

***Strategies for Discovering
Mechanisms of Mind using fMRI:***

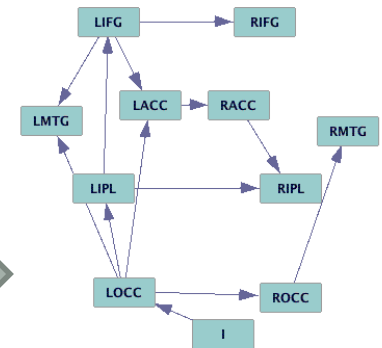
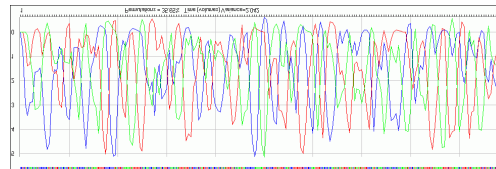
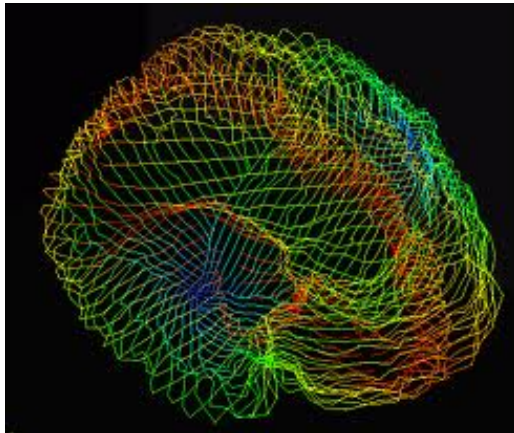
6 NUMBERS

Joseph Ramsey, Ruben Sanchez Romero
and Clark Glymour

The Numbers

- 20
- 50
- 5024
- 9205
- 8388
- 500

fMRI and Mechanism



From indirect signals of local changes in blood oxygenation in small (2 mm^3) brain regions (voxels) during a cognitive task

...Infer...

The CAUSOME—the causal relations between those regions in performing that task.

Why?

- By contrasting these mechanisms where there is a task with “resting state” mechanisms where there is no task, one could try to discover sets of causal relations specific to kinds of cognitive tasks.
- Which one hopes would lead to
 - A profound refinement of “real estate” neuropsychology
 - An alignment of psychological descriptions with neural processes
 - Identification of processing disturbances in neuroatypicals

Typical Task Experiments

- These are data sets from Russ Poldrack's Open fMRI project.
- Experiment 1 Subjects inflate balloons, judging whether to keep inflating them (risk).
- Experiment 3 Subjects judge nonsense words as rhyming or not.
- Experiment 5: Subjects are given a stake and bets (real money) which they can accept or reject.

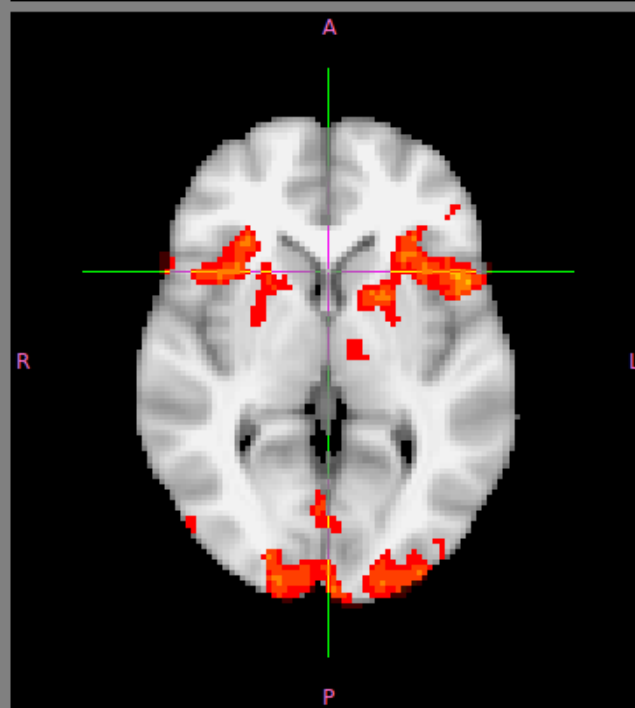
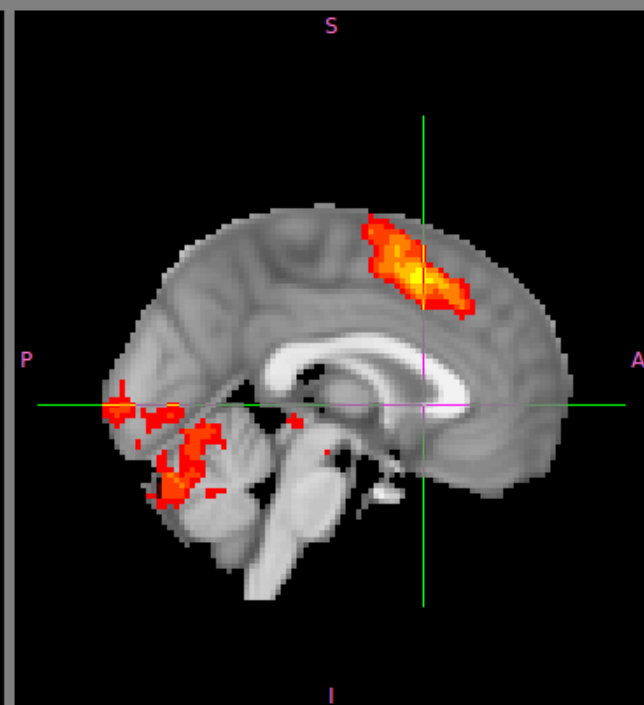
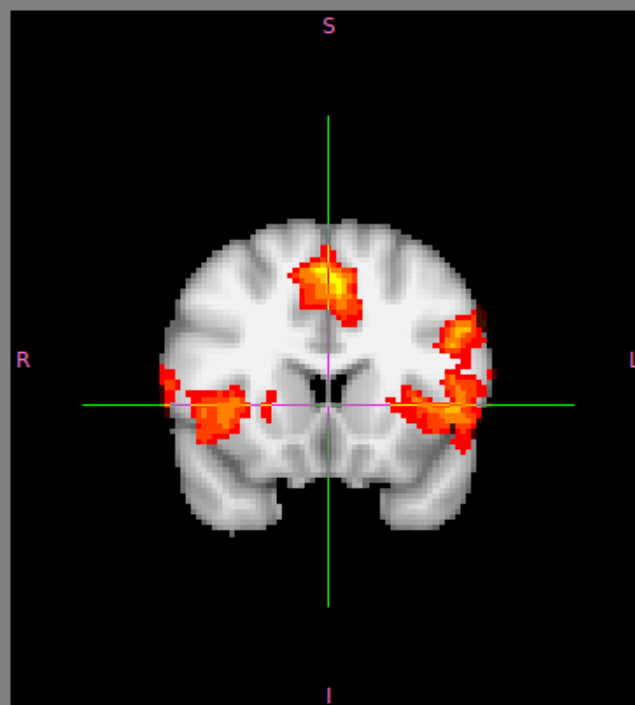
Usual Procedure

1. Obtain BOLD time series from a large region of the brain (e.g., whole brain or cortex).
2. Cluster time series into “regions of interest” (ROIs) by anatomy, statistical clustering, or eyeballing.
3. Aggregate the BOLD measurements at a time for voxels within a cluster into a single measure (e.g., the average)
4. Apply a search procedure to estimate the causal relations among the clusters.

ROIs for Experiment 3:

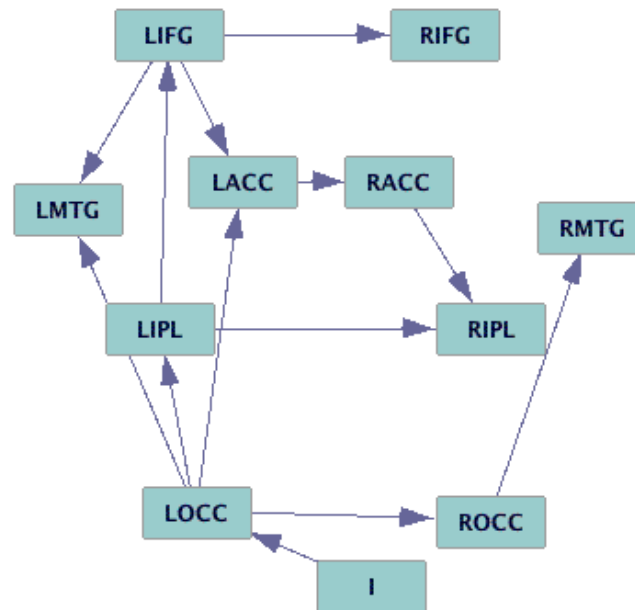
I
LOCC, ROCC
LIPL, RIPL
LMTG, RMTG
LACC, RACC
LIFG, RIFG

Not all visible
in this image.



Causal Search Output from Exp. 3

- With IMaGES search procedure



Search Procedures

- Guess
- Inverse covariance (partial correlation)
- PC
- FCI
- GES, IMaGES
- Granger
- LiNGAM
- LISREL (GIMME)
- Non-Gaussian scores
- Friston

Guess

- Look at the ROIs and the data and guess what the causal connections are.
- This method is in fact used!

Inverse covariance (partial correlation)

- Calculate the covariance matrix of your data.
- Invert it (possibly with L_1 penalty)
- Zero entries are $X \perp\!\!\!\perp Y \mid Z$ where $Z = \text{all other variables}$.
- Problem: $X \rightarrow Z \leftarrow Y$
 - X and Y are uncorrelated.
 - But if you condition on Z, they're correlated!
 - So you think there's an edge where there's not!
 - Marries parents!
- Note that just *thresholding* the covariance matrix leads to extra edges too!
 - $X \rightarrow Y \rightarrow Z$: X is correlated with Z, so you get the edge $X \text{---} Z$!

PC

- Assume the true model is a directed acyclic graph (DAG)
- Assume there are no unrecorded common causes and i.i.d. sampling:
- Under these assumptions, the adjacency search finds correct adjacencies (undirected causal connections) in the large sample limit.
 - This works well for fMRI.
 - Scales up to high dimensional problems.
- PC can process multiple subjects by using a *multi-subject conditional independence test*.
- The orientation search does not work so well for fMRI.

FCI

- Like PC but does *not* assume that there are no unmeasured common causes.
 - There may be latent variables!
- Uses the same adjacency search as PC.
 - But the adjacencies are interpreted differently: not all of them represent direct causal paths
- Does not scale up to high dimensions but a modification (RFCI: *Really* Fast Causal Inference) scales up better. How much better?

GES, IMaGES

- GES (Meek; Chickering) is a *Bayesian score search*
- Under essentially the same assumptions as PC, correct in the large sample limit.
- IMaGES is a multi-subject version of GES.
 - Average the BIC scores!
 - Penalty term to counteract the effects of Indirect measurement creating spurious associations of measured variables.
 - Works well for fMRI (Ramsey et al., 2010, 2011)
- Scales up so-so.

Granger Causality

- Asks, which variables can I condition on to best predict future states of variables given current and past states of variables?
- Regress Y_t on (X_{t-1}, Y_{t-1}) . Take significant X predictors to be causes of Y .
- The *lags* are critically important for this method. In particular, it is important that the lags be uniform across the data set, a condition which doesn't hold well for fMRI data. (Though work has been done to adjust for this.)

LINGAM

- “Linear Non-Gaussian Acyclic Model”
- Runs Independent Components Analysis (ICA)
- Associates ICA output with variables by reordering the matrix into a lower triangle.
- Infers causal connections *and* directions in one fell swoop.

LISREL (GIMME)

- Will defer; Kathleen Gates will talk about this.
- Basically, models the causal structure over ROIs as a linear, Gaussian system.
- Amazingly effective for the case of *non-stationary connections*—that is when the coefficient for $X \rightarrow Y$ changes over time.

Orientation Using Non-Gaussian Scores

- *Given* adjacencies, non-Gaussian scoring can be effective at orienting edges.
- Hyvärinen and Smith (2013): Estimate direction between X, Y from signs of skews, e.g.. $X^2Y - Y^2X$
- Other methods due to Ramsey et al.:
- R1, R2 assume sums of independent variables are closer to Gaussian than the summands.
- R3 uses an information theoretic measure based on a non-Gaussianity score.
- R4 adapts an independent components method for cyclic graphs (Lacerda, et al., 2008).

Friston Method

- In the context of the Dynamical Causal Modeling paradigm, K. Friston proposes a method that scores a series of models and reports the model with the best score.
- This is possible due to a very fast scoring procedure.
- However, the number of possible models over a set of variables increases super-exponentially.
- So only models over a very few number of variables can be considered.
- Problem: Exhaustive search versus Search.

Testing Search Methods

- Simulate
- Simulate
- Simulate
- Torture animals and measure humans.

One Torture Test

- (Dawson et al., 2013) Compared causal relations between visual cortex regions of macaque monkeys with causal relations between analogous areas in the human cortex estimated by several search methods from fMRI.
 - Only adjacencies identified.
 - PC most accurate.

Lots of Simulation Data

- (Smith, et al., 2011) tested identifications of simple graphical causal relations from simulated fMRI in 28 conditions.
- Smith et al. tested 35 search methods. None succeeded in identifying causal directions accurately.
- Their code (Smith and Woolrich mainly), based on the best available biophysical model of the generation of the BOLD signal, can be used to simulate data from much more complex, higher-dimensional models.

Smith Study Conclusions

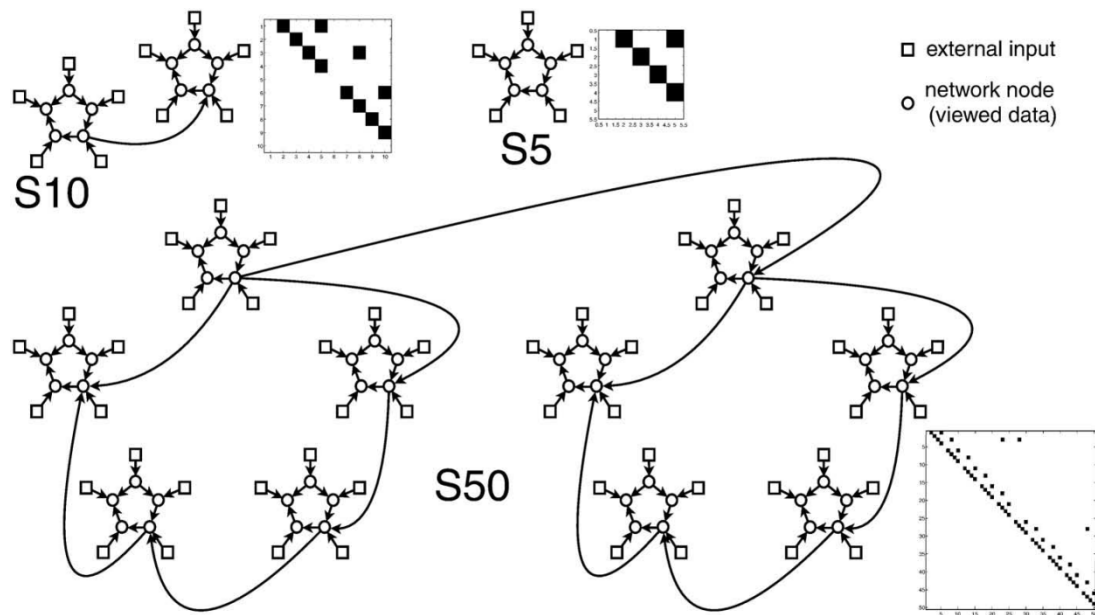
- “Bayes net methods” effective at finding adjacencies—PC, GES, FCI, etc.
- Inverse covariance methods effective at finding adjacencies (because there were only a few unshielded colliders in the Smith models)
- Granger methods, in all versions tested, ineffective at finding adjacencies (but still frequently used).
- LiNGAM ineffective at finding adjacencies.
- ***No methods effective at finding orientations.***
 - *(Non-Gaussian methods other than LiNGAM not tested.)*
 - *(Except maybe Patel.)*

20

- Maximum number of variables (Regions of Interest) in any published attempt to infer causal mechanisms from empirical fMRI data.

50

- The number of simulated graphs in the largest model in the Smith study.



5,024

5054 = The number of 6mm^3 voxels in the cortex.

9 = The number of variables in the graphical causal models estimated by Poldrack for the pseudorhyme experiment.

9205

- The number of directed edges found for one single subject data set from experiment 1
- Alpha = 0.01
- Calculate covariance matrix C over 5054 voxels
- Run PC adjacency search over C
- Orient using R3
- Side note: GES/IMaGES is not currently scalable to 5054 variables; I've scaled it up to about 800 variables...

8388

- The number of directed edges found for one single subject dataset from experiment 3
- Alpha = 0.05
- Calculate covariance matrix C over 5054 voxels
- Run PC adjacency search over C
- Orient using $R3$.
- Many missing (unrecorded) voxels, maybe should run RFCI...future work...can RFCI be scaled up that far?

500

- Why Trust The Estimates?
- Point 1: Simulation with 500 variables with Smith code yields 90% precision; 50% recall
- Point 2: In experiment 3, PC + R3 identifies the stimulus variable as a cause, not an effect, of neural variables.
- Point 3: In experiment 3, PC + R3 agrees with experts' opinion of directions of influence.

Details for 500 variable simulation.

		R1	R3	EB	Tanh	Skew	Rskew	Patel	SkewE	RSkewE
Average	AP	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	AR	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	PTPA	0.76	0.94	0.75	0.7	0.96	0.97	0.72	0.9	0.9
Group	AP	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
	AR	1	1	1	1	1	1	1	1	1
	PTPA	0.99	1	0.77	0.77	1	1	0.76	0.94	0.95

AP = Adjacency precision ($\#$ true positive adjacencies / ($\#$ true positive adjacencies + $\#$ false positive adjacencies)), AR = adjacency recall ($\#$ true positive adjacencies / ($\#$ true positive adjacencies + $\#$ false negative adjacencies)), PTPA = Precision for true positive adjacencies ($\#$ true positive orientations / $\#$ true positive adjacencies), averaged over 5 individual subjects and for the 5 subject analyzed groupwise, for the 500 variable, 621 edges simulation described in the text. For the individual subject case 6% of the estimated edges were false positives and for the group case 2% of the estimated edges were false positives.

How Is It Possible?

- PC for adjacencies
- Non-Gaussian scores for directions

Programming challenges:

- Scaling up to 5054 variables.
 - Seems doable for PC, non-Gaussian orientation.
 - RFCI?
 - GES?
 - Others?
- This is not the limit of aspiration—we want to do 40,000 or more voxels!

Why the Low Recall for Single Subject?

- With PC there's a tradeoff between computational complexity and information retrieved.
- Choice is to find the largest effects only or to have the computer never return.
 - For PC, a choice of alpha level.
 - Higher alpha levels imply more false positive edges!
 - Also, small sample size, 160.

What Are the Problems?

- Stability of PC under resampling
- Disagreements in estimates from different non-Gaussian scoring procedures
- Disagreements in estimates between subjects
- Use of multiple subject data simultaneously
- And:
 - Feedback relations (addressed by one non-Gaussian method, R4)
 - Cancellation of correlations by positive and negative influences
 - Latent variables
 - Non-stationary time series (see GIMME!)

What Are the Potential Clinical Applications?

- Catherine Hanson will talk about some of that, I hope.

Thanks to many people...

- Steve Hanson
- Catherine Hanson
- Russ Poldrack
- Stephen Smith
- Patrick Hoyer
- Aapo Hyvärinen
- Cosma Shalizi
- Peter Spirtes
- Richard Scheines
- Marloes Matuis
- Kathleen Gates
- *and many more! Sorry if I didn't list you!*

Some few references...

Dawson, D. A., Cha, K., Lewis, L. B., Mendola, J. D., and Shmuel, A. (2013). Evaluation and calibration of functional network modeling methods based on known anatomical connections. *NeuroImage* 67:331-343.

Hyvärinen, A., and Smith, S. M. (2013). Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *The Journal of Machine Learning Research*, 14:111-152.

Lacerda, G., Spirtes, P., Ramsey, J., Hoyer, P. (2008). Discovering cyclic causal models by independent components analysis. *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*.

Ramsey, J., Hanson, S., Hanson, C., Halchenko, Y., Poldrack, R., Glymour, C. (2010). Six problems for causal inference from fMRI. *NeuroImage*, 49:1545-1558.

Ramsey, J., Hanson, S., and Glymour, C. (2011). Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. *NeuroImage*, 58(3):838-848.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J., and Woolrich, M. W. (2011). Network modeling methods for FMRI. *NeuroImage*, 54(2):875-891.

