Inferring mental states from neuroimaging data

Russell Poldrack

Departments of Psychology and Neurobiology

Imaging Research Center
University of Texas at Austin
This article was written by Marco Iacoboni, Joshua Freedman and Jonas Kaplan of the University of California, Los Angeles, Semel Institute for Neuroscience; Kathleen Hall Jamieson of the Annenberg Public Policy Center at the University of Pennsylvania; and Tom Freedman, Bill Knapp and Kathryn Fitzgerald of FKF Applied Research.

IN anticipation of the 2008 presidential election, we used functional magnetic resonance imaging to watch the brains of a group of swing voters as they responded to the leading presidential candidates. Our results reveal some voter impressions on which this election may well turn.

Our 20 subjects — registered voters who stated that they were open to choosing a candidate from either party next November — included 10 men and 10 women. In late summer, we asked them to answer a list of questions about their political preferences, then observed their brain activity for nearly an hour in the scanner at the Ahmanson Lovelace Brain Mapping Center at the University of California, Los Angeles. Afterward, each subject filled out a second questionnaire.
“In response to images of Democratic candidates, men exhibited activity in the medial orbital prefrontal cortex, indicating emotional connection and positive feelings.”

“Images of Fred Thompson led to increased activity in the inferior frontal cortex, a brain structure associated with empathy.”

“Subjects who had an unfavorable view of John Edwards responded to pictures of him with feelings of disgust, evidenced by increased activity in the insula, a brain area associated with negative emotions.”
LETTER; Politics and the Brain
Published: November 14, 2007
To the Editor:

"This Is Your Brain on Politics" (Op-Ed, Nov. 11) used the results of a brain imaging study to draw conclusions about the current state of the American electorate. The article claimed that it is possible to directly read the minds of potential voters by looking at their brain activity while they viewed presidential candidates.

For example, activity in the amygdala in response to viewing one candidate was argued to reflect "anxiety" about the candidate, whereas activity in other areas was argued to indicate "feeling connected." While such reasoning appears compelling on its face, it is scientifically unfounded.

As cognitive neuroscientists who use the same brain imaging technology, we know that it is not possible to definitively determine whether a person is anxious or feeling connected simply by looking at activity in a particular brain region. This is so because brain regions are typically engaged by many mental states, and thus a one-to-one mapping between a brain region and a mental state is not possible.

For example, rather than simply providing a brain marker of anxiety levels, as the article assumed, we know that the amygdala is activated by arousal and positive emotions as well. Such problems of interpretation with brain imaging studies can be avoided only by careful experimental design, and, as with any scientific data, the peer review process is critical to understanding whether the data are sound or based on faulty methodology.

Unfortunately, the results reported in the article were apparently not peer-reviewed, nor was sufficient detail provided to evaluate the conclusions.

As cognitive neuroscientists, we are very excited about the potential use of brain imaging techniques to better understand the psychology of political decisions. But we are distressed by the publication of research in the press that has not undergone peer review, and that uses flawed reasoning to draw unfounded conclusions about topics as important as the presidential election.

Adam Aron, Ph.D., University of California, San Diego
David Badre, Ph.D., Brown University
Matthew Brett, M.D., University of Cambridge
John Cacioppo, Ph.D., University of Chicago
Chris Chambers, Ph.D., University College London
Roshan Cools, Ph.D., Radboud University, Netherlands

Steve Engel, Ph.D., University of Minnesota
Mark D'Esposito, M.D., University of California, Berkeley
Chris Frith, Ph.D., University College London
Eddie Harmon-Jones, Ph.D., Texas A&M University
John Jonides, Ph.D., University of Michigan
Brian Knutson, Ph.D., Stanford University

Liz Phelps, Ph.D., New York University
Russell Poldrack, Ph.D., University of California, Los Angeles
Tor Wager, Ph.D., Columbia University
Anthony Wagner, Ph.D., Stanford University
Piotr Winkielman, Ph.D., University of California, San Diego
Do you really love your iPhone?

The New York Times
The Opinion Pages

You Love Your iPhone. Literally.

By MARTIN LINDSTROM
Published: September 30, 2011

“Earlier this year, I carried out an fMRI experiment to find out whether iPhones were really, truly addictive, no less so than alcohol, cocaine, shopping or video games. In conjunction with the San Diego-based firm MindSign Neuromarketing, I enlisted eight men and eight women between the ages of 18 and 25. Our 16 subjects were exposed separately to audio and to video of a ringing and vibrating iPhone...most striking of all was the flurry of activation in the insular cortex of the brain, which is associated with feelings of love and compassion. The subjects’ brains responded to the sound of their phones as they would respond to the presence or proximity of a girlfriend, boyfriend or family member. In short, the subjects didn’t demonstrate the classic brain-based signs of addiction. Instead, they loved their iPhones.
To the Editor:

“You Love Your iPhone. Literally,” by Martin Lindstrom (Op-Ed, Oct. 1), purports to show, using brain imaging, that our attachment to digital devices reflects not addiction but instead the same kind of emotion that we feel for human loved ones.

However, the evidence the writer presents does not show this.

The brain region that he points to as being “associated with feelings of love and compassion” (the insular cortex) is active in as many as one-third of all brain imaging studies.

Further, in studies of decision making the insular cortex is more often associated with negative than positive emotions.

The kind of reasoning that Mr. Lindstrom uses is well known to be flawed, because there is rarely a one-to-one mapping between any brain region and a single mental state; insular cortex activity could reflect one or more of several psychological processes.

We find it surprising that The Times would publish claims like this that lack scientific validity.

RUSSELL POLDRACK
Austin, Tex., Oct. 3, 2011

The writer is a professor of psychology and neurobiology at the University of Texas at Austin. His letter was signed by 44 other neuroscientists.
Does reverse inference work?

\[ p(\text{process}|\text{act}) = \frac{p(\text{process}) \cdot p(\text{act}|\text{process})}{p(\text{act})} \]
Insula activation is weakly selective

Some voxels active in as many of 20% of studies

Yarkoni et al., 2011
Reverse inference

- Informal reverse inference provides relatively weak evidence

TICS, 2006
Formalizing reverse inference

- How can we more formally test the predictive ability of fMRI?
- Answer: statistical methods for prediction
- Machine learning/statistical learning/pattern recognition
Creating meta-analytic brain maps

  - Automatically extracts activation tables from fMRI papers for 17 journals
  - Current database has 5809 papers
  - Good accuracy
    - 84% sensitivity, 97% specificity against SumsDB manual database
- Meta-analytic maps created for each paper
  - 10mm sphere placed at each focus

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>57</td>
<td>-6</td>
</tr>
<tr>
<td>33</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>42</td>
<td>6</td>
<td>51</td>
</tr>
<tr>
<td>24</td>
<td>-3</td>
<td>57</td>
</tr>
</tbody>
</table>
Automated meta-analysis of the term "working memory"

Analysis details

- # of studies: 363 [view]
- % active voxels: 4.6%

Selected location

- Posterior probability: 69%
- Coords (x,y,z): -28 4 56

Search again: working memory

Thumbnails of brain images with highlighted regions of interest.
Automated meta-analysis

A Term-based search

“Pain”

Related studies

Mechanisms of Directed An fMRI Investigation of Placebo-Induced Changes in fMRI in the Anticipation and Experience of Pain

1. Negative affect
2. Positive affect
3. Pain

Automated coordinate extraction

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>-23</td>
<td>18</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>-40</td>
<td>0</td>
<td>-16</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>-41</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>33</td>
<td>2</td>
</tr>
</tbody>
</table>

Meta-analysis

\[ P(\text{Pain} | \text{Activation}) \]

B Forward inference

Pain

Working Memory?

Emotion?

Pain?

Reverse inference

C Classification

Working mem. Emotion Pain

\[ P = 78\% \quad P = 64\% \quad P = 87\% \]

Select highest probability

“Pain”

Yarkoni et al., 2011, *Nature Methods*
Automated meta-analysis

Yarkoni et al., 2011, *Nature Methods*
Classification of cognitive states

- Given 2+ terms, can determine which is most likely given the data.
- Naive Bayes classifier: assumes that all features (voxels) are independent; selects the most probable class.
- Can apply this to any activation map—studies, individual subjects, etc.

Classification of new studies

- Cross-validated classification of all studies in database
- Select 25 high-frequency terms
- Pairwise classification: how well can we distinguish between each pair of terms?

Figure 8: Pairwise binary classification accuracy for single label prediction

But the SVM does have a slight edge over the logistic regression model. As was expected, the ensemble model of the three is the clear winner, although it only marginally beats the SVM.

Use majority vote of naive Bayes, \textit{l1}-regularized logistic regression, and linear SVM

Madhura Parikh,
Subhashini Venugopalan
Sanmi Koyejo
Automating reverse inference

Table 2. Pearson correlations between searchlight classification map and NeuroSynth term-based reverse inference activation maps

<table>
<thead>
<tr>
<th>Term</th>
<th>Correlation (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.1451</td>
</tr>
<tr>
<td>Working</td>
<td>0.1159</td>
</tr>
<tr>
<td>Numerical</td>
<td>0.1157</td>
</tr>
<tr>
<td>Letter</td>
<td>0.1081</td>
</tr>
<tr>
<td>Attention</td>
<td>0.1062</td>
</tr>
<tr>
<td>Correct</td>
<td>0.1060</td>
</tr>
<tr>
<td>Cue</td>
<td>0.0995</td>
</tr>
<tr>
<td>Preparatory</td>
<td>0.0970</td>
</tr>
<tr>
<td>Load</td>
<td>0.0959</td>
</tr>
<tr>
<td>Hand</td>
<td>0.0924</td>
</tr>
</tbody>
</table>

The 10 most highly correlated terms are listed. From Yarkoni et al. (26).

Helfinstein et al, 2014, PNAS
What about individual subjects?

- Can we identify cognitive states in individual (new) subjects?

- Difficult, because:
  - No opportunity for training
  - Data is of a fundamentally different type
  - Tested in samples of subjects from working memory, emotion, and pain studies

- Can we predict source study type?

Classifying individual subjects

Yarkoni et al, 2011, Nature Methods
Using classification to understand mental structure

WM: working memory
TS: Task switching
RS: Response selection
RI: Response inhibition
CC: Cognitive control
BI: Bilingual language

Towards meta-analytic testing of cognitive theories

Model 1

- inhibition
  - accuracy on antisaccade task
  - SSRT on stop signal task

- updating
  - accuracy on tone counting task
  - 2-back versus 0-back accuracy

Model 2

- executive function

Observed covariance
Neural computations underlying action-based decision making in the human brain

Cortical substrates for exploratory decisions in humans

Nucleus Accumbens D2/3 Receptors Predict Trait Impulsivity and Cocaine Reinforcement

Authors: K.W., A.R., and J.P.O.
designed research; K.W. performed research; A.R. and J.P.O. analyzed data; A.R. and J.P.O. wrote the paper.

Author contributions: K.W., A.R., and J.P.O. designed research; K.W. performed research; A.R. and J.P.O. analyzed data; A.R. and J.P.O. wrote the paper.

Kerry Watanabe, Andrew Romer, and John P. O'Doherty designed research; performed research; analyzed data; and wrote the paper.

Kerry Watanabe, Andrew Romer, and John P. O'Doherty designed research; performed research; analyzed data; and wrote the paper.

K.W., A.R., and J.P.O. designed research; K.W. performed research; A.R. and J.P.O. analyzed data; A.R. and J.P.O. wrote the paper.

K.W., A.R., and J.P.O. designed research; K.W. performed research; A.R. and J.P.O. analyzed data; A.R. and J.P.O. wrote the paper.

K.W., A.R., and J.P.O. designed research; K.W. performed research; A.R. and J.P.O. analyzed data; A.R. and J.P.O. wrote the paper.

K.W., A.R., and J.P.O. designed research; K.W. performed research; A.R. and J.P.O. analyzed data; A.R. and J.P.O. wrote the paper.
Topic Mapping

- Each document has a loading on each topic
  - On average, each document loads on ~6.5 topics
- Used ACE to extract activation coordinates for all 5,809 papers
- Perform voxelwise chi-square test with FDR correction to examine association between topics and activation

<table>
<thead>
<tr>
<th>Topic</th>
<th>Documents</th>
<th>Activation Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>emotion</td>
<td>&quot;...amygdala...emotion...negative...&quot;</td>
<td></td>
</tr>
<tr>
<td>negative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>unpleasant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Poldrack et al., 2012, *PLOS Comp Biol*
Topic 61 (442 docs): memory working_memory maintenance visual_working_memory spatial_working_memory manipulation episodic_buffer retention rehearsal retrieval

Poldrack et al., 2012, PLOS Comp Biol
Topic 3 (389 docs): memory episodic_memory recall learning verbal_memory association encoding risk visual_memory working_memory

Poldrack et al., 2012, PLOS Comp Biol
Topic 106 (391 docs): movement coordination motor_control feedback planning integration goal context knowledge learning

Poldrack et al., 2012, PLOS Comp Biol
Clustering disorders by topic maps

Poldrack et al., 2012, *PLOS Comp Biol*
Mega-analysis of fMRI data

26 tasks, 482 images from 338 subjects

Poldrack et al., 2013,
Frontiers in Neuroinformatics
Classification results

Whole-brain with linear SVM: 48% accuracy

Poldrack et al., 2013, *Frontiers in Neuroinformatics*
Conclusions

• Cognitive neuroscience needs to get formal about describing the mental processes that are being mapped to brain function

• Much interesting structure can be extracted using text mining, but ultimately progress will require manual annotation by domain experts

• Ontologies plus databases will provide the means to ask whether the claims of psychology regarding mental architecture are respected by the brain
Acknowledgments

UCLA
- Robert Bilder
- Don Kalar
- Fred Sabb
- D. Stott Parker

UT Austin
- Jeanette Mumford
- Sanmi Koyejo
- Tal Yarkoni

Carnegie-Mellon
- Niki Kittur

Get involved!
www.cognitiveatlas.org

James S. McDonnell Foundation

National Institute of Mental Health
NSF