100 contiguous voxels, (b) with an FWE threshold of $p<.05$ and 100 contiguous voxels. White squares represent predefined mappings between ACT-R's resources and the brain. Crosshairs are centered at the most significant voxel, except for the manual, which is centered on the most significant voxel in the cluster in the motor cortex and for the visual, which was moved down 13 mm to enable comparison to the predefined region of ACT-R. The middle column shows the stimulus function that was entered into the GLM, averaged per condition and trial, the right column the measured BOLD response in the located area.

cognitive model (as compared to the more low-level mathematical models that have been used previously; e.g., Daw et al., 2006; Gläscher and O'Doherty, 2010; Hampton et al., 2006; Haruno and Kawato, 2006; Kim et al., 2006; O'Doherty et al., 2007; Wunderlich et al., 2009). Instead of focusing on one process, using a high-level model integrated in a cognitive architecture allows for analyzing all cognitive processes that are involved in the task, and thus localizing multiple resources in one experiment. Furthermore, because the analysis is

Table 3
ANOVA results of the area under the curve in the located regions. Interaction is the interaction between Subtraction Difficulty and Text-Entry Difficulty.

Source	F(1,9)	p	η_p^2
Problem state			
Subtraction	121.24	<.001	.93
Text-entry	23.07	<.001	.72
Interaction	<1	–	–
Declarative memory			
Subtraction	52.25	<.001	.85
Text-entry	15.76	.003	.64
Interaction	<1	–	–
Vision			
Subtraction	43.51	<.001	.83
Text-entry	<1	–	–
Interaction	4.66	.059	.34
Manual			
Subtraction	6.81	.028	.43
Text-entry	<1	–	–
Interaction	2.06	.185	.19

based on a cognitive model, the fMRI results are grounded in the theoretical framework the model was built on, providing a functional explanation of the results (Gläscher and O'Doherty, 2010; O'Doherty et al., 2007).

The experiment was set up to test a hypothesis related to the problem state resource. Because activity in the model is related to *changes* of the problem state resource, the located region represents these changes, and not storage of problem states per se. The analysis showed that the model's problem state activity corresponded best to a region focused in the inferior parietal lobule, but also included parts of the superior parietal lobule and the intraparietal sulcus. The intraparietal sulcus has been linked previously to ACT-R's problem state resource (e.g., Anderson, 2005; Anderson et al., 2003, 2005; Sohn et al., 2005). While it is most often referred to in connection with spatial working memory, it is also known to be involved in object and verbal working memory (e.g., LaBar et al., 1999; Smith et al., 1998; Wager and Smith, 2003), functions that are attributed to the problem state resource.

Use of declarative memory also correlated with activity in the identified problem state region, but the best fitting area was the inferior frontal gyrus, slightly anterior to the standard ACT-R region for declarative memory (e.g., Anderson, 2007; Anderson et al., 2008). This region is known to be involved in memory retrieval (e.g., Cabeza et al., 2002; Fletcher and Henson, 2001; Wagner et al., 2001). More interestingly, if we lower the significance threshold, it becomes clear that both areas are part of a larger fronto-parietal network, a network that is often implicated in working memory research (e.g., Collette et al., 2006; Collette and Van der Linden, 2002). While both declarative memory and problem state activity are associated with functions of working memory, the current analysis implies that retrieving information is done via an area around the inferior frontal gyrus, while maintaining and updating working memory are performed in the parietal regions.

Model-based fMRI versus traditional analysis methods

We have shown that model-based fMRI makes it possible to directly locate the neural correlates of model resources, and therefore allows for fine-grained exploratory fMRI analyses. However, does it perform better than traditional methods? While there is no traditional method that allows for direct localization of model components, we used two existing methods to localize the over-additive interaction effect that was predicted by the model's problem state resource (see the BOLD prediction in Fig. 5(a), middle column), and compared the results of these methods to the results of the model-based method. First, we used a classical cognitive subtraction approach to find an interaction effect; when that failed we tried a parametric method that is somewhat similar to the model-based method.

For the classical cognitive subtraction analysis we defined a stimulus function for each condition in the experiment. Subsequently, the four stimulus functions were convolved with a hemodynamic response function and entered into a general linear model. We then tested for an over-additive interaction effect of Subtraction and Text-Entry Difficulty by contrasting the difference between the *hard-hard* condition and the *hard-easy* condition against the difference between the *easy-hard* and the *easy-easy* condition (i.e. (*hard-hard* – *hard-easy*) – (*easy-hard* – *easy-easy*); e.g., Friston et al., 2007). The results are shown in Fig. 6(a): no voxels crossed the FWE significance threshold of .05. Thus, the traditional cognitive subtraction method was unable to find a region that showed the predicted interaction effect, and could not be used to locate the neural correlates of the problem state resource.

We then used a method that is more similar to model-based fMRI: a parametric analysis (e.g., Büchel et al., 1996; Cohen, 1997). For this method we defined one stimulus function for all conditions in the experiment, and then specified a parametric model, with 0 for the *easy-easy* condition, 1 for the *hard-easy* and *easy-hard* conditions, and 3 for the *hard-hard* condition. This method is comparable to the model-based method, except that the amplitudes of the different conditions have to be specified by the researcher, instead of the model providing these estimates (note that the model also predicts a detailed pattern within each trial and differences between participants, see Section 4.2 below).

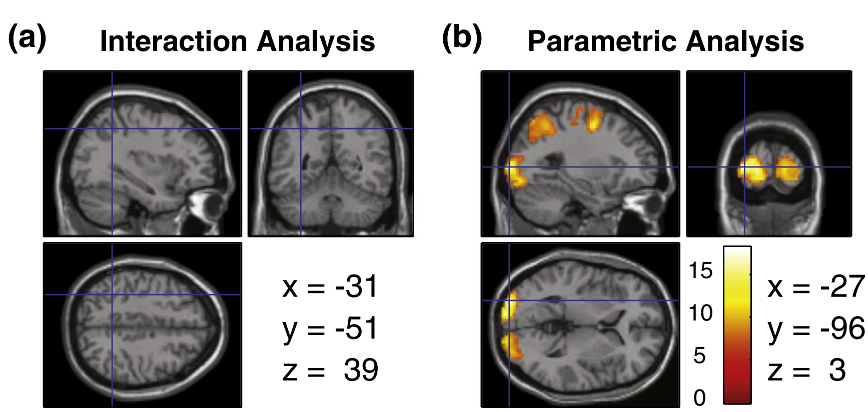

Fig. 6. Results of traditional fMRI analysis methods, for (a) a classical interaction analysis, and (b) a parametric modulation analysis (see the main text for details). The results were thresholded with (a) FWE $p<.05$ and (b) FWE $p<.01$ and 100 contiguous voxels.

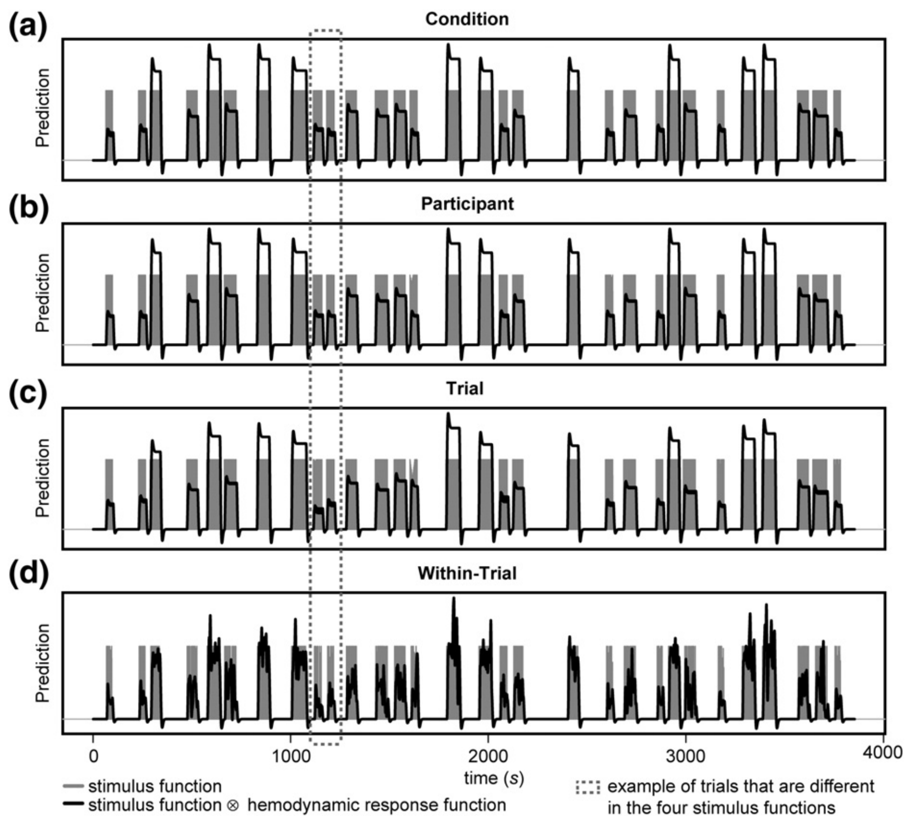

Fig. 7. Example of four different stimulus functions, shown for one session of one participant for the problem state resource. These stimulus functions were used to test which properties of the stimulus functions were important for the model-based analysis. Stimulus function (a) only contains effects of condition, (b) of condition and participant, (c) of condition, participant, and trial, and (d) of condition, participant, trial, and of cognitive resource usage within a trial. While it is clear that (d) is different from (a)–(c), the differences between (a)–(c) are minor.

The results are shown in Fig. 6(b): while the problem state region was found, visual and motor areas also showed significant activation, with the most significant voxel being located in the visual cortex. Thus, the parametric method yielded less specific results than the model-based method, and, moreover, indicated a seemingly incorrect region (compared to previous results, e.g., Anderson et al., 2008).

Model-based fMRI thus outperformed these two traditional methods. There are obviously other possibilities to analyze the data, but, to our knowledge, none of these options can directly show the neural correlates of resources in a model. Additionally, model-based fMRI allows for locating multiple resources in one experiment, without having to adapt the experimental design for each resource, which is the case for traditional methods. In the next section we performed a detailed analysis to investigate what enabled the model-based fMRI analysis to locate the model's resources.

What makes the model-based method powerful?

To arrive at the reported results, the model-based method uses a computational cognitive model to look in a more informed way at fMRI data. The model-based stimulus functions not only contain differences between conditions (as in classical fMRI analyses), but also differences per participant and trial, and even a detailed temporal pattern within each trial (Fig. 4). To assess what drives the results, we compared a series of models that incorporate increasing levels of detail. First, we constructed four different stimulus functions for each cognitive resource (Fig. 7). Please note that we used stimulus function 7(d) for the model-based analysis described above; (a)–(c) are only used for investigating which level of detail drives the results. The first stimulus function, 7(a), contains differences per condition, but was the same for all trials in a condition and all participants. The second stimulus function, 7(b), also contained differences per participant,[7] while the third and the fourth stimulus functions in addition contained effects of trial.[8] The fourth stimulus function differed from the third with respect to the temporal detail within a trial: the first three stimulus functions are smooth, while the fourth has a very detailed temporal structure[9] (i.e., this is the stimulus function that was used for the model-based analysis reported earlier in this paper). Fig. 7 illustrates the four different stimulus functions (note that here a stimulus function is shown for one resource, with different levels of detail; in contrast, Fig. 4 shows four stimulus functions for four different resources).

These stimulus functions were then entered one by one into linear-mixed-effects models (e.g., Baayen et al., 2008) with the stimulus function as a fixed effect and participant as a random effect. For each cognitive resource we constructed two LMEs, one to fit the stimulus function to the BOLD response in the best fitting voxel, and one to fit the average BOLD response in the 100 best fitting voxels. All stimulus functions correlated significantly with the data. The results, in the form of the log-likelihood of the fit, are listed in Table 4 (best fitting voxel) and Table 5 (average of the 100 best fitting voxels). As

[7] Effects of participant are caused by differences in speed between participants: some participants are slower in, for example, the *easy-easy* condition, resulting in a lower BOLD prediction (resource activity is spread more over the trial). However, all trials of the same condition of one participant have the same predicted BOLD amplitude.

[8] Effects of trial originate in the response times on a particular trial: for instance, for a quick trial with the same number of key-presses as a slow trial, a higher BOLD response is predicted for the manual motor resource.

[9] This temporal pattern stems from the time course of cognitive resource usage within a trial, as explained in the 'fMRI — model-based method' section.

Table 4
Log-likelihood of linear-mixed-effects models indicating which properties of the stimulus functions are important in the model-based analysis, for the best fitting voxel per cognitive resource. Each stimulus function is more detailed than the previous one, e.g., the stimulus function 'Participant' also includes effects of condition. See the main text for details.

Stimulus function	Problem state	Declarative memory	Vision	Manual
Condition	−36,743	−14,463	**−44,476**	−33,163
Participant	**−36,733**	−14,459	−44,480	−33,139
Trial	−36,771	**−14,444**	−44,492	**−33,138**
Within-trial	−37,029	−14,701	−44,848	−33,676

can be seen in these tables, for none of the resources was it helpful to include the temporal pattern within a trial. For declarative memory and the manual resource the model made correct predictions on a trial-by-trial level, for the problem state resource and vision on a condition level (for the 100 best fitting voxels; for the problem state resource on a participant level for the best fitting voxel). Based on this we can conclude that it is useful to include trial-by-trial differences in the model-based stimulus functions, but not the temporal pattern within a trial (either because the model predictions are not precise enough within a trial, or because the data is too noisy). Thus, the strength of the model-based analysis lies in the predicted amplitude levels per trial. By providing the analysis with these precise a priori estimates of the amplitudes of the BOLD response, the model allows for a better identification of the regions involved in the task than is possible with classical fMRI analysis methods.

ACT-R

While the model-based analysis can be used with different kinds of models, the current paper focuses on a model implemented in the ACT-R architecture. ACT-R only predicts when and for how long a cognitive resource is used, and not the intensity with which a resource is used. For instance, it is conceivable that an automatized movement takes as much time as a novel movement, but less effort. While this can be seen as a shortcoming, accounting for intensity would introduce extra free parameters, weakening the current predictive power of the model, as we would have to fit them post-hoc. On the other hand, if those new parameters were to explain a significant portion of the variance in the experimental data, they would increase the generalizability of the model (see e.g., Pitt and Myung, 2002).

If we compare the model-based fMRI method to the standard method of relating ACT-R models to neuroimaging data using predefined regions (Anderson, 2007; Anderson et al., 2008; see Borst et al., 2010a for a region-of-interest analysis of the current task), two things become clear: First, the areas that were located with the model-based analysis are very close to the predefined regions normally associated with ACT-R, which are shown as white squares in Fig. 5. Only the visual area is different: the located region overlaps with V1, while the predefined ACT-R region is located in the fusiform gyrus. Thus, it seems that lower level vision actually fits better with the ACT-R predictions than the slightly higher-level visual processing of the fusiform gyrus. Second, the strength of the model-based method lies in its exploratory nature. Using this method, we can not only validate cognitive models, but also determine which brain regions are involved in complex tasks.

Table 5
Log-likelihood of linear-mixed-effects models indicating which properties of the stimulus functions are important in the model-based analysis, for the average of the 100 best fitting voxels per cognitive resource. Each stimulus function is more detailed than the previous one, e.g., the stimulus function 'Participant' also includes effects of condition. See the main text for details.

Stimulus function	Problem state	Declarative memory	Vision	Manual
Condition	**−28,398**	−11,537	**−39,221**	−29,227
Participant	−28,402	−11,534	−39,231	−29,214
Trial	−28,426	**−11,516**	−39,257	−29,209
Within-trial	−28,823	−11,849	−39,736	−29,650

Conclusion

In this paper we have shown that the technique of model-based fMRI can be used in combination with a high-level symbolic process model. The model-based analysis method uses the results of a computational cognitive model to look in a more informed way at fMRI data: it shows areas in the brain that correlate with activity of the model. This method is especially useful for cognitive functions that are hard to discriminate in a pure subtraction-based design, for instance working memory storage and updating: These processes go hand-in-hand, which makes it difficult to find experimental conditions with one process but without the other. However, when a good model is available, these processes would yield different stimulus functions, which could in turn lead to different regions in the fMRI analysis, or at least different focal points in networks of activity. Because the model-based analysis works by refining the stimulus function, the method can be used with all modeling techniques that yield information that is more detailed than the condition structure of an experiment, which is used as the stimulus function in classical fMRI analyses.

Appendix A. Behavioral results outside the scanner

Here we report the results of the pilot experiment that we ran outside the fMRI scanner. This experiment has exactly the same setup as the experiment that is reported in the main text.

Participants

Twenty students of Carnegie Mellon University participated in the experiment (11 women, average age 20.6, range 18–23). All participants had normal or corrected-to-normal vision and normal hearing. Informed consent as approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh was obtained before the experiment. Participants received US$ 10 for performing the experiment.

Results

Outliers in reaction times were eliminated by means of a two-step procedure. First, response times faster than 250 ms and slower than 10,000 ms were removed. Then, data exceeding 3 standard deviations from the mean per condition per participant were excluded. Overall, 2.0% of the data was discarded. Accuracy data were transformed using an arcsine transformation before performing ANOVAs.

Fig. A1(a) shows the response times on the text-entry task (left) and the subtraction task (right); Table A1 contains the results of the ANOVAs. The interaction between Subtraction Difficulty and Text-Entry Difficulty was significant in both tasks. Furthermore, Subtraction Difficulty had a significant effect on the response times of the text entry task, while both Subtraction Difficulty and Text Entry Difficulty had a significant effect on the response times of the subtraction task.

Fig. A1(b) shows the accuracies, in which the interaction effect between Subtraction Difficulty and Text-Entry Difficulty reached significance for the text entry task, but not for the subtraction task. All main effects reached significance.

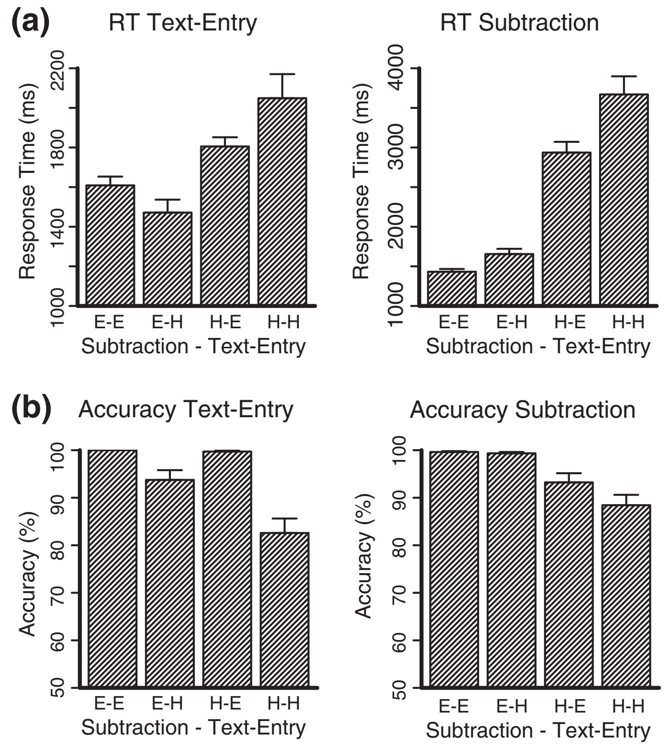

Fig. A1. Behavioral results in the pilot experiment. RT = response time, E/E = *easy-easy*, E/H = *easy-hard*, etc., error bars indicate standard error.

Table A1
ANOVA results of the pilot experiment. Interaction is the interaction between Subtraction Difficulty and Text-Entry Difficulty.

Source	Response times			Accuracy		
	F(1,19)	*p*	η_p^2	F(1,19)	*p*	η_p^2
Text-entry task						
Subtraction	69.47	<.001	.79	16.64	<.001	.47
Text-entry	<1	–	–	50.72	<.001	.73
Interaction	19.84	<.001	.51	17.92	<.001	.49
Subtraction task						
Subtraction	139.4	<.001	.88	62.98	<.001	.77
Text-entry	32.18	<.001	.63	4.29	.052	.18
Interaction	22.73	<.001	.54	2.85	.108	.13

References

Anderson, J.R., 2005. Human symbol manipulation within an integrated cognitive architecture. Cogn. Sci. 29, 313–341.
Anderson, J.R., 2007. How Can the Human Mind Occur in the Physical Universe? Oxford University Press, New York.
Anderson, J.R., Albert, M.V., Fincham, J.M., 2005. Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. J. Cogn. Neurosci. 17 (8), 1261–1274.
Anderson, J.R., Fincham, J.M., Qin, Y., Stocco, A., 2008. A central circuit of the mind. Trends Cogn. Sci. 12 (4), 136–143.
Anderson, J.R., Qin, Y., Sohn, M.H., Stenger, V.A., Carter, C.S., 2003. An information-processing model of the BOLD response in symbol manipulation tasks. PB&R 10, 241–261.
Anderson, J.R., Schooler, L., 1991. Reflections of the environment in memory. Psychol. Sci. 2 (6).
Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59 (4), 390–412.
Borst, J.P., Taatgen, N.A., 2007. The costs of multitasking in threaded cognition. Proceedings of the Eighth International Conference on Cognitive Modelling, pp. 133–138.
Borst, J.P., Taatgen, N.A., Stocco, A., Van Rijn, H., 2010a. The neural correlates of problem states: testing fMRI predictions of a computational model of multitasking. PLoS One 5 (9), e12966.
Borst, J.P., Taatgen, N.A., Van Rijn, H., 2010b. The problem state: a cognitive bottleneck in multitasking. J. Exp. Psychol. Learn. Mem. Cogn. 36 (2), 363–382.
Büchel, C., Wise, R.J., Mummery, C.J., Poline, J.B., Friston, K.J., 1996. Nonlinear regression in parametric activation studies. Neuroimage 4 (1), 60–66.
Cabeza, R., Dolcos, F., Graham, R., Nyberg, L., 2002. Similarities and differences in the neural correlates of episodic memory retrieval and working memory. Neuroimage 16 (2), 317–330.
Cabeza, R., Nyberg, L., 2000. Imaging cognition II: an empirical review of 275 PET and fMRI studies. J. Cogn. Neurosci. 12 (1), 1–47.
Cohen, M.S., 1997. Parametric analysis of fMRI data using linear systems methods. Neuroimage 6 (2), 93–103.
Collette, F., Hogge, M., Salmon, E., Van der Linden, M., 2006. Exploration of the neural substrates of executive functioning by functional neuroimaging. Neuroscience 139 (1), 209–221.
Collette, F., Van der Linden, M., 2002. Brain imaging of the central executive component of working memory. Neurosci. Biobehav. Rev. 26 (2), 105–125.
Cooper, R., 2007. The role of falsification in the development of cognitive architectures: insights from a Lakatosian analysis. Cogn. Sci. 31, 509–533.
Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., Dolan, R.J., 2006. Cortical substrates for exploratory decisions in humans. Nature 441 (7095), 876–879.
Fletcher, P.C., Henson, R.N., 2001. Frontal lobes and human memory: insights from functional neuroimaging. Brain 124 (5), 849–881.
Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.), 2007. Statistical Parametric Mapping. The Analysis of Functional Brain Images. Academic Press.
Garavan, H., 1998. Serial attention within working memory. Mem. Cognit. 26 (2), 263–276.
Gläscher, J.P., O'Doherty, J.P., 2010. Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. WIREs Cogn. Sci. 1 (4), 501–510.
Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J. Neurosci. 26 (32), 8360–8367.
Haruno, M., Kawato, M., 2006. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. Neural Netw. 19 (8), 1242–1254.
Kim, H., Shimojo, S., O'Doherty, J.P., 2006. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol. 4 (8), e233.
LaBar, K.S., Gitelman, D.R., Parrish, T.B., Mesulam, M., 1999. Neuroanatomic overlap of working memory and spatial attention networks: a functional MRI comparison within subjects. Neuroimage 10 (6), 695–704.
Logothetis, N., 2008. What we can do and what we cannot do with fMRI. Nature 453 (7197), 869–878.
McElree, B., 2001. Working memory and focal attention. J. Exp. Psychol. Learn. Mem. Cogn. 27 (3), 817–835.
Newell, A., 1990. Unified Theories of Cognition. Harvard University Press, Cambridge, MA.

O'Doherty, J.P., Hampton, A., Kim, H., 2007. Model-based fMRI and its application to reward learning and decision making. Ann. N. Y. Acad. Sci. 1104, 35–53.
Pitt, M., Myung, I., 2002. When a good fit can be bad. Trends Cogn. Sci. 6 (10), 421–425.
Salvucci, D.D., Taatgen, N.A., 2008. Threaded cognition: an integrated theory of concurrent multitasking. Psychol. Rev. 115 (1), 101–130.
Salvucci, D.D., Taatgen, N.A., 2011. The Multitasking Mind. Oxford University Press, New York.
Salvucci, D.D., Taatgen, N.A., Borst, J.P., 2009. Toward a unified theory of the multitasking continuum: from concurrent performance to task switching, interruption, and resumption. Human Factors in Computing Systems: CHI 2009 Conference Proceedings. ACM Press, New York, pp. 1819–1828.
Smith, E.E., Jonides, J., Marshuetz, C., Koeppe, R.A., 1998. Components of verbal working memory: evidence from neuroimaging. P. Natl. Acad. Sci. U. S. A. 95 (3), 876–882.
Sohn, M.H., Goode, A., Stenger, V.A., Jung, K.J., Carter, C.S., Anderson, J.R., 2005. An information-processing model of three cortical regions: evidence in episodic memory retrieval. Neuroimage 25 (1), 21–33.
Stocco, A., Anderson, J.R., 2008. Endogenous control and task representation: an fMRI study in algebraic problem-solving. J. Cogn. Neurosci. 20 (7), 1300–1314.
Wager, T.D., Smith, E.E., 2003. Neuroimaging studies of working memory: a meta-analysis. Cogn. Affect. Behav. Neurosci. 3 (4), 255–274.
Wagner, A.D., Maril, A., Bjork, R.A., Schacter, D.L., 2001. Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. Neuroimage 14 (6), 1337–1347.
Wunderlich, K., Rangel, A., O'doherty, J.P., 2009. Neural computations underlying action-based decision making in the human brain. P. Natl. Acad. Sci. U. S. A. 106 (40), 17199–17204.