cogs 105 this week

BIGDATA

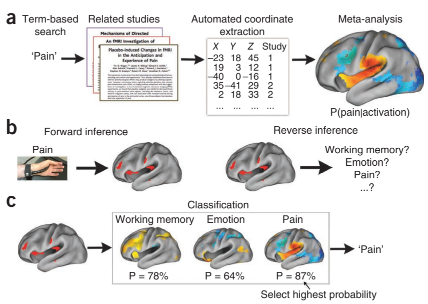
today: neurosynth

BIGDATA

culture (Google Ngram)

the brain (NeuroSynth)

BIGDATA



the brain (NeuroSynth)

NeuroSynth

The development of noninvasive neuroimaging techniques such as functional magnetic resonance imaging (fMRI) has spurred rapid growth of literature on human brain imaging in recent years. In 2010 alone, more than 1,000 fMRI articles had been published[1]. This proliferation has led to substantial advances in our understanding of the human brain and cognitive function; however, it has also introduced important challenges. In place of too little data, researchers are now besieged with too much. Because individual neuroimaging studies are often underpowered and have relatively high false positive rates[2–4], multiple studies are required to achieve consensus regarding even broad relationships between brain and cognitive function. It is therefore necessary to develop new techniques for the large-scale aggregation and synthesis of human neuroimaging data[4–6].

Method

- Looked for neuroimaging papers that contained a list of words of interest (e.g., pain, memory, etc.).
    - Word had to occur at a frequency of 1 per 1,000 words (.001%).
- They wrote a “content identifier” technique (discussed in last class) to scrape out the brain activation data from tables in those papers.
- They (basically) correlate the words with what areas of the brain tend to be activated.

Database Status

386455 activations reported in 10903 studies

Interactive, downloadable meta-analyses of 3342 terms

Functional connectivity and coactivation maps for over 150,000 brain locations

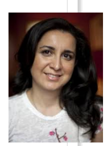
http://neurosynth.org/

Example Paper

- Here is what they are “scraping” using their content identifier (see big-data strategies from last class).

Learning to Sample: Eye Tracking and fMRI Indices of Changes in Object Perception

Lauren L. Emberson[1] and Dima Amso[2]



Abstract

■ We used an fMRI/eye-tracking approach to examine the mechanisms involved in learning to segment a novel, occluded object in a scene. Previous research has suggested a role for effective visual sampling and prior experience in the development of mature object perception. However, it remains unclear how the naive system integrates across variable sampled experiences to

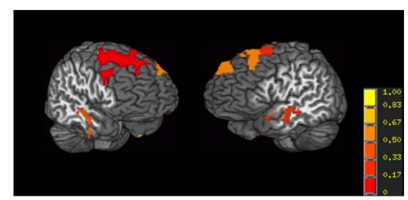
not incorporate the Target Object. We found that, relative to the Control condition, participants in the Training condition were significantly more likely to change their percept from “disconnected” to “connected,” as indexed by pretraining and posttraining test performance. In addition, gaze patterns during Target Scene inspection differed as a function of variable object

Coordinate System

Table 2. Main Effect of Scene Type

		Coordinates		
Side	*Areas (Paired > Target)*	*X*	*Y*	*Z*
R	Cuneus	18	−75	24
R	Fusiform gyrus	40	−66	−13
R	Inferior parietal lobule	39	−42	26
R	Inferior temporal gyrus	51	−61	−10
R	Middle occipital gyrus	40	−77	−10
R	Middle temporal gyrus	59	−49	−4
R	Parahippocampal gyrus	32	−21	−24
R	Posterior cingulate	16	−6	11
R	Preceuneus	24	−49	31
L	Caudate nucleus	−23	−34	2
L	Declive	0	−57	−11
L	Fusiform gyrus	−40	−66	−11
L	Inferior parietal lobule	−41	−54	46
L	Middle occipital gyrus	−31	−61	2
L	Parahippocampal gyrus	−24	−42	2
L	Posterior cingulate	−24	−61	18
L	Precuneus	24	−76	35
L	Superior parietal lobule	−31	−61	57
L	Thalamus	−21	−27	5

Corrected to $p < .05$.



journal articles contain these x,y,z coordinates where significant brain activity was detected in an experiment

Veracity?

- Last class we discussed veracity: How do we know that Yarkoni et al. extracted the tables correctly? What if some papers have their data extracted with incorrect locations, or incorrect activations, etc.?
- Again, coupling big data with other tests of reliability (as in the English lexicon/dictionary example from Ngram):
    - They checked another database in which some of the authors of the studies had entered their data.

Amazing Things

- We can do **forward vs. reverse inference** about how brain area activation relates to cognitive function.
- We can do cognitive state detection by **using brain areas to classify** which cognitive state is taking place.

Amazing Things

- We can do **forward vs. reverse inference** about how brain area activation relates to cognitive function.
- We can do cognitive state detection by **using brain areas to classify** which cognitive state is taking place.

Forward vs. Reverse

- **Forward** inference:
    - Calculating the probability that you detect activation in a given brain area *given* that a term is mentioned in the paper.
    - For example:
      P(activation in left frontal | "language" occurs)
    - We can calculate these probabilities because NeuroSynth has so much data in it.

Conditional Probability

(cream, sugar)
(cream)
(cream, sugar)   5 coffees ordered
(cream)
(.)

P(cream) = 4/5 = .8

P(sugar) = 2/5 = .4

P(sugar | cream) = P(sugar, cream) / P(cream)

= (2/5) / .8 = .4 / .8 = .5

Conditional Probability

(dACC, pain)
(APFC, pain)
(dACC, language)
(APFC, pain)
(APFC, emotion)
…

forward

P(dACC | pain) = P(dACC, pain) / P(pain)

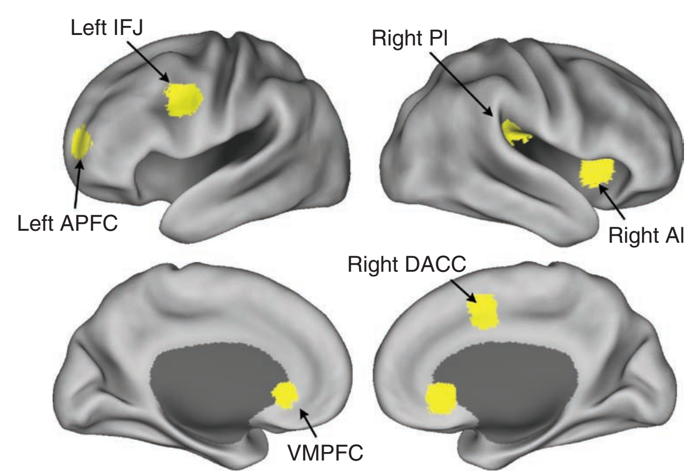
reverse

P(pain | dACC) = P(dACC, pain) / P(dACC)

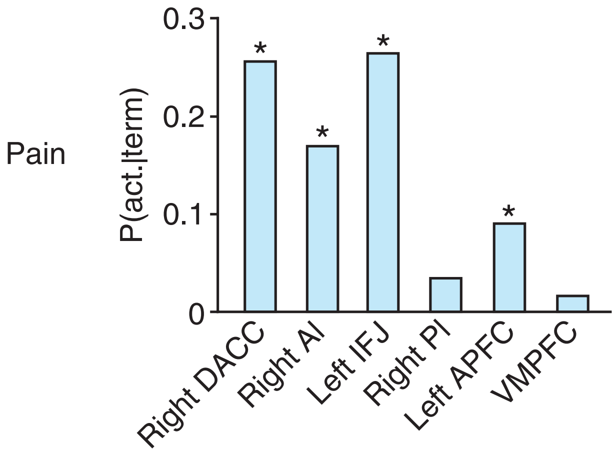
Forward vs. Reverse

- **Reverse** inference:
  - Calculating the probability that a term is mentioned in the paper *given* that you see activation in a given brain area.
  - For example:
    P("language" occurs | left front lobe activation)
  - We can calculate these probabilities because NeuroSynth has so much data in it.

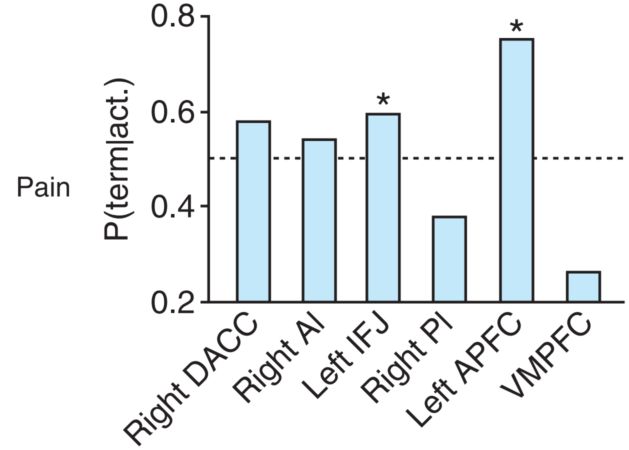
threshold of 0.05; (whole-brain false discovery rate, (*q*) = 0.05). DACC, dorsal anterior cingulate cortex (stereotactic coordinates in Montreal Neurological Institute space: +2, +8, +50); AI, anterior insula (+36, +16, +2); IFJ, inferior frontal junction (−50, +8, +36); PI, posterior insula (+42, −24, +24); APFC, anterior prefrontal cortex (−28, +56, +8); VMPFC, ventromedial prefrontal cortex (0, +32, −4). Dashed lines indicate even



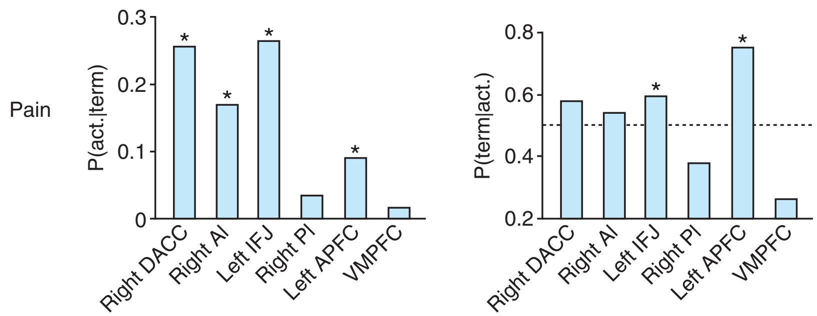
Forward



Reverse

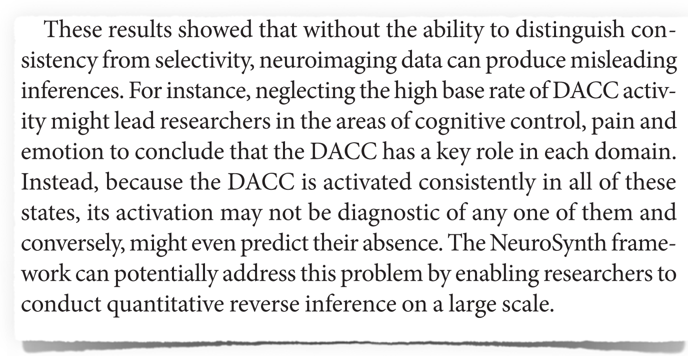


????



Paradox

- Why would forward and reverse inference generate different patterns of results?



These results showed that without the ability to distinguish consistency from selectivity, neuroimaging data can produce misleading inferences. For instance, neglecting the high base rate of DACC activity might lead researchers in the areas of cognitive control, pain and emotion to conclude that the DACC has a key role in each domain. Instead, because the DACC is activated consistently in all of these states, its activation may not be diagnostic of any one of them and conversely, might even predict their absence. The NeuroSynth framework can potentially address this problem by enabling researchers to conduct quantitative reverse inference on a large scale.

dACC = dorsal anterior cingulate cortex

Amazing Things

- We can do **forward vs. reverse inference** about how brain area activation relates to cognitive function.
- We can do cognitive state detection by **using brain areas to classify** which cognitive state is taking place.

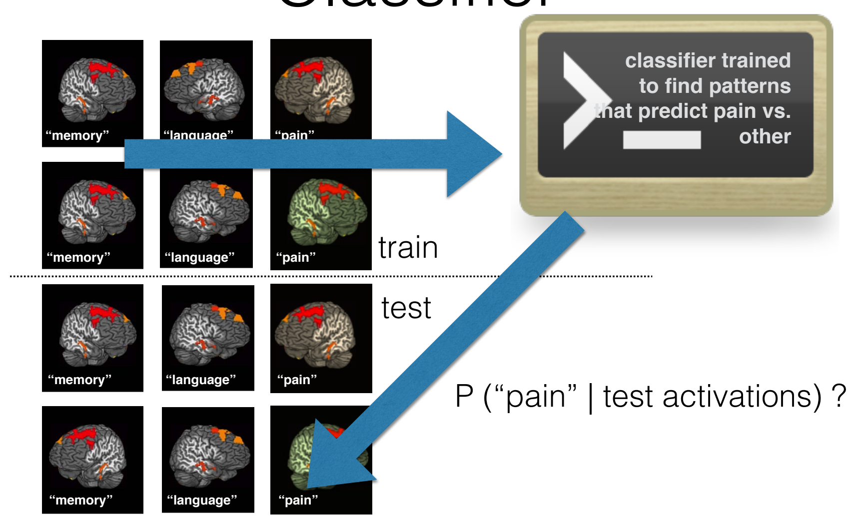
Amazing Things

- We can do **forward vs. reverse inference** about how brain area activation relates to cognitive function.
- We can do cognitive state detection by **using brain areas to classify** which cognitive state is taking place.

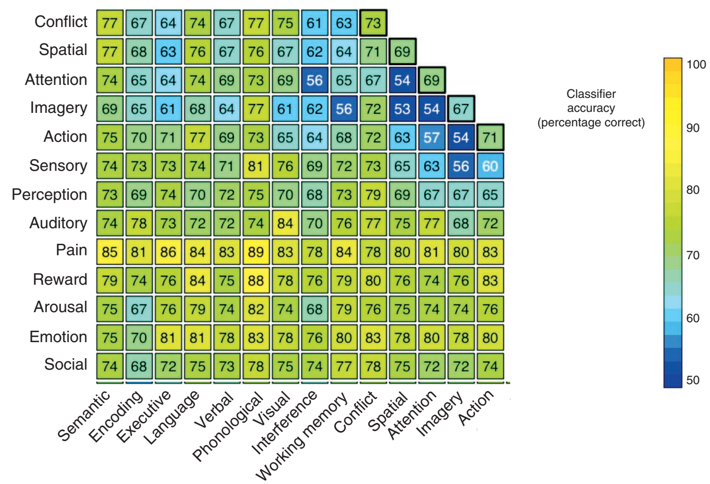
Classifiers

- The precise details of the method in Yarkoni et al. are outside the scope of our interest here, but it is still important to know something basic about their method: "classifiers."
- Big data are being used with **machine learning** techniques: techniques that can train computer systems to do intelligent (or seemingly intelligent) things, like **classify**.

Classifier



Classification of Cognitive States



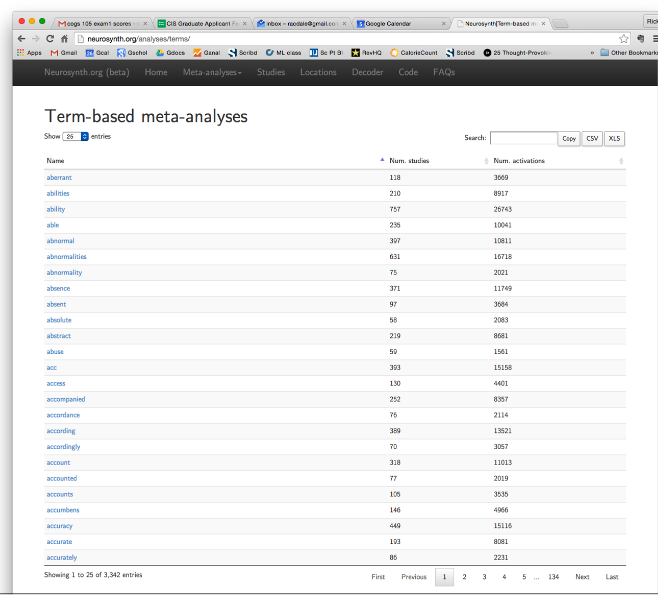
Conclusion from Classification

Second, we decoded broad psychological states in a relatively open-ended way in individual subjects; this was, to our knowledge, the first application of a domain-general classifier that can distinguish a broad range of cognitive states based solely on prior literature. The ability to decode brain activity without previous training data or knowledge of the 'ground truth' for an individual is particularly promising. Our results raise the prospect that legitimate 'mind reading' of more nuanced cognitive and affective states might eventually become feasible with additional technical advances. However, the present NeuroSynth implementation

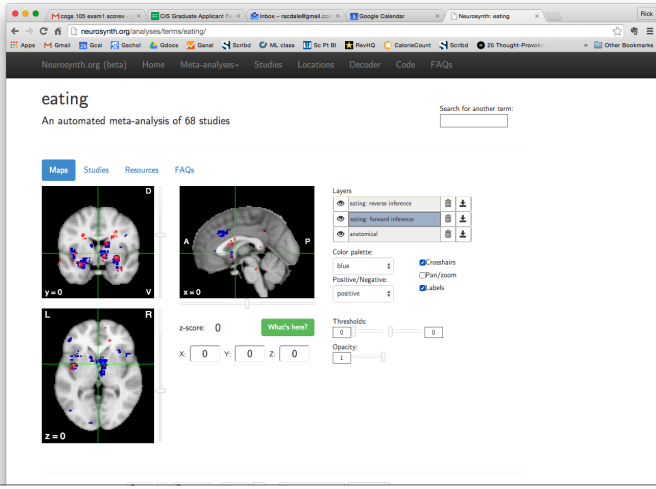
Amazing Things

- We can do **forward vs. reverse inference** about how brain area activation relates to cognitive function.
- We can do cognitive state detection by **using brain areas to classify** which cognitive state is taking place.

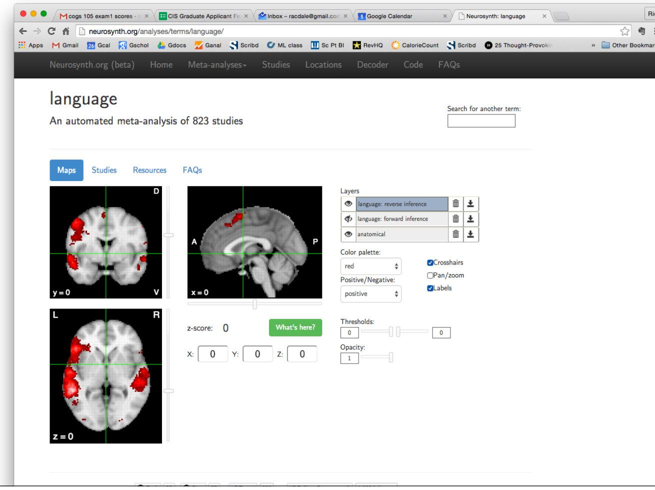
http://neurosynth.org/analyses/terms/



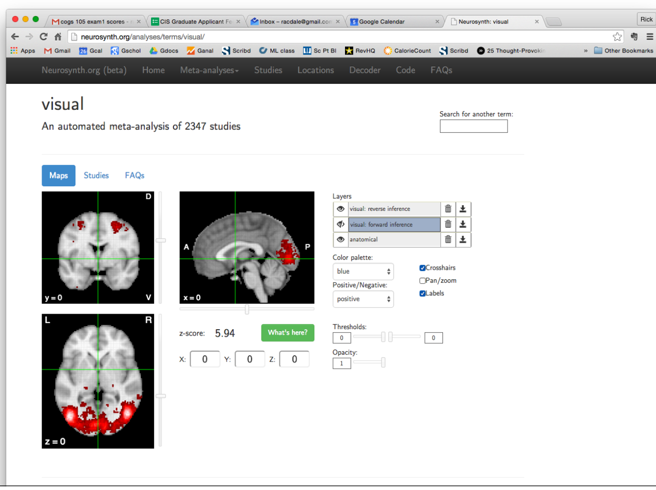
Eating (68 studies)



Language (823 studies)

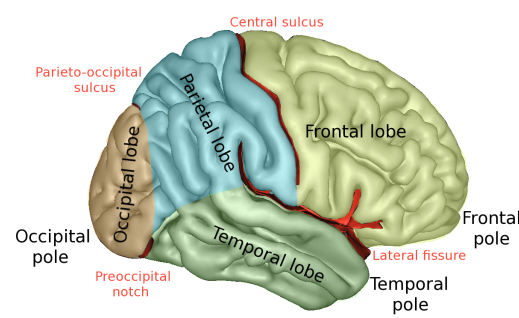


Visual (2,347 studies)



For Lab Next Week

- We will not concern ourselves with highly specific brain areas (though you are welcome to consider them if you know them); instead let's focus on exploration based on very high-level systems neuroscience: the lobes.



Big Data

- Today: the basics, and culturomics.
- Thursday: NeuroSynth.
- Next week: language analysis and modeling.