Tracing Problem Solving in Real Time: fMRI Analysis of the Subject-paced Tower of Hanoi

John R. Anderson, Mark V. Albert, and Jon M. Fincham

Abstract

■ Previous research has found three brain regions for tracking components of the ACT-R cognitive architecture: a posterior parietal region that tracks changes in problem representation, a prefrontal region that tracks retrieval of task-relevant information, and a motor region that tracks the programming of manual responses. This prior research has used relatively simple tasks to incorporate a slow event-related procedure, allowing the blood oxygen level-dependent (BOLD) response to go back to baseline after each trial. The research described here attempts to extend these methods to tracking problem solving in a complex task, the Tower of Hanoi, which involves many complex steps of cognition and motor actions in rapid succession. By tracking the activation patterns in these regions, it is possible to predict with intermediate accuracy when participants are planning a future sequence of moves. The article describes a cognitive model in the ACT-R architecture that is capable of explaining both the latency data in move generation and the BOLD responses in these three regions. ■

INTRODUCTION

The research here is an attempt to use established fMRI techniques to reveal more about the real-time dynamics of human thought. The Tower of Hanoi (TOH) is a task involving a complex sequence of decisions that can be difficult to untangle with traditional behavioral methods. Also, standard exploratory approaches for rapid event-related fMRI cannot be applied, as the events are not independent. To avoid some of the limitations of these exploratory approaches, we use a confirmatory approach with brain regions that have proven important in more isolated tasks, and will show how they are involved in a more complex task.

In a series of studies, we have identified three left cortical regions that are intimately involved in various symbolic tasks. Some of these studies (Qin, Anderson, Silk, Stenger, & Carter, 2004; Anderson, Qin, Sohn, Stenger, & Carter, 2003) involved solving of algebraic equations, whereas others involved isomorphs of algebra that allowed better control of information-processing demands (Anderson, Qin, Stenger, & Carter, 2004; Qin, Sohn, et al., 2003) and study of the learning of such tasks in adult populations that already know standard algebra. Other studies have involved memory experiments where participants retrieved facts from a mental database of facts (Sohn, Goode, Stenger, Jung, et al., 2005; Sohn, Goode, Stenger, Carter, & Anderson, 2003). Across these experiments, we have consistently found the involvement of two regions that we interpret as reflecting abstract information processing. One is a left parietal region (see Figure 1) whose activity reflects changes to the problem representation and the second is a left prefrontal region whose activity reflects retrieval of stored information, such as arithmetic facts. These regions are similarly active when participants are solving standard algebraic equations involving actual numbers as well as performing isomorphic string manipulations. We developed an information-processing model in the ACT-R architecture (Anderson, Bothell, et al., 2004; Anderson & Lebiere, 1998) that was able to predict the blood oxygen level-dependent (BOLD) responses in these regions. One modality-specific area that we found was a motor region. It was necessarily involved because participants had to respond with their fingers but there was also some evidence (Anderson, Qin, et al., 2004) that it was involved in rehearsal activities in anticipation of giving the response. However, except in the case of implicit rehearsal, the magnitude of its activity did not vary with the factors that controlled the cognitive complexity of the task. These factors did, however, affect the activation in the parietal and prefrontal regions.

Our identification of these regions is consistent with research in other laboratories. The relevant motor region is in the left hemisphere, consistent with participants responding with their right hands. Others (Dehaene, Piazza, Pinel, & Cohen, 2003; Reichle, Carpenter, & Just, 2000) have found the left parietal region to reflect imagery and a number of researchers (Cabeza, Dolcos, Graham, & Nyberg, 2002; Donaldson, Peterson,

Carnegie Mellon University, Pittsburgh, PA

Figure 1. Representation of the three regions of interest that we have related to modules in the ACT-R architecture. Left Prefrontal: BA 45/46 (centered on $x = -40, y = 21, z = 21$); Left Motor: BA 3/4 (centered on $x = -37, y = -25, z = 47$); Left Parietal: BA 39/40 (centered on $x = -23, y = -64, z = 34$).
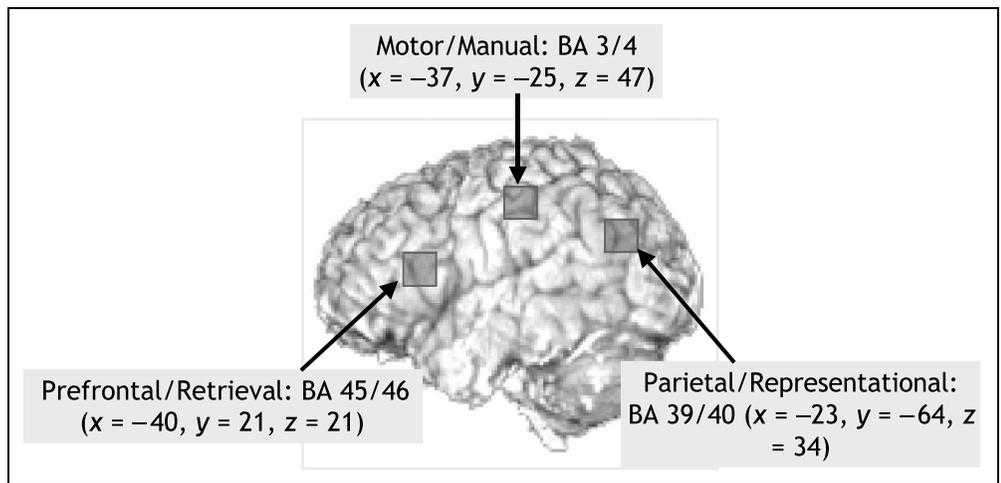
Ollinger, & Buckner, 2001; Fletcher & Henson, 2001; Wagner, Maril, Bjork, & Schacter, 2001; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001; Lepage, Ghaffar, Nyberg, & Tulving, 2000; Buckner, Kelley, & Peterson, 1999) have found a strong memory response in this prefrontal region.

Our research has involved developing information-processing models for the cognitive tasks, fitting these to the latency data, and then using these models to predict the exact form of the BOLD response found in each of these regions. Later in the article we will describe the details of this methodology. For now, we note that all of our previous experiments have used a slow event-related methodology in which we randomly presented separate trials. However, this approach cannot be taken in a more realistic setting, as when we are looking at complex problem solving involving a number of interdependent steps of cognition that occur at a pace determined by the problem solver. One of the goals of this research is to extend this methodology to complex problem solving where the sequences of actions are neither spaced nor independent. The TOH was chosen as a particularly well-understood exemplar of such problem solving.

The Tower of Hanoi Task and the Grid of Pittsburgh

Since its introduction as a task to study planning from the information-processing perspective by Simon (1975), the TOH has been a prototype task in the study of high-level cognition and problem-solving behavior (Altmann & Trafton, 2002; Anderson & Douglass, 2001; Kotovsky, Hayes, & Simon, 1985; Simon & Hayes, 1976). The TOH and the somewhat similar Tower of London (TOL) have been widely used in studies of patient populations as well (Owen, Downs, Shahakian, Polkly, & Robbins, 1990; Shallice, 1982), whereas most recent

neuroimaging studies have used the TOL (Newman, Carpenter, Varma, & Just, 2003; Dagher, Owen, Boecker, & Brooks, 1999; Baker, Rogers, Owen, 1996; Owen, Doyon, Petrides, & Evans, 1996).

Fincham, Carter, van Veen, Stenger, and Anderson (2002) conducted a functional magnetic resonance imaging (fMRI) study on the TOH. To track what was involved on each step, they slowed the pace of problem solving to one move every 16 sec, including an 8-sec distracting task (unconstrained, participants make about one move every 2 sec). The study found that a number of regions, including the prefrontal and parietal regions, responded in proportion to how much planning preceded a move. The current studies will use this methodology but without the artificial delays inserted into the problem solving. Thus, this experiment is similar to fast event-related imaging (Dale & Buckner, 1997), where the hemodynamic functions from one move are superimposed upon another move. However, unlike the typical rapid event-related experiment, these steps are not independent trials. Interpreting the BOLD signal requires that we use a cognitive model that will identify the cognitive demands at various points in the performance of the task. We will adapt the model for TOH developed by Anderson and Douglass (2001) to account for the behavioral data.

The TOH problem and an isomorph we call the Grid of Pittsburgh (GOP) are illustrated in Figure 2. In the TOH there are three pegs and disks of differing size. The goal is to change the start configuration to the goal configuration. Figure 2A illustrates such a problem. The constraints on solving the problem are:

1. Only one disk can be moved at a time.
2. Only the top disk on any peg can be moved.
3. Only a smaller disk can be placed on a larger disk.

As the example in Figure 2 illustrates, one need not just have problems that involve moving a tower of disks
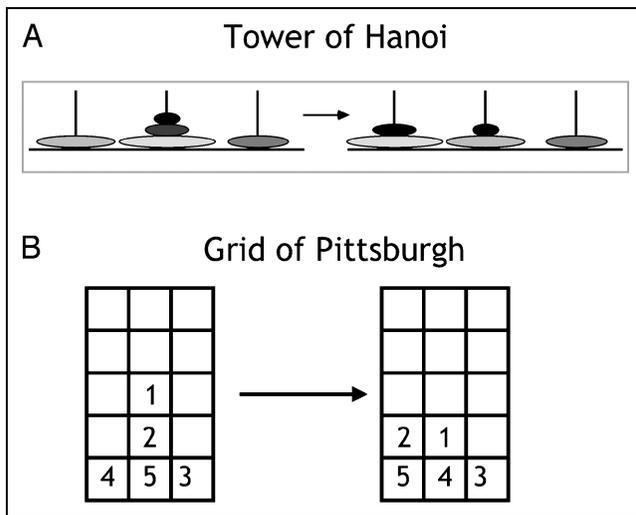
Figure 2. (A) A TOH problem and (B) a GOP problem.

to another peg. Rather, one can create problems that involve transforming arbitrary configurations into other arbitrary configurations. This yields a very large space of problems. Interestingly, the problem in Figure 2A is actually identical in terms of subgoal structure (explained later) to the traditional problem of moving a five-disk tower on one peg to a five-disk tower on another peg.

The GOP has the same logical structure where columns represent pegs and disk sizes are indicated by digits. The GOP was developed by Fincham et al. (2002) for imaging purposes because it minimizes eye movements. Participants in Experiment 1 memorized the goal configuration and only look at the current state, which can be kept within the fovea. Eye movements can cause scanner artifacts in protocols that scan through the frontal eye fields. Participants are initially trained on the TOH and then transferred to using the GOP.

In both Fincham et al. (2002) and Anderson and Douglass (2001), we taught participants to use a variant of what Simon (1975) called the sophisticated perceptual strategy. The strategy can be specified as follows:

1. Find the largest disk not in its goal position and make the goal to get it in that position. This is the initial "goal move" for purposes of the next two steps. If all disks are in their goal positions, the problem is solved.

2. If there are any disks blocking the goal move, find the largest blocking disk (either on top of the disk to be moved or at the destination peg) and make the new goal to move this blocking disk to the other peg (i.e., the peg that is neither the source nor destination of the goal move disk). The previous goal move is stored as the parent goal of the new goal move. Repeat this step with the new goal move.

3. If there are no disks blocking the goal move perform the goal move and

(a) If the goal move had a parent goal retrieve that parent goal, make it the goal move, and go back to step 2.

(b) If the goal had no parent goal, go back to step 1.

This strategy is guaranteed to solve all TOH problems; however, this strategy will not always find the shortest solution in the case where one is moving between two nontower configurations (Hinz, 1992). Should one forget subgoals at any point, one can simply recalculate the goal structure. However, as documented by Anderson and Douglass (2001), participants do retain subgoal information from one move to another, largely avoiding the need to recalculate.

As we will describe, participants practiced a great number of TOH problems before getting into the fMRI scanner. We discovered that by this point in time they had memorized solution sequences for moving towers of two and three disks from one peg to another. This enabled participants to plan sequences of seven moves (for moving towers of three) and three moves (for moving towers of two) at single points. Use of such macro operators facilitated analysis by helping to separate specific event contributions to the total BOLD signal. Below we describe how the sophisticated perceptual strategy combined with the learned tower sequences yielded a solution to this complex problem.

EXPERIMENT 1

Figure 3 illustrates a typical problem from Experiment 1, and some critical states (with three towers) along the way to a solution. Each of the problems we will analyze had this same abstract structure that involved 28 moves with isomorphic key states at the same points. Participants were kept on these 28-move sequences; should they make a move that deviated the problem state transitioned to the next move in the sequence in any case. The different 28-move problems they solved dif-
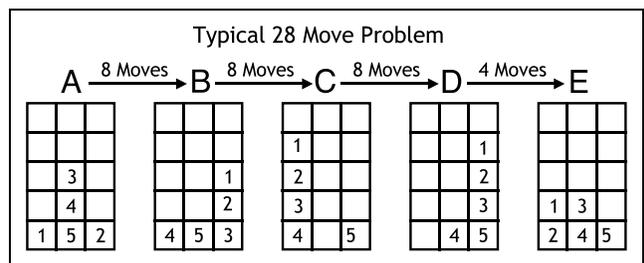
Figure 3. A typical problem from Experiment 1. The letters (A)–(E) denote distinct states of the problem.

fered from one another in terms of the start state. Nonetheless, for all such problems, the first eight moves (from A to B in Figure 3) involved going from an arbitrary state to a state where a three-tower has been created and Disk 4 no longer blocked the move of the Disk 5. The next eight moves (B to C) removed the tower of three disks that now blocked Disk 5 and moved Disk 5. The next eight moves (C to D) removed the tower of three disks that blocked the move of Disk 4 to its destination and moved Disk 4. Finally, there are four final moves (D to E) to get to the ultimate goal. Table 1 contains an analysis of the cognitive demands associated with each of these 28 moves according to the ACT-R model for the task. Encoding and updating the state of the problem, checking for blocks, and creating goals are all demands placed on the representational system that we associate with the posterior parietal region. Both storing (part of creating goals in Table 1) and retrieving subgoals are memory operations that we associate with the left prefrontal regions. Finally, each move will involve a demand on the motor region for the programming of the specific finger presses.

Table 1. Analysis of the Cognitive Demands Associated with each Move in Figure 3

Move 1 (6 encodings, 4 goals created). The sophisticated planning strategy directs one to focus on getting Disk 5 to the right peg, but this is blocked by Disk 4 and so a subgoal is created to move it out of the way to the left peg, but this is blocked by Disk 3 and so a subgoal is created to move this to the right peg, but this is blocked by Disk 2 and so a subgoal is set to move this to the left peg, but this is blocked by Disk 1 and so a subgoal is created to move this to the middle peg. This move can be made. Note in achieving getting to this move, four goals had to be encoded and committed to memory, and six disk locations had to be encoded (all five from the current state and the location of Disk 5 in the goal state).

Move 2 (1 updating, 1 check). The new position of Disk 1 needs to be updated and Disk 2 moved to its destination on the left peg after a check is made to determine it can be made.

Move 3 (1 updating, 1 retrieval, 1 check). The position of Disk 2 is updated. Retrieve the goal of moving Disk 3 to the right peg. A check determines that it is blocked by Disk 1 and Disk 1 is moved out of the way.

Move 4 (1 updating, 1 check). The position of Disk 1 is updated and a check determines that Disk 3 can be moved.

Move 5 (1 updating, 1 retrieval, 1 check, 1 goal created). The position of Disk 3 is updated and the goal to move Disk 4 is retrieved. A check determines that it is blocked by a two-tower and subgoal is created to move that tower. Disk 1 is moved as the first move in the plan.

Moves 6 and 7 (1 updating each). The remaining two moves of the two-tower plan are executed. Each begins with an updating of the result of the prior move.

Move 8 (1 updating, 1 retrieval). The position of Disk 1 is updated and the goal to move Disk 4 is retrieved. A check determines that it can be moved. We arrive at the state illustrated in B of Figure 3.

Move 9 (1 updating, 1 encoding, 1 retrieval, 1 check, 1 goal created). The position of Disk 4 is updated and the goal of moving Disk 5 is retrieved. A check determines it is blocked by a three-tower and a subgoal is created to move the three-tower to the left peg. The five goals is re-encoded and Disk 1 is moved as the first move in the plan.

Moves 10–15 (1 updating each). The remaining six moves of the three-tower plan are executed. Each begins with an update of the result of the prior move.

Move 16 (1 updating, 1 retrieval). The position of Disk 1 is updated and the goal to move Disk 5 is retrieved. It is moved and we arrive at the state illustrated in C of Figure 3.

Move 17 (1 updating, 1 encoding, 1 check, 2 goals created). The position of Disk 5 is updated and there are no more goals. A comparison between the current and goal configuration determines that Disk 4 is out of place and a subgoal is created to move that to the middle peg. A check determines that it is blocked by a three-tower and a subgoal is created to move the three-tower to the right peg. Disk 1 is moved as the first move in the plan.

Moves 18–23 (1 updating each). The remaining six moves of the three-tower plan are executed. Each begins with an updating of the result of the prior move.

Move 24 (1 updating, 1 retrieval). The position of Disk 1 is updated and the goal to move Disk 4 is retrieved. It is moved and we arrive at the state illustrated in D of Figure 3.

Move 25 (2 encodings, 1 check, 1 goal created). The last move is encoded and there are no more goals. A comparison between the current and goal configuration determines that Disk 3 is out of place and a subgoal is created to move the two-tower followed by the move of Disk 3.

Moves 26–28 (1 updating each). The remaining three moves of the plan are executed. Each begins with the encoding of the prior move. We arrive at the state illustrated in E of Figure 3.

Participants indicated their moves by issuing a two-finger press sequence. The first finger indicated the column from which the disk was to be taken and the second indicated the column to which the disk was supposed to be moved. Each brain scan was 1.5 sec and after a participant made a move the problem state was updated at the beginning of the next scan; therefore, participants could not average more than one move per 1.5 sec.

Results

Figure 4 displays the mean latency for each of the 28 moves measured as time from the state change to the completion of the second finger press indicating the move. The average error rate was 6.1% and its correlation with latency was $r = .935$. Figure 4 also displays the predictions of the theory described in the introduction to this experiment. These predictions depended on the setting of a number of parameters: As in past experiments (e.g., Anderson, Qin, et al., 2004; Sohn, Goode, Stenger, Carter, et al., 2003), we have assumed 0.2 sec per representational operation. We assumed 0.5 sec for the finger presses per move (there are two finger presses and the ACT-R model, based on Meyer & Kieras, 1997 Epic, assigns 0.3 sec to program the first finger press and 0.2 sec for the second finger press). As in past efforts, we then estimated the time of the memorial operations (retrievals, encodings) that would give the best fit to the data; this results in an estimate of .22 sec for the memorial operations and the fit to the data displayed in Figure 4. Generally, the fit is quite good except for the first point, which is greatly underpredicted. Anderson and Douglass (2001) similarly found a first move time much longer than would be expected.
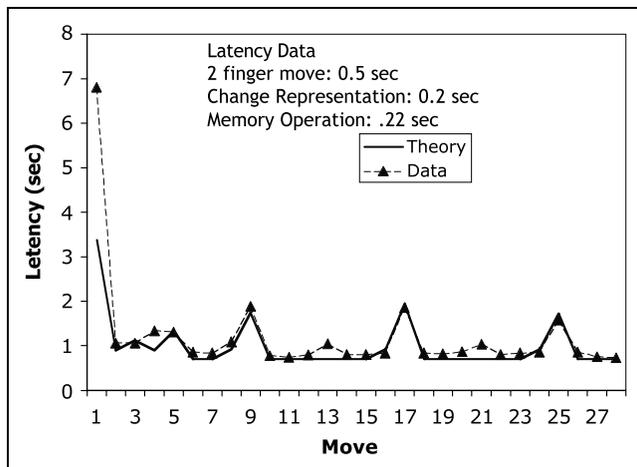
Figure 4. Experiment 1: Mean observed and predicted times to make a move as a function of the position of the move in the problem sequence.

We assume it reflects the time to orient to the task. Anderson and Douglass simply estimated an extra parameter to reflect this orientation. We did not do the same here because we are only interested in the representational, memorial, and manual times for fitting the BOLD response. It is also worth noting that there are slight rises in the latencies at Moves 13 and 21, which are at the points in the seven-moves sequences where a two-tower is moved. This suggests that participants had not altogether memorized the seven-moves sequences and were sometimes planning the last three moves for the two-tower. A similar discrepancy in latencies will appear in Experiments 2 and 3. However, it was so small that we did not think it worth complicating the model to accommodate it. The Altmann and Trafton (2002) and Anderson and Douglass (2001) models actually predict substantial spikes here but this was for data from less practiced subjects who presumably had not collapsed the seven moves into single operators.

Figure 5 illustrates the BOLD response obtained in the three regions of interest over the 28 moves of the problem. It plots the BOLD response on the first scan associated with the new problem state that is presented as the prompt for the next move. Thus, for instance, there is a point associated with the seventh move for the prefrontal region with an x-coordinate of 19.99 sec and a y-coordinate of 0.63%. The x value reflects the average midpoint in time of the first scan associated with the seventh move and the y value reflects the mean percent above baseline in this region on that scan. The BOLD response is measured as percent difference from the baseline defined by the average of the three scans before the presentation of the problem. The functions appear to have risen from zero by the beginning of the second move and then display some degree of variability over the course of the 28 moves. To measure the reliability of these effects, we performed two statistical tests—first, a t test of whether the mean response from Scans 2 to 28 was above zero and second, an F test of whether the variation on the scans associated with Moves 2 through 28 was significant. The effects for the motor region are highly significant [$t(7) = 4.74$, $p < .005$; $F(26,182) = 6.02$; $p < .0001$] as are the effects for the parietal region [$t(7) = 6.29$, $p < .0005$; $F(26,182) = 2.06$; $p < .005$]. However, the apparent effects associated with the prefrontal region are not significant [$t(7) = 0.86$; $F(26,182) = 1.08$].

There are two features to be stressed about (Figure 5). First, the parietal and motor responses are distinguished from the prefrontal by their consistently high BOLD responses. In contrast, the prefrontal response tends to go back down to baseline when there are the periods of no memory demands associated with the seven-moves sequences. Second, during activation, the rises and falls of the parietal and motor BOLD responses are mirror opposites of each other. To get a measure of their opposite tendencies, we needed to ignore the first
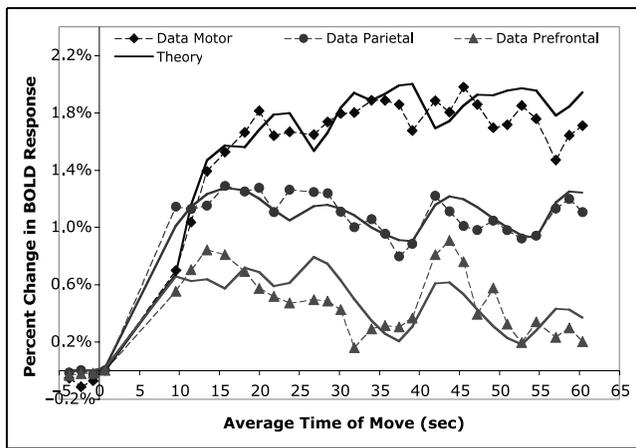
Figure 5. Experiment 1: Mean observed and predicted BOLD responses for the regions in Figure 1 as a function of mean time of move in the problem solution. The BOLD responses are calculated as percent change from the baseline established by the premove scans.

points where both functions are rising from zero, and therefore, we focused on the last 24 moves at which point the parietal and frontal have reached their first peak. The correlation between the parietal and motor BOLD responses over the last 24 moves is $r = -.44$ $[t(22) = 2.30; p < .05,$ two-tailed]. During the periods when the seven-move plans for three towers are being executed, there is relatively little representational activity but the manual actions occur at a rapid pace. On the other hand, during these same last 24 moves, the parietal and prefrontal are positively correlated as $r = .49$ $[t(22) = 2.64; p < .05,$ two-tailed], reflecting the fact that times of representational change tend also to be times of memorial operation.

Figure 5 also illustrates our efforts to predict the BOLD functions in these regions. These predictions are based on the proportion of time that the manual, representational, and memorial modules are engaged. Figure 6 is an attempt to illustrate how this proportion of engagement varies over the course of the experiment. The scale at the bottom is the average point in time at which the move occurred. We have divided the duration of operations by the duration of that move. Thus, the first move took an average of 8.85 sec (this is actually the time from when the problem was first presented until it was updated to reflect the first move). There are 10 representational operations (6 encodings and 4 goals) taking 0.2 sec each and 4 memory operations (the 4 goals) taking 0.22 sec each. Therefore, the proportion of engagement for the representational module is $10 \times 0.2/8.85 = 0.23$ and the memorial proportion is $4 \times 0.22/8.85 = 0.10$. We assume that the motor density is restricted to the last 2.5 sec of this move and so it is $0.5/2.5 = 0.20$.[1] The other values are similarly calculated.

Functions giving proportion of engagement, $P(t)$, like those in Figure 6, can be convolved with hemodynamic

functions, $H(t)$ to come up with predicted BOLD functions, $B(t)$, for the regions of interest:

$$B(t) = \int_0^t P(x) \times H(t - x)dx$$

The standard hemodynamic function used (e.g., Glover, 1999; Cohen, 1997; Dale & Buckner, 1997; Boyton, Engle, Glover, & Heeger, 1996) is a gamma:

$$H(t) = m \times (t/s)^a \times e^{-(t/s)}$$

where m is the magnitude of the response and s is a time scale. The function peaks at time $a \times s$. The parameter a determines the shape of the function such that the larger a is the more narrowly the function will be distributed around its peak.

Figure 5 displays our predicted and observed BOLD functions for the three regions of interest. We have plotted the response for the three scans before problem presentation (which defines baseline of 0%) and for the first scan of each move. The parameters and measures of fit are given in Table 2. The exponent a was set at 1 for all three regions to make the BOLD functions for the different regions comparable and because we did not have the temporal resolution needed to estimate the exact shape of the BOLD function. The overall quality of fit can be measured as a χ^2 distribution with degrees of freedom (75). This 75 is the sum of 25 degrees of freedom for each of the three regions (28 observations for the 28 moves minus 3 parameters). The total χ^2 for Experiment 1 is 87.84, which is not significant. So we can conclude that the model does account for the systematic variance in the data.
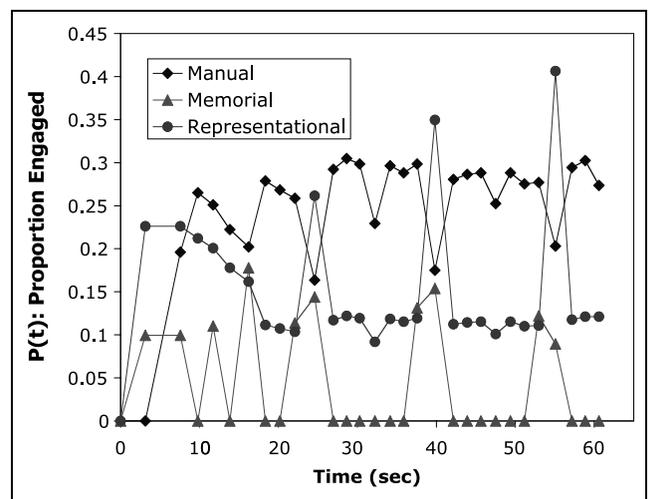
Figure 6. Engagement functions for Experiment 1: Proportion of time different cognitive functions are occupied during different periods of solving a problem.

Table 2. Parameters Used in Fitting the Experiments and Goodness-of-Fit Measures

	Experiment 1	Experiment 2	Experiment 3
Motor			
Scale (*s*)	2.126	1.618	2.973
Exponent (*a*)	1	1	1
Magnitude (*m*)	3.38%	2.30%	0.60%
Correlation	0.979	0.989	0.985
χ^2	43.81	16.43	25.38
df	25	28	28
Parietal			
Scale (*s*)	4.355	3.551	4.119
Exponent (*a*)	1	1	1
Magnitude (*m*)	1.59%	1.58%	0.84%
Correlation	0.977	0.972	0.987
χ^2	25.09	44.25	17.76
df	25	28	28
Prefrontal			
Scale (*s*)	3.704	1.505	2.483
Exponent (*a*)	1	1	1
Magnitude (*m*)	2.50%	0.69%	0.62%
Correlation	0.786	0.689	0.734
χ^2	18.94	35.58	30.74
df	25	28	26

Table 3a displays the correlations among the functions in Figure 5. The parietal is strongly correlated with both the motor and the prefrontal, but the prefrontal and motor are only as correlated with the motor as would be predicted by their correlation with the parietal. This is true in both the data and the theory. The correlation between motor and parietal is driven by the fact that they both rise substantially from zero and stay at this level throughout the task (but show opposite trends at this asymptote). The parietal and prefrontal are strongly correlated because they tend to display the same dips and peaks after the initial moves.

It is worth noting here how our approach of using both latency and brain imaging data provides converging tests of the underlying theory. Basically, in fitting the latency data (Figure 4), we are testing the theory and establishing a set of estimates of the timing and duration of the representational, motor, and memorial components. We can then take these estimates and make predictions about the BOLD responses (Figure 5)

seen in the corresponding regions. This allows us to further test the theory and interpret a complex BOLD signal.

To determine whether there were regions that our confirmatory analysis missed, we did an exploratory analysis focusing on Moves 9–24. These moves are after the BOLD functions in Figure 5 have reached near asymptote and are beginning to fluctuate with task structure. They involve the eight-moves sequences from (B) to (C) and from (C) to (D) in Figure 3. We looked for voxels that seemed responsive to the planning steps that were occurring on Moves 9 and 17. We used a contrast of the average activity on Scans 10 and 11 and Scans 18 and 19 after these moves with the average activity on Scans 14 and 15 and Scans 22 and 23, which are far from planning moves. Using this statistical contrast, we looked for regions that showed 30 contiguous voxels significant at $p = .01$. By this definition, only two regions were significant. One was a left premotor region (Talairach coordinates $x = -32$, $y = -6$, $z = 54$) and the other was a left parietal region ($x = -12$, $y = -74$, $z = 34$). Both are similar to regions found by Fincham et al. (2002) and the parietal region is close to our predefined parietal region. Their behavior over the 28 moves correlates strongly with each other ($r = .892$) and with our predefined parietal region ($r = .912$ for correlation with exploratory premotor and $r = .953$ for correlation with exploratory parietal). The premotor correlation perhaps reflects planning of the motor sequence.

EXPERIMENT 2

The first experiment required participants to remember the goal state for the problems. To avoid confounding results of the prefrontal region with memory demands for the goal state, we adopted a procedure developed by Fincham (in preparation) to use color coding to indicate the goal as part of the problem representation. There is a "hot-to-cold" coding of the goal locations for the disks with red, orange, yellow, green, and blue for the largest (5) to smallest (1) disks, respectively. Thus, the cell in which Disk 5 should go was colored red, the cell in which Disk 4 should go was colored orange, and so forth. Also, as Figure 7 indicates, the participants in this experiment solved 31-move problems. Each problem had its own random start state and end state but all were isomorphic in their logical structure to this 31-move problem. Figure 8 illustrates the demands of this task in a format comparable to Figure 6.

Results

Figure 9 displays the mean latency for each of the 31 moves. The average error rate was 5.7% and its correlation with latency was $r = .974$. Figure 9 also displays the

Table 3. Intercorrelations among the Predicted and Observed BOLD Responses

		Data Motor	Data Parietal	Data Frontal	Theory Motor	Theory Parietal	Theory Frontal
(a) Experiment 1							
Data	Motor	1.000	.839	.524	.979	.873	.551
Data	Parietal		1.000	.727	.801	.977	.848
Data	Frontal			1.000	.429	.727	.786
Theory	Motor				1.000	.845	.453
Theory	Parietal					1.000	.803
Theory	Frontal						1.000
(b) Experiment 2							
Data	Motor	1.000	.890	−.485	.985	.908	−.401
Data	Parietal		1.000	.125	.883	.987	.142
Data	Frontal			1.000	−.781	−.625	.734
Theory	Motor				1.000	.899	−.594
Theory	Parietal					1.000	−.047
Theory	Frontal						1.000
(c) Experiment 3							
Data	Motor	1.000	.890	−.485	.985	.908	−.401
Data	Parietal		1.000	.125	.883	.987	.142
Data	Frontal			1.000	−.781	−.625	.734
Theory	Motor				1.000	.899	−.594
Theory	Parietal					1.000	−.047
Theory	Frontal						1.000

predictions of the theory based on the model described in the introduction to this experiment. In fitting ACT-R to these data, we used the same parameters as in Figure 2 for Experiment 1.

Figure 10 illustrates the BOLD response obtained in these three regions over the 31 moves of the problem.
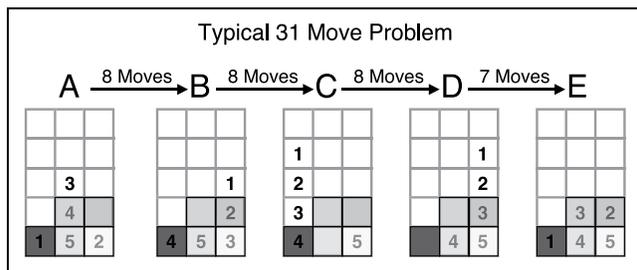
Figure 7. A typical problem from Experiment 2. The letters (A)–(E) denote distinct states of the problem. The shading indicates the color coding of disk destination: darkest to lightest indicates blue to red, which indicates destinations for Disks 1 through 5.

Using the same conventions as Figure 5, it plots the BOLD response on the first scan associated with each move. The BOLD response is measured as the percent difference from the baseline defined by the three scans before the presentation of the problem (also plotted in Figure 10). We performed two statistical tests—first, a t test of whether the mean response was above zero and second, an F test of whether the variation on the scans associated with Moves 2 through 31 were significant. The effects for the motor region are highly significant $[t(7) = 5.57, p < .001; F(29,203) = 4.76, p < .0001]$ as are the effects for the parietal region $[t(7) = 3.72, p < .01; F(29,203) = 5.17, p < .0001]$. The apparent effects associated with the prefrontal region are marginal $[t(7) = 2.06, p < .1; F(29,203) = 1.77, p < .05]$. Again, after the initial rise, the motor and the parietal appear to be mirror images of one another with a negative correlation of −.57 for the last 24 moves $[t(22) = 3.24, p < .005,$ two-tailed] and the parietal and prefrontal are positively correlated at $r = .92$ $[t(22) = 11.02, p < .0001,$ two-tailed].
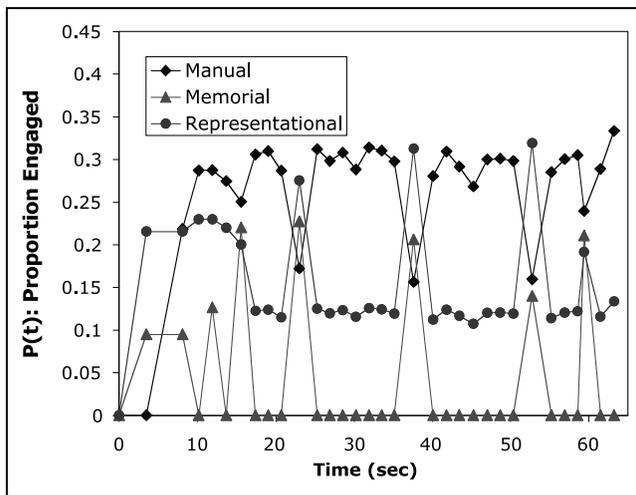
Figure 8. Engagement functions for Experiment 2: Proportion of time different cognitive functions are occupied during different periods of solving a problem.

Figure 10. Experiment 2: Mean observed and predicted BOLD responses for the regions in Figure 1 as a function of mean point of move in the problem solution. The BOLD responses are calculated as percent change from the baseline established by the premove scans.

Figure 10 also illustrates our efforts to predict the BOLD functions in these regions. These predictions are based on the densities in Figure 8 of motor, representational, and memorial operations as the problem is solved. The parameters and measures of fit are given in Table 2. The overall quality of fit can be measured as a χ^2 distribution with degrees of freedom (84) equal to three times the difference between the number of moves (31) and parameters (3). This χ^2 is 97.76, which is not significant. Thus, we can conclude that the model does account for the systematic variance in the data. Table 3b displays the correlations among the functions in Figure 10, replicating the pattern in Table 3a for the previous experiment.

Figure 9. Experiments 2 and 3: Mean observed and predicted times to make a move as a function of the position of the move in the problem sequence.

As in Experiment 1, we performed an exploratory analysis focusing on Moves 9–24 to determine if there were any regions that our confirmatory analysis had missed. As in Experiment 1, this analysis revealed a left premotor region (centered at Talairach coordinates $x = -30$, $y = -6$, $z = 54$) and a left posterior parietal region (centered at Talairach coordinates $x = -30$, $y = -66$, $z = 38$). The intercorrelation of these two regions was strong ($r = .972$) and both correlated strongly with the predefined parietal region ($r = .985$ for the exploratory parietal region and $r = .964$ for the exploratory premotor region).

EXPERIMENT 3

Although the previous two experiments used a fast-paced procedure, there still were artificial pauses because the state updates were synched with the beginnings of scans. In the third experiment, we removed these delays. Otherwise, the procedure was identical to that in Experiment 2 including the sample size of eight participants.

Figure 9 displays the mean latency for each of the 31 moves for Experiments 2 and 3. The difference in mean times between the two experiments was not significant [$t(14) = 0.93$]. We subjected the scanning data to the same analyses as the previous, noting the BOLD response on the first scan was associated with each move. Sometimes the same scan was associated with more than one move. Figure 11 plots the engagement functions for this experiment. These tend to be higher than the engagement functions for past experiments because there is no extra time between moves.

Figure 12 plots BOLD response for each move as a function of the mean time of that move. t tests and F
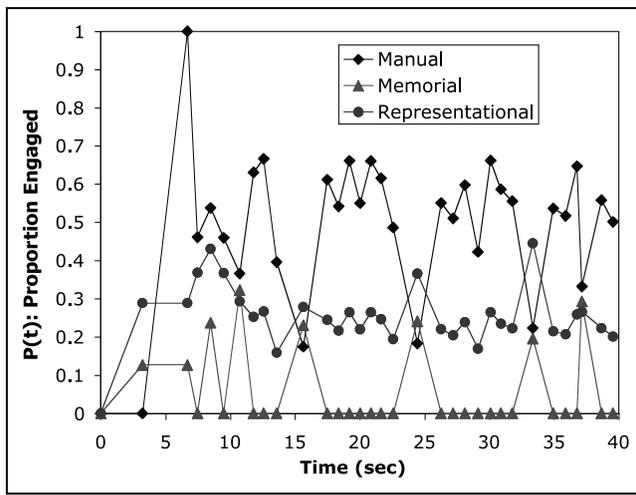
Figure 11. Engagement functions for Experiment 2: Proportion of time different cognitive functions are occupied during different periods of solving a problem.

tests similar to the previous experiment were performed. The effects for the motor region are highly significant [$t(7) = 3.30$, $p < .01$; $F(29,203) = 20.71$, $p < .0001$] as are the effects for the parietal region [$t(7) = 3.06$, $p < .05$; $F(29,203) = 4.81$, $p < .0001$]. The effects associated with the prefrontal region are again somewhat marginal [$t(7) = -1.23$, $p < .1$; $F(29,203) = 2.44$, $p < .0005$]. Again there is a negative correlation between the parietal and motor over the last 24 moves [$r = -.56$; $t(22) = 3.21$; $p < .005$, two-tailed] and a positive correlation between the parietal and prefrontal [$r = .82$; $t(22) = 6.83$; $p < .0001$, two-tailed]. Note that, unlike the previous experiments, the function for the prefrontal region actually starts higher on Scan 1 than it is on later scans. We have dealt with this by adding a constant to the BOLD responses such that the lowest value for the BOLD function (Scan 17 at 23.8 sec) is 0 and scaling the other values appropriately. Although not shown in the figure, the BOLD response actually goes up in the prefrontal area after the end of the trial, reaching 0.37% four scans after the problem has been solved. Then it goes down to where it is at the beginning of the next problem and it continues to fall off some from there. This interproblem rise in the prefrontal response was not observed in the previous experiments. The statistics in Tables 2 and 3 for the prefrontal region in this experiment were calculated excluding the data for the first two moves.

As previously, we performed an exploratory analysis focusing on Moves 9–24 to determine if there were any regions that our confirmatory analysis had missed. Again, we found a left premotor (Talairach coordinates $x = -33$, $y = -6$, $z = 53$) region and a parietal region (although there was some left activation, the center of the strongest effect was right; $x = 33$, $y = -63$,

$z = 33$). The correlation between these two regions was not as strong ($r = .800$) as in the previous experiments. The exploratory parietal correlated strongly with the predefined parietal region ($r = .944$) but the correlation was relatively weak ($r = .599$) between the exploratory premotor region and the predefined parietal region.

DISCUSSION

This research used regions that have provided reliable fMRI indicants of information processing in slow event-related paradigms to track cognition in a rapid, complex problem-solving task. The results are consistent with respect to the motor and parietal regions. As expected, these regions appear to provide mirror images of each other with motor activity decreasing and parietal activity increasing at points that require considerable intellectual engagement. The result with respect to the motor region, while confirming the overall methodology, is what one would expect: After points where the participant is making their responses less rapidly, the BOLD response tends to go down.

The response in the parietal region is interesting because it suggests that we might be able to track planning in tasks where participants are not making overt motor responses. As Figures 5, 10, and 12 show, latency of response is a reasonable behavioral indicant of when planning is taking place. However, the TOH with its many moves is rather unusual for problem-solving tasks; for instance, many mathematical problem-solving tasks can involve a lot of thought occasionally marked by the creation of a plan and without any accompanying motor activity. For instance, in geometry theorem proving, a participant may set upon the plan of proving two triangles congruent, spend a lot of time engaged in
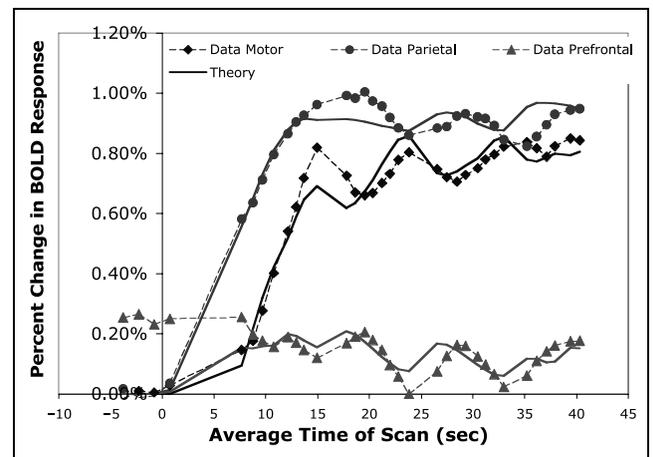
Figure 12. Experiment 3: Mean observed and predicted BOLD responses for the regions in Figure 1 as a function of mean point of move in the problem solution. The BOLD responses are calculated as percent change from the baseline established by the premove scans.

finding ways to achieve the congruence, and only then output anything. Could we tell by parietal activation when the most intense planning was taking place? We tried to address this question for our TOH data where we know both from task analysis and latency of response when intense planning is taking place. Could we have detected such planning points if we did not have access to the latency markers? We addressed this question with respect to Experiments 2 and 3, where participants solved a great many more problems than Experiment 1.

The first question we addressed was whether we could detect the onset of planning associated with the beginning of a TOH problem. The BOLD statistic that seemed best able to predict this was the magnitude of rise in the BOLD response in the parietal region. If the BOLD response increased by a threshold amount over the next five scans (7.5 sec), we declared this a planning event. We set the threshold so that we successfully classified 50% of the first scans associated with the first move as the onset of planning events. To achieve this, we used threshold increases of 0.90% for Experiment 2 and 0.78% for Experiment 3. The question of relevance is how often would we have false alarms by these thresholds. That is, how often at later points when participants are not engaged in such intense planning would the parietal rise have surpassed this threshold? Table 4a presents the percent false alarms for the first scans associated with Moves 4–29[2] and the resulting d' measures of discriminability for the two experiments. As Table 4a indicates, we would suffer false alarm rates between 5% and 10% and achieve a d' of about 1.5. As a comparison, we calculated in Table 4b how well this

could be done using the latencies of these moves. The latency statistic that seemed most discriminating was to calculate the mean difference between a move and the average of the two successor moves. The threshold differences for 50% planning events was 5.57 sec for Experiment 2 and 4.35 sec for Experiment 3. Table 4b also reports the false alarm rates we had on other moves and d' measures. As can be seen, latencies yielded false alarms around 1% and d' measures around 2.5. Thus, it is clear that the behavioral measure is much better than the BOLD measure, but it does seem that the brain measure would give us a pretty good test in cases where participants are not giving a continuous stream of latency measures.

As more ambitious matter, we asked how well we could do at detecting the subplanning episodes that were occurring on Moves 9, 17, and 25. We addressed how well we could discriminate these moves among the 24 moves from Move 6 to Move 29, a period when things are relatively stable. For the BOLD contrast, we looked at the rise over the next three scans (4.5 sec after Moves 9, 17, and 25)—a smaller time window than in the previous contrast because the planning at these points is relatively short. Table 4c shows the results we got using thresholds to give 50% hits. The false alarm rates are calculated for the 15 moves (6, 7, 11–15, 19–23, and 27–29), which are more than one move removed from the critical move. The d' was just under 1 for Experiment 2 but just above 0 for Experiment 3. The reason for the poor performance in Experiment 3 is probably that the rapid moves give little time for the BOLD function to rise or fall. Table 4d also shows the discrimination we can get using move latency. Although this is likewise

Table 4. Ability to Detect Planning Events Using Parietal Activation versus Latency

Detecting Onset of Planning						
(a) BOLD Difference				(b) Latency Difference		
	Experiment 2 0.90%	Experiment 3 0.78%			Experiment 2 5.85 sec	Experiment 3 4.35 sec
Hits	49.8%	49.9%		Hits	49.8%	49.9%
False Alarms	5.8%	8.4%		False Alarms	0.2%	1.2%
d'	1.57	1.38		d'	2.81	2.26

Detecting Onset of Subplanning						
(c) BOLD Difference				(d) Latency Difference		
	Experiment 2 0.31%	Experiment 3 0.06%			Experiment 2 1.06 sec	Experiment 3 0.79 sec
Hits	50.2%	50.2%		Hits	50.1%	50.2%
False Alarms	18.7%	42.8%		False Alarms	2.4%	3.9%
d'	0.89	0.19		d'	1.98	1.77

reduced from detection of the onset planning, making the more subtle discriminations of subplan moves can still be discriminated on the basis of latency with a d' of nearly 2.

In conclusion, activity in the parietal region offers some potential for predicting when a person is planning. However, as the planning episodes become brief and as the interludes between planning episodes become brief, parietal activity loses its predictive ability.

METHODS

Subjects

Eight right-handed English-speaking subjects enrolled in each experiment. The sex–age distributions were 4 men, 4 women, ages 20–25 years for Experiment 1; 4 men, 4 women, ages 18–23 for Experiment 2; and 4 men, 4 women, ages 19–28 for Experiment 3. IRB approval was obtained from both Carnegie Mellon University and the University of Pittsburgh. All participants gave informed consent in accordance with Carnegie Mellon and University of Pittsburgh guidelines.

Experiment 1

3-day Protocol

The protocol for the experiment involved a very similar methodology to that used in Fincham et al. (2002), which in turn involves elements of the methodology described in Anderson and Douglass (2001). The experiment depends heavily on having subjects who are consistently using the same procedure for the problems. Therefore, the experiment involved two days of training before participants went into the scanner on the third day. The following were the activities on each day:

Day 1. Participants were instructed on the sophisticated perceptual strategy and given 14 problems as in Anderson and Douglass (2001). Then they were given an additional 14 problems involving the GOP representation. These initial 28 problems involved explicit posting of subgoals with a mouse-based interface. This was then followed by 28 more GOP problems where participants just moved disks (numbers) and indicated their moves by finger presses in a data glove. Moves were indicated by two finger presses—selecting the peg (column) from and to which the disk was to be moved. The index, middle, and ring fingers of the right hand were mapped on the pegs in a spatially compatible form.

Day 2. Participants were much faster on the second day and did not have the introductory instructions, and so it was possible to give them more practice. They received 28 tower problems and 14 grid problems to review the sophisticated perceptual strategy. There

were also 40 problems in which they practiced working with the data glove towards a memorized goal rather than one presented on the screen. They practiced with four different memorized goals, each for 10 problems. The last 10 involved the goal they would be working towards in the scanner but these were simpler than the problems they would see in the scanner.

Day 3. They performed eight blocks in the scanner, each block 5 min long. Each block began with a small warm-up problem and then two 28-move problems like the one illustrated in Figure 3. Thus, we obtained from each participant 16 observations of their performance on 16 different isomorphs of the 28-move problems.

Scanning procedure. Event-related fMRI data were collected by using a gradient echo-planar image (EPI) acquisition on a Siemens 3-T Allegra Scanner. The imaging parameters were TR = 1500 msec, TE = 30 msec, RF flip angle = 55°, FOV = 210 mm, matrix size = 64 × 64 (3.125 × 3.125 mm per pixel), slice thickness = 3.2 mm, slice gap = 0 mm, and 26 axial slices per scan with the AC–PC on the 20th slice from the superior. In addition, structural images of two-dimensional T1 spin-echo were acquired at the same slice location and spatial definition as EPI.

The presentation of the next state after a move was synched to the 1.5-sec stepping of the scans. This meant that some time could occur between the execution of the move and the updating of the screen. Should the wrong next move be made, the problem state was updated to the correct next state and participants were given an additional 3 sec (2 scans) to study the new state. Participants had already experienced these timing conventions in their data glove problems on Days 1 and 2.

Experiments 2 and 3

As participants did not have to practice the goal states, the procedure was simplified into a 2-day experiment:

Day 1. The 28-problem strategy learning introductory sequence was given. Then participants were instructed on how to use the glove with the colored goal state. The subjects then practiced in a setting equivalent to their scanner experience the next day. This involved doing the full procedure described below for Day 2.

Day 2. This was performed in the scanner and involved eight blocks with 8 min/block. Except for smaller warm-up problems to begin each block, participants solved 31-move problems with a color-coded goal state. In Experiment 2, the state did not update and responses could not be made until the next fMRI scan began. In Experiment 3, the state updated and responses were given as soon as a move was made. The other difference between Experiments 2 and 3 was

that there were eight 1.5-sec scans between problems in Experiment 2 and 11 scans in Experiment 3.

Notes

1. In contrast, the representational and memorial activities should be distributed throughout the interval.
2. We excluded Moves 2 and 3 because the BOLD function still tends to be rising from the initial planning and we excluded Moves 30 and 31 because the latency statistic (to be described) is not defined for these moves.

REFERENCES

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26,* 39–83.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111,* 1036–1060.

Anderson, J. R., & Douglass, S. (2001). Tower of Hanoi: Evidence for the cost of goal retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1331–1346.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

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. *Psychonomic Bulletin and Review, 10,* 241–261.

Anderson, J. R., Qin, Y., Stenger, V. A., & Carter, C. S. (2004). The relationship of three cortical regions to an information-processing model. *Cognitive Neuroscience, 16,* 637–653.

Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S. J., & Robbins, T. W. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia, 34,* 515–526.

Boyton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience, 16,* 4207–4221.

Buckner, R. L., Kelley, W. M., & Petersen, S. E. (1999). Frontal cortex contributes to human memory formation. *Nature Neuroscience, 2,* 311–314.

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,* 317–330.

Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *Neuroimage, 6,* 93–103.

Dagher, A., Owen A. M., Boecker, H., & Brooks, D. J. (1999). Mapping the network for planning: A PET study of the Tower of London task. *Brain, 122,* 1973–1987.

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5,* 329–340.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20,* 487–506.

Donaldson, D. I, Petersen, S. E., Ollinger, J. M., & Buckner, R. L. (2001). Dissociating state and item components of recognition memory using fMRI. *Neuroimage, 13,* 129–142.

Fincham, J. M. (in preparation). Cognitive modeling and fMRI: An integrated approach toward understanding mechanisms of complex skill performance.

Fincham, J. M., Carter, C. S., van Veen, V., Stenger, V. A., & Anderson, J. R. (2002). Neural mechanisms of planning: A computational analysis using event-related fMRI. *Proceedings of the National Academy of Sciences, 99,* 3346–3351.

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain, 124,* 849–881.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9,* 416–429.

Hinz, A. M. (1992). Shortest paths between regular states of the Tower of Hanoi. *Information Sciences, 63,* 173–181.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17,* 248–294.

Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of National Academy of Sciences, U.S.A., 97,* 506–511.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104,* 2–65.

Newman, S. D., Carpenter, P. A., Varma, S., & Just, M. A. (2003). Frontal and parietal participation in problem-solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia, 41,* 1668–1682.

Owen, A. M., Downs J. J., Shahakian, B. J., Polkly, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following fontal lobe lesions in man. *Neuropsychologia, 28,* 1021–1034.

Owen, A. M., Doyon, J., Petrides, M., & Evans, A. C. (1996). Planning and spatial working memory: A positron emission topography study in humans. *European Journal of Neuroscience, 8,* 353–364.

Qin, Y., Anderson, J. R., Silk, E., Stenger, V. A., & Carter, C. S. (2004). The change of the brain activation patterns along with the children's practice in algebra equation solving. *Proceedings of National Academy of Sciences, U.S.A., 101,* 5686–5691.

Qin, Y., Sohn, M.-H., Anderson, J. R., Stenger, V. A., Fissell, K., Goode, A., & Carter, C. S. (2003). Predicting the practice effects on the blood oxygenation level-dependent (BOLD) function of fMRI in a symbolic manipulation task. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 4951–4956.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural basis of strategy and skill in sentence–picture verification. *Cognitive Psychology, 40,* 261–295.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, 199,* 199–209.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7,* 268–288.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology, 8,* 165–190.

Sohn, M.-H., Goode, A., Stenger, V. A, Carter, C. S., & Anderson, J. R. (2003). Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 7412–7417.

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,* 21–33.

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,* 1337–1347.

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338.