Deep sampling and individual differences in fMRI

Catherine Walsh, Ph.D. and Tyler Morgan, Ph.D.
Section on Functional Imaging Methods and Functional MRI Core Facility, NIMH
FMRIF Summer Neuroimaging Course
July 30, 2026

1

Deep sampling and **individual differences** in fMRI

Catherine Walsh, Ph.D. and Tyler Morgan, Ph.D.
Section on Functional Imaging Methods and Functional MRI Core Facility, NIMH
FMRIF Summer Neuroimaging Course
July 30, 2026

2

Outline

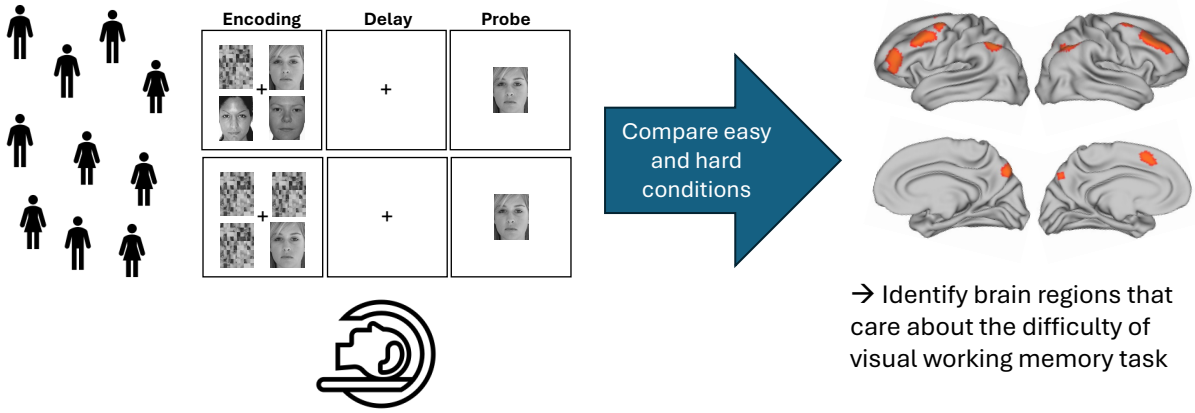
- **What** are individual differences?
- **Why** study individual differences?
- **How** to individual differences?

3

What are individual differences?

4

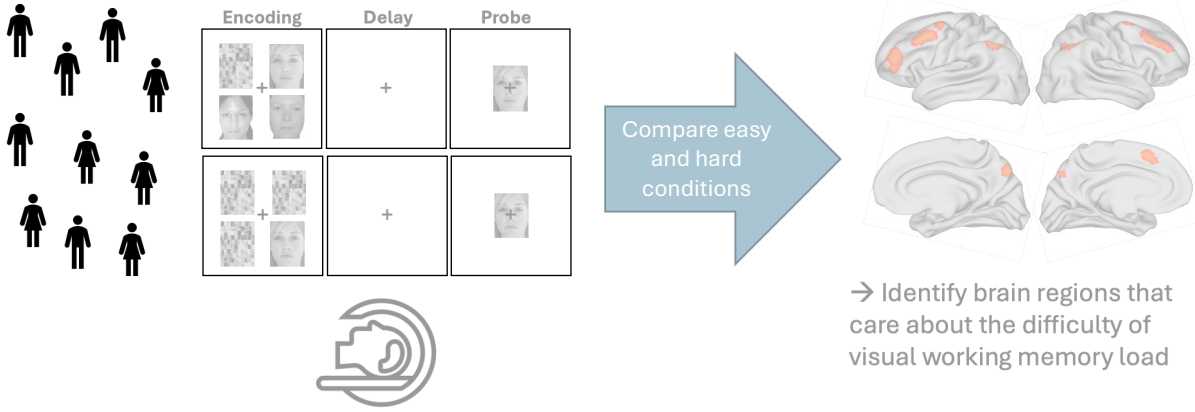
Traditional fMRI (group-level) analyses



→ Identify brain regions that care about the difficulty of visual working memory task

5

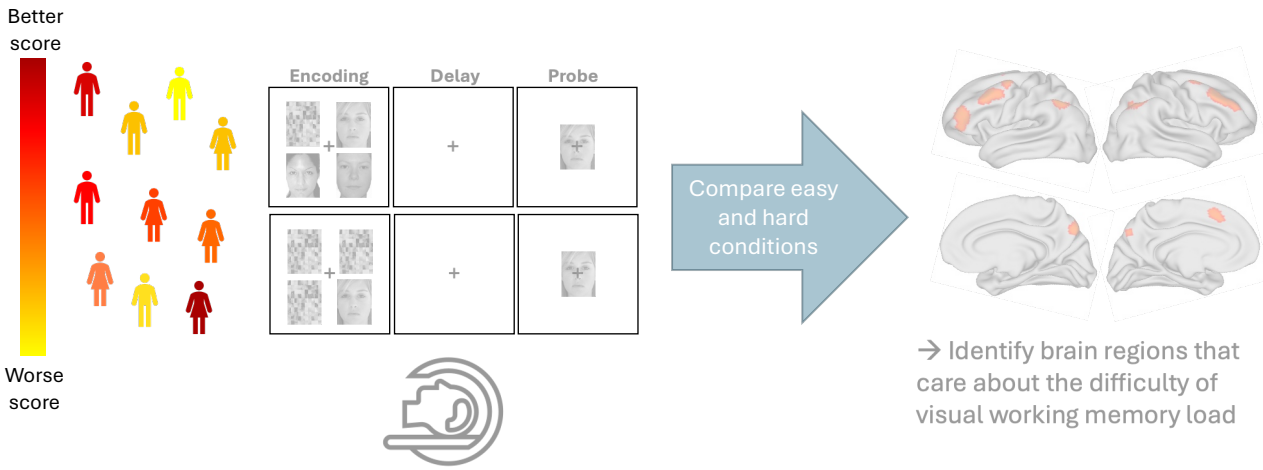
Traditional fMRI (group-level) analyses



→ Identify brain regions that care about the difficulty of visual working memory load

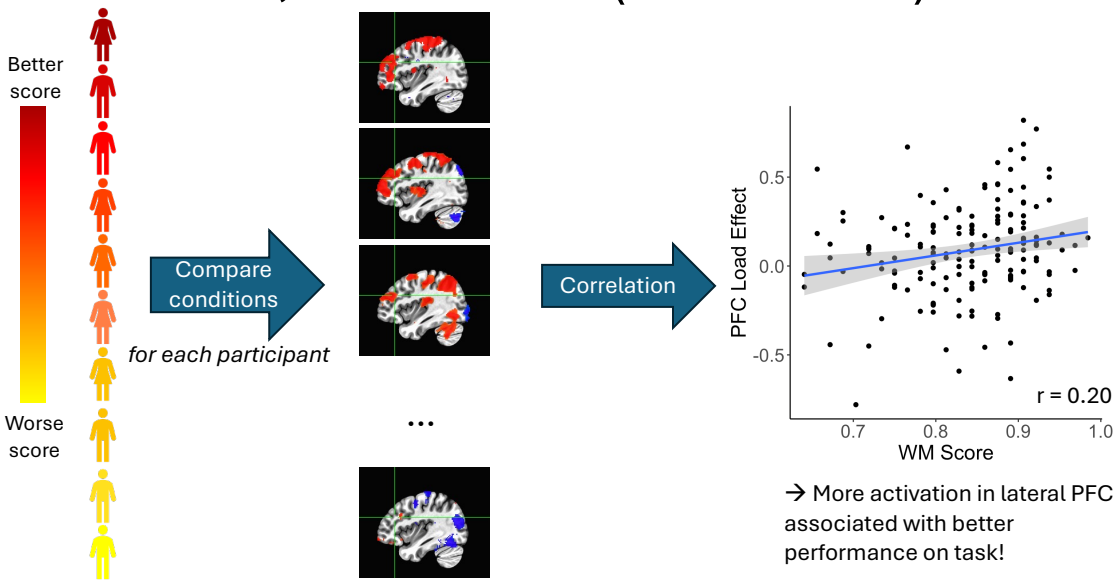
6

But wait, there's more (information!)



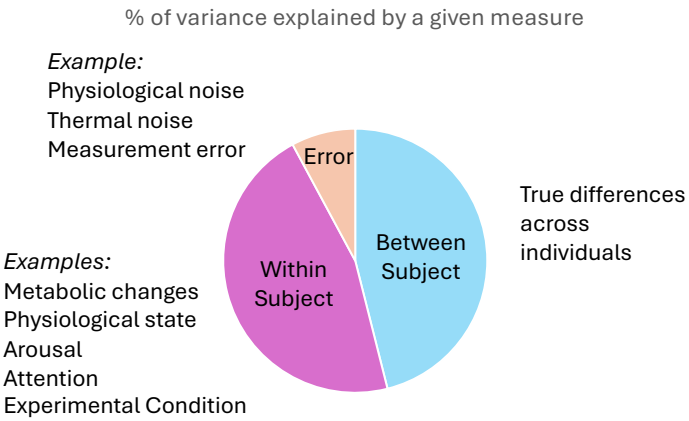
7

But wait, there's more (information!)



8

But wait, there's more (information!)



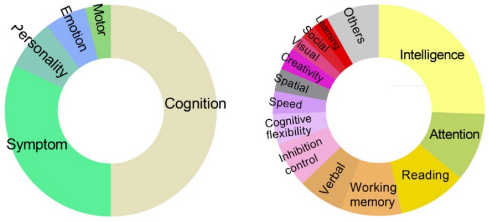
Between and within subject variance can both be interesting targets of prediction, they just are asking different questions!

9

Why study individual differences?

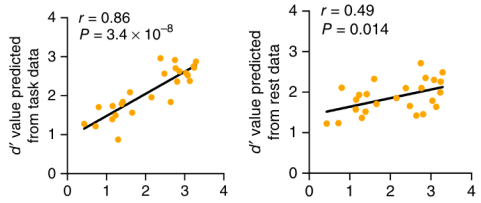
10

Predicting cognition



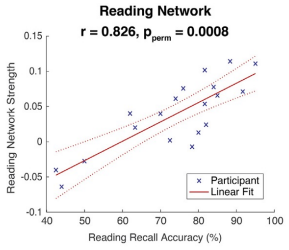
Sui et al., Biol Psych 2022

Sustained Attention



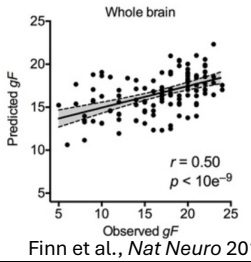
Rosenberg et al., *Nat Neuro* 2015

Reading Ability



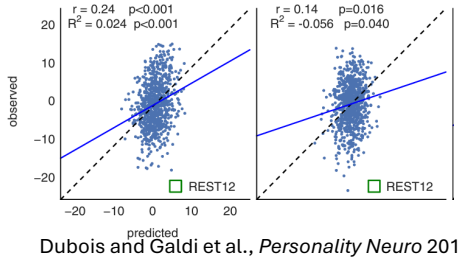
Jangraw et al., *NeuroImage* 2018

Fluid Intelligence



Finn et al., *Nat Neuro* 2015

Openness and Conscientiousness



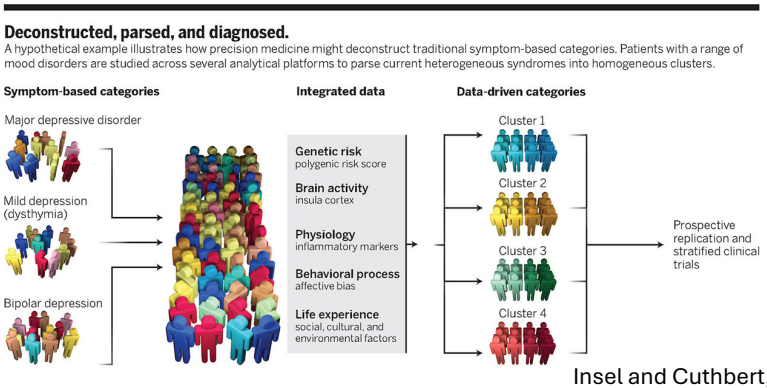
Dubois and Galdi et al., *Personality Neuro* 2018

11

Psychiatric symptoms as dimensional, not categorical

Research Domains Criteria (RDoC)

Goal: "to develop, for research purposes, new ways of classifying mental disorders based on dimensions of observable behavior and neurobiological measures"



Insel and Cuthbert, Science 2015

So what?

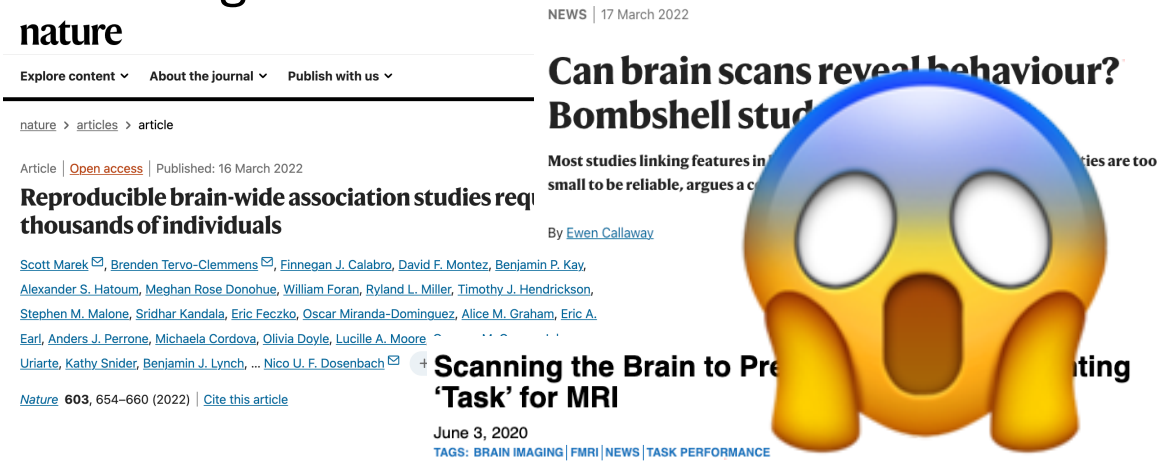
- Better understanding of disease pathophysiology
- Track disease progression
- Develop new (more effective) treatments
- Understand who will respond to which treatments

12

But wait!

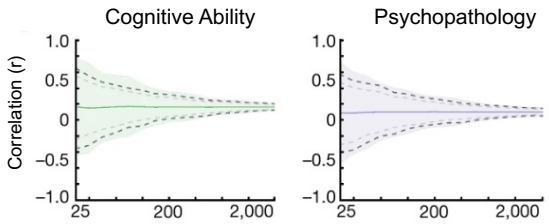
13

Are most individual differences even meaningful?!?



14

We need thousands of individuals?!?

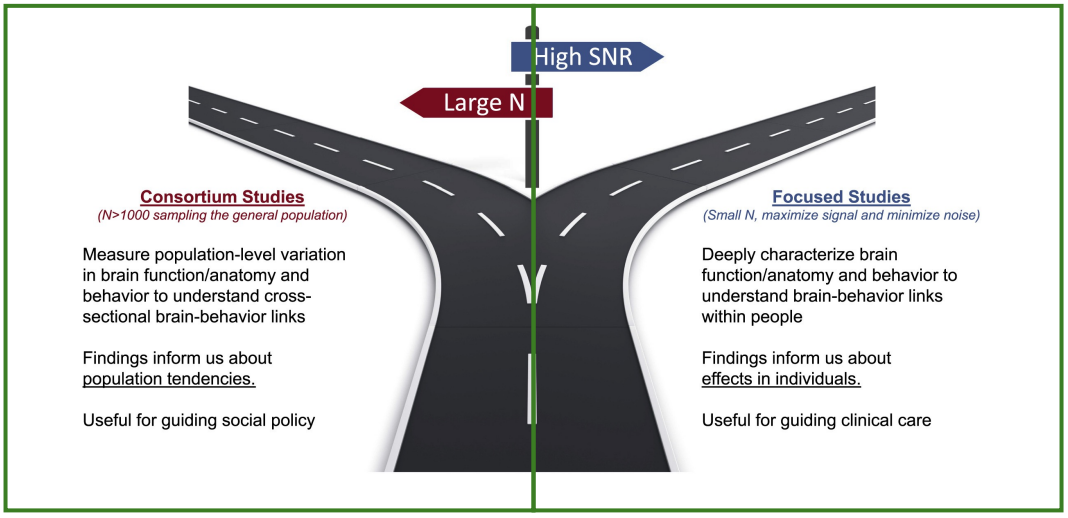


→ Widely varying effect sizes at small sample sizes – even producing completely opposite results!

Marek et al., *Nature* 2022

15

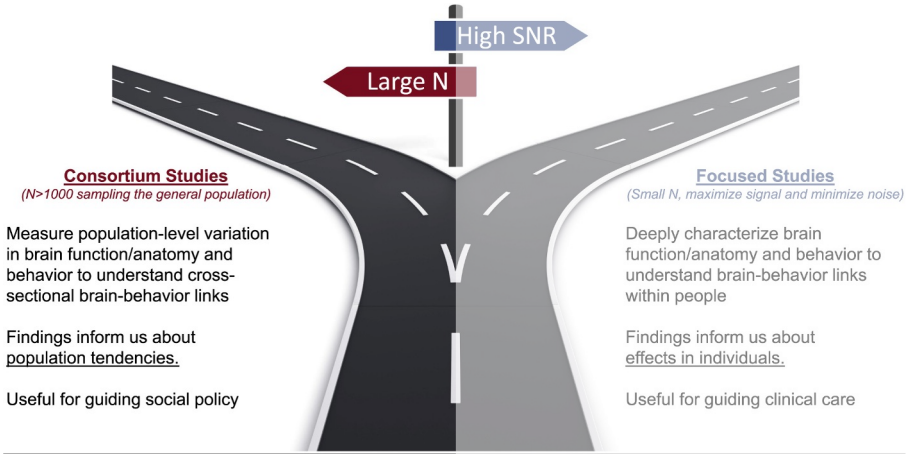
Two paths forward



Gratton et al., *Neuron* (2022)

16

Two paths forward



Gratton et al., *Neuron* (2022)

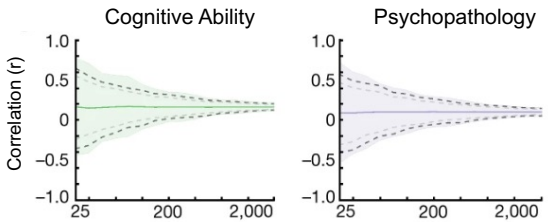
17

How to study* individual differences?

*robustly

18

Collect lots of subjects!



Marek et al., *Nature* 2022

Benefits of a larger sample size:

- Correlations get more reliable!
- Wider distribution of phenotypes
- Allows us to use more robust statistical/machine learning methods like cross validation

Problems with using a larger sample size:

- Expensive
- Time consuming
- Hard to recruit patients

19

So what if I can't recruit thousands of subjects?

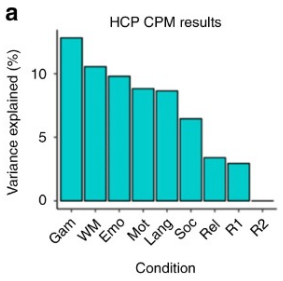
→ Consider the measures you use: task vs rest

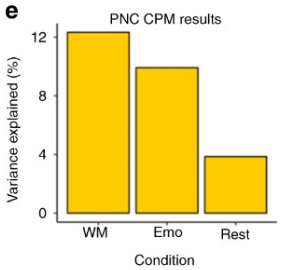
Many studies use resting state fMRI to predict behavior

- Relatively reliable across session ("trait-like")
- Reflects functional networks that are a "backdrop" to anything that happens during task
- Easy to measure – no task to learn, requires relatively little scan time (~15 minutes)
- Many big, open datasets include it!

BUT: is rest the best for prediction?

Maybe not.



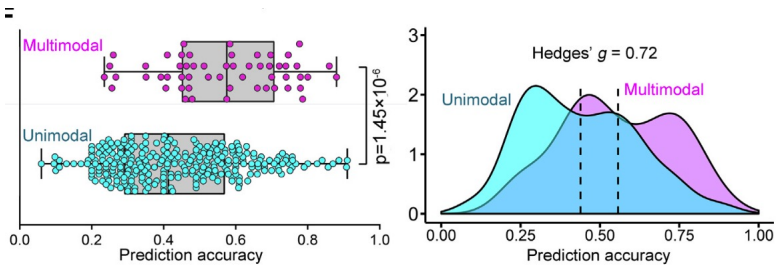


Greene et al., *Nat Comms* 2018

20

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: task vs rest (or both!?)



Sui et al., *Biol Psych*. 2020

→ Integrating across neuroimaging features can improve prediction performance and leverage unique aspects of brain structure and function to better characterize individual differences

21

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: self-report vs cognitive task

Take for example: face blindness (prosopagnosia)

Option 1: Self-report measures

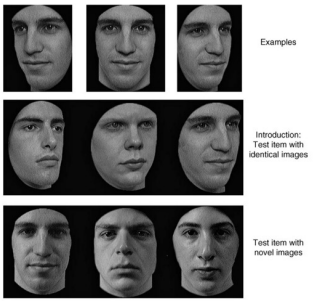
20 item prosopagnosia index (PI20)

1	My face recognition ability is worse than most people
2	I have always had a bad memory for faces
3	I find it notably easier to recognize people who have distinctive facial features
4	I often mistake people I have met before for strangers
5	When I was at school I struggled to recognize my classmates

Shah et al., *R. Soc. Open sci*, 2015

Option 2: Data from cognitive tasks

Cambridge Face Memory Test



Duchaine et al., Neuropsychologia 2006

22

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: self-report vs cognitive task

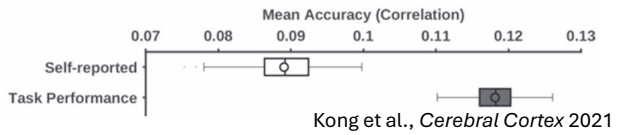
Take for example: face blindness (prosopagnosia)

Option 1: Self-report measures

- Pros:
  - Easy to administer
  - Might be more relevant for psychiatric conditions
  - Potentially more stable within an individual
- Cons:
  - Potential for bias
  - Relies on participant's ability to introspect

Option 2: Data from cognitive tasks

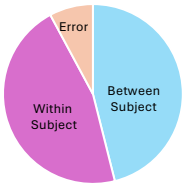
- Pros:
  - Less potential for bias
  - Able to do in scanner
  - Easier to predict?
- Cons:
  - Might have less relevance for "biomarkers"
  - Potentially less generalizable



Kong et al., *Cerebral Cortex* 2021

23

So what if I can't recruit thousands of subjects?



→ Consider the measures you use: optimize sources of variance

Between and within subject variance can both be interesting targets of prediction, they just are asking different questions!

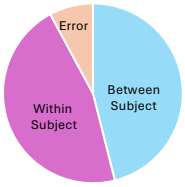
→ Different questions necessitate different tasks

Within Subject Effects
(Group/Condition Differences)

$$t = \frac{Difference\ in\ conditions}{Error}$$

24

So what if I can't recruit thousands of subjects?



→ Consider the measures you use: optimize sources of variance

Between and within subject variance can both be interesting targets of prediction, they just are asking different questions!

→ Different questions necessitate different tasks

Within Subject Effects (Group/Condition Differences)

$$t = \frac{Difference\ in\ conditions}{Variance\ between\ individuals + Error}$$

Between Subject Effects (Individual differences)

$$ICC = \frac{Variance\ between\ individuals}{Error}$$

→ Tasks that are optimized for within-subjects effects may be poorly optimized for between-subjects effects

See Hedge et al., *Beh. Res. Methods* 2017 for more info/math

25

So what if I can't recruit thousands of subjects?

→ Consider what measures you use: maximize reliability

Key assumption of individual differences: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**

Trial 1 ✓
Trial 2 ✓
Trial 3 ✗
Trial 4 ✓
Trial 5 ✓
Trial 6 ✗

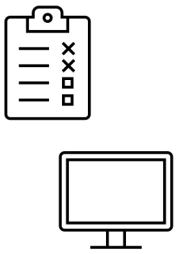
Overall accuracy: 0.66

26

So what if I can't recruit thousands of subjects?

→ Consider what measures you use: maximize reliability

Key assumption of individual differences: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**



Trial 1 ✓ Trial 2 ✓
Trial 3 ✗ Trial 4 ✓
Trial 5 ✓ Trial 6 ✗

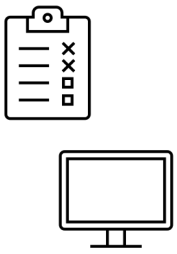
Accuracy 1: 0.66 Accuracy 2: 0.66

27

So what if I can't recruit thousands of subjects?

→ Consider what measures you use: maximize reliability

Key assumption of individual differences: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**



Trial 1 ✓ Trial 2 ✓
Trial 3 ✗ Trial 5 ✓
Trial 4 ✓ Trial 6 ✗

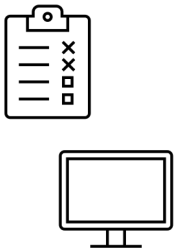
Accuracy 3: 0.66 Accuracy 4: 0.66

28

So what if I can't recruit thousands of subjects?

→ Consider what measures you use: maximize reliability

Key assumption of individual differences: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**



Trial 5 ✓ Trial 2 ✓
Trial 6 ✗ Trial 1 ✓
Trial 4 ✓ Trial 3 ✗

Accuracy 5: 0.66 Accuracy 6: 0.66

Split-halves reliability: the task you're using is internally consistent (i.e. task is measuring the same construct all the way through)

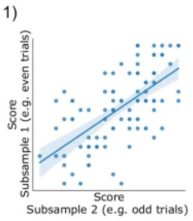
→ *Do all tasks show acceptable levels of split-halves reliability??*

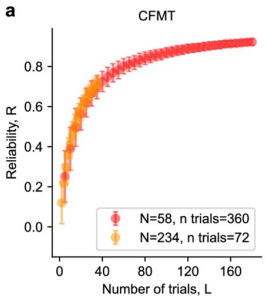
29

So what if I can't recruit thousands of subjects?

→ Consider what measures you use: maximize reliability

Key assumption of individual differences: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**





→ Number of trials it takes to reach an "acceptable" level of reliability varies across task

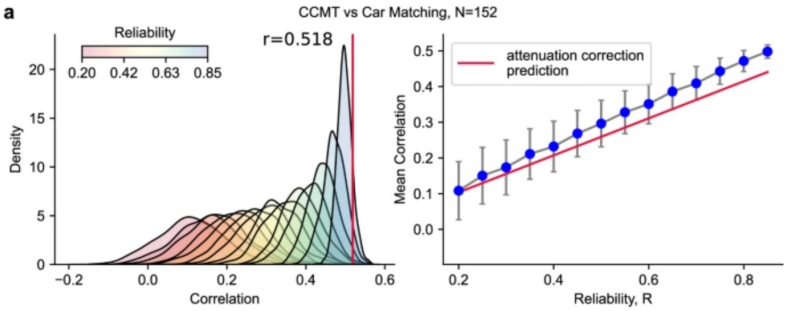
Kadlec, *Walsh* et al., *Comms Psych* 2024

30

So what if I can't recruit thousands of subjects?

→ Consider what measures you use: maximize reliability

Key assumption of individual differences: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**



→Less reliable tasks show attenuated correlations, even when they are truly related

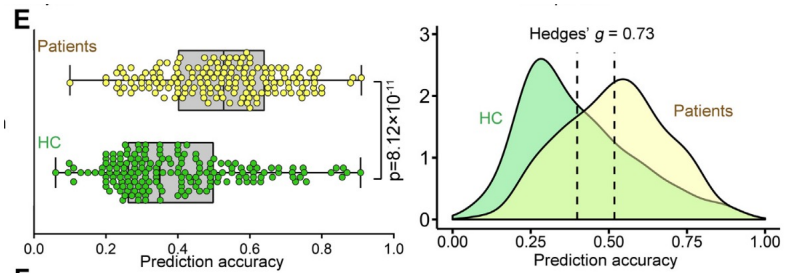
Kadlec, *Walsh* et al., *Comms Psych* 2024

*** Mathematically the same for brain-behavior, behavior-behavior, brain-brain, etc ***

31

So what if I can't recruit thousands of subjects?

→ Use an appropriate sample



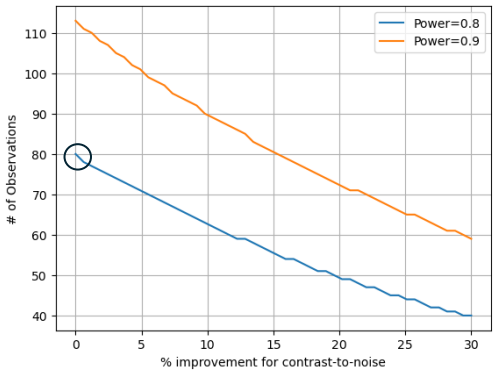
Sui et al., *Biol Psych.* 2020

Consider: if you're trying to find a biomarker for a specific psychiatric phenotype in a sample of healthy volunteers, there might not be enough between-subject psychiatric variance for a model to pick up on

32

So what if I can't recruit thousands of subjects?

→ Improve your data quality



Examples:

- Use different acquisition parameters
- Reduce head motion
- Better preprocessing methods

A 10% improvement in contrast-to-noise could mean a statistical power of 0.8 is possible with 63 vs 80 subjects

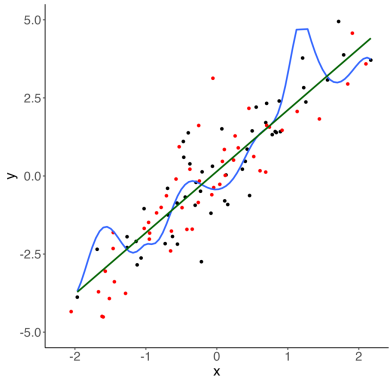
Slide adapted from Dan Handwerker

33

So what if I can't recruit thousands of subjects?

→ Use better statistical methods: Cross-Validation

x = randomly drawn from normal distribution
y = 2*x + noise

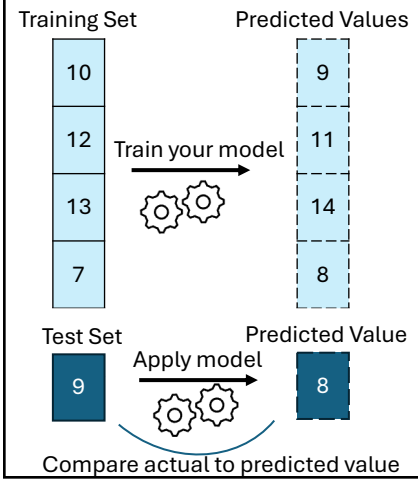


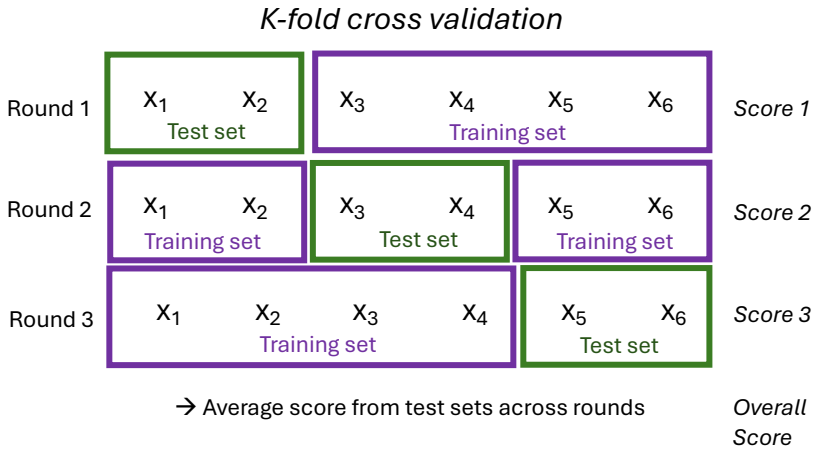
→ Creating your model in your entire sample can lead to fitting to sample specific characteristics (i.e. noise) and lead you to think you're doing better than you actually are

34

So what if I can't recruit thousands of subjects?

→ Use better statistical methods: Cross-Validation

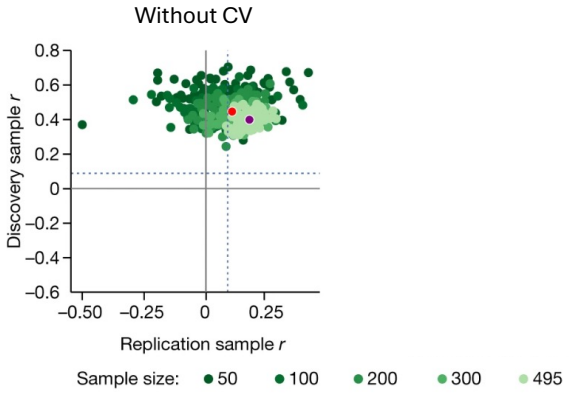




35

So what if I can't recruit thousands of subjects?

→ Use better statistical methods: Cross-Validation



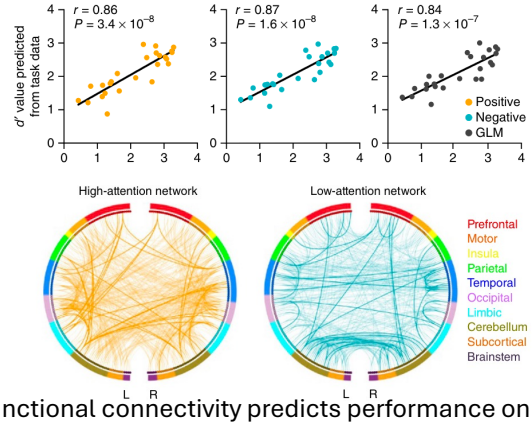
→ Without CV, effects are overestimated in discovery sample
→ CV can un-bias effect estimates

Spisak et al., *Nature* 2023

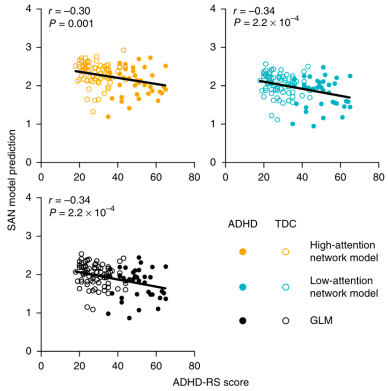
36

So what if I can't recruit thousands of subjects?

→ Use better statistical methods: Generalize to independent sample



Functional connectivity predicts performance on sustained attention task in healthy volunteers



Model can also successfully predict ADHD scores in an independent sample

Rosenberg et al., *Nat Neuro* 2015

37

Wrapping up: individual differences

- What are individual differences?
  - Statistical association between some sort of brain measure and some sort of behavioral measure
- Why should we care about individual differences?
  - They give us new insights into nuances behind neural processes and support the move towards brain-based psychiatry
- How can we robustly study individual differences in (large N) studies?
  - Think critically about which measures you're going to collect
  - Improve data quality of both the brain and behavior data
  - Use appropriate statistical/machine learning methods like cross validation and generalization in an independent sample

38

Acknowledgements

Section on Functional Imaging Methods and Functional MRI Core Facility

- Peter Bandettini
- Tyler Morgan
- Dan Handwerker
- Javier Gonzalez Castillo
- Pete Molfese
- Burak Akin
- Josh Faskowitz
- Sharif Kronemer
- Marly Rubin
- Marlene Smith

Weizmann Institute of Science

- Jan Kadlec
- Michal Ramot
- Jesse Rissman

Center for Cognitive Neuroscience

Questions? catherine.walsh@nih.gov

39

NIH National Institutes of Health
Turning Discovery Into Health

Shifting gears toward "Deep Sampling"

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

40

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Shifting gears toward “Deep Sampling”

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

41

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Shifting gears toward “Deep Sampling”

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

42

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Shifting gears toward “Deep Sampling”

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

43

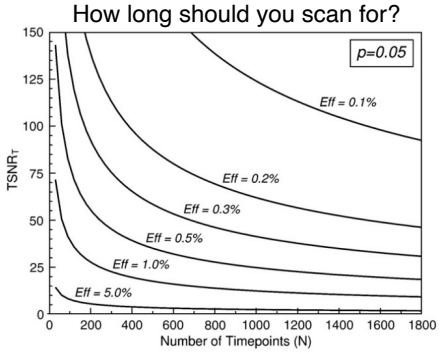
NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

What is Deep Sampling in fMRI?

- Most brain processes are complex, so it takes a lot of data to try to understand them.
- How much data is required depends on multiple factors:
  - TSNR - Signal strength compared to noise
  - eff - Effect size
  - erfc/P - Desired level of confidence
  - R - Proportion of time mapping the response

How long should you scan for?



$$N = \frac{2}{R(1-R)}\left(\frac{\text{erfc}^{-1}(P)}{(\text{TSNR})(\text{eff})}\right)^2$$

Murphy, et al. 2007

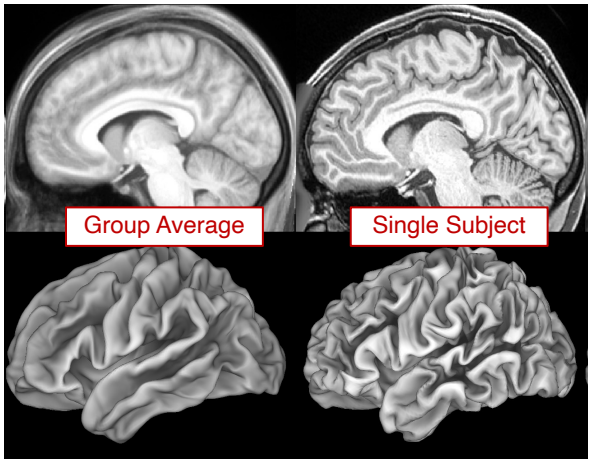
44

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

What is Deep Sampling in fMRI?

- Some noise sources are related to MRI itself, while some are related to subject variation.
- Many studies have tried to gain insight based on limited sampling from many individuals
  - This approach has a lot of power to uncover individual differences and population effects
- However, averaging across individuals results in the loss of individual details
  - Includes spatial blurring, physiological differences, hemodynamic differences, etc.



WU-Minn Human Connectome Project Consortium

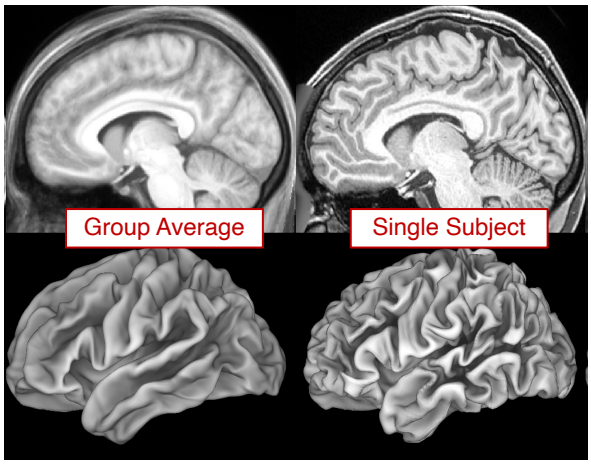
45

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

What is Deep Sampling in fMRI?

- So, then what is Deep Sampling?
  - ***It's just a study that collects multiple scanning sessions per subject…***



WU-Minn Human Connectome Project Consortium

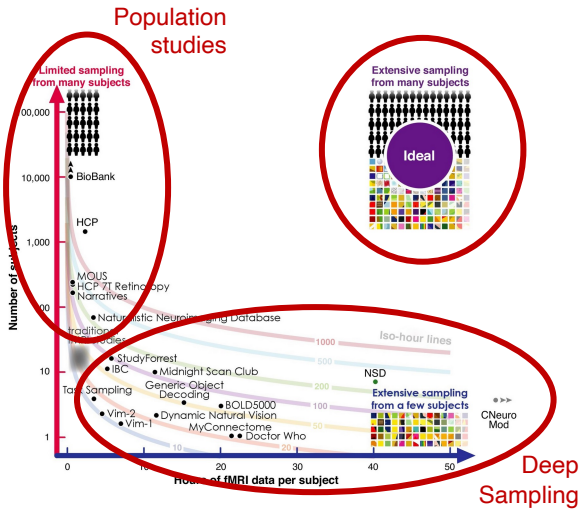
46

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

What is Deep Sampling in fMRI?

- Deep sampling aims to avoid issues arising from cross-subject averaging by collecting extensive fMRI datasets from only a few subjects
- Ideally, we would always record lots of data from many subjects, but in the real world, this is very difficult…



Naselaris, et al. 2021

47

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Shifting gears toward “Deep Sampling”

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

48

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Shifting gears toward “Deep Sampling”

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

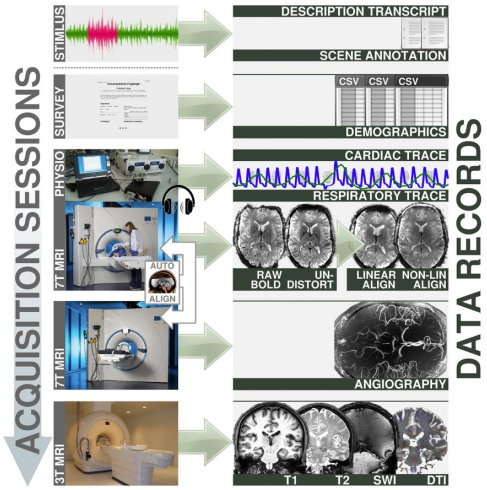
49

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Complex natural stimulation with audio

- Hanke (et al.) acquired functional data from subjects listening to the audio track from *Forest Gump* (considered a form of resting state)
- Supplementary data included survey, physio, and anatomical measures (T1, T2, SWI, DTI, MRA, etc.)



Hanke, et al. 2014

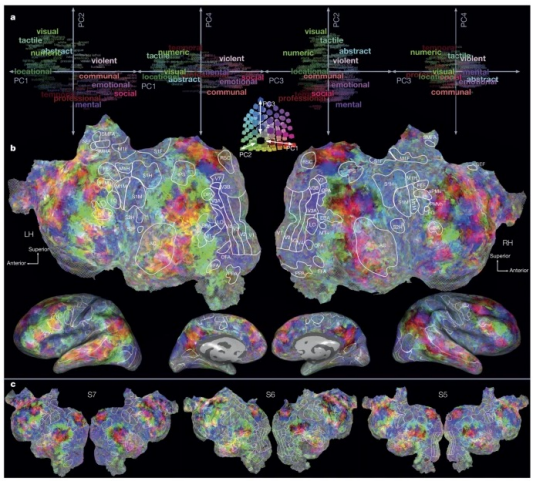
50

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Complex natural stimulation with audio

- Huth (et al.) used a similar design to develop models of speech semantics
- Uncovered separate “concept” preferences in different portions of cortex



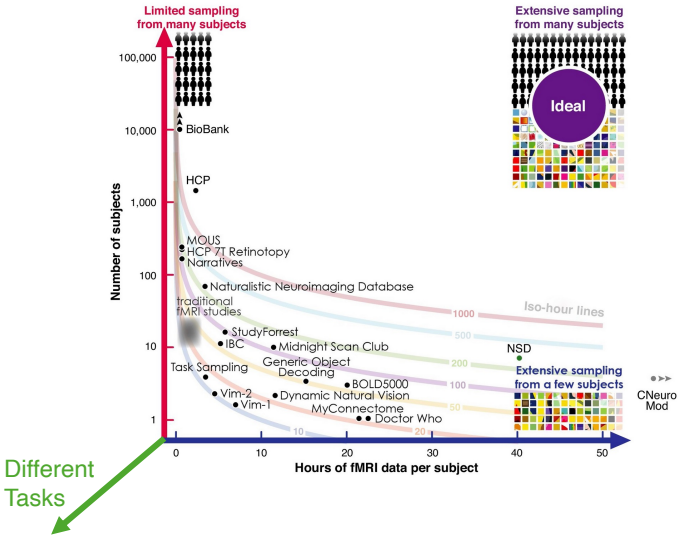
Huth, et al. 2016

51

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Is the amount of data the only thing important thing?



Naselaris, et al. 2021

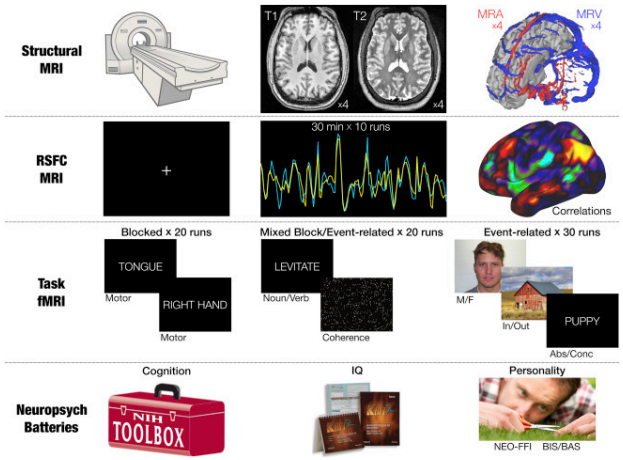
52

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Multiple tasks to better understand brain activity

- The Midnight Scan Club aimed to combined deeply sampled resting state activity with neuropsych batteries and task fMRI
- This design allows them to better explain connectivity profiles but incorporating single-subject data from multiple domains
- It also enables analysis of individual differences



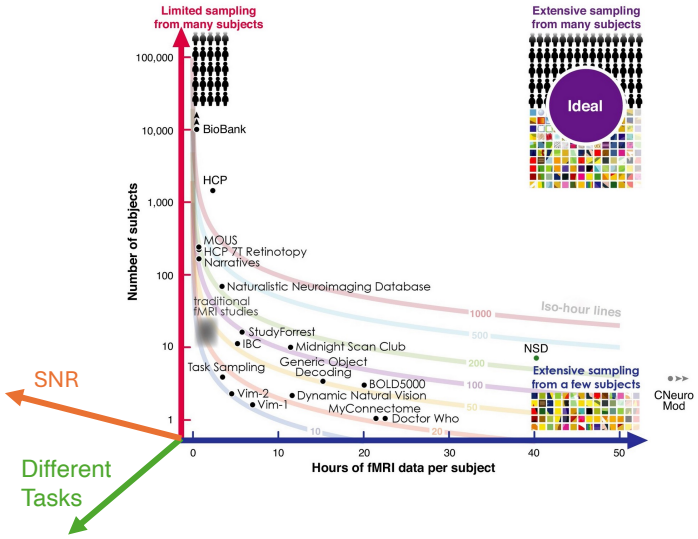
Gordon, et al. 2017

53

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Is the amount of data the only thing important thing?



Naselaris, et al. 2021

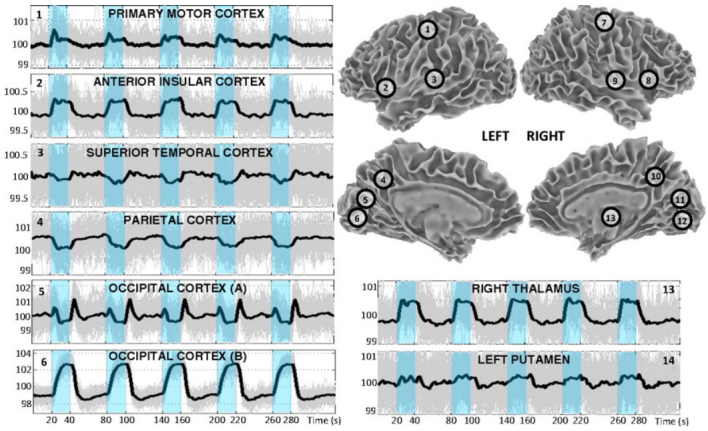
54

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Cortical responses are complicated

- Gonzalez-Castillo (et al.) looked at an incredibly simple paradigm—a block design flashing checkerboard with attention task
- They found that temporal responses varied widely across cortex (but were time-locked to stimuli)



Gonzalez Castillo, et al. 2012

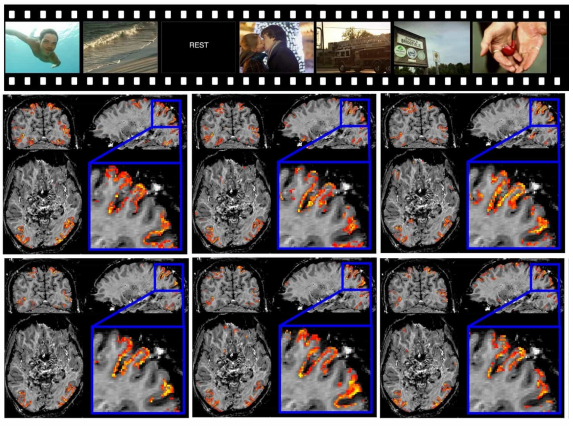
55

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Complex natural stimulation with audio and video

- "The Kenshu Dataset" expands stimuli to video and audio by repeatedly showing the Human Connectome Project's MOVIE1 stimulus set
- This project showed that Deep Sampling is an important aspect of whole brain datasets aimed at layer-specific connectivity profiles



Koiso, et al. 2023

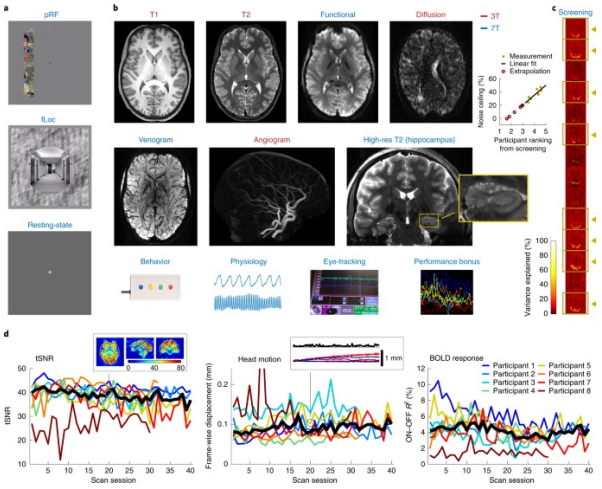
56

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Building better computational models of the brain

- The Natural Scenes Database collected an incredibly large dataset of subjects viewing static natural images (70000 images to 8 subjects over 30-40 sessions each)
- These data allow computational modelers to better understand how to brain encodes visual information and interfaces with large machine learning models



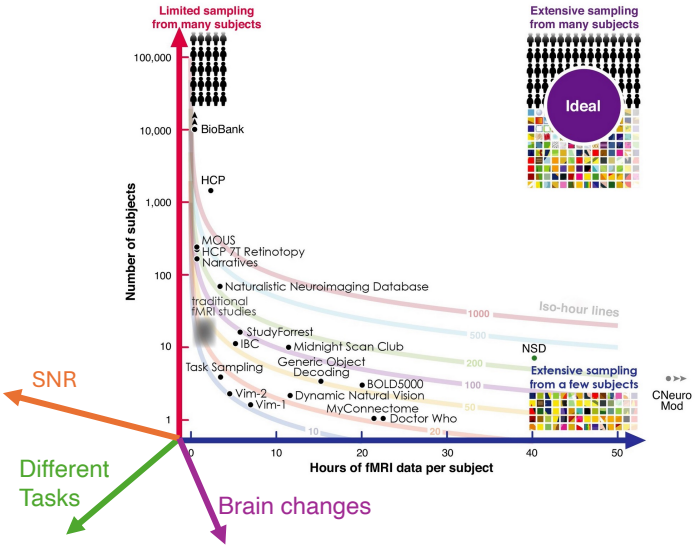
Allen, et al. 2022

57

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Is the amount of data the only thing important thing?



Naselaris, et al. 2021

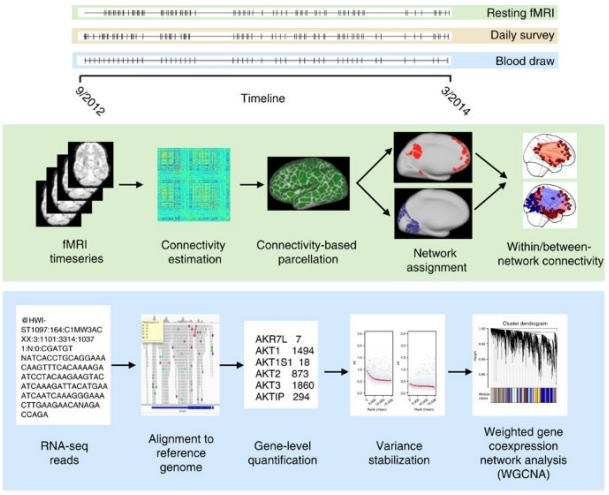
58

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Are we ever scanning the same brain?

- The MyConnectome project scanned one individual approximately 100 times over the course of one year.
- They collected surveys, blood draws, and genetics to better understand how connectivity is influenced by a person's daily life



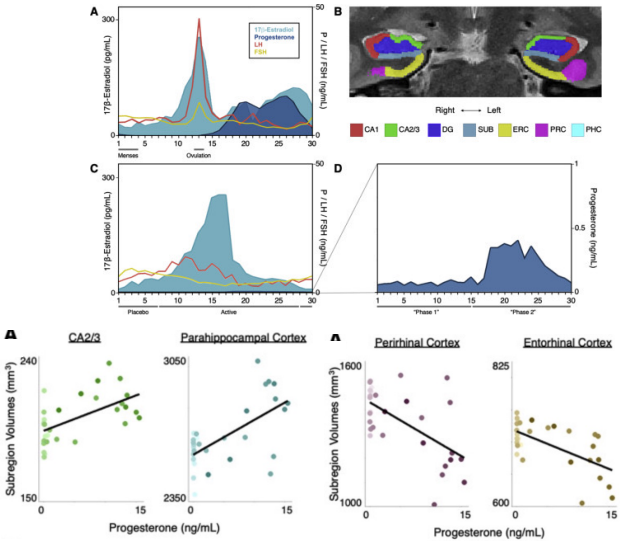
59

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Are we ever scanning the same brain?

- Taylor (et al.) examined how our brains change based on hormone cycles by collecting anatomical, resting state scans, and blood draws over the course of one individual's menstrual cycle
- One thing they found was that hippocampal sub-region volumes change over the cycle and are related to levels of progesterone in the blood
- They also did a follow up when the subject was on birth control (reduces progesterone levels) to confirm that these changes were related to the hormone.



Taylor, et al. 2020

60

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Shifting gears toward “Deep Sampling”

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

61

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Shifting gears toward “Deep Sampling”

1. What is Deep Sampling in fMRI?
2. Some Examples
3. Considerations in Deep Sampling

62

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

Okay, so you want to do a Deep Sampling study…

- Overall, there are 4 principles for Deep Sampling:
  1. Design well: Experiments should be motivated by theory and elicit specific cognitive processes. Statistical power should be maximised. Protocol should ensure participants do the task
  2. Scan more: Lots of data, but also T1-weighted and T2-weighted anatomical scans, venograms, angiograms, diffusion-weighted MRI, quantitative MRI
  3. Optimise quality: Optimization of quality, including the experimental protocol (stimulus design and trial timing), data acquisition (scanning parameters, image quality, and participant monitoring), data preparation (preprocessing, quality control, and across-session alignment
  4. Share effectively: Make sure data are curated and documented very well

Kupers, et al. 2024

63

NIH National Institutes of Health
Turning Discovery Into Health

FMRIF NIH

What is Deep Sampling in fMRI?

- Since collecting deep sampling projects can be expensive and time consuming, it is important to consider how to best cover multiple possible hypotheses with their experimental designs
- It can also be helpful to partially overlap with other large datasets to help link different areas of scientific inference.

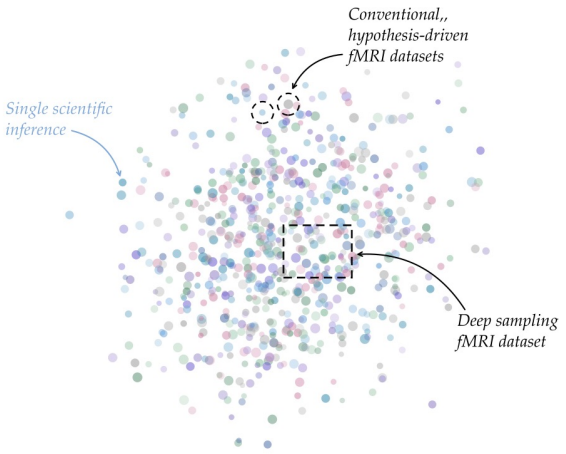


Figure adapted from Kupers, et al. 2024

64