Using Neural Imaging to Inform the Instruction of Mathematics

John R. Anderson
Shawn Betts
Jennifer L. Ferris
Jon M. Fincham
Carnegie Mellon University

Framework for Today's Talk

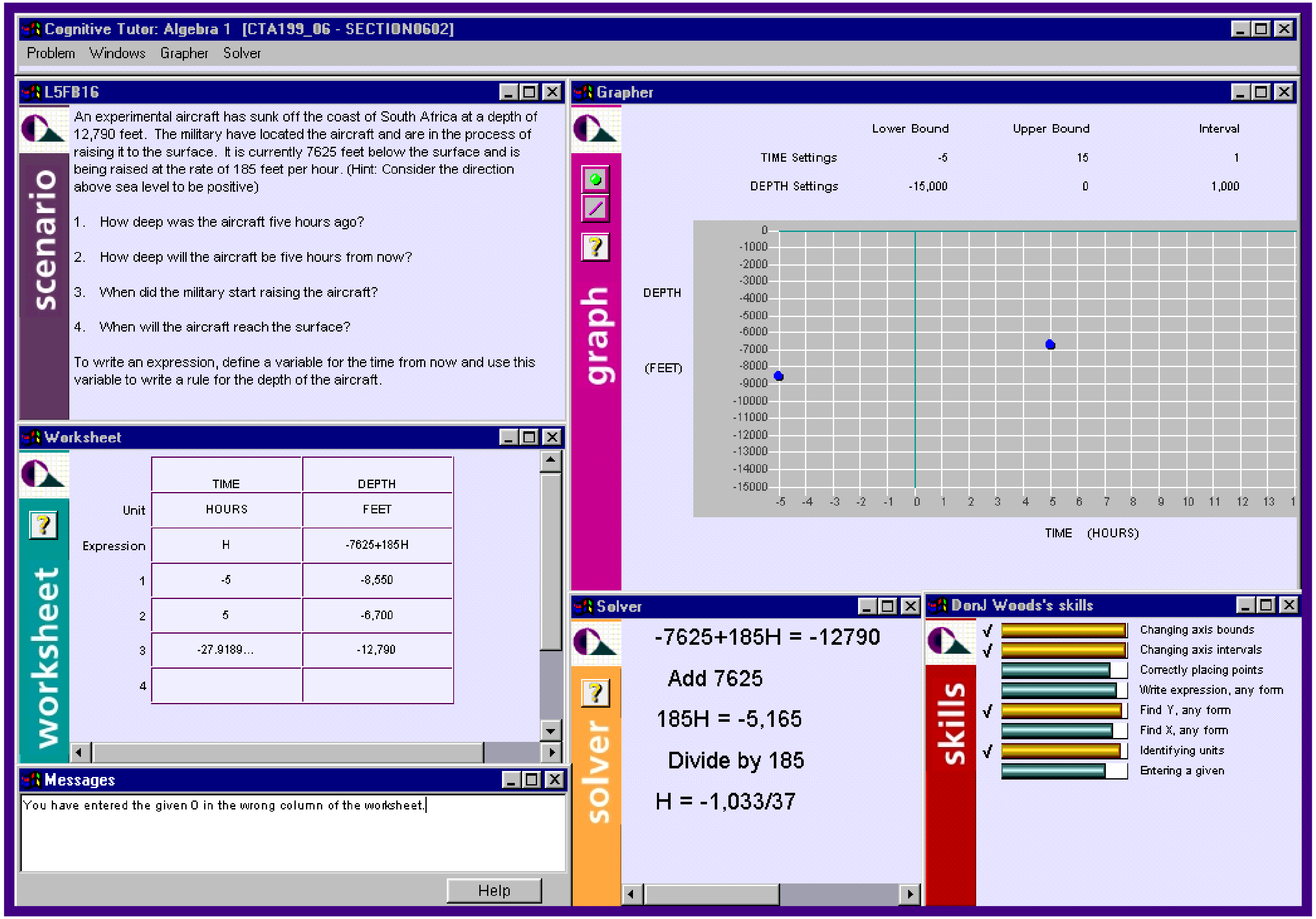
Our Research Program:

- Develop a cognitive architecture (**ACT-R**) of how people perform complex cognitive tasks.
- Within that architecture develop detailed models of how students learn mathematics.
- Build instructional systems (**Cognitive Tutors**) that are based on these models.
- Have the instructional experiments inform the cognitive architecture.

Today's Talk:

- Describe how we have brought ACT-R and fMRI brain imaging together in the context of Cognitive Tutors.

The Algebra Tutor



Currently teaches about 500,000 students in the United States

Cognitive Tutors

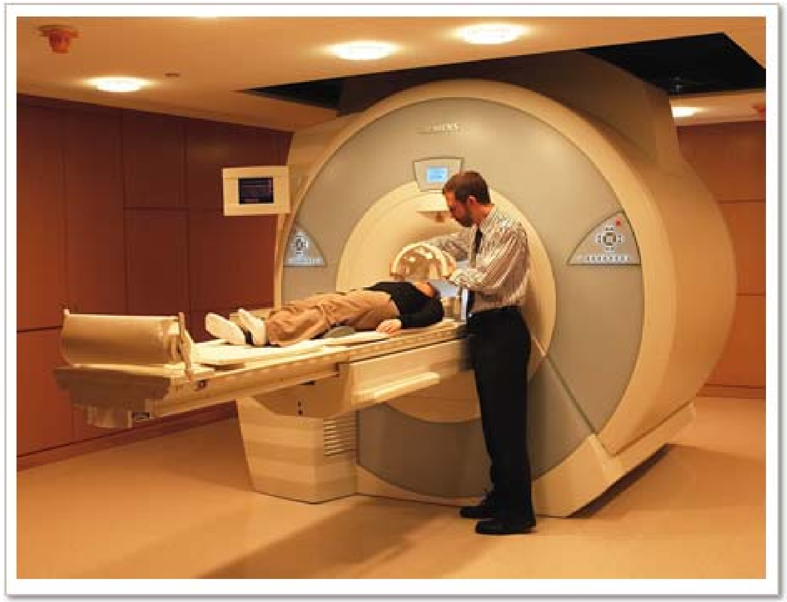
- **Cognitive Model:** A system that can solve problems in the various ways students can

$3(2x - 5) = 9$

$6x - 15 = 9$ ✓ | $2x - 5 = 3$ ✓ | $6x - 5 = 9$ ✗

- **Model Tracing:** Follows student through their individual approach a problem -> context-sensitive instruction
- **Knowledge Tracing:** Assesses student's knowledge growth -> individualized activity selection and pacing

Brain Imaging and Cognitive Tutors



- Cognitive Tutors work by using cognitive models to interpret the student's behavior.
- Cognitive Tutors are limited by the crude nature of the cognitive models and the difficulty of diagnosis using the behavioral event stream.
- One contribution of brain imaging would be to improve the sophistication of the underlying cognitive models.
- Another contribution of brain imaging would be to help diagnose **when** a student is thinking **what**.

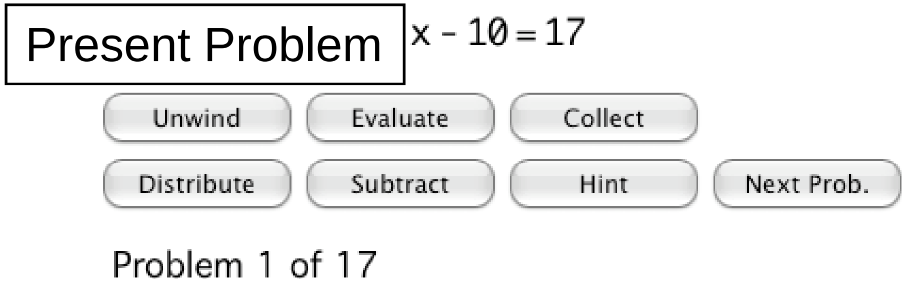
The Experimental Tutor

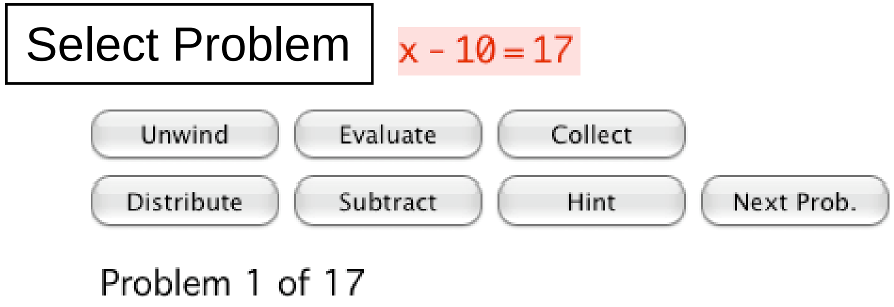
- We have developed a experimental tutoring system based on the Algebra 1 curriculum in Foerster (1990) for solving linear equations.
- The tutoring system is minimalist for the purposes of studying students in an fMRI scanner, but involves basic instruction, error feedback, and help on request.
- We have also developed a data-flow isomorph of this system which can be used with adults.
- Results are very similar for children and adults.

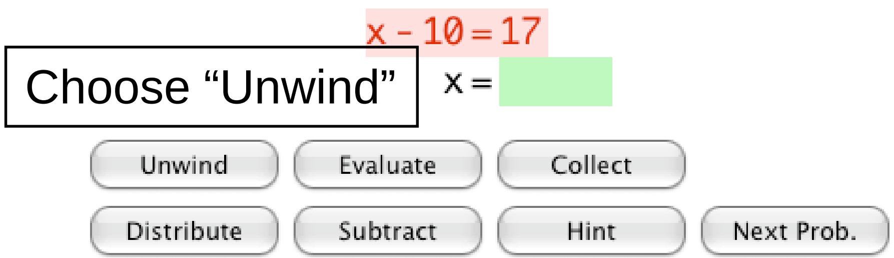
Experiment

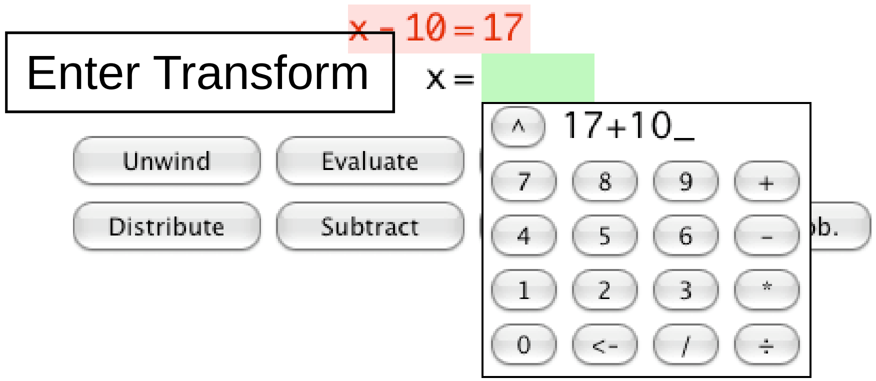
- Goal 1: Discriminate between on task and off task behavior.
- Goal 2: Identify problem student is solving and where they are in that problem.
- Students goal through the curriculum in 5 sessions on Days 0 - 4 and then do similar material on Day 5 as they did on Day 1.
- They are scanned on Days 1 and 5.
- To create off-task moments we insert periods of n-back at reasonable points of transition in the equation solving.
- Because we have detailed computer logs we have a pretty good definition of ground truth -- where they actually are.

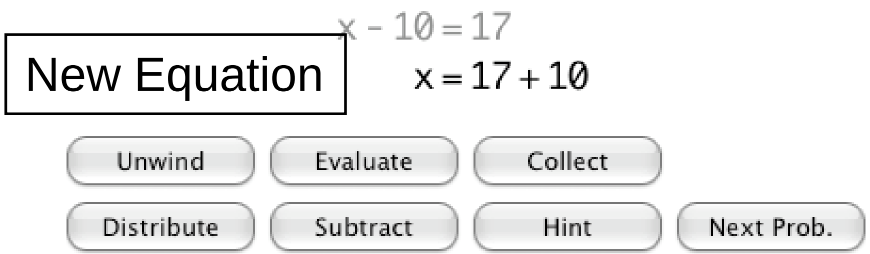
Earliest Material:Transformation Phase









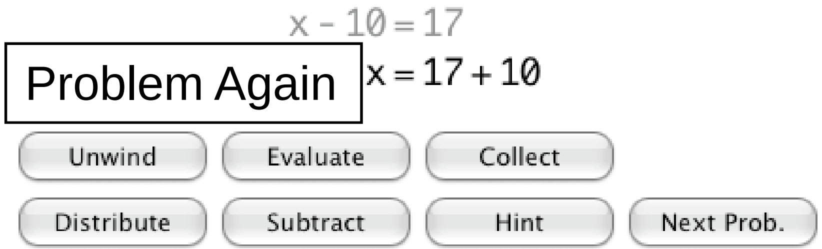


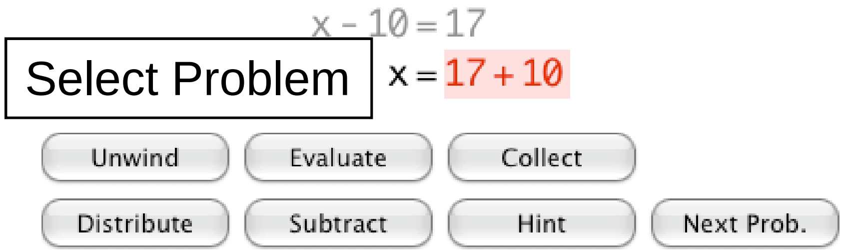
Note:
All actions are with a mouse

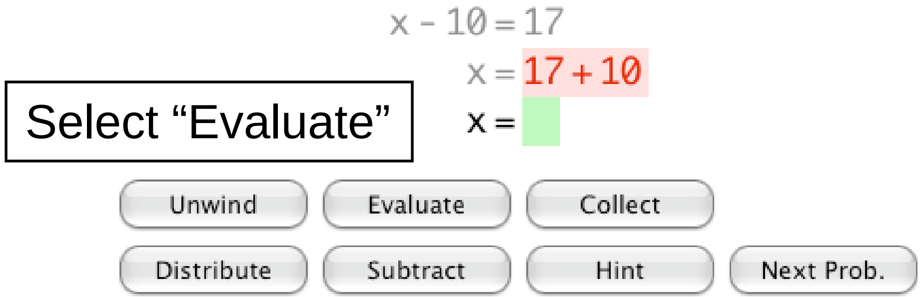
Phase 2: Enforced Distraction: n-back: Detect Repeated Letters

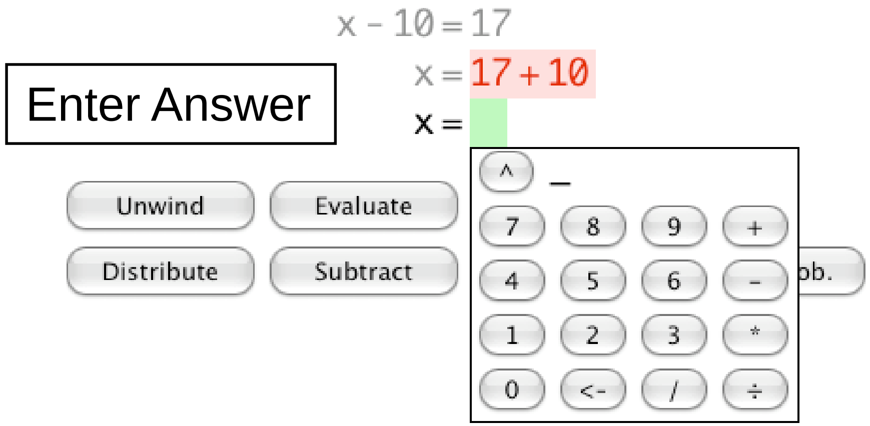
R
S
S
D
Y
F
F
J
U
R
L
L
P

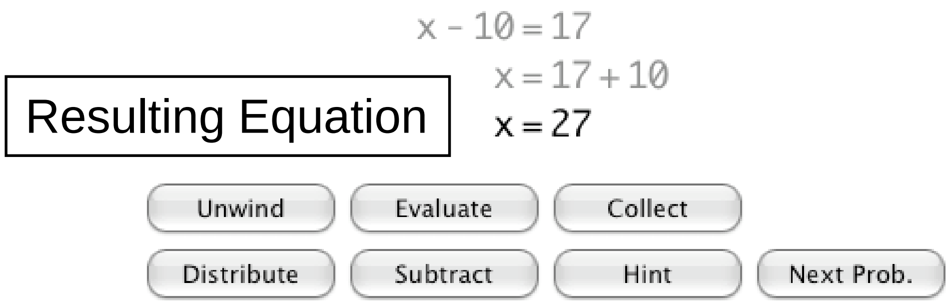
Earliest Material: Evaluation Phase











- Followed by more n-back & transition to next problem

An Example of Mind Reading

- Representative example: first student going through his first 8-minute sequence of problems alternating with n-back.
- We are going to see every 2 seconds what the student sees on the screen and what the algorithm predicts the student is seeing on the screen given their imaging data.
- Green will indicate material that involves mathematical problem solving according to our model and red will indicate off-task time.
- This is a sketch in powerpoint of a movie I would like to make of the actual task to illustrate the potential of this methodology.

Minute 1

New Scan every 2 seconds

Student	Predict
+	+
+	+
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	X - 10 = 17
X - 10 = 17	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	X = 17 + 10
X = 17 + 10	X = 17 + 10
X = 17 + 10	X = 17 + 10
X = 17 + 10	X = 17 + 10
X = 17 + 10	X = 17 + 10
X = 17 + 10	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back

Minute 2

Note	Student	Predict
Start with Fixation	n-back	n-back
Red indicates Off Task	n-back	X = 27
Equation appears on 3rd scan	X = 27	X = 27
Green indicates On Task	X = 27	X = 27
	X = 27	+
New Problem	+	+
Prediction Early	+	X + 4 = 13
Back in Synch	X + 4 = 13	X + 4 = 13
	X + 4 = 13	X + 4 = 13
	X + 4 = 13	X + 4 = 13
	X + 4 = 13	X + 4 = 13
	X + 4 = 13	X + 4 = 13
	X + 4 = 13	X + 4 = 13
	X + 4 = 13	X + 4 = 13
First Scan Prediction is wrong	n-back	n-back
Prediction Correct Again	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	X = 13 - 4
	X = 13 - 4	X = 13 - 4
	X = 13 - 4	X = 13 - 4
	X = 13 - 4	X = 13 - 4
	X = 13 - 4	n-back
	X = 13 - 4	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	X = 9

Minute 3

Student	Predict
n-back	+
X = 9	+
+	5 * X = 90
+	5 * X = 90
5 * X = 90	5 * X = 90
5 * X = 90	5 * X = 90
5 * X = 90	5 * X = 90
5 * X = 90	5 * X = 90
5 * X = 90	5 * X = 90
5 * X = 90	5 * X = 90
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
X = 90 / 5	X = 90 / 5
X = 90 / 5	X = 90 / 5
X = 90 / 5	X = 90 / 5
X = 90 / 5	X = 90 / 5
X = 90 / 5	X = 90 / 5
X = 90 / 5	n-back
X = 90 / 5	n-back
X = 90 / 5	n-back
X = 90 / 5	n-back
n-back	n-back
n-back	X = 18
n-back	X = 18
n-back	X = 18
n-back	+

Minute 4

Student	Predict
n-back	+
X = 18	X / 3 = 21
+	X / 3 = 21
+	X / 3 = 21
X / 3 = 21	X / 3 = 21
X / 3 = 21	X / 3 = 21
X / 3 = 21	X / 3 = 21
X / 3 = 21	X / 3 = 21
X / 3 = 21	X / 3 = 21
X / 3 = 21	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	X = 21 * 3
n-back	X = 21 * 3
X = 21 * 3	X = 21 * 3
X = 21 * 3	X = 21 * 3
X = 21 * 3	X = 21 * 3
X = 21 * 3	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	X = 63
X = 63	+
+	+
+	X - 7 = 16
X - 7 = 16	X - 7 = 16

Minute 5

Student	Predict
X - 7 = 16	X - 7 = 16
X - 7 = 16	X - 7 = 16
X - 7 = 16	X - 7 = 16
X - 7 = 16	X - 7 = 16
X - 7 = 16	X - 7 = 16
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	X = 16 + 7
X = 16 + 7	X = 16 + 7
X = 16 + 7	X = 16 + 7
X = 16 + 7	X = 16 + 7
X = 16 + 7	X = 16 + 7
X = 16 + 7	X = 16 + 7
X = 16 + 7	X = 16 + 7
n-back	X = 16 + 7
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
X = 23	X = 23
X = 23	+
+	+
+	X - 32 = 95
X - 32 = 95	X - 32 = 95
X - 32 = 95	X - 32 = 95
X - 32 = 95	X - 32 = 95
X - 32 = 95	X - 32 = 95

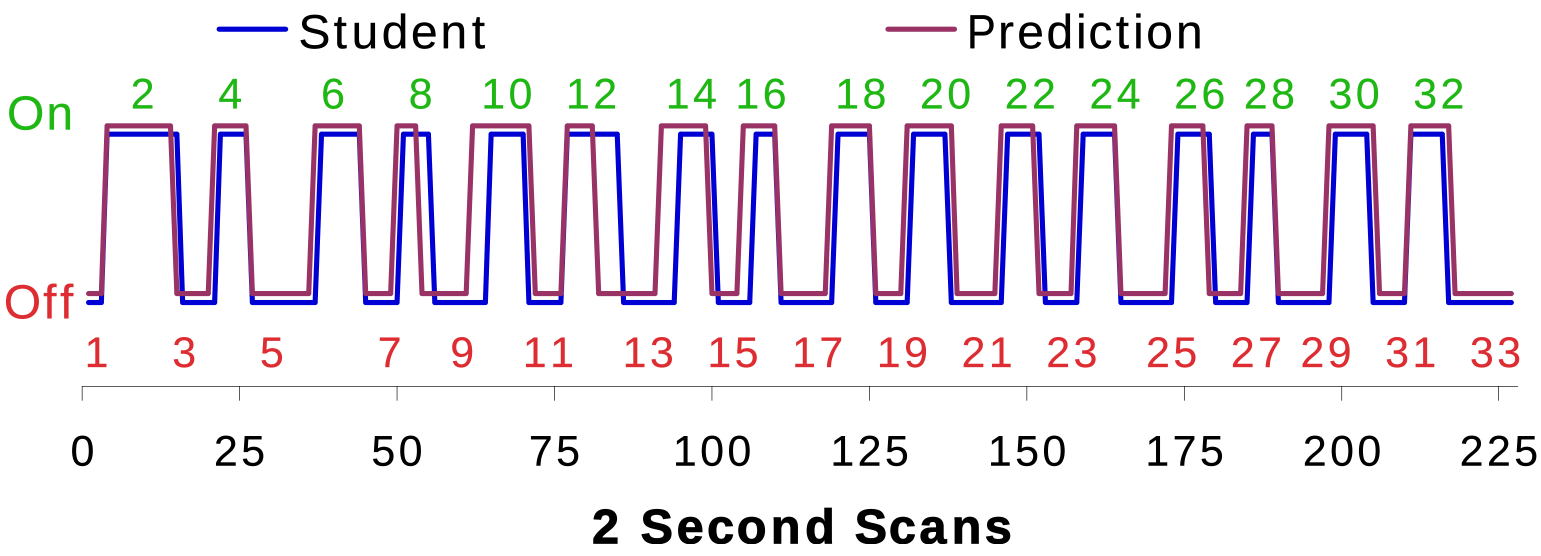
Minute 6

Student	Predict
X - 32 = 95	X - 32 = 95
X - 32 = 95	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	X = 95 +32
X = 95 +32	X = 95 +32
X = 95 +32	X = 95 +32
X = 95 +32	X = 95 +32
X = 95 +32	X = 95 +32
X = 95 +32	X = 95 +32
X = 95 +32	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	n-back
n-back	X = 127
X = 127	+
+	+
+	X + 54 =74
X + 54 = 74	X + 54 =74
X + 54 = 74	X + 54 =74
X + 54 = 74	X + 54 =74
X + 54 = 74	X + 54 =74
X + 54 = 74	X + 54 =74
X + 54 = 74	n-back
n-back	n-back

Minute 7	Student	Predict
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	X = 74 - 54
	X = 74 - 54	X = 74 - 54
	X = 74 - 54	X = 74 - 54
	X = 74 - 54	X = 74 - 54
	X = 74 - 54	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	X = 20
	X = 20	X = 20
	X = 20	+
	+	+
	+	X + 91 = 87
	X + 91 = 87	X + 91 = 87
	X + 91 = 87	X + 91 = 87
	X + 91 = 87	X + 91 = 87
	X + 91 = 87	X + 91 = 87
	X + 91 = 87	X + 91 = 87
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back

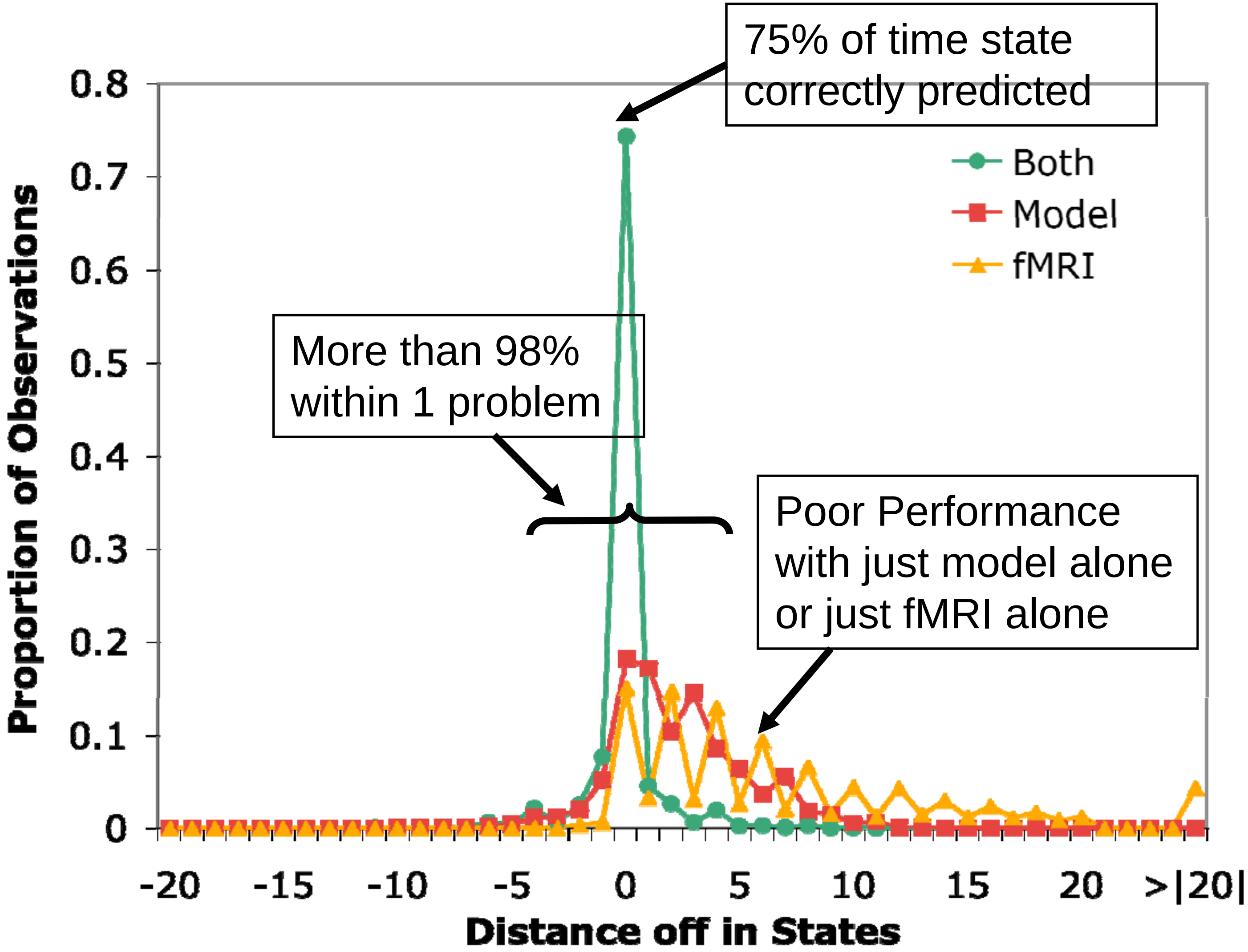
Minute 8	Student	Predict
	X = 87 - 91	X = 87 - 91
	X = 87 - 91	X = 87 - 91
	X = 87 - 91	X = 87 - 91
	X = 87 - 91	X = 87 - 91
	X = 87 - 91	X = 87 - 91
	X = 87 - 91	X = 87 - 91
	n-back	X = 87 - 91
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	n-back	n-back
	X = -4	n-back
	Done	X = -4
	Done	Done
	Done	Done
	Done	Done

Statistics of this Example



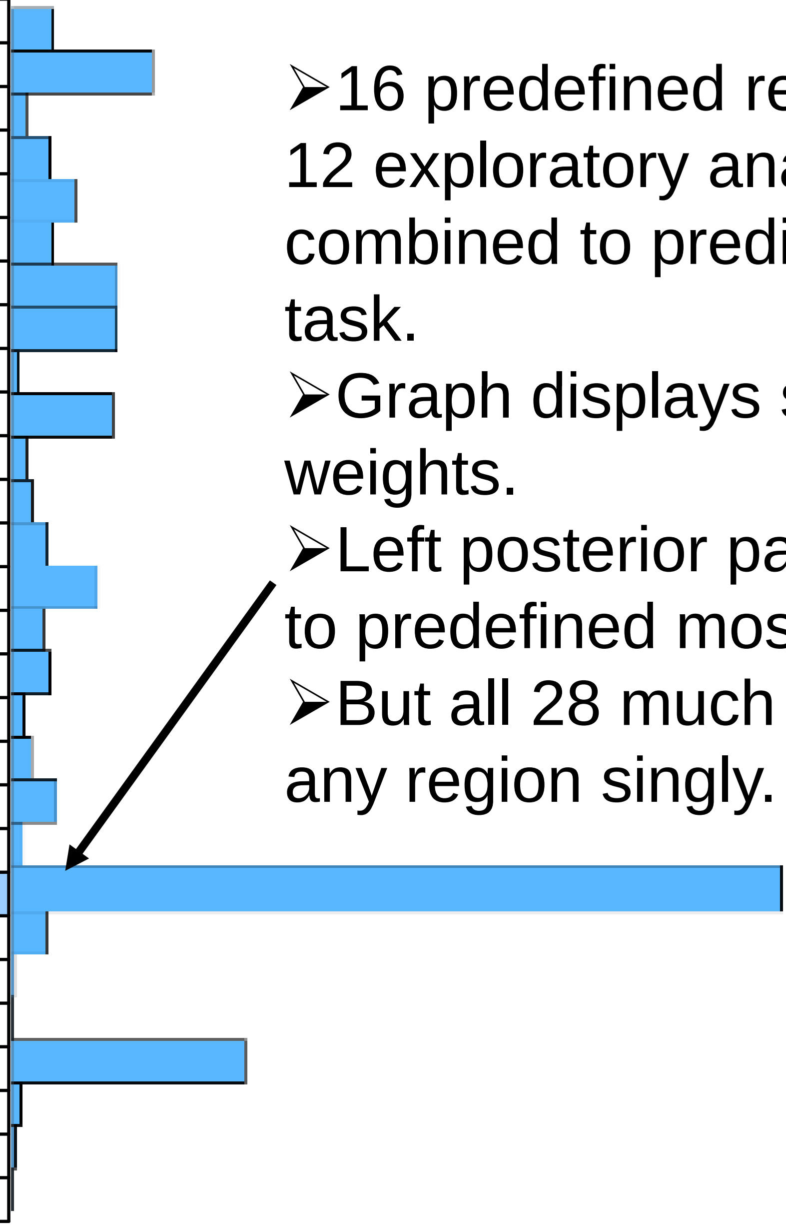
- Student takes 227 scans to go through 33 states to solve 8 problems.
- Prediction is never off by more than 1 state and this happens on 32 of the 227 scans.
- I will try to explain how we combine a cognitive model and fMRI data to obtain this result.
- But first lets see whether we really need both the cognitive model and fMRI data.

You Need Both a Model & fMRI Statistics on 210 Blocks



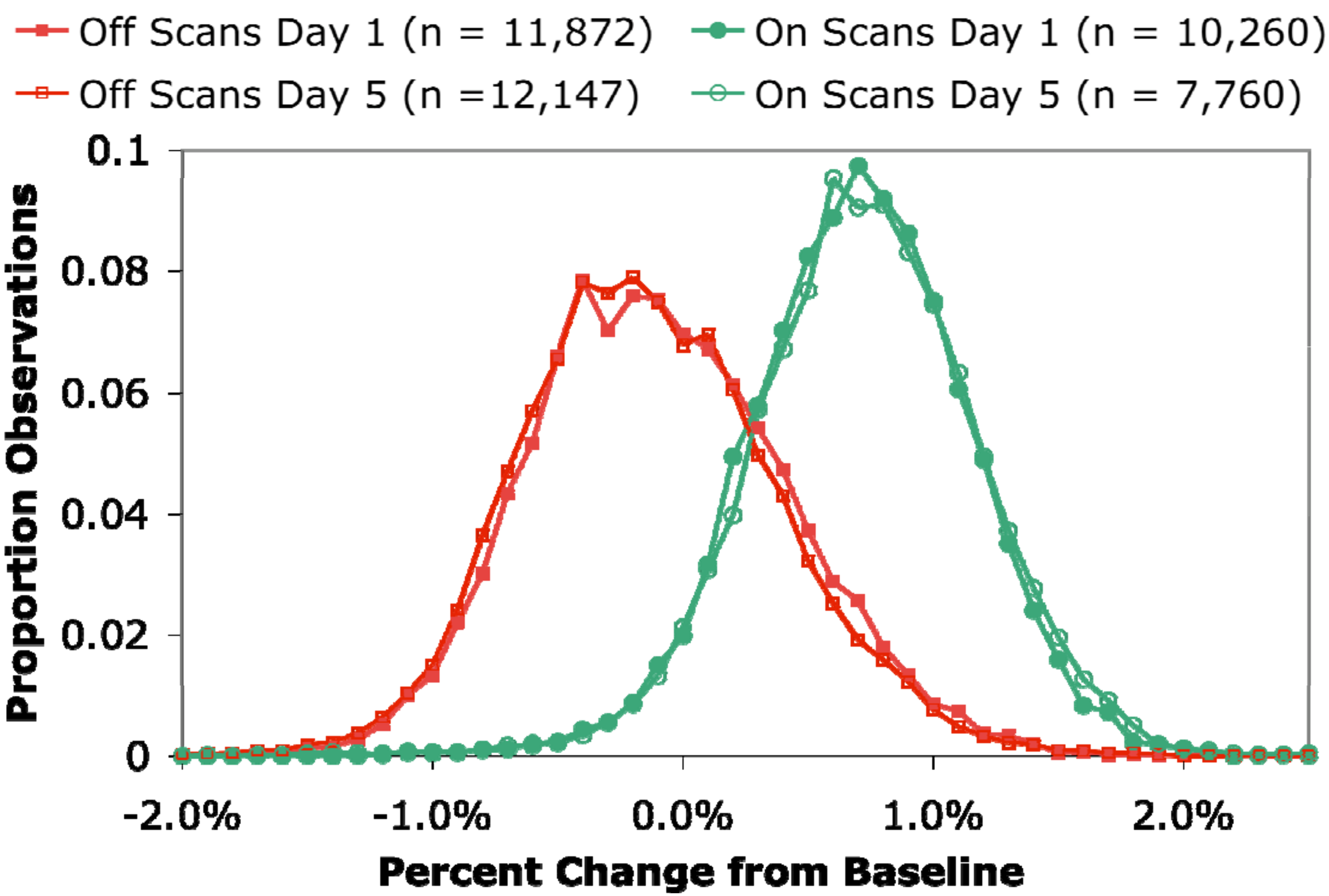
fMRI Analysis

Region	x	y	z
1. P R Manual	41	-20	50
2. P L Manual	-41	-20	50
3. P R ACC	7	10	39
4. P L ACC	-7	10	39
5. P R Vocal	43	-14	33
6. P L Vocal	-43	-14	33
7. P R PPC	23	-63	40
8. P L PPC	-23	-63	40
9.P R LIPFC	43	23	24
10. P L LIPFC	-43	23	24
11. P R Caudate	14	10	7
12. P L Caudate	-14	10	7
13. P R Auditory	46	-22	9
14. P L Auditory	-46	-22	9
15. P R Fusiform	42	-61	-9
16. P L Fusiform	-42	-61	-9
17. E R Premotor	32	1	58
18. E R PFC	44	23	36
19. E R Ang Gyrus	37	-48	43
20. E L PFC	-46	24	32
21. E L PPC	-29	-60	42
22. E R PPC	13	-73	50
23. E R Orb Frontal	47	46	0
24. E L Orb Frontal	-26	48	-6
25.E R Occ Gyrus	23	-89	-2
26. E L Occ Gyrus	-20	-90	-5
27. E L Occ Gyrus	-19	-77	-13
28. E R Cerebellum	28	-61	-18



- 16 predefined regions with 12 exploratory analysis combined to predict On and Off task.
- Graph displays squared weights.
- Left posterior parietal close to predefined most predictive.
- But all 28 much better than any region singly.

Performance with Combined Signal



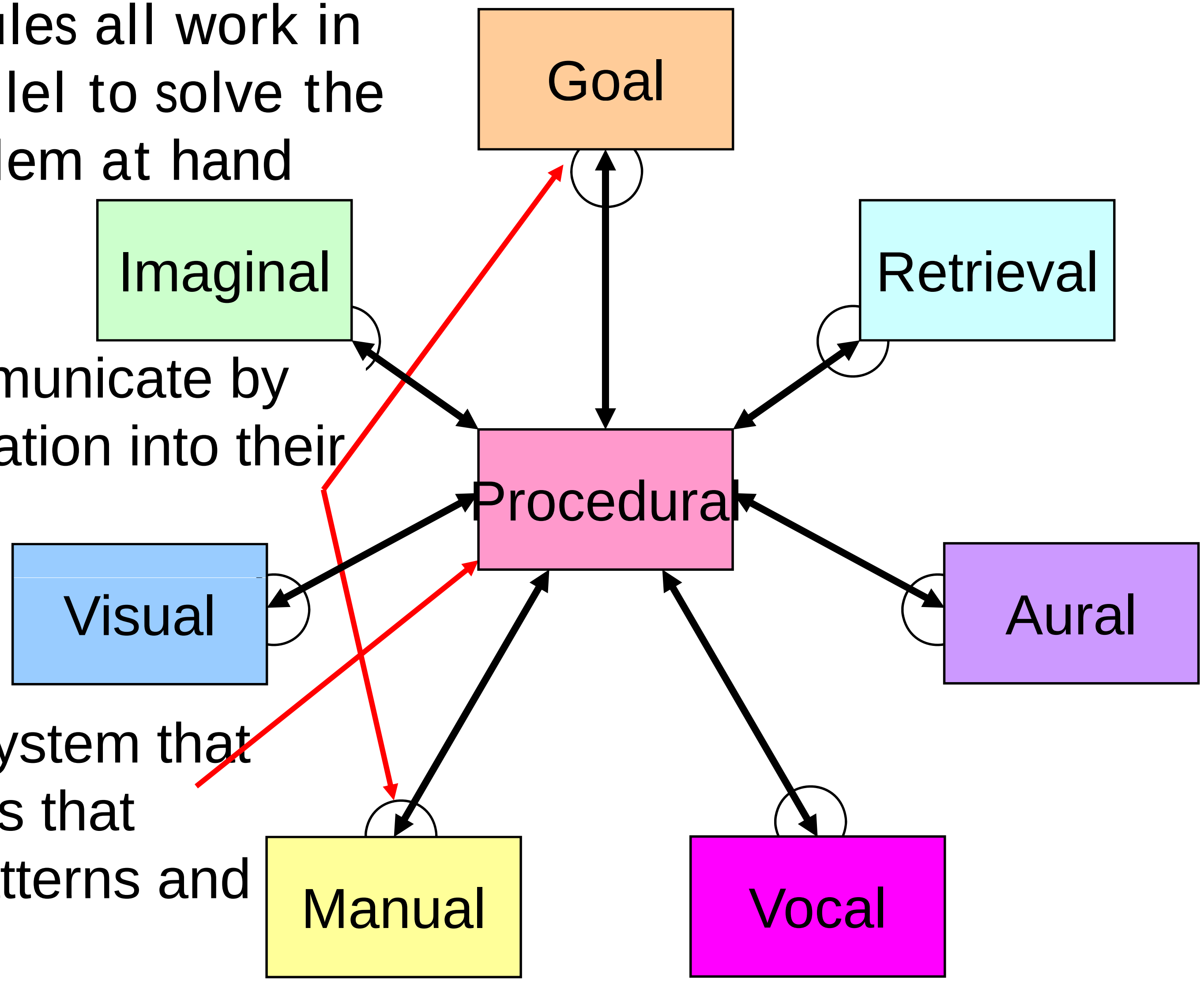
- ➢Combined activity offers moderate discrimination of ON versus OFF task.
- ➢Regions and weights defined on Day 1 data generalize to Day 5.
- ➢However, fMRI by itself it offers poor basis for identifying where student is in the sequence of problems. We need a cognitive model.

Modules in ACT-R Cognitive Architecture

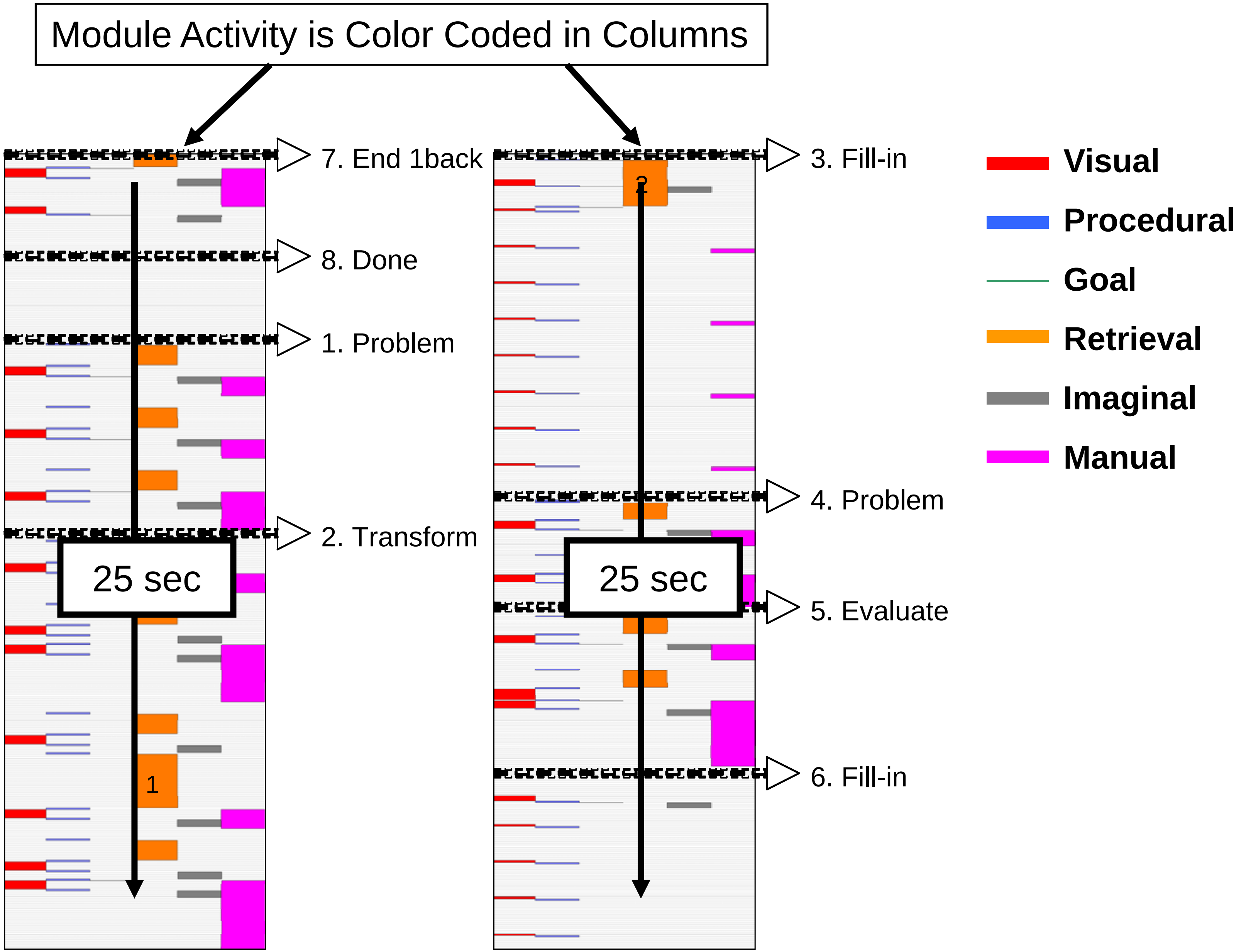
Modules all work in parallel to solve the problem at hand

Modules communicate by putting information into their buffers.

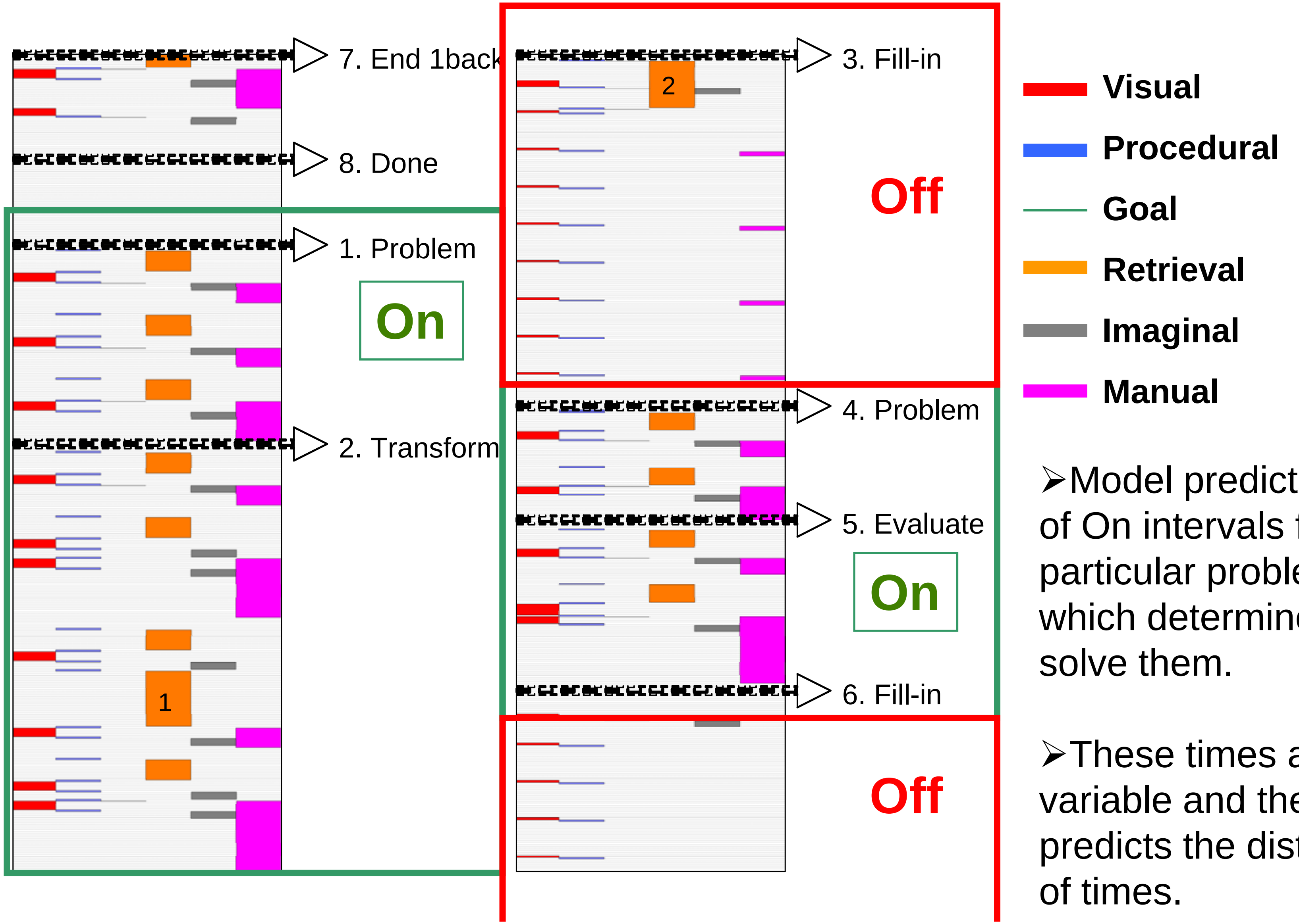
Production system that contains rules that recognize patterns and react



50 Sec of Model Interaction with Algebra Tutor



On and Off Periods in ACT-R Model Interaction



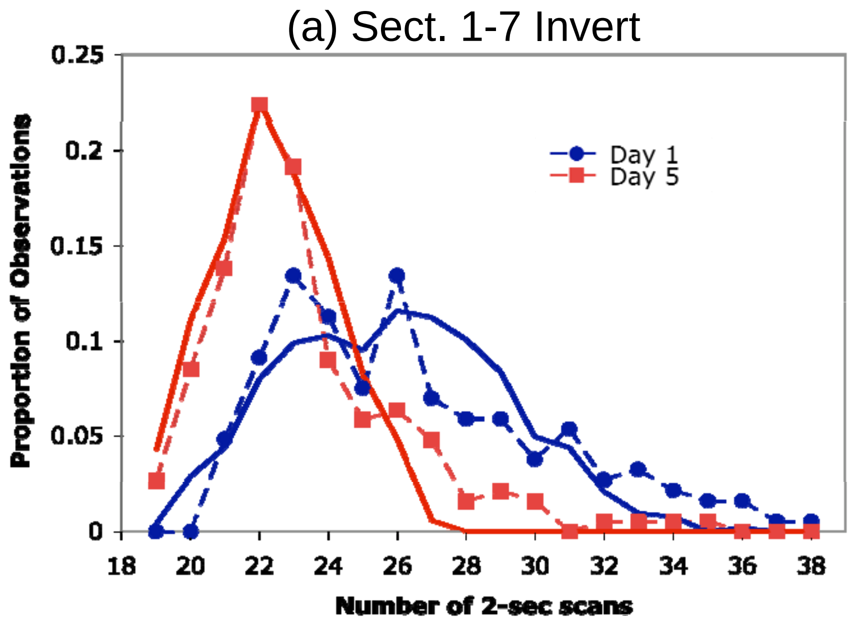
- ➢Model predicts length of On intervals for particular problems which determines time to solve them.
- ➢These times are variable and the model predicts the distribution of times.

Time to Solve Problems

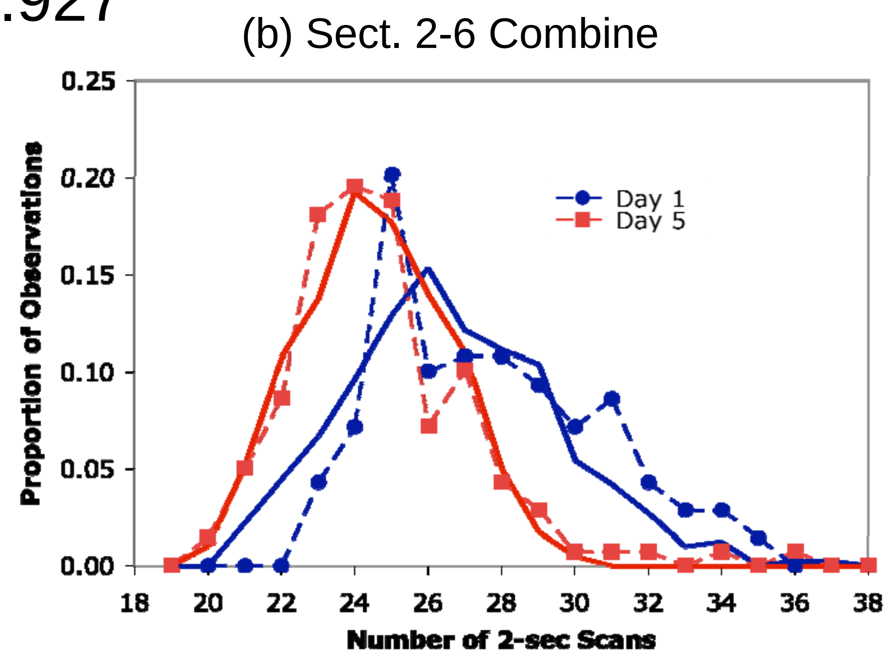
➢The model predicts the distribution of times to solve problems on both days and for various sections of material.

➢Because of learning in ACT-R, parameters defined on Day1 successfully predict problem times on Day 5.

➢However using these distributions alone offers poor basis for identifying where student is in the sequence of problems. We need fMRI.



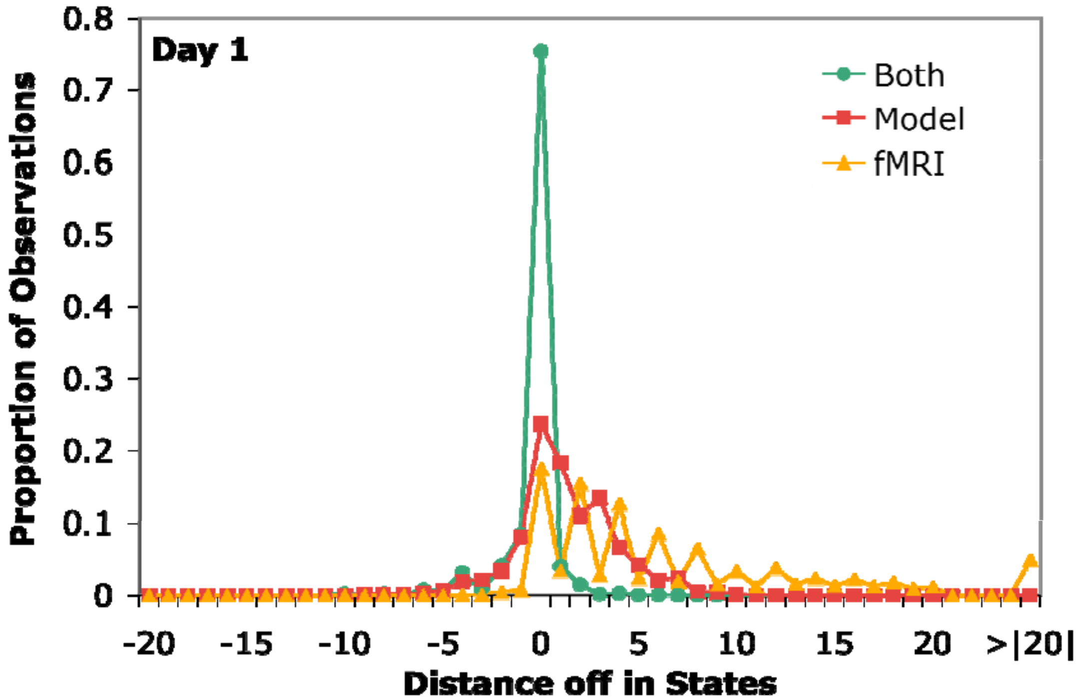
r = .927



Combining Model and fMRI: Hidden Markov Model (HMM) Algorithms

- We are looking for an interpretation of the m scans in a block that contains n problems as a linear sequence of 4n+1 states.
- The number of such interpretations is $\frac{(m-1)!}{(4n)(m-4n-1)!}$
- The probability of any of these interpretations can be calculated from the probabilities of interval lengths for a state and likelihood of the signal magnitudes as ON and OFF task.
- Future actions only depend on the current state. Since the states are not directly observable and their durations are variable the model a hidden semi-Markov process.
- Dynamic programming algorithms associated with hidden Markov models can efficiently compute the probability that the student is in any state on a particular scan.

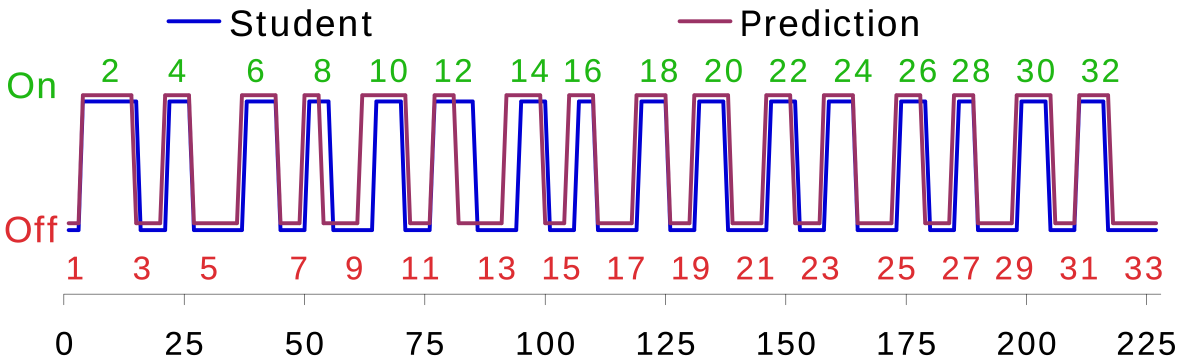
Both Model and fMRI are Needed



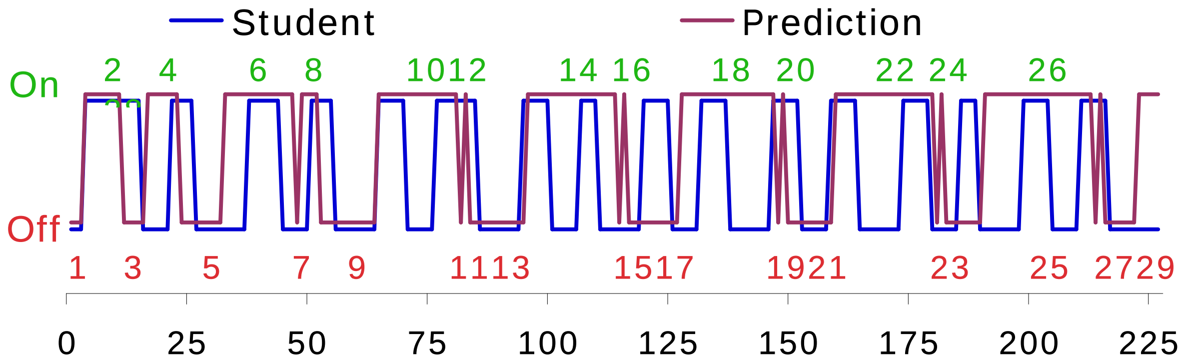
- ➢Our prediction for any scan is the most probable state according to the HMM algorithm.
- ➢This is achieved by using the distribution of lengths of state intervals from the model and the distribution of ON and OFF signal magnitudes from fMRI.
- ➢Model-only predictions are obtained by using just one distribution for both ON and OFF signal magnitudes, thus negating the fMRI contribution.
- ➢fMRI-only predictions are obtained by making all interval lengths equally probable, thus negating the model contribution.

Comparison on First Example Block

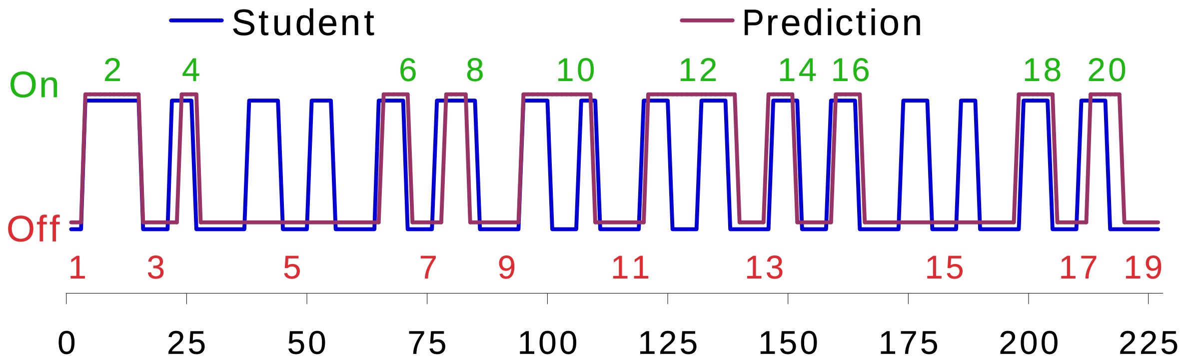
Model + fMRI:
Identifies all 33 States



Model Only:
Identifies Only 29 States

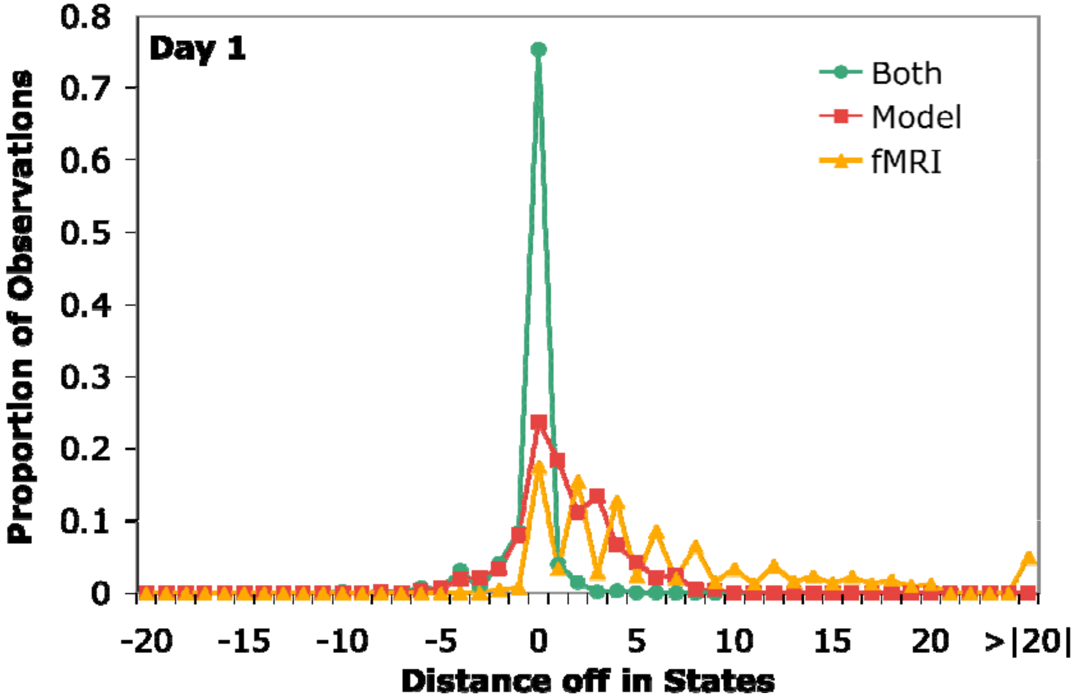


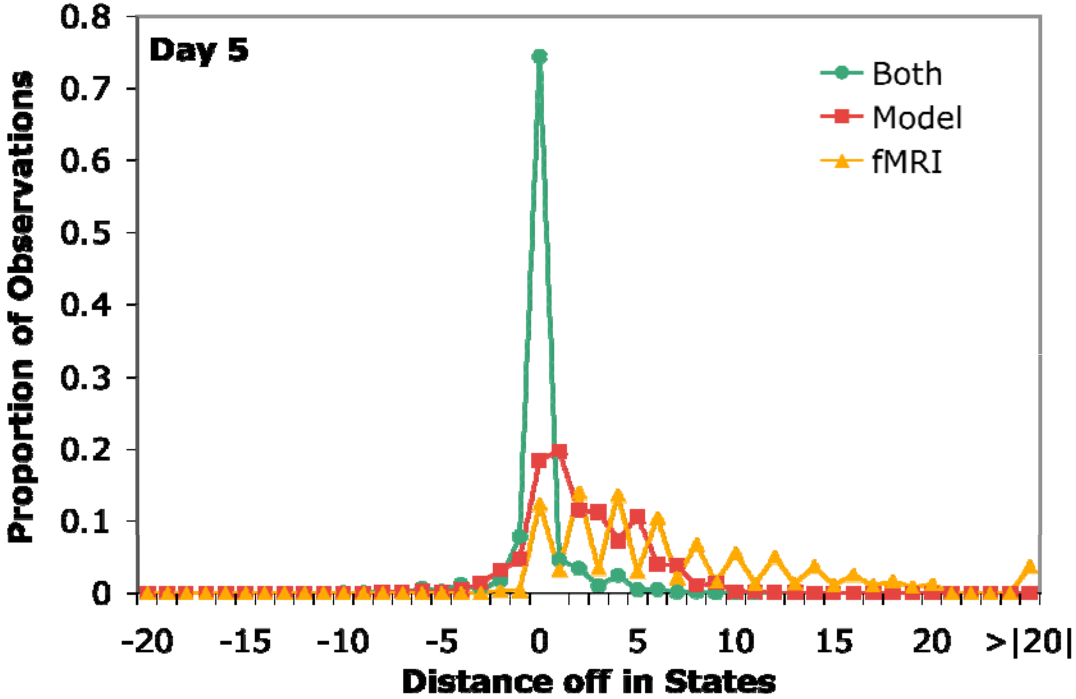
fMRI Only:
Identifies Only 20 States



Predictions Generalize from Day 1 to Day 5

- ➢Regions, weights, and model parameters were estimated from Day 1.
- ➢There is the danger of overfitting in our claimed success for Day 1 data.
- ➢They can be used to predict with no further estimation Day 5.
- ➢Learning effects in the model predicts a speed up for Day 5 which matches students.





Day 5 Exemplifies Model Tracing in Cognitive Tutors

- The parameters were all estimated from group data. Typically there is not enough data about a single student to produce reliable parameter estimations.
- The parameters are estimated from a situation (Day 1) different than the one on which they are used (Day 5). This is not typical in most mind-reading applications where one data set is split into a training and test set.
- Even though the parameters estimated from average behavior in a different situation they are nonetheless used to interpret what a particular student is thinking at the moment.
- A learning model allows one to adjust the expectations to reflect the progress of the student.

Conclusions

- It is possible to meaningfully and rigorously relate brain imaging data to a detailed model of student-tutor interactions.
- Brain imaging data can guide model development which would lead to better cognitive tutors.
- It is possible to use brain imaging data to to diagnose student problem solving in real-time with considerable accuracy.
- The method can be fit to one data set and generalize to another data set.
- While this demonstration uses only brain imaging data, real applications would want to integrate imaging data with behavioral information.

Thank You!