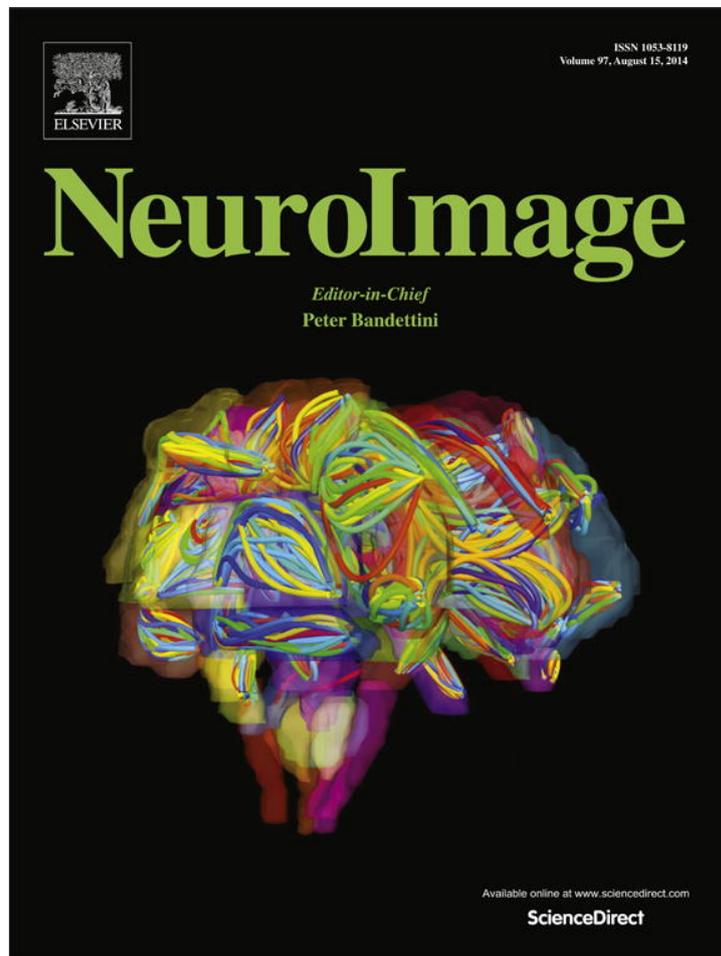


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

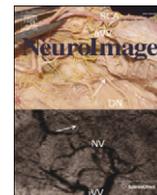
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect

NeuroImage

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

Discovering the structure of mathematical problem solving

John R. Anderson^{a,1}, Hee Seung Lee^b, Jon M. Fincham^a

^a Department of Psychology, Carnegie Mellon University, USA

^b Department of Education, Yonsei University, Republic of Korea



ARTICLE INFO

Article history:

Accepted 8 April 2014

Available online 16 April 2014

Keywords:

fMRI

Hidden Markov model

Mathematical problem solving

Multivariate pattern analysis

ABSTRACT

The goal of this research is to discover the stages of mathematical problem solving, the factors that influence the duration of these stages, and how these stages are related to the learning of a new mathematical competence. Using a combination of multivariate pattern analysis (MVPA) and hidden Markov models (HMM), we found that participants went through 5 major phases in solving a class of problems: A Define Phase where they identified the problem to be solved, an Encode Phase where they encoded the needed information, a Compute Phase where they performed the necessary arithmetic calculations, a Transform Phase where they performed any mathematical transformations, and a Respond Phase where they entered an answer. The Define Phase is characterized by activity in visual attention and default network regions, the Encode Phase by activity in visual regions, the Compute Phase by activity in regions active in mathematical tasks, the Transform Phase by activity in mathematical and response regions, and the Respond phase by activity in motor regions. The duration of the Compute and Transform Phases were the only ones that varied with condition. Two features distinguished the mastery trials on which participants came to understand a new problem type. First, the duration of late phases of the problem solution increased. Second, there was increased activation in the rostrolateral prefrontal cortex (RLPFC) and angular gyrus (AG), regions associated with metacognition. This indicates the importance of reflection to successful learning.

© 2014 Elsevier Inc. All rights reserved.

Introduction

The past decade has seen a considerable growth in the understanding of the neural basis of certain aspects of mathematics. The greatest amount of research has gone into understanding the role of various parietal regions in basic arithmetic tasks and their role in normal and abnormal development (e.g., Ansari and Dhital, 2006; Arsalidou and Taylor, 2011; Butterworth et al., 2011; Castelli et al., 2006; Molko et al., 2003). Dehaene's (1997) triple-code theory identifies three regions as critical to the representation of number: the horizontal intraparietal sulcus that processes numerical quantity, the angular gyrus that is involved in the verbal processing of numbers, and the fusiform gyrus that processes number form. In addition, the prefrontal cortex is particularly involved in more advanced tasks involving topics like algebra, geometry, or calculus (e.g., Krueger et al., 2008; Qin et al., 2004). One prefrontal region of interest is the lateral inferior prefrontal cortex that is involved in retrieval of arithmetic facts and semantic facts (Danker and Anderson, 2007; Dehaene et al., 1999; Menon et al., 2000). More dorsal and more anterior prefrontal regions become engaged as the problem solving gets more complicated (Wintermute et al., 2012).

Most of this past research has looked at the execution of well-established procedures. The current research investigated how mathematical knowledge becomes “alive” and extends to solving novel problems. We taught participants a new mathematical skill (which is really just equation-solving in disguise) and then challenged them to extend what they had learned to novel transfer problems. In order to identify when the key cognitive events occurred we needed to develop new methods that deal with the variability in complex mathematical problem solving.

A complex skill like algebra problem solving involves a rich mixture of perceptual, cognitive, and motor activities. For instance, when manipulating an equation in traditional paper and pencil mode, a student has to scan past lines of equations, identify the next critical transformation, determine what the new equation will be, and then write that equation. In more modern computer interfaces and tutoring systems, handwriting can be removed but there still are the same basic steps with computer gestures replacing handwriting. This complexity and mixture of activities makes it difficult to identify when the critical cognitive events are taking place. This paper will show that it is possible to analyze individual trials and identify the critical events by combining multivariate pattern analysis (MVPA—e.g., Norman et al., 2006; Pereira et al., 2009) and Hidden Markov Model (HMM) algorithms (Rabiner, 1989). The MVPA recognizes the mental states and the HMM recognizes the sequence of states.

E-mail address: ja+@cmu.edu (J.R. Anderson).

¹ Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA.

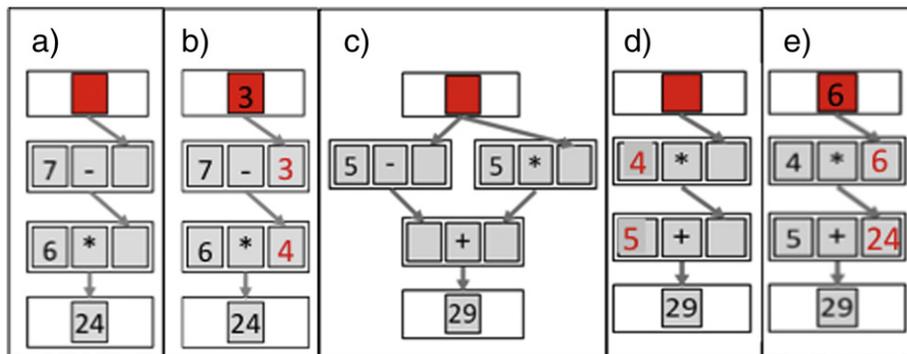


Fig. 1. Data flow graphs where an unknown number flows down from the top box. Red denotes numbers added to the diagram by the participant. (a) A simple Propagate problem equivalent to $6 * (7 - x) = 24$; (b) The solution for (a) where 4 is entered because $6 * 4 = 24$ and 3 is entered because $7 - 3 = 4$; (c) A Linearize problem, equivalent to $(5 - x) + (5 * x) = 29$, with two paths which must be converted into Propagate form; (d) The Propagate equivalent of (c) since $5 - x + 5x = 4x + 5$; (e) The solution for (d) since $5 + 24 = 29$ and $4 * 6 = 24$.

Our past work (Anderson et al., 2010, 2012a, 2012b) trained an HMM and an MVPA classifier on the states with one set of data and then tested it with another set. This required that the states be known in advance and marked in the training set. More recently, Anderson and Fincham (2013) showed that this approach could be extended to discover what the states were solely based on the data, without needing the state structure to be specified in advance. Their demonstration involved using only imaging data, but the current paper will show that this approach becomes more powerful when the imaging data are merged with behavioral data. These methods can analyze problem-solving episodes that involve up to 60 s of mixed activities and identify the few key moments in the episodes where the most critical cognitive events are happening.

This paper is divided into four parts. Part 1 describes an imaging experiment studying how participants learn new mathematical problem solving skills. Part 2 explains the MVPA–HMM method and describes the states that it discovers for this experiment. Part 3 uses the inferred states to gain a deeper understanding of task performance at both behavioral and neural levels. Finally, Part 4 interprets our results and their implications.

A study of mathematical learning and transfer

We have developed a data-flow isomorph of school algebra that has allowed us to study college students learning algebra all over again (Brunstein et al., 2009; Lee et al., 2011). Because it is a laboratory invention and not a real mathematics topic, we have been free to explore a range of instructional variation that might not be appropriate or ethical for students learning real algebra. The data for this paper come from an experiment (Lee et al., in press²) that involved a contrast between learning by discovery and learning by direct instruction. Participants learned how to solve the problems in one session outside the scanner and then had to transfer this knowledge to solving new, challenging problems in the scanner. Discovery participants were somewhat slower in mastering the material in the learning session but there were no differences (behavioral or imaging) between instruction and discovery participants in the transfer session. Lee et al. analyzed the instructional effects in learning and their disappearance in transfer. Here we are interested in analyzing the common processes by which participants approached these transfer problems and will pool the data from the different instructional groups.

Fig. 1 shows examples of the data-flow structures that we used. They consist of a set of boxes containing tiles with numbers or operators. Arrows connect boxes to tiles. In data-flow representations a number

flows from a top box through a set of arithmetic operations to a bottom box. If that number is unknown, the data-flow structure is equivalent to an algebraic equation with a single variable. For instance, Fig. 1a is the data-flow equivalent of $6 * (7 - x) = 24$. The task is to determine what values to fill into the empty tiles in the boxes. For a linear structure like Fig. 1a, the values can be determined by simply “propagating” the number up from the bottom and performing the arithmetic operations. The solution (as illustrated in Fig. 1b) involves placing 4 in the empty tile above the bottom box (since $6 * 4 = 24$), then placing 3 in the empty tile above it (because $7 - 3 = 4$), and finally placing 3 in the top box (equivalent to solving as $x = 3$). Most participants find solving these problems by this propagation strategy easy and intuitive (one participant described it as similar to playing Sudoku). However, when problems cannot be solved by this simple propagation strategy, participants tend to have difficulty understanding the problem structure and figuring out a procedure for solving the problem.

One class of difficult problems involves the unknown value flowing down multiple paths. Fig. 1c illustrates a simple case of such a problem, which is equivalent to solving an equation with multiple appearances of the variable. The diagram in Fig. 1c is equivalent to the algebraic expression, $(5 - x) + (5 * x) = 29$. In the diagram an unknown value flows down into the two tiles in a box below, which are summed to produce a result of 29. Because two paths converge in a single result, the propagation strategy does not work. The way to solve this problem within the rules of the system is to transform the graph in Fig. 1c into the linear form in Fig. 1d (equivalent to $4x + 5 = 29$), where this simple propagation procedure is possible again as illustrated in Fig. 1e. This transformation step, called *linearization*, is a major conceptual hurdle in this artificial curriculum. It corresponds to collection of variables and constants in regular algebra, which in combination with distribution causes some difficulty when regular algebra is taught in school.

The most difficult step in a Linearize problem is determining the values to enter into the linearized form—for instance, the 4 and the 5 in Fig. 1d. Participants in this experiment had spent the first day, outside of the scanner, mastering this linearization step on relatively simple problems like Fig. 1c. On the second day, they went into the scanner and solved Linearize problems that posed new challenges. Fig. 2 shows two examples of such challenging problems. They would see the problem on the left with the multiple boxes highlighted that had been replaced by a linear structure on the right. Their task was to enter into the two blue tiles the numbers that would make the left and right structures equivalent.

The major experimental manipulation in the transfer section involved the type of problem participants were asked to solve. The problems were either

1. Graphic problems: These involved more complex graph structures than participants had solved up until this point. Fig. 2a illustrates a

² This paper and entire experimental materials are available at <https://www.dropbox.com/sh/bya83pytbixzff/OLMbG0OVX4>.

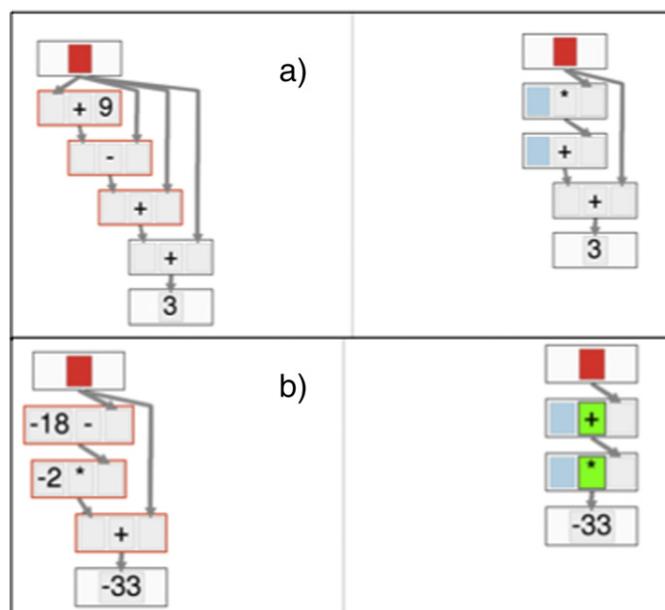


Fig. 2. Transfer problems where the participant must fill in the blue tiles on the right to create an equivalent graph to the one on the left: (a) An example of a Graphic transfer problem: the left is the equivalent of $((((x + 9) - x) + x) + x) = 3$ and it should be transformed by filling in the right to be the equivalent of $((1x + 9) + x) = 3$. (b) An example of an Algebraic transfer problem: the left is the equivalent of $((-2 * (-18 - x)) + x) = -33$ and it should be transformed by filling in the right to be the equivalent of $(3 * (12 + x)) = -33$.

Graphic problem. This particular problem involves combining multiple loops—something participants had not seen before. To make the right graph equivalent, participants would need to enter a 1 and a 9 into the blue tiles in the right side of Fig. 2a.

- Algebraic problems: These involved performing some transformation of the answer from its normal form. The change was highlighted in green. In the example in Fig. 2b, where previously the right structure had involved a multiplication flowing into an addition, now an addition flows into a multiplication. Participants would have entered the equivalent of $3x + 36$ for Fig. 2b in the learning session, but the changed graph structure on the right half of Fig. 2b requires that they enter the equivalent of $3 * (x + 12)$. Thus, rather than entering 3 and 36 they would now enter 3 and 12.

Graphic problems could be solved even if participants did not understand why the right and left graphs were equivalent but rather just understood how to combine numbers in the left graph. In contrast, the Algebraic problems cannot be solved without understanding the equivalence of the two graphs. Thus, Graphic problems were a test of how well participants understood how to combine the numbers on the left, while Algebraic problems were also a test of their understanding of the equivalence of the two structures. Even though Algebraic problems were physically less complicated (averaged fewer boxes), we expected that Algebraic problems would be more difficult because of their extra conceptual requirements.

There were 4 types of Graphic problems and 4 types of Algebraic problems. Participants went through 8 blocks of 8 problems in which they saw one instance of each type. The first time participants saw each problem type they would only succeed if they could extend what they had been taught to the new problem type. After each problem the correct answer was shown without an explanation. By studying the correct answer, participants had an opportunity to figure out how the problem type was solved.

Procedure

³ Analyses will include data from 39 participants⁴ in the scanning session on the second day. Participants interacted with this system by means of a mouse. They moved the mouse to a tile where they want to enter a value, then a keypad of numbers came up on the screen, they clicked on the digits to create the value to enter into the tile, and they moved the mouse out of the tile when they were done. They made numerous false starts such as moving to a tile and then choosing not to enter anything or going back and changing what they entered. This record of mouse actions provides a rich behavioral data source to coordinate with the fMRI data.

Each trial in the scanner had the following phases:

- 3 s of fixation.
- The problem was presented and participants had up to 60 s to solve the problem. This is the portion of the trial that will be analyzed.
- If their answer was correct, participants saw the solution confirmed for 2 s. If it was incorrect or timed out, they had 10 s to study the correct solution without any explanation.
- 3 s of fixation.
- 12 s of repetition detection. During repetition detection, letters appeared on the screen at a rate of 1 per 1.25 s, and participants were instructed to click a match button on the screen whenever the same letter appeared twice in a row. This task served to distract the participants from the main Linearize task and return brain activity to a relatively constant level.

In addition, the block began with a fixation phase and a repetition detection phase. Thus each trial was surrounded by this baseline sequence.

Image analysis

Images were acquired using gradient echo-echo planar image (EPI) acquisition on a Siemens 3 T Verio Scanner using a 32 channel RF head coil, with 2 s repetition time (TR), 30 ms echo time (TE), 79° flip angle, and 20 cm. field of view (FOV). The experiment acquired 34 axial slices on each TR using a 3.2 mm thick, 64×64 matrix. This produces voxels that are 3.2 mm high and 3.125×3.125 mm². The anterior commissure–posterior commissure (AC–PC) line was on the 11th slice from the bottom scan slice. Acquired images were pre-processed and analyzed using AFNI (Cox, 1996; Cox and Hyde, 1997). Functional images were motion-corrected using 6-parameter 3D registration. All images were then slice-time centered at 1 s and co-registered to a common reference structural MRI by means of a 12-parameter 3D registration and smoothed with an 6 mm full-width-half-maximum 3D Gaussian filter to accommodate individual differences in anatomy.

In complex tasks like this we have found it useful to perform MVPA on whole brain activity. However, as a step of dimension reduction and to accommodate variations in anatomy over participants that may not be dealt with in co-registration, we work with relatively large regions. A total of 408 regions were created by evenly distributing $4 \times 4 \times 4$ voxel cubes over the 34 slices of the 64×64 acquisition matrix.⁵ Between-region spacing was 1 voxel in the x- and y-directions in the axial plane and one slice in the z-direction. Some of these regions

³ The detailed methods are described in Lee et al. (in press).

⁴ There were 40 participants in the original experiment but one participant was excluded who showed mean percent BOLD changes twice any other participant and had more than 10% scans with changes greater than 5%. This was done to avoid distorting the group PCA.

⁵ Averaging has the danger of losing information that depends on precise anatomical location. Simulations have shown that this can have substantial effects on functional connectivity analyses (Smith et al., 2011), although our method is not a functional connectivity analysis.

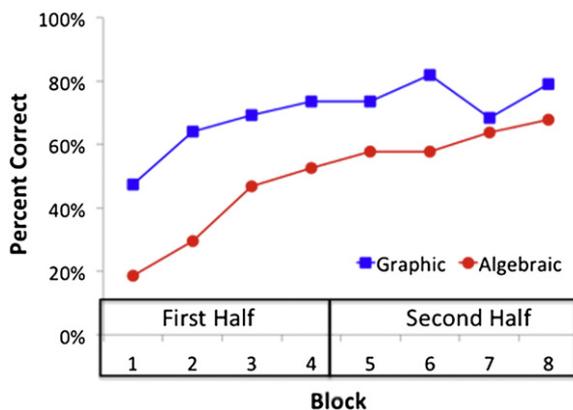


Fig. 3. Mean accuracy for Graphic and Algebraic problems as a function of experimental block.

show an excess of extreme values for some participants, probably reflecting differences in anatomy. These were regions mainly on the top and bottom slices as well as some regions around the edge of the brain. Difficulties in getting consistent signals at edges reflect limitations in the co-registration of different brains to the reference brain and limitations in the corrections for slight head movements. Anderson and Fincham (2013) eliminated these edge regions to produce a final set of 290 regions that we also use here.

The BOLD response is calculated as the percent change from a linear baseline defined from first scan (beginning of fixation before problem onset) to last scan (beginning of fixation before next problem). This is deconvolved with a hemodynamic response function to produce an estimate of the underlying activity signal. The hemodynamic function is the SPM difference of gammas (Friston et al., 2011: $\gamma(6,1) - \gamma(16,1)/6$). A Wiener filter (Glover, 1999) with a noise parameter of .1 was used to deconvolve the BOLD response into an inferred activity signal on the scan. To an approximation, this results in shifting the BOLD signal to the left by 2 scans (4 s). We have used this simple scan shift in mind-reading studies (e.g., Anderson et al., 2010, 2012a, 2012b) where there is not a constant baseline activity interspersed at regular intervals.

Behavioral results

Fig. 3 shows the learning trends over the 8 blocks of the experiment. As expected, Graphic problems have significantly higher accuracy than Algebraic problems ($F(1,38) = 32.93, p < .0001$). There is also a significant improvement over trials ($F(1,38) = 124.16, p < .0001$, comparing first half versus second half). There is a significant type-by-half interaction ($F(1,38) = 8.43, p < .01$), such that the difference between Graphic and Algebraic problems is reduced in the second half of the experiment.

Scanner difficulties resulted in the loss of 2 blocks of imaging data for 1 participant and 1 block for 3 participants, leaving $(39 \text{ participants} \times 8 \text{ blocks} - 5 \text{ lost blocks}) \times 8 \text{ problems} = 2520 \text{ trials}$. Table 1 shows the counts of correct, error, and time-out trials for the two problem types. The imaging analyses will be restricted to corrects and errors because of the difficulty in interpreting time-outs and because only 15 of the 39 participants had time-outs in both the Graphic and Algebraic conditions (6 had no time-outs in either condition, 15 no time-outs in the

Table 1
Number of trials and average number of scans.

	Graphic	Algebraic
Correct	n = 856 14.3 scans	n = 606 13.5 scans
Error	n = 326 15.7 scans	n = 531 15.9 scans
Time-out	n = 46 30 scans	n = 91 30 scans

Graphic condition only, and 3 no time-outs in the Algebraic condition only). An analysis of variance on corrects and errors finds no significant difference in time for Graphic and Algebraic problems ($F(1,38) = .03$), slower times for errors ($F(1,38) = 39.69, p < .0001$), and a significant interaction between the two factors ($F(1,38) = 4.83, p < .05$) such that the difference between corrects and errors is larger for Algebraic problems.

Mouse gestures provide another source of behavioral data and they will contribute substantially to the state analysis. To correctly solve a problem participants need to enter each of the two blue tiles in the right graph (see Fig. 2), key a number of characters, and exit. Although they need to enter and exit a tile only once they average 1.68 visits per tile. They also change their answers on 19% of the trials, resulting in the keying of more characters than the minimum. The extra visits and self-corrections indicate some vacillation on the part of participants as to what to do and what the answer is.

While such behavioral data provides some information, they largely leave open a characterization of what is happening during a trial. Next we describe how the MVPA-HMM approach can identify the unique structure of each trial and so reveal what is happening.

Model discovery by combining fMRI and mouse gestures⁶

Anderson and Fincham (2013) described an HMM-MVPA approach to model discovery that used only fMRI activation patterns. Here we provide an alternative description of the same process, adapted for this experiment. First, we will describe the parameter estimation process for a fixed number of states and then we will describe how we determine the number of states.

Estimating parameters for a fixed number of states

We conceive of the participant as going through a sequence of mental states during the solution of one of these problems. These states are “hidden” in the sense that we do not see them but only see their consequences in the brain signals and behavioral data. Fig. 4 illustrates how we infer the characteristics of these states from the data. The data input into the HMM-MVPA is a set of trials consisting of a variable numbers of scans, where each scan has a set of independent variables. In our case, as described in the Methods, these variables consist of the first 20 PCA dimensions (which capture 73% of the variance in brain activation) and 6 binary variables that reflect mousing activity. The mousing data provide behavioral data not available in Anderson and Fincham and prove to be very valuable. As we will describe, the mousing data provide a “co-training” signal (Blum and Mitchell, 1998) to use with the imaging data and provide a “ground truth” to judge the contribution of the other components of the effort.

As explained in the Appendix A to this paper, for any number of states, the HMM-MVPA methodology estimates a set of parameters that maximize the probability of this data set. For each state, three types of parameters are estimated:

1. Brain signature: 20 mean values for the 20 PCA dimensions in that state.
2. Mousing signature: 6 probabilities for the 6 types of mouse actions during a scan in that state.
3. Temporal signature: A gamma distribution (defined by 2 parameters), which gives the probability that the state will last for a number of scans including 0 scans (which means the state is skipped).

The HMM-MVPA produces a second product besides the estimated parameters. For each scan of each trial, given the data and maximum-likelihood parameters, it provides a state occupancy description, which is the probability that each scan was in each state. As we will

⁶ The data sets and Matlab analysis functions for Figs. 4–11 are available at <https://www.dropbox.com/sh/bya83pftybixzzf/OLMbG00VX4>.

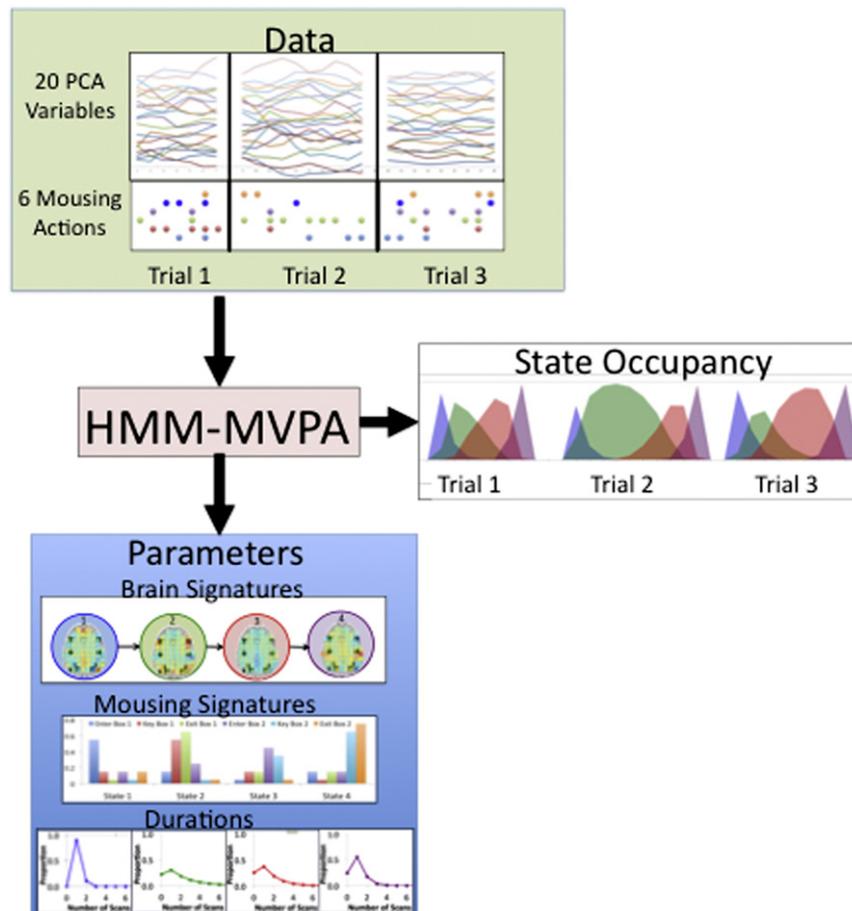


Fig. 4. An illustration of HMM–MVPA producing a 4–state model. Input to the data are scans organized within trials. Each scan consists of the 20 PCA values and 6 binary variables indicating whether a mousing action has occurred. Parameters are estimated for each state: brain signatures reflecting patterns of activations over the 290 regions, mousing signatures giving probabilities of each action, and distributions of state durations. Also calculated are state occupancies, the probabilities that the participant is in each state on a particular scan. Parameters are estimated to provide best fitting description of brain activity, duration, and mousing behavior. The resulting data and parameters can be combined to get a description of each trial as probabilities that scans are in one of the 4 states.

describe, these estimates can be aggregated to provide estimates of mean duration in each state for each trial. These state durations prove to be very informative and will provide most of the information for the third section of this paper.

The “HMM” component of this process involves use of dynamic programming techniques to achieve efficiency in the computation. The “MVPA” component refers to the estimation of the brain signature, as we will now describe. An expectation maximization process iterates on its parameter estimates until it finally converges on the maximum likelihood estimates. On each iteration, it takes the current estimate of state occupancy and uses this as the training signal for performing a linear discriminant analysis (LDA) to find the features that define the states. While we are able to take advantage of the structure of the PCA factors and apply a simple, constrained LDA (see Appendix A), any MVPA technique could be used that delivered a probability of the image data given the technique’s classification of the scans into states.

Determining the number of states

This approach will find the best fitting model for any number of states. To determine the best number of states, N , we use a leave-one-out cross-validation (LOOCV) methodology. This approach estimates the maximum-likelihood parameters for all but one of the participants and then uses these parameters to calculate the likelihood of the data from the last participant. This likelihood measures the success of using these parameters to predict the data of the left-out participant. This process is rotated through all the participants and so can calculate the

predicted log-likelihood of the data for each participant assuming N states. The data of the all-but-one participants can always be fit better with more states because there are more parameters, but there is no guarantee that more states will predict the data of the left-out participant better. If using more states is just overfitting, the predicted log-likelihoods will not be better. As in Anderson and Fincham, we will use a sign test to determine the best number of states to see if the number of participants correctly predicted is more than would be expected by chance. An N -state model is justified if it fits significantly more participants than any model with fewer states. More generally, a model with more parameters is to be preferred over a model with fewer parameters if it fits significantly more participants.

Fig. 5 displays the result of a search for the optimal number of states in the current data set, plotting separately the results using just the image data, just the mouse actions, or both. It plots average gain per participant in log-likelihood relative to a 1-state model for each number of states. In the case of using both sources of information, 15 states results in highest log-likelihood gain. Nine states is the closest competitor. Fifteen states results in better predictions than 9 states for 29 of the 39 participants ($p < .005$) with a mean log-likelihood gain of 17.08.⁷

The best number of states is 7 when only using activations and 12 when only fitting mouse actions. The 15-state solution, by using the constraints of both data sources, better predicts either data source than the solutions that just focus on one data source. That solution

⁷ A log-likelihood difference of 17 means the participant’s data is more than 20,000,000 times more likely with the 15-state solution.

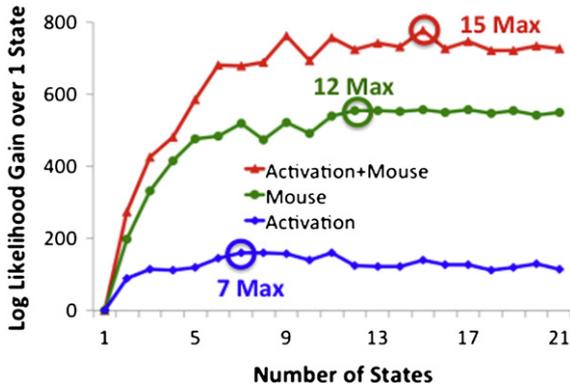


Fig. 5. Mean log-likelihood gain (per participant) over 1 state for models involving different numbers of states. Plotted separately are the results using just fMRI activation, just mouse gestures, or both. The state with the maximal log-likelihood is indicated for each state.

better predicts the activation patterns for 27 of the 39 participants than the 7-state model just fit to activations, with a mean log-likelihood gain of 13.52. Similarly, it better predicts the mousing patterns for 22 of the 39 participants than the best-fitting 12-state solution for mouse gestures only, with a mean log-likelihood gain of 3.40. Thus, either data source can be predicted better by taking advantage of the constraints provided by the other data source. As we will show, the 15-state solution has a quite interpretable structure. The structures of the 7-state activation solution or the 12-state mouse solutions are less interpretable. This is an example of co-training where the two data sources are better than either alone.

As indicated by the height of the curve for the mouse-actions-only relative to the curve for the activation-only, the mouse actions provide a very regular signal that is captured by the state analysis. Fig. 6 displays the probability of mouse actions in various states and reveals that there are three adjacent states (6–8) that are involved in moving into the first tile, entering the value, and exiting the tile. The final three states (13–15) represent the same sequence for the second tile. While the mousing data is noisy and these are not the only states where mouse actions occur, there is enough consistency from the mousing data that these 6 states anchor the rest of the analysis.

Determining the number of brain signatures

The use of LOOCV enables us to find the least complex model that captures the systematic variance common across subjects in the data.

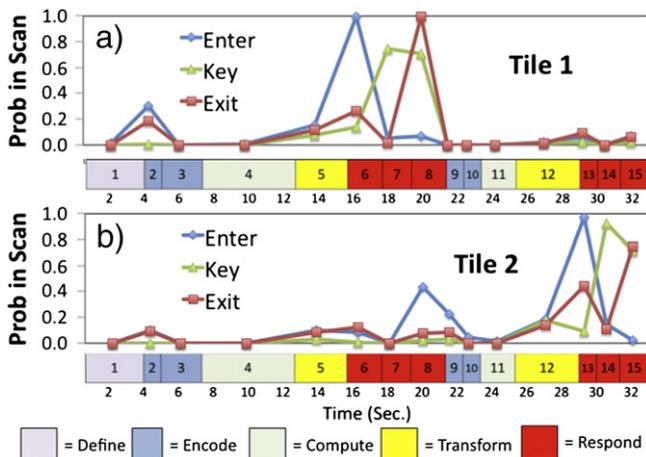


Fig. 6. (a) Probability of a mouse action in Tile 1 during a scan for each state. (b) Probability of a mouse action in Tile 2 during a scan for each state. The length of the boxes illustrates the mean duration of each state. The boxes are color coded to indicate common activation patterns.

Fig. 5 shows that the best model is a 15-state model. It is important to note that each state in the discovered model is defined by a conjunction of features: the temporal, mousing and brain signatures discussed earlier. A priori, there is no reason to expect that each feature type will be unique in each state. This leaves open the possibility of further constraining the discovered model by identifying and exploiting commonalities within feature types across states. Our focus here is simplifying the 15-state model by minimizing the number of distinct brain signatures so that we may gain a better understanding of the problem solving. Fig. 7 summarizes two critical steps along the search we carried out toward generating a simpler 15-state model:

- It seemed reasonable that the brain signatures for the two sequences of 3 response states would be the same since the mousing patterns were similar except for the tile involved. Beyond just these response states we found that the preceding 4 states had similar brain signatures suggesting that participants were going through similar processes for each tile before they began mousing. This led to the *Wrap* model where the brain signatures for states 2–8 repeated for states 9–15. To determine whether we were losing any predictable variance in reducing the 15 brain signatures to 8, we compared the original *Full* model to the *Wrap* model in a LOOCV. Since the *Full* model is more complex we can apply the same LOOCV logic and see whether its complexity is justified. The *Full* model was not significantly better than the *Wrap* model, only predicting 21 of the 39 participants better. Thus the *Wrap* model is to be preferred in the search to find the least complex model that captures the systematic variance in the data.
- Inspecting the correlation between brain signatures in the *Wrap* model, we found that the brain signatures in the response states (6, 7, 8, 13, 14, 15) all showed high inter-correlation (mean inter-correlation .75) and that there was also a substantial inter-correlation (.6) between the signatures in states 2, 3, 9 and 10. Therefore, we constructed a *Final* model in which states 6–8 and 13–15 had the same brain signature and states 2, 3, 9 and 10 had another common brain signature. LOOCV comparison with either of the more complex models indicated that their complexity did not result in better predictions and so it is to be preferred to the *Wrap* model and the *Full* model.

While there may be some other collapsing of the original 15 brain signatures that is better than the *Final* model, there is no further collapsing of these 5 signatures that yields a simpler model that would not be rejected in LOOCV.⁸ Therefore, we will use this model in further discussion as the simplest known set of brain signatures to capture the activity in the task. To summarize, this *Final* model is a 15-state model that has been constrained to have 5 unique brain signatures. As we will explain shortly, we have labeled these five signatures (and consequently their associated states) as the “phases” Define, Encode, Compute, Transform, and Respond.

Fig. 6 illustrates this *Final* model, with the states represented by boxes, color coded to indicate those that shared a common brain signature. The lengths of the boxes reflect the mean durations of the states Fig. 8 shows examples of the state occupancies using our 5-phase naming convention on three trials (these three trials had equal duration of 40 s or 20 scans). Example (a) is the most prototypical where there is an initial rise of the Define phase followed by two passes of the Encode, Compute, Transform, and Respond phases as the participant fills in each of the tiles. For extreme contrast, example (c) shows a pattern where many of the states in the second pass are skipped and the participant goes almost immediately on to filling the second tile. As these examples

⁸ The number of ways to collapse n signatures into fewer is the Bell exponential number (Brualdi, 2004) which is approximated by $(n + 2)! / (3 * 2^n)$ and grows exponentially with n, the number of signatures. Thus, it was feasible to search all 52 collapsings of the 5-signature *Final* model, but not all 1,382,958,545 collapsings of the original 15-signature *Full* model.

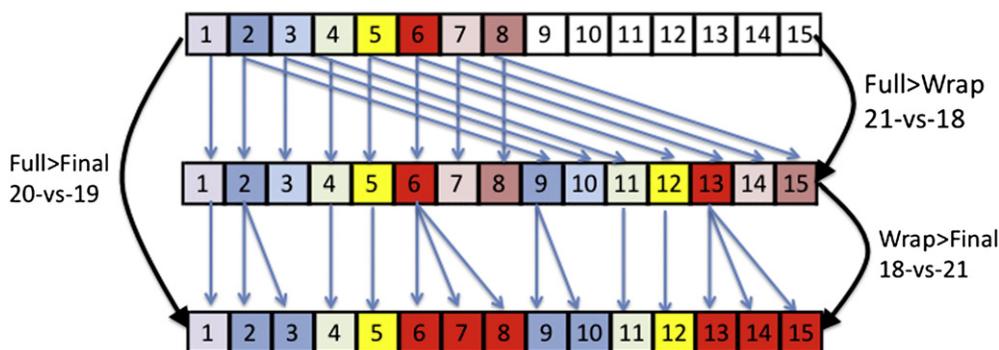


Fig. 7. An illustration of the collapsing of brain signatures in the identification of a simpler description of the 15-state Full model: In the Wrap model the brain signatures for states 2–8 repeat for states 9–15 reducing the number of brain signature parameters from $15 \times 20 = 300$ to $8 \times 20 = 160$. In the Final model adjacent states 2&3, 6–8, 9&10, and 13–15 share brain signatures reducing the number of parameters from 160 to $5 \times 20 = 100$.

illustrate, each trial involves the 5 phases defined by the 5 signatures, but that these phases break up differently into states on individual trials. Table 2 gives the inter-correlations among the 5 brain signatures.

The brain signatures associated with the phases define complex multivariate patterns of activation over the brain. While these patterns capture significant predictable variance across participants it need not follow that a particular state show increased activation in any state using a univariate test. Nonetheless, to help understand what these patterns are, we investigated what regions would show significant differences in activation across states. A whole brain analysis was conducted to determine which regions were active for each of the five phases. The data were modeled using a general linear model (GLM). The design matrix consisted of 6 model regressors and a baseline model of an order-4 polynomial to account for general signal drift. The 6 model regressors corresponded to the 5 state occupancy probabilities over all trials (e.g., Fig. 8) and the feedback period for each trial. The design matrix regressors were constructed by convolving the state occupancy probabilities and the feedback boxcar function with the

standard SPM hemodynamic response function (Friston et al., 2011). The GLM yielded 6 voxel-wise betas for each participant. For each of the 5 phases of interest, group level statistical maps were created by performing a voxel-wise *t*-test (2-tailed) that the corresponding mean beta weight across all participants was significantly different from zero. Voxels were thresholded at the significance level $p < 0.000001$ uncorrected. Figs. 9a–e shows those voxels that are significant in each of the 5 phases. Warm colors reflect positive activity and cool colors reflect deactivation. As can be seen there are regions that both activate and deactivate in the different phases.

Reflecting the overall correlation in brain activation, there is considerable overlap of regions among phases, but there are differences. Fig. 9f provides one illustration of these differences and how they contribute to the brain signatures associated with the phases. Of those regions that are significantly positive ($p < 0.000001$) over the entire problem-solving period, Fig. 9f notes which areas were more active in which phases.

As we will discuss, sometimes these activation differences result in significant univariate contrasts and sometimes they do not (Jimura and Poldrack, 2012). Our goal here is to try to illustrate what might be contributing to the multivariate patterns that are driving the ability of phases and thus their corresponding states to capture predictable variance in the image data. Nonetheless, it is of interest to know when the activity in a state is significantly greater than the activity in any other state. Fig. 9f separates those regions that meet this criterion from the other regions that are just more active.

Overall, the Phase 1 is most active followed by Phases 2 and 5, followed by Phases 3 and 4. Considering these overall activation differences, the coloring in Fig. 9f uses the following conventions: the blue regions are significantly more active ($p < .01$) in Phase 1 than any other phase. Of the remaining regions, those in green are more active in Phase 2 than Phase 1 (the darker green significantly so) and significantly more active in Phase 2 than any of Phase 3–5. Similarly, those regions in yellow are more active in Phase 5 than Phase 1 and significantly more active than Phases 2, 3 or 4. Those in red are more active in Phase 3 than any other, although most are not significantly so. The orange regions similarly are more active in Phase 4 than any other region but there are virtually no such regions. However, a phase specified by being less than other regions is just as well defined in an MVPA as a phase defined by being more active.

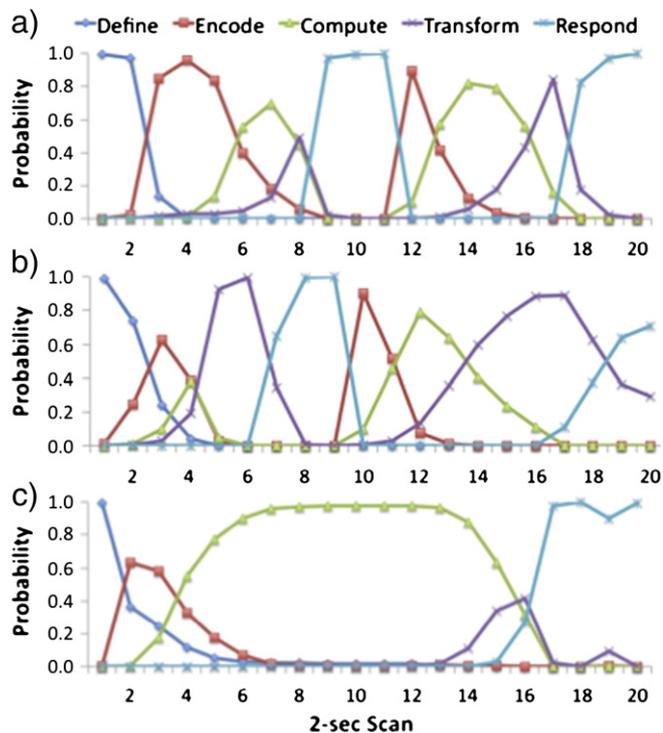


Fig. 8. Example trials showing the probability of scans being occupied by states with different brain signatures.

Table 2
Correlations among brain signatures.

	Define	Encode	Compute	Transform
Encode	0.291			
Compute	−0.253	0.305		
Transform	−0.737	−0.384	0.094	
Respond	−0.265	−0.524	−0.525	0.212

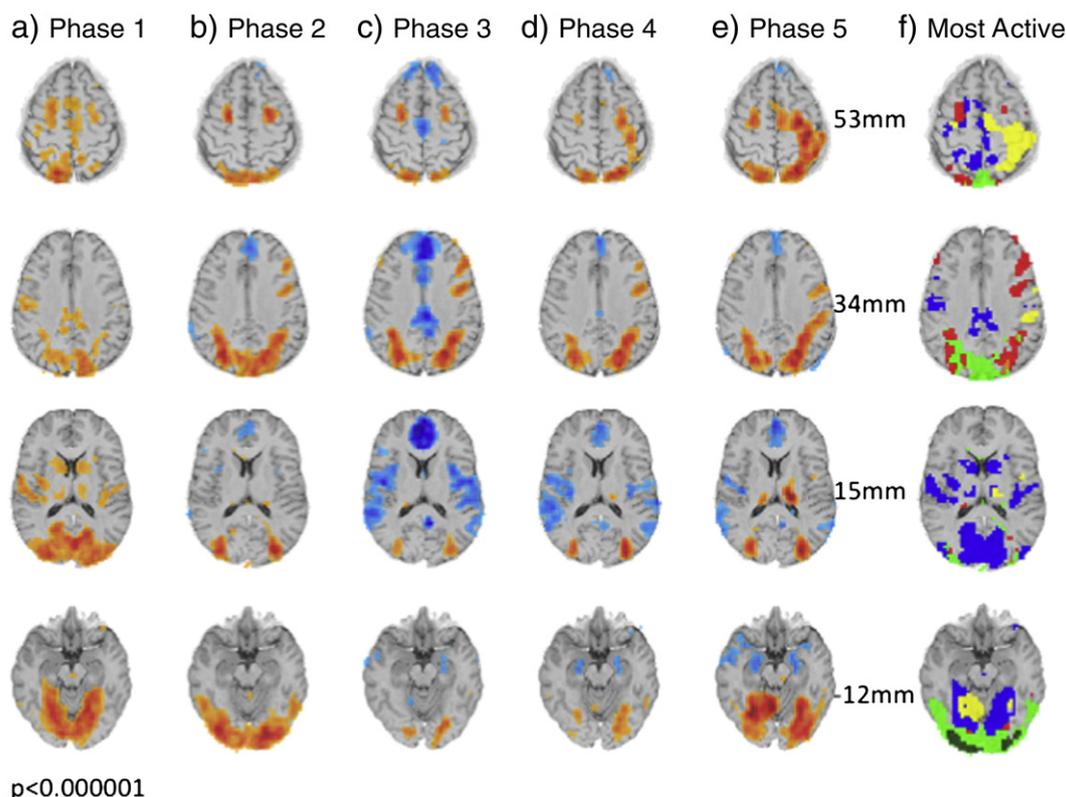


Fig. 9. (a)–(e) Regions showing significant ($p < .000001$) activation or deactivation in each of the phases. (f) Most active phases for the positive regions: blue denote Phase 1 most active, green Phase 2, red Phase 3, orange Phase 4, and yellow phase 5. The blue phase 1 regions are significantly more active than any other state. The dark green Phase 2 regions are similarly more significant than any other state. The other regions displayed in part (f) do not meet this statistical criterion. (radiological convention: image left = participant's right).

The five phases

Considering these activation patterns, the mousing patterns (Fig. 6), and the logic of the task, we suggest the following identification of the phases.

Define phase

The first phase is the only phase to show no significant negatively responding regions. The blue regions in Fig. 9f are not only more active than other states, but also significantly so. Among the positive regions in this phase are bilateral lingual gyrus, precuneus and caudate nucleus, typically involved in orienting attention, gaze and visual search (Fairhall et al., 2009; Fan et al., 2005; Hillen et al., 2013; Mayer et al., 2004). Other positive regions are the posterior cingulate and hippocampal regions that are often considered default mode regions (Buckner et al., 2008). We think that during this state participants are searching for the loop that they will have to eliminate (i.e., a set of boxes highlighted on the left graphs in Fig. 2). They also need to identify any special requirements of the resulting structure on the right. Thus, in this phase they are defining the problem that they will have to solve.

Encode phase

Among the regions that show highest activation in this phase are fusiform areas that are associated with recognition of fine visual detail. We take this as evidence that this is the phase during which the participants are encoding the numbers and operators to produce their answer. Much of the visual area is significantly more active in the Encode Phase than the Define phase—one of the few regions to be significantly greater than the Define Phase in any other phase.

Compute phase

The third phase involves parietal and prefrontal regions that are typically active during studies that involve arithmetic and algebraic computation (e.g., Anderson, 2005; Arsalidou and Taylor, 2011; Dehaene et al., 1999; Menon et al., 2000). They include the horizontal portion of the intraparietal sulcus associated with the representation of quantity and left prefrontal regions associated with retrieval of semantic information. This suggests that during this period participants are computing the values of the constant term and the coefficient term.

Transform phase

The fourth phase involves a combination of the regions that are active in the third and fifth phase, but virtually no region that is more active than both of these phases. Despite this overlap, Table 2 shows that the 20 factors that define its brain signature are either uncorrelated or negatively correlated with the other brain signatures. We think that this is the when participants are performing any structural transformations of the answer. As Fig. 6 reveals, occasionally participants perform mousing actions in this phase.

Respond phase

The fifth phase is primarily when the participant is entering the answer and shows greatest activation in regions in the left hemisphere regions associated with controlling the right hand.

Comparison with Anderson and Fincham

Another way of understanding the phases of this experiment is to compare them with the problem-solving phases obtained in Anderson and Fincham (2013). To briefly review that study, participants saw

visually simple equations involving a new operator. The challenge was that these equations had novel features that required participants develop a new plan for solving them. Anderson & Fincham identified a sequence of 4 states, which involved encoding the problem, planning a solution strategy, performing the arithmetic calculations, and then outputting the answer. They named these states Encoding, Planning, Solving, and Responding. These states were defined with respect to activation patterns in the same 290 regions as this experiment and so we can compare their states with the current phases. Because the brain signatures are complex multivariate patterns, we facilitate the comparisons by projecting each of these 290 activations down to three dimensions.

Fig. 10 provides an illustration of this subspace and the dimensions that define it. We have set the origin (0, 0, 0) to be the activity pattern in the Define phase in this experiment and set the axes to be informative. The activity pattern in the other states or phases can be created by adding to this pattern the activity patterns of the three vectors weighted by the coordinates of their points (The resulting images correlate an average .989 with the images that are directly reconstructed from the brain signatures.). The three vectors all have mean negative activity, meaning that they decrease the activity that is at the Define origin, but they do add positive activity in specific regions. The Visual Vector that defines the vertical dimension adds positive activity in areas associated with vision and visual attention. This vector is weighted positively for points in the current experiment and negatively for Anderson & Fincham, reflecting the fact that the current experiment is more demanding visually (simple equations vs. data flow graphs). The left graph in Fig. 10 shows the positions of the points in three dimensions and on the right, to facilitate comparison between the experiments, in two dimensions (x and y axes). As the x-coordinate increases, the Right-Hand Vector is weighted, which has activation in

the left motor region that controls the responding hand. As the y-coordinate increases, the Calculate Vector is weighted which adds activation in left prefrontal regions typically found active in mathematical problem solving (e.g., Anderson, 2005).

The similarity among states and phases can be seen in the 2-dimensional graph in the top right of Fig. 10. The Encode phase of this experiment is close to the Encoding state of Anderson and Fincham in the x and y dimensions. Both points are also closest to the Define origin, reinforcing our interpretation of these three states. Similarly, the Respond phase from this experiment is closest to the Responding state in Anderson & Fincham. The Compute and Transform phases in this experiment are closest to the Planning and Solving states of Anderson & Fincham. The points that involve more routine calculations (Compute and Solving) are higher on the Calculate Vector. The points that are closer to emitting a response (Solving and Transform) are higher on the Right-Hand Vector. In summary, the brain signatures in this experiment are capturing similar dimensions of activity as in Anderson and Fincham (2013).

Predicting mousing actions as ground truth

This phase-state model of the problem solving was obtained by maximizing the predictability of the data, where that predictability was measured in terms of the log-likelihood of the nth participant's data given parameters estimated from n - 1. While this is a well-justified statistical criterion, it is an "internal" measure of the regularity of the data. It would be useful to validate the phase-state solution on its ability to predict some "external" ground truth. For this purpose we used the 6 mousing actions—could we predict when participants would issue specific mousing actions from their brain imaging data?

To have a reference point we trained LDA classifiers to predict mousing actions from the 20 PCA activation values. Again we used LOOCV—

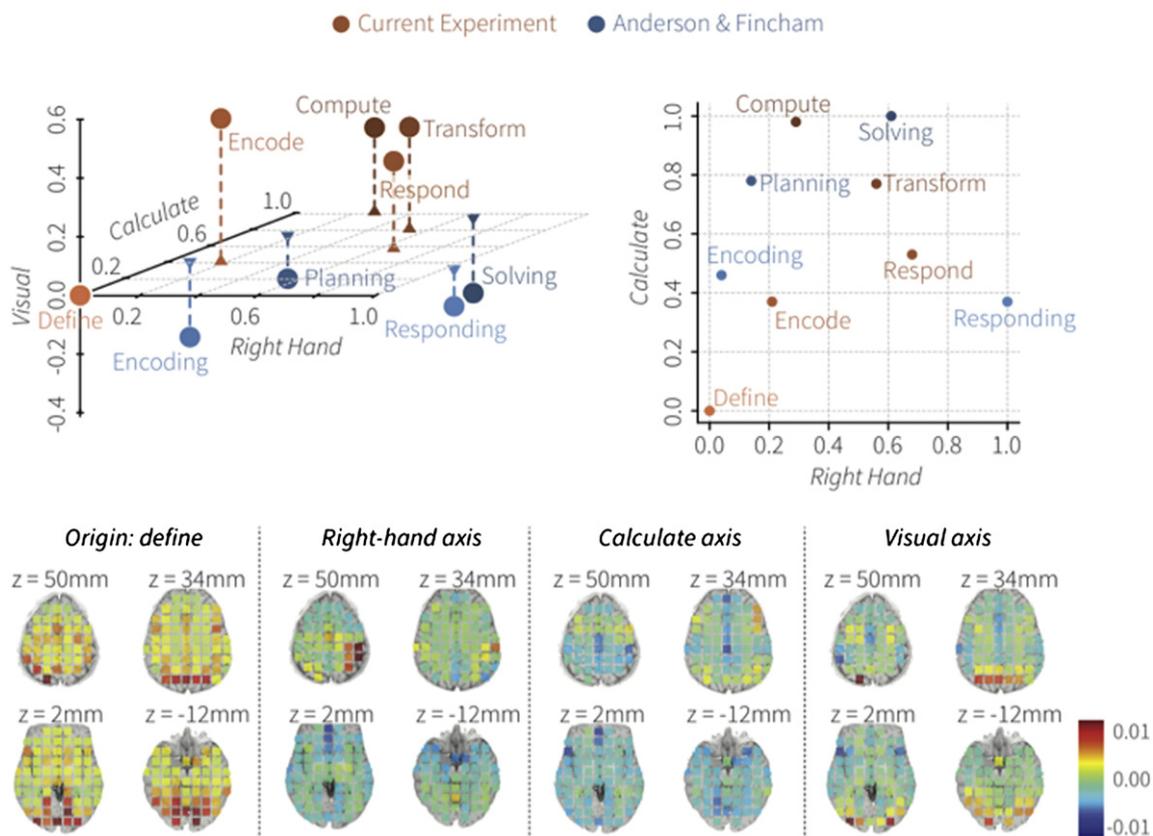


Fig. 10. Left top: Three-dimensional Comparison of the Phases of this experiment with the states of Anderson and Fincham (2013). Right top: Two-dimensional projection. Bottom: Four representative slices are shown to illustrate the origin and vectors defining the major dimensions of difference. The z coordinates for a brain slice (radiological convention: image left = participant's right) is at x = y = 0 in Talairach coordinates.

Table 3
Prediction of mousing data using various predictors.

			Tile 1			Tile 2			Average
			Enter	Key	Exit	Enter	Key	Exit	
<i>(a) d-primes</i>									
Activations	Brain signatures	Gamma	0.86	1.15	0.80	0.88	1.45	1.02	1.00
		Uniform	1.24	1.67	1.23	1.14	2.18	1.53	1.44
State occupancies	Null signatures	Gamma	1.08	1.45	1.08	1.10	2.18	1.46	1.33
		Uniform	0.59	0.84	0.69	1.10	2.06	1.52	1.03
			0.65	1.06	0.71	1.08	2.09	1.43	1.06
<i>(b) Accuracy</i>									
Activations	Brain signatures	Gamma	0.64	0.70	0.65	0.65	0.74	0.69	0.68
		Uniform	0.82	0.85	0.82	0.76	0.91	0.89	0.84
State occupancies	Null signatures	Gamma	0.77	0.82	0.77	0.74	0.89	0.87	0.81
		Uniform	0.67	0.69	0.68	0.77	0.91	0.90	0.77
			0.58	0.58	0.56	0.73	0.90	0.87	0.70
<i>(c) Hits</i>									
Activations	Brain signatures	Gamma	0.64	0.70	0.65	0.65	0.74	0.69	0.68
		Uniform	0.85	0.87	0.85	0.78	0.93	0.93	0.87
State occupancies	Null signatures	Gamma	0.79	0.84	0.79	0.75	0.91	0.90	0.83
		Uniform	0.69	0.70	0.69	0.79	0.93	0.95	0.79
			0.58	0.55	0.54	0.73	0.91	0.91	0.70
<i>(d) False alarms</i>									
Activations	Brain signatures	Gamma	0.37	0.31	0.38	0.36	0.27	0.33	0.34
		Uniform	0.42	0.30	0.43	0.36	0.27	0.50	0.38
State occupancies	Null signatures	Gamma	0.42	0.36	0.42	0.35	0.24	0.45	0.37
		Uniform	0.46	0.38	0.42	0.40	0.30	0.55	0.42
			0.34	0.20	0.27	0.34	0.26	0.48	0.31

we trained the classifier on all the scans from $n - 1$ participants, where each training instance consisted of the 20 activation PCA values for that scan (predictor variables) and the binary category of whether a particular mousing action occurred on that scan. Table 3 gives the percent correct and breaks this down into hits (correctly predicting a mousing action will occur during a scan) and false alarms (incorrectly predicting a mousing action will occur during a scan). To have a measure that incorporates the tradeoff between hits and false alarms, we will focus on a d-prime measure of discriminability (Wickens, 2001). This classifier averaged d-primes of about .9 for entering and exiting a tile and 1.3 for keying in the tile. These accuracies are all well above zero and highly significant—we were able to predict every mouse action for every participant with a d-prime better than 0. So, perhaps not surprising, there are signals in the brain data as to the occurrence of these motor activities.

The interesting question is whether the phase-state model is able to improve on this. We took the gamma distributions and brain signatures but not the mousing signatures from our final model and re-estimated probabilities of occupancy for the 15 states. Thus, we are using the brain signatures and gamma distributions to perform a transformation of the activation values used in the prior analysis. Critically, this transformation had no access to the mousing data that we are trying to predict or the mousing parameters; we are seeing what the benefit is of co-training with mousing data on estimation of the other parameters. This effort compresses our original 20 dimensions per scan down to 15 probabilities that each scan is in a particular state. Since these probabilities sum to 1, we just have 14 predictor variables. Again, we trained LDA classifiers. In every case the model d-primes are better predictors than the activation d-primes, with the smallest difference being for entering tile 2 where it predicts 29 of the 39 participants better ($t(38) = 3.15, p < .005$, 2-tailed for a test of the size of the difference). With respect to the overall average, it predicts 34 of the 39 participants better ($t(38) = 6.97, p < .0001$). So,

the state analysis is capturing regularity in the data relevant to this externally defined ground truth.⁹

The remaining lines in Table 3 try to separate the contribution of the gamma distributions for state durations and the brain signatures. We tried various combinations of replacing the means that defined the brain signatures with zeros (the mean overall activity for each dimension) or replacing the gamma distribution with a uniform distribution that made all durations equiprobable. Replacing the gamma distributions with a non-informative distribution led to a significant decrease in d-prime (average: $t(38) = 6.71, p < .0001$). However, replacing the brain signatures with non-informative signatures led to a much larger decrease (difference between brain-signature-only average and gamma-distribution-only average: $t(38) = 12.90, p < .0001$).

Analysis of participant behavior on individual trials

The estimation process delivers a probability that a participant is in each state for each scan (“Estimates of state occupancy” in Fig. 4 and examples in Fig. 8). We can sum these probabilities to get an estimated duration in each state during that trial. One can take these state durations and treat them as 15 dependent measures that decompose the overall trial times or one can aggregate these 15 times into the 5 times corresponding the 5 phases. Analysis of these times proves to be quite informative.

Part (a) of Fig. 11 shows the distribution of the total amount of time spent in the five phases. While individual states may be skipped, some time is usually spent in every phase. Nonetheless, trials vary considerably in how long the phases last. One might think that these phases

⁹ The weights learned by the classifiers for the 6 mousing actions correlates highly with mousing signatures (estimated probabilities of each action in each state) with a mean correlation of .905.

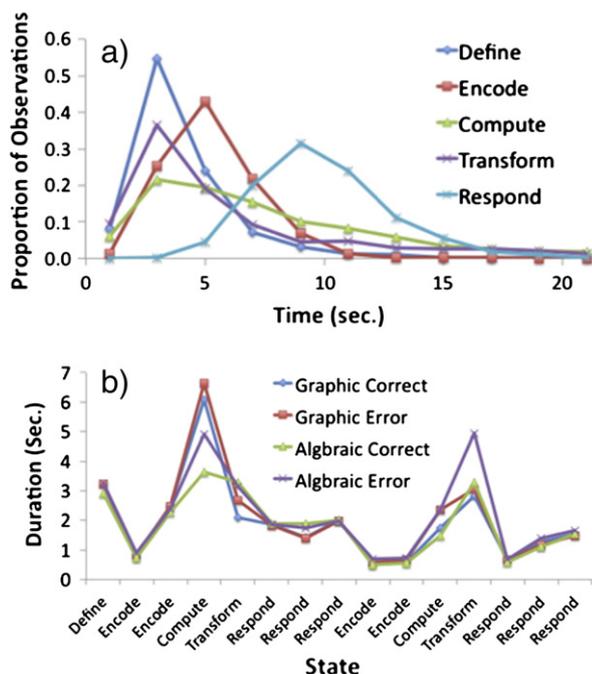


Fig. 11. (a) Distribution of total time spent in the 5 phases of a trial. Displayed are the proportions of observation in 2-second bins as a function of the mean time of that bin. (b) Mean time in states for correct and error Graphic and Algebraic trials.

would be correlated on a trial-to-trial basis with long trials having all long-phases. However, there is very little trial-to-trial correlation between the durations of the phases: The mean correlation between phase durations is .06, with a range from -0.05 to $.21$ for different pairs of phases. Given the variability in both phase duration and how the phases are divided into states, the structure of the data would be lost if we averaged trials as a function of scan. It is a major accomplishment of the semi-Markov model that it can determine the trial-by-trial realignment of the scans so that the states and phases can be identified.

Part (b) of Fig. 11 shows the time in various states for the four conditions defined by crossing the Algebraic-Graphic distinction with Correct-Error distinction. The major differences among the 4 types of trials occur in the Compute and Transform states. Therefore, we focused on these states and performed a $2 \times 2 \times 2 \times 2$ analysis of variance where the factors were (a) which tile followed the states (States 4 & 5 precede filling Tile 1 vs. States 11 & 12 precede filling Tile 2), (b) State Category (Compute vs. Transform), (c) Type of problem (Graphic vs. Algebraic), and (d) Correctness. There is not a significant effect of Problem Type ($F(1,38) = 0.09$) or State Category ($F(1,38) = 1.64$), but participants were significantly slower on errors ($F(1,38) = 32.76$; $p < .0001$) and for the states associated with the first tile ($F(1,38) = 11.48$; $p < .005$).

There were two highly significant 2-way interactions, which are summarized in Table 4. Table 4a shows the interaction between state and tile ($F(1,38) = 71.26$; $p < .0001$) where participants spend more time computing than transforming for Tile 1 but more time transforming than computing for Tile 2. This probably reflects trials where participants complete their calculation before Tile 1, but postpone struggling with the transformations until reaching Tile 2. Table 4b shows the interaction between Type and State ($F(1,38) = 40.07$; $p < .0001$) where participants spend more time computing for Graphic problems but more time on transforming for Algebraic problems. Algebraic problems pose simpler computations but require the non-trivial transformations of the computed answer. In addition to the 2-way interactions in Table 4, the only other significant effect is a 4-way interaction involving all the factors

Table 4
Mean time (s) that produce the significant 2-way interactions.

	Compute	Transform
(a)		
Tile 1	5.31	2.80
Tile 2	2.39	3.53
(b)		
Graphic	4.20	3.09
Algebraic	2.67	3.67

($F(1,38) = 8.95$, $p < .005$). This reflects the fact that Algebraic errors are distinguished by long times for the second transformation step, while correct Algebraic problems are distinguished by short times for the first computation step.

Ignoring the somewhat obscure 4-way interaction, these effects on state duration make sense and indicate a good deal about the course of the problem solving. It is worth emphasizing that this state analysis does not know about the different types of trials, but still finds effects that vary substantially with whether the problem is correct or not and whether the problem is Graphic or Algebraic. This indicates the power of the state analysis to localize important effects in a long (averaging about 30 s) episode of problem solving.

Analysis of learning across trials

Informally, it was apparent that some of the transfer problems posed no challenge to some participants and they knew how to solve them from the outset. Other problems were so confusing to some participants that they never figured out how to solve them. Finally, there was a third, most interesting, class of problems that participants mastered over the course of the experiment. To formalize this characterization, we adapted the knowledge-tracing algorithm that has been used with intelligent tutors (Corbett and Anderson, 1994). With it we were able to identify these three categories of problems and when mastery occurred for the third category.

The model assumes that each of the 8 problem types for each participant is in one of two states at any point of time: Mastered or Unmastered. There is a probability, $P(\text{prior})$, of coming into the experiment with that problem type mastered and a probability, $P(\text{master})$, of mastering an unmastered problem type on a trial with that item. Whether participants get problems correct or not is informative about whether they know how to solve a problem type or not, but not perfectly diagnostic. There is a probability $P(\text{slip})$ that participants will make errors on problem types that they have mastered. This is particularly likely because of computational slips. There is also a probability $P(\text{guess})$ that someone will be correct on an unmastered problem type. While such guessing probabilities can be quite high in some learning experiments, it is not too likely that one would get an answer correct by chance in this experiment.

This model is formally a 2-state hidden Markov model with a probability of transitioning from the Unmastered to the Mastered state and a probability of being correct in each state. Unlike the Markov model we have been using, this model describes transitions between instances of a problem type rather than between states within problems. It also uses only information about whether the problem is correct or not. The maximum-likelihood estimates of the 4 underlying probabilities are

$$\begin{aligned}
 P(\text{prior}) &= .400 \\
 P(\text{master}) &= .216 \\
 P(\text{slip}) &= .168 \\
 P(\text{guess}) &= .037
 \end{aligned}$$

In fitting this model to the data we get posterior probabilities that each trial is mastered. The learning data are sufficiently regular that most of the trials can be diagnosed as being in the Mastered or

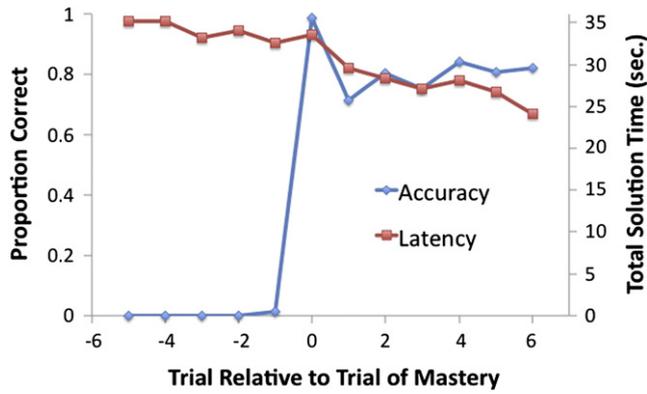


Fig. 12. Accuracy and latency on trials relative to the trial of mastery.

Unmastered states with substantial confidence.¹⁰ The high confidence of these classifications has two consequences for use of the imaging data: First, the neural imaging results can offer little to improve classification. Second, on the other hand, we can be fairly confident in these behaviorally based classifications in pursuing their neural imaging correlates.

Fig. 12 shows accuracy and solution time relative to the trial of Mastery (defined as when estimated mastery probability exceeds .5). Accuracy is very high (99.3%) on the trial of mastery because a correct response tends to cinch a mastery diagnosis. What is of interest is accuracy before and after this trial. It can be seen that there is no trend for accuracy to increase before the mastery trial but accuracy does increase afterwards. Latency shows a fairly constant decrease (.63 s per trial before mastery and .88 s per trial after mastery). Thus, participants show a continuing trend to get faster at the task but the trial of mastery marks a qualitative change in accuracy.

There are 312 learning case histories (39 participants * 8 types). Of these 103 are labeled as mastered on the first trial and may reflect problems that the participant already knew how to solve. There are an additional 40 problems that are never labeled as mastered. Finally and of most interest, there are 169 problems that are mastered over the course of the experiment. As a reflection of the greater difficulty of Algebraic problems, only 28% of those that start out mastered are Algebraic, 54% of those that are mastered over the course of the experiment are Algebraic, and 85% of those that are never mastered are Algebraic. Of the 2319 non-time-out trials, 284 involve problem types that are never mastered, 1228 involve problem types that are mastered over the course of the experiment, and 807 involve problem types that start out mastered. Our focus is on the 1228 cases that are mastered over the course of the experiment. All participants contribute to this category. Taking a posterior probability of .5 as the threshold for mastery, we can classify these 1228 cases into three categories:

1. Pre-Mastery (361 trials): Trials with probability less than .5
2. Mastery (169 trials): First trial where probability exceeds .5
3. Post-Mastery (698 trials): Trials after the Mastery trial.

Every case will have one or more Pre-Mastery trials, one Mastery trial, and, unless it was mastered in the last block one or more Post-Mastery trials. The Markov model's estimated mastery probability was .119 for Pre-Mastery Trials, .914 for Mastery Trials, and .978 for Post-Mastery Trials.

Fig. 13 shows the duration of the 5 problem-solving phases in terms of position relative to the mastery point. An analysis of variance reveals significant effects of phase ($F(4,152) = 46.55, p < .0001$), position relative to mastery ($F(2,76) 23.64, p < .0001$), and a significant interaction between the two ($F(8,304) = 4.55, p < .0001$). To further analyze the effects of

¹⁰ 18.4% of the trials involve cases whose probability of being mastered on that trial is less than .05, another 9.6% are between .05 and .25, only 4.5% are in the grey zone between .25 and .75, 5.3% are between .75 and .95, and 60.2% have mastery probabilities greater than .95.

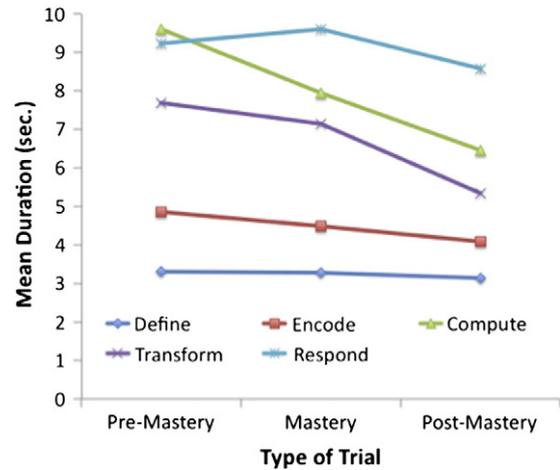


Fig. 13. Mean time in the problem-solving phases as a function of position relative to mastery.

position and their interaction, the durations of each phase at the three mastery positions were summarized with two orthogonal contrasts—a linear trend (Pre-Mastery minus Post-Mastery) and a quadratic component (Mastery Trial minus average of Pre and Post Mastery). All phases but the Define Phase shows significant drops (linear trends) from Pre-Mastery to Post-Mastery trials (Define: .17 s, $t(38) = 1.61, p > .10$; Encode: .77 s, $t(38) = 4.56, p < .0001$; Compute: 3.16 s, $t(38) = 3.98, p < .0005$; Transform: 2.36 s, $t(38) = 4.42, p < .0001$; Respond: .65 s, $t(38) = 2.93, p < .01$). Further, all pair-wise comparisons among magnitudes of these drops are significant except that between Encode and Respond and that between Compute and Transform. Only the Respond Phase shows a significant quadratic component ($t(38) = 2.68, p < .05$)—there is a .39 s increase going from Pre-Mastery to Mastery trial. The Transform Phase also shows some indication of a similar deviation from linearity, but this is not significant ($t(38) = 1.15$).

In summary, the phases speed up as the problem type is mastered, with the Compute and the Transform Phases showing the greatest decrease in duration. In addition, the Respond phase shows a slow down on the trial of mastery. This slow down may indicate reflection on the correct answer that is being produced. Participants sometimes correct the answers they enter before submitting them. There was a greater tendency for participants to change their answers on Mastery trials (quadratic trend: ($t(38) = 2.09; p < .05$: 29.0% changes answers on Mastery trials versus 20.8% changed answers on Pre-Mastery and 18.5% on Post-Mastery).

We performed an exploratory analysis of the imaging data, contrasting mean activation on Pre-Mastery trials, Mastery trials, and Post-Mastery trials. Only the scans from these trials were analyzed and three separate regressors were created for the problem-solving periods of these three types of trials by convolving the boxcar functions for these periods with the standard SPM hemodynamic response function (Friston et al., 2011). Regions were identified that showed significant linear trends (Post-Mastery minus Pre-Mastery) and regions that showed significant quadratic trends (Mastery minus average of Pre and Post-Mastery).¹¹ Fig. 14 shows those regions displaying a linear trend (part a) and those displaying a quadratic trend (part b) and Table 5 provides information on these regions. All the linear trends are in the direction of increased activation on Post-Mastery trials. These regions include motor, visual, and subcortical areas that are part of the basal ganglia thalamic loop. We think the increased activation is a consequence of the faster processing as a problem type is mastered. The participants still have to perform the same motor actions and visual

¹¹ The threshold was that the regions consisted of at least 44 voxels having $p < .005$ for a brain-wise alpha of .05.

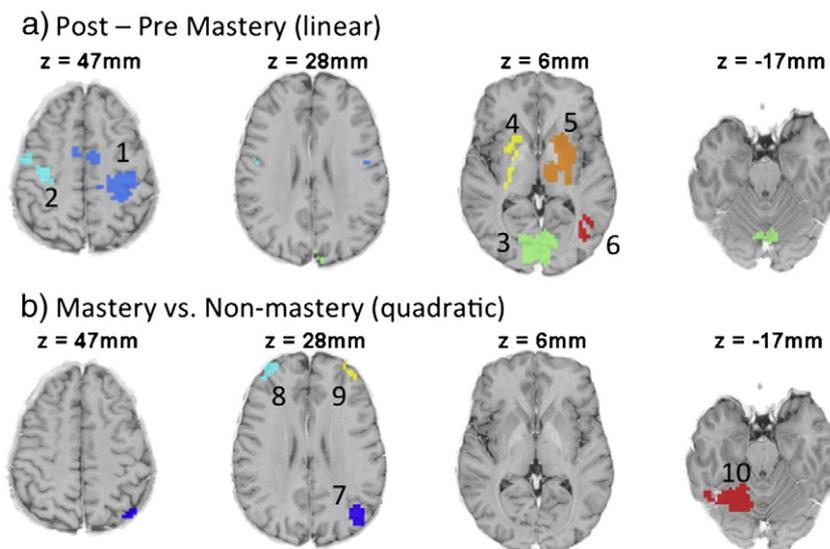


Fig. 14. Activation effects relative to point of mastery: (a) regions showing significant linear trend; (b) regions showing significant quadratic effects. Different regions are color coded and the numbering refers to entries in Table 5. The z coordinates for a brain slice (radiological convention: image left = participant's right) is at $x = y = 0$ in Talairach coordinates.

encodings but do so in less time when a problem type is mastered. (The activation estimates reflect activity per unit time.)

Significant quadratic trends are found in left and right rostrolateral prefrontal cortex (RLPFC), the left angular gyrus (AG), and visual areas overlapping the right fusiform. All these regions show increased activation on the trial of mastery. Lee et al. (in press) found that average amount of activation in the right RLPFC on the first trial predicted percent correct on later trials. In our analysis the first trial is never classified as a mastery trial and so our results add to the earlier result. The two results provide independent evidence that increased activation in the RLPFC is associated with mastery of the material.

Conclusions

By combining mouse movements and fMRI activity, the MVPA–HMM approach was able to decompose trials that took up to 60 s into 5 aggregate phases of activity. Except for the initial Define phase, each of these phases broke up into a number of states defined by the mousing behavior and whether the processing involved the first or second tile. It was not possible to discover these phases and states using only fMRI data or only mousing data. This structure became apparent only when both sources of data were combined.

Table 5
Regions showing significant trends relative to points of mastery.

Region of interest	Brodmann Area(s)	Coordinates (x,y,z)	Voxel count
<i>(a) Linear trend</i>			
1. L Pre/Post central gyrus and anterior cingulate	6/4/3/32/24	−18, −14, 54	726
2. R Pre central gyrus	6	41, −7, 45	172
3. Medial cuneus/lingual gyrus	17/18/19	0, −78, 10	550
4. R putamen/lentiform nucleus/thalamus		22, −4, 10	109
5. L putamen/lentiform nucleus/thalamus		−20, −8, 8	346
6. L middle/inferior temporal/occipital	37/19	−42, −61, 8	82
<i>(b) Quadratic trend</i>			
7. L angular gyrus/inferior parietal lobule	7/39	−41, −64, 43	131
8. R superior/middle frontal gyrus	10	33, 53, 15	77
9. L superior/middle frontal gyrus	10	−34, 53, 16	49
10. R fusiform/inferior occipital gyrus	37/19/18	30, −64, −14	288

Solving these problems involves a complex mixture of basic processes including perceptual, representational, memory, motor, and control processes. Many of these processes are in play at any point in the experiment but a phase reflects a period of time when the amounts of the different processes are constant. Reflecting the fact that the phases involve overlapping processes, the activation patterns in Figs. 9a–e are rather similar. However, the HMM–MVPA analysis finds independent dimensions of variation produced by increases and decreases in regional activation for a phase (illustrated in Fig. 9f). These increases and decreases reflect differences in the proportion of the different processes in the mixture for that phase.

The existence of the initial Define Phase was a surprise and has not shown up in prior research on mathematical problem solving. Many of these brain areas active in this phase are involved in orienting of attention, gaze and visual search and often these same areas are engaged throughout problem solving. The default mode network regions that activated in this phase do not usually co-activate with task-positive regions (however, see Gerlach et al., 2011; Spreng et al., 2010). The overall pattern of activation suggests that during this phase participants are taking in the complex diagram that has just appeared and determining where the difficulty in the solution will be. We suspect it has not appeared in other studies of mathematical problem solving because the structure of the problems were predictable and occurred over and over again across trials.

The Encode and Respond Phases involved predictable increases in visual and motor regions. The Encode Phase had an optional first state where participants moved the mouse to the tile where they were going to enter the answer. Participants also had the option of encoding all the material before calculating the value for the first tile or postponing some encoding until it was needed for the second tile. Because of these variations, it would have been particularly hard to identify the periods of time where participants are encoding without the MVPA–HMM approach.

The ordering of the 6 Respond states is required by the task—by definition the first tile is dealt with before the second, and each tile must first be entered, then the answer keyed, and the tile exited. However, 78% of the trials involved some exception to this minimal mousing pattern. For instance, 15% of the trials involved entering Tile 2 before deciding to enter the first answer into Tile 1,¹² 46% involved repeated visits to Tile 1, and 45% repeated visits to Tile 2. 19% of trials involved changing

¹² Recall that Tile 1 is defined as the first tile into which participants key an answer.

an answer after initially keying it. Behavioral variations like these make it difficult to interpret complex problem solving, but the HMM–MVPA approach can deal with them.

The Compute and Transform Phases tended to activate the same cognitive regions, with the Transform phase showing somewhat less activation in regions associated with calculating and more activity in regions associated with motor activity (participants are moving the mouse in the Transform phase). Participants varied from trial to trial in how much of these phases they chose to complete before entering a value into Tile 1. These were the phases that showed the differences between Graphic and Algebraic problems. The Graphic problems had a longer Compute Phase while Algebraic problems had a longer Transform Phase (Table 4b). Also the durations of these phases were sensitive to problem correctness.

Participants started out not knowing how to solve 67% of the transfer problem types, but over the course of the experiment came to understand how to solve over 80% of these initially unmastered problems types. As they repeatedly saw a problem type, they sped up in most of the phases. However, the specific trial where they mastered a problem type was distinguished by a slowdown in later phases of the problem solution.

The RLPFC and AG seem the most critical regions for purposes of predicting mastery of these challenging problems. Activation in these regions increased on mastery trials. Lee et al. (in press) found increased activity in the RLPFC on the first trial was predictive of future success. The RLPFC and AG have been found to be involved in reflection in another mathematical problem solving task (Anderson et al., 2011; Wintermute et al., 2012). Anderson and Fincham (2013) found the RLPFC and AG most active when planning a solution. Considerable other research relates the RLPFC to reflective functions. For instance, the RLPFC is engaged upon feedback in episodic memory experiments (e.g., Reynolds et al., 2006), when considering intentions in prospective memory (e.g., Benoit et al., 2012), when reasoning about higher-order relationships and analogies (e.g., Volle et al., 2010; Wendelken et al., 2008), and when reflecting on task performance (Fleming et al., 2012).

While it is not surprising that the RLPFC is engaged by challenging mathematical problems, the engagement of the AG is somewhat unexpected. One prominent hypothesis is that the AG is involved in retrieval of arithmetic facts (Dehaene et al., 2003). However, the angular gyrus often deactivates during difficult mental arithmetic and has been related to default mode activity (Grabner et al., 2013). Outside of the mathematics domain, the AG has most often been associated with higher-level language comprehension, such as understanding of metaphors (e.g., Desai et al., 2011; Rapp et al., 2012). It seems more likely that its engagement in challenging mathematical problem solving is related to its comprehension function than retrieval of basic arithmetic facts.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100109 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Correspondence may be directed to John Anderson, Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213 (ja+@cmu.edu). We would like to thank Jelmer Borst, Julie Fiez, Aryn Pyke, and Katerina Velanova for their comments on the paper. We would also like to thank Jelmer Borst for his help in constructing the illustrations in Fig. 10.

Appendix A

As in Anderson and Fincham (2013), we assume that participants go through N states in solving one of these problems. Each state reflects a period of constant brain activity and (new to this analysis) constant mousing behavior. Each state can vary in its length. A discrete version of a gamma distribution is used to characterize the probability that a

state will last a particular number of scans. Given that each scan is 2 s in length, the probability of spending exactly m scans in state i is given as:

$$G(m|v_i, a_i) = \int_{2m-1}^{2m+1} \text{gamma}(t|v_i, a_i) dt$$

where v_i and a_i are the shape and scale parameters of the gamma distribution for the i th state. This allows for cases where a state lasts 0 scans. The probability of this is the probability of a duration less than 1 s. In such cases the model skips that state and moves on to the successor state. Allowing such skipped states is critical in explaining brief trials (for instance, 3 trials out of the 2319 lasted 1 or 2 scans).

The 290 megavoxels produce 290 signal estimates for each scan. To reduce dimensionality and to deal with the non-independence among regions we perform a principal component analysis (PCA). We use the first 20 factors, a number which seems to result in the best state identification. These 20 factors are normalized to have a mean of 0 and standard deviations of 1 over all of the scans in the experiment. Each state i will be associated with a set M_i of 20 means μ_{ik} for these factors. Since the factors from the PCA are approximately distributed as independent normals, we calculate the probability of a set F_j of observed factor values f_{jk} for scan j in state i as:

$$Pr_{image}(F_j|M_i) = \prod_{k=1}^{20} \text{Normal}(f_{jk}, \mu_{ik}, 1)$$

The 20 factors means for a state are referred to as its *brain signature*.

We encoded the mouse actions into 6 categories defined by whether the action involved moving the mouse into the tile, keying a character into the tile, or exiting a tile and whether the action involved Tile 1 or Tile 2. Tile 1 is defined as the first tile into which participants key a character. We represented the mousing behavior during each scan j in terms of binary variables b_{jk} which have value 1 if there is one or more mouse actions of type k during that scan and 0 otherwise. Estimating a set of probabilities p_{ik} for a type k mouse action for state i , the probability of the observed 6 mouse actions B_j in scan in state i will have the probability

$$Pr_{mouse}(B_j|P_i) = \prod_{k=1}^6 p_{ik}^{b_{jk}} * (1-p_{ik})^{(1-b_{jk})}$$

The 6 probabilities for a state are referred to as its *mousing signature*.

We can combine the probabilities specified above to assign a probability to an *interpretation* of a trial. An interpretation of a trial of m scans as N states will involve an assignment of m_1 scans to state 1, then m_2 scans to state 2, and so on through to m_N scans to state N , such that $m = m_1 + m_2 + \dots + m_N$. The probability of such an assignment is a product of the probabilities of states of that length and the probabilities of the factors and mouse actions in those states:

$$p(m_1, m_2, \dots, m_N) = \prod_i^N G(m_i|v_i a_i) * \prod_{j=1}^{m_i} [Pr_{image}(F_j|M_i) * Pr_{mouse}(B_j|P_i)]$$

This is the probability of just one interpretation of the trial. The total probability of the trial is the sum of the probabilities of all possible ways of assigning the scans to a sequence of states. The number of possible interpretations grows rapidly as the number of scans and states increases, but the sum of these probabilities can be efficiently calculated using the dynamic programming techniques associated with hidden Markov models. The current case is semi-Markov because durations in the states are variable and so we use software derived from the explicit duration Markov implementation of Yu and Kobayashi (2003, 2006). This software uses expectation maximization to estimate a set of parameters that will maximize the probability of all the data, which is the product of the probabilities of the individual trials. Since these probabilities get very small, we will speak of log-likelihoods throughout the paper.

References

- Anderson, J.R., 2005. Human symbol manipulation within an integrated cognitive architecture. *Cogn. Sci.* 29, 313–342.
- Anderson, J.R., Fincham, J.M., 2013. Discovering the sequential structure of thought. *Cogn. Sci.* 38, 322–353.
- Anderson, J.R., Betts, S., Ferris, J.L., Fincham, J.M., 2010. Neural imaging to track mental states while using an intelligent tutoring system. *Proc. Natl. Acad. Sci. U. S. A.* 107, 7018–7023.
- Anderson, J.R., Betts, S., Ferris, J.L., Fincham, J.M., 2011. Cognitive and Metacognitive Activity in Mathematical Problem Solving: Prefrontal and Parietal Patterns. *Cognitive, Affective, and Behavioral Neuroscience* 11, 52–67.
- Anderson, J.R., Betts, S.A., Ferris, J.L., Fincham, J.M., 2012a. Tracking children's mental states while solving algebra equations. *Hum. Brain Mapp.* 33, 2650–2665.
- Anderson, J.R., Fincham, J.M., Yang, J., Schneider, D.W., 2012b. Using brain imaging to track problem solving in a complex state space. *NeuroImage* 60, 633–643.
- Ansari, D., Dhital, B., 2006. Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: an event-related functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 18, 1820–1828.
- Arsalidou, M., Taylor, M.J., 2011. Is $2 + 2 = 4$? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage* 54, 2382–2393.
- Benoit, R.G., Gilbert, S.J., Frith, C.D., Burgess, P.W., 2012. Rostral prefrontal cortex and the focus of attention in prospective memory. *Cereb. Cortex* 22, 1876–1886.
- Blum, A., Mitchell, T., 1998. Combining labeled and unlabeled data with co-training. *Proceedings of the eleventh annual conference on Computational learning theory. ACM*, pp. 92–100.
- Brualdi, R.A., 2004. *Introductory Combinatorics*, 4th edition. Pearson Prentice Hall.
- Brunstein, A., Betts, S., Anderson, J.R., 2009. Practice enables successful learning under minimal guidance. *J. Educ. Psychol.* 101, 790–802.
- Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38.
- Butterworth, B., Varma, S., Laurillard, D., 2011. Dyscalculia: from brain to education. *Science* 332, 1049–1053.
- Castelli, F., Glaser, D.E., Butterworth, B., 2006. Discrete and analogue quantity processing in the parietal lobe: a functional MRI study. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4693–4698.
- Corbett, A.T., Anderson, J.R., 1994. Knowledge tracing: modeling the acquisition of procedural knowledge. *User Model. User-Adap. Inter.* 4, 253–278.
- Cox, R., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.
- Cox, R., Hyde, J.S., 1997. Software tools for analysis and visualization of fMRI data. *NMR Biomed.* 10, 171–178.
- Danker, J., Anderson, J., 2007. The roles of prefrontal and posterior parietal cortex in algebra problem solving: a case of using cognitive modeling to inform neuroimaging data. *NeuroImage* 35, 1365–1377.
- Dehaene, S., 1997. *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press, New York.
- Dehaene, S., Spelke, E., Pinel, R., Stranescu, R., Tsivkin, S., 1999. Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974.
- Dehaene, S., Piazza, M., Pinel, P., Cohen, L., 2003. Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20 (3–6), 487–506.
- Desai, R.H., Binder, J.R., Conant, L.L., Mano, Q.R., Seidenberg, M.S., 2011. The neural career of sensory-motor metaphors. *J. Cogn. Neurosci.* 23, 2376–2386.
- Fairhall, S.L., Indovina, I., Driver, J., Macaluso, E., 2009. The brain network underlying serial visual search: comparing overt and covert spatial orienting, for activations and for effective connectivity. *Cereb. Cortex* 19 (12), 2946–2958.
- Fan, J., McCandliss, B.D., Fossella, J., Flombaum, J.I., Posner, M.I., 2005. The activation of attentional networks. *NeuroImage* 26 (2), 471–479.
- Fleming, S.M., Huijgen, J., Dolan, R.J., 2012. Prefrontal contributions to metacognition in perceptual decision making. *J. Neurosci.* 32, 6117–6125.
- Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.), 2011. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press.
- Gerlach, K.D., Spreng, R.N., Gilmore, A.W., Schacter, D.L., 2011. Solving future problems: default network and executive activity associated with goal-directed mental simulations. *NeuroImage* 55 (4), 1816–1824.
- Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* 9, 416–429.
- Grabner, R.H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., 2013. The function of the left angular gyrus in mental arithmetic: evidence from the associative confusion effect. *Hum. Brain Mapp.* 34, 1013–1024.
- Hillen, R., Günther, T., Kohlen, C., Eckers, C., van Ermingen-Marbach, M., Sass, K., Heim, S., 2013. Identifying brain systems for gaze orienting during reading: fMRI investigation of the Landolt paradigm. *Front. Hum. Neurosci.* 7.
- Jimura, K., Poldrack, R.A., 2012. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia* 50 (4), 544–552.
- Krueger, F., Spampinato, M.V., Pardini, M., Pajevic, S., Wood, J.N., Weiss, G.H., Landgraf, S., Grafman, J., 2008. Integral calculus problem solving: an fMRI investigation. *Neuroreport* 19, 1095–1099.
- Lee, H.S., Anderson, A., Betts, S., Anderson, J.R., 2011. When does provision of instruction promote learning? In: Carlson, L., Hoelscher, C., Shipley, T. (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Austin, TX, pp. 3518–3523.
- Lee, H.S., Fincham, J., Betts, S., Anderson, J.R., 2014. An fMRI investigation of instructional guidance in mathematical problem solving. *Trends Neurosci. Educ.* (in press).
- Mayer, A.R., Dorfinger, J.M., Rao, S.M., Seidenberg, M., 2004. Neural networks underlying endogenous and exogenous visual-spatial orienting. *NeuroImage* 23 (2), 534–541.
- Menon, V., Rivera, S.M., White, C.D., Glover, G.H., Reiss, A.L., 2000. Dissociating prefrontal and parietal cortex activation during arithmetic processing. *NeuroImage* 12, 357–365.
- Molko, N., Cachia, A., Riviere, D., Mangin, J.F., Bruandet, M., Le Bihan, D., Cohen, L., Dehaene, S., 2003. Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. *Neuron* 40, 847–858.
- Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430.
- Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45, S199–S209.
- 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. *Proc. Natl. Acad. Sci.* 101, 5686–5691.
- Rabiner, R.E., 1989. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286.
- Rapp, A.M., Mutschler, D.E., Erb, M., 2012. Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage* 63, 600–610.
- Reynolds, J.R., McDermott, K.B., Braver, T.S., 2006. A direct comparison of anterior prefrontal cortex involvement in episodic retrieval and integration. *Cereb. Cortex* 16, 519–528.
- Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Woolrich, M.W., 2011. Network modelling methods for FMRI. *NeuroImage* 54 (2), 875–891.
- Spreng, R.N., Stevens, W.D., Chamberlain, J.P., Gilmore, A.W., Schacter, D.L., 2010. Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage* 53 (1), 303–317.
- Volle, E., Gilbert, S.J., Benoit, R.G., Burgess, P.W., 2010. Specialization of the rostral prefrontal cortex for distinct analogy processes. *Cereb. Cortex* 20, 2647–2659.
- Wendelken, C., Nakhachenko, D., Donohue, S.E., Carter, C.S., Bunge, S.A., 2008. Brain is to thought as stomach is to ???: investigating the role of rostralateral prefrontal cortex in relational reasoning. *J. Cogn. Neurosci.* 20, 682–693.
- Wickens, T.D., 2001. *Elementary Signal Detection Theory*. OUP, USA.
- Wintermute, S., Betts, S.A., Ferris, J.L., Fincham, J.M., Anderson, J.R., 2012. Brain networks supporting execution of mathematical skills versus acquisition of new mathematical competence. *PLoS One* 7.
- Yu, S.-Z., Kobayashi, H., 2003. An efficient forward-backward algorithm for an explicit duration hidden Markov model. *IEEE Signal Proc. Lett.* 10, 11–14.
- Yu, S.-Z., Kobayashi, H., 2006. Practical implementation of an efficient forward-backward algorithm for an explicit duration hidden Markov model. *IEEE Trans. Signal Process.* 54, 1947–1951.