Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS

Yu Huang[1], Xinyu Liu[1], Ryan Krueger[1], Tyler Santander[2], Xiaosu Hu[1], Kevin Leach[1], Westley Weimer[1]

[1]University of Michigan
[2]University of California, Santa Babara

May 29, 2019

Medical Imaging for Software Engineering

- Objectively understanding the subjective cognitive processes of software engineering is important

Medical Imaging for Software Engineering

- Objectively understanding the subjective cognitive processes of software engineering is important
    - Self-reporting

Medical Imaging for Software Engineering

- Objectively understanding the subjective cognitive processes of software engineering is important
  - Self-reporting
  - Pedagogy

Medical Imaging for Software Engineering

- Objectively understanding the subjective cognitive processes of software engineering is important
    - Self-reporting
    - Pedagogy
    - Technology transfer

Medical Imaging for Software Engineering

- Objectively understanding the subjective cognitive processes of software engineering is important
  - Self-reporting
  - Pedagogy
  - Technology transfer
  - Programming expertise

Medical Imaging for Software Engineering

- Objectively understanding the subjective cognitive processes of software engineering is important
    - Self-reporting
    - Pedagogy
    - Technology transfer
    - Programming expertise
- Medical imaging is quite rare in SE
    - Only 9 papers at main conferences in SE starting from 2014

Medical Imaging for Software Engineering

Debugging

- BrainImg '18
- ISSRE '16

Readability

- ICPC '18

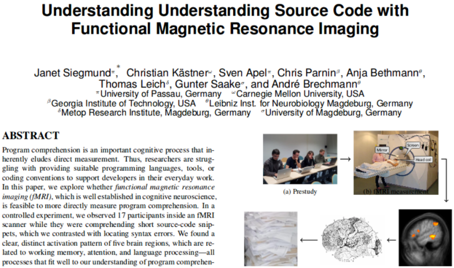
Code Review

- ICSE '17

Comprehension

- ESEM '18
- FSE '17
- SNPD '14
- ICSE Companion '14
- ICSE '14

Understanding Understanding Source Code



Understanding Understanding Source Code with Functional Magnetic Resonance Imaging

Janet Siegmund[π],[*] Christian Kästner[ω], Sven Apel[π], Chris Parnin[β], Anja Bethmann[θ],
Thomas Leich[δ], Gunter Saake[σ], and André Brechmann[θ]
[π]University of Passau, Germany [ω]Carnegie Mellon University, USA
[β]Georgia Institute of Technology, USA [θ]Leibniz Inst. for Neurobiology Magdeburg, Germany
[δ]Metop Research Institute, Magdeburg, Germany [σ]University of Magdeburg, Germany

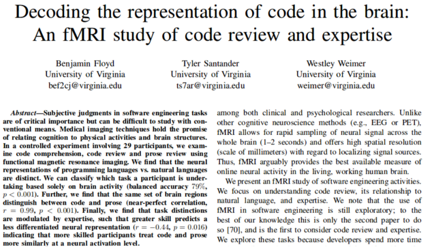
ABSTRACT

Program comprehension is an important cognitive process that inherently eludes direct measurement. Thus, researchers are struggling with providing suitable programming languages, tools, or coding conventions to support developers in their everyday work. In this paper, we explore whether *functional magnetic resonance imaging (fMRI)*, which is well established in cognitive neuroscience, is feasible to more directly measure program comprehension. In a controlled experiment, we observed 17 participants inside an fMRI scanner while they were comprehending short source-code snippets, which we contrasted with locating syntax errors. We found a clear, distinct activation pattern of five brain regions, which are related to working memory, attention, and language processing—all processes that fit well to our understanding of program comprehen-

Medical Imaging for Software Engineering

Debugging	**●BrainImg '18** **●ISSRE '16**
Readability	**●ICPC '18**
Code Review	**●ICSE '17**
Comprehension	**● ESEM '18** **● FSE '17** **●SNPD '14** **●ICSE Companion '14** **● ICSE '14**

Understanding **"Understanding Understanding Source Code"**



Decoding the representation of code in the brain: An fMRI study of code review and expertise

Benjamin Floyd
University of Virginia
bef2cj@virginia.edu

Tyler Santander
University of Virginia
ts7ar@virginia.edu

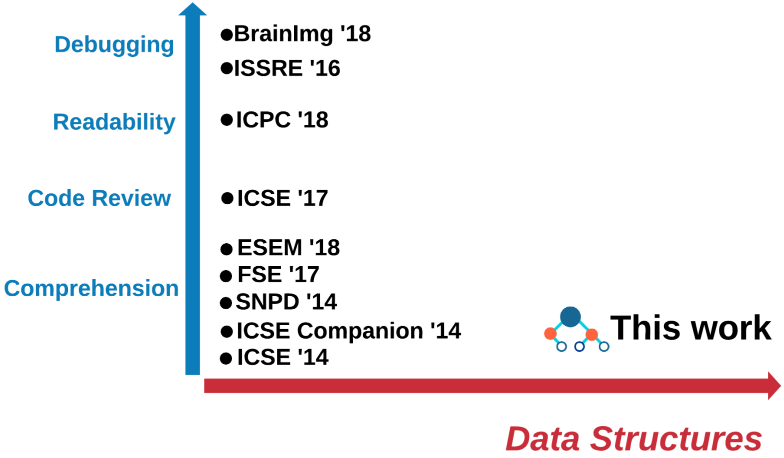
Westley Weimer
University of Virginia
weimer@virginia.edu

***Abstract*—Subjective judgments in software engineering tasks are of critical importance but can be difficult to study with conventional means. Medical imaging techniques hold the promise of relating cognition to physical activities and brain structures. In a controlled experiment involving 29 participants, we examine code comprehension, code review and prose review using functional magnetic resonance imaging. We find that the neural representations of programming languages vs. natural languages are distinct. We can classify which task a participant is undertaking based solely on brain activity (balanced accuracy 79%, $p < 0.001$). Further, we find that the same set of brain regions distinguish between code and prose (near-perfect correlation, $r = 0.99$, $p < 0.001$). Finally, we find that task distinctions are modulated by expertise, such that greater skill predicts a less differentiated neural representation ($r = -0.44$, $p = 0.016$) indicating that more skilled participants treat code and prose more similarly at a neural activation level.**

among both clinical and psychological researchers. Unlike other cognitive neuroscience methods (e.g., EEG or PET), fMRI allows for rapid sampling of neural signal across the whole brain (1–2 seconds) and offers high spatial resolution (scale of millimeters) with regard to localizing signal sources. Thus, fMRI arguably provides the best available measure of online neural activity in the living, working human brain.

We present an fMRI study of software engineering activities. We focus on understanding code review, its relationship to natural language, and expertise. We note that the use of fMRI in software engineering is still exploratory; to the best of our knowledge this is only the second paper to do so [70], and is the first to consider code review and expertise. We explore these tasks because developers spend more time

Medical Imaging for Software Engineering



Debugging

- BrainImg '18
- ISSRE '16

Readability

- ICPC '18

Code Review

- ICSE '17

Comprehension

- ESEM '18
- FSE '17
- SNPD '14
- ICSE Companion '14
- ICSE '14

This work

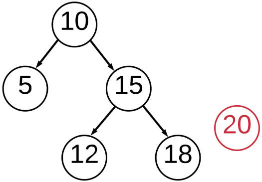
Data Structures

High-level Question

- How do human brains represent data structures? Is it more like text or more like 3D objects?

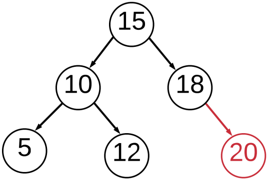
High-level Question

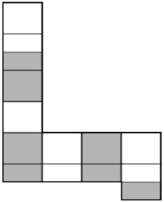
- How do human brains represent data structures? Is it more like text or more like 3D objects?



High-level Question

- How do human brains represent data structures? Is it more like text or more like 3D objects?

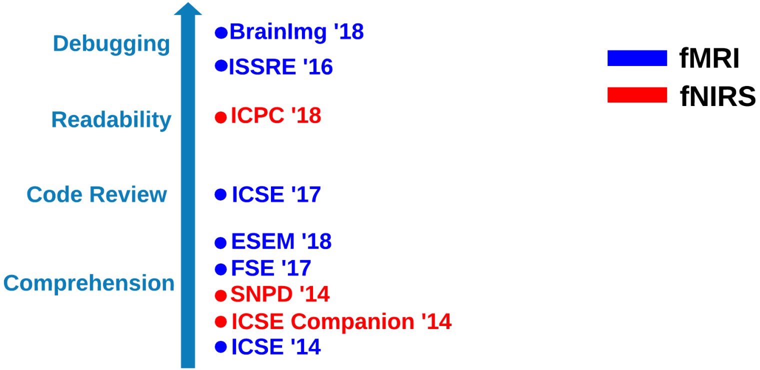




Spatial Ability: Mental Rotations

- The determination of spatial relationships between objects and the mental manipulation of spatially presented information
- Measured by mental rotation tasks: 3D objects
- Related to success in STEM

Another Glance: Medical Imaging in Software Engineering



Debugging	🔵 BrainImg '18 🔵 ISSRE '16
Readability	🔴 ICPC '18
Code Review	🔵 ICSE '17
Comprehension	🔵 ESEM '18 🔵 FSE '17 🔴 SNPD '14 🔴 ICSE Companion '14 🔵 ICSE '14

🟦 fMRI
🟥 fNIRS

fMRI vs. fNIRS

- **F**unctional **M**agnetic **R**esonance **I**maging
- **F**unctional **N**ear-**I**nfra**R**ed **S**pectroscopy
- Measure brain activities by calculating the blood-oxygen level dependent (BOLD) signal
  - Your brain needs energy but does not store it
  - We can track where oxygen is consumed
  - Contrasts-based experiments

fMRI vs. fNIRS

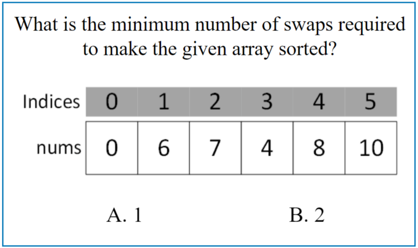
- **F**unctional **M**agnetic **R**esonance **I**maging
  - Magnets
  - Strong penetration power
  - Lying down in a magnetic tube: **cannot move**
- **F**unctional **N**ear-**I**nfra**R**ed **S**pectroscopy
  - Light
  - Weak penetration power
  - Wearing a specially-designed cap: **more freedom of movement**

Outline

1. Medical imaging in software engineering and motivation
2. fMRI vs. fNIRS
3. Experimental Design
4. Results
5. fMRI vs. fNIRS for Software Engineering
6. Conclusion

Experimental Design: 2 Tasks

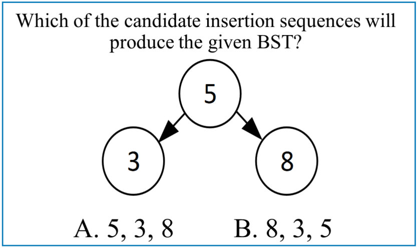
- Data Structure manipulations
  - List/Array operations
  - Tree operations



What is the minimum number of swaps required to make the given array sorted?

Indices	0	1	2	3	4	5
nums	0	6	7	4	8	10

A. 1 B. 2



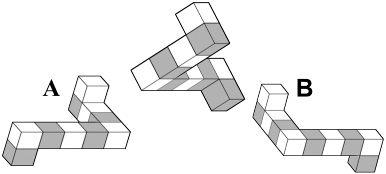
Which of the candidate insertion sequences will produce the given BST?

A. 5, 3, 8 B. 8, 3, 5

Experimental Design: 2 Tasks

- Data Structure manipulations
  - List/Array operations
  - Tree operations
- Mental rotations: 3D objects

Which object is the same as the original object, aside from its orientation?

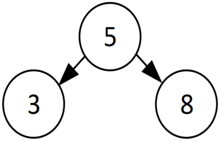


What is the minimum number of swaps required to make the given array sorted?

Indices	0	1	2	3	4	5
nums	0	6	7	4	8	10

A. 1 B. 2

Which of the candidate insertion sequences will produce the given BST?



A. 5, 3, 8 B. 8, 3, 5

Experiment Setup and Data

- 76 Participants: 70 valid *
    - fMRI: 30
    - fNIRS: 40
    - Two hours for each participant: 90 stimuli

*De-identified data is public: https://web.eecs.umich.edu/~weimerw/fmri.html

Experiment Setup and Data

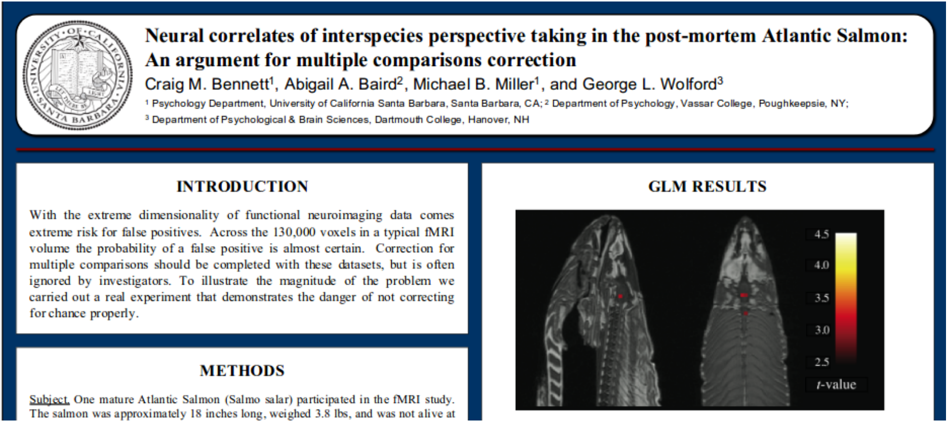
- 76 Participants: 70 valid *
  - fMRI: 30
  - fNIRS: 40
  - Two hours for each participant: 90 stimuli
  - Big human study!

*De-identified data is public: https://web.eecs.umich.edu/~weimerw/fmri.html

Data Analysis: Be Careful!

Data Analysis: Be Careful!

- Dead fish is thinking?!



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY; [3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we carried out a real experiment that demonstrates the danger of not correcting for chance properly.

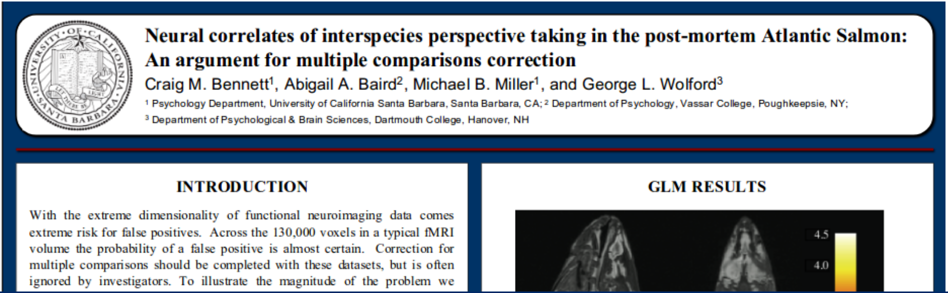
METHODS

Subject. One mature Atlantic Salmon (Salmo salar) participated in the fMRI study. The salmon was approximately 18 inches long, weighed 3.8 lbs, and was not alive at

GLM RESULTS

Data Analysis: Be Careful!

- Dead fish is thinking?!
- fMRI and fNIRS use the same high-level 3-step analysis approach



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY;
[3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

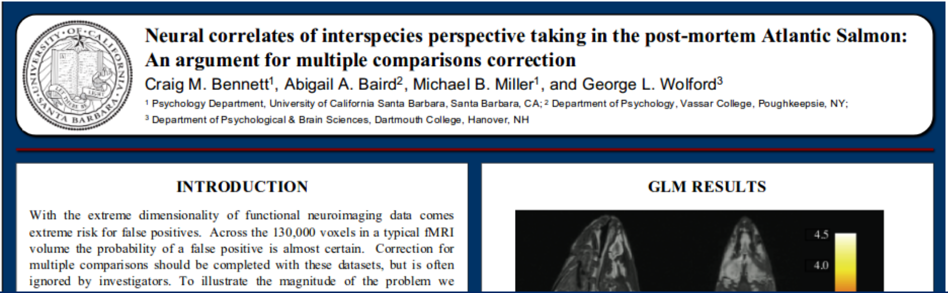
With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we

GLM RESULTS

Data Analysis: Be Careful!

- Dead fish is thinking?!
- fMRI and fNIRS use the same high-level 3-step analysis approach

Preprocessing



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY;
[3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

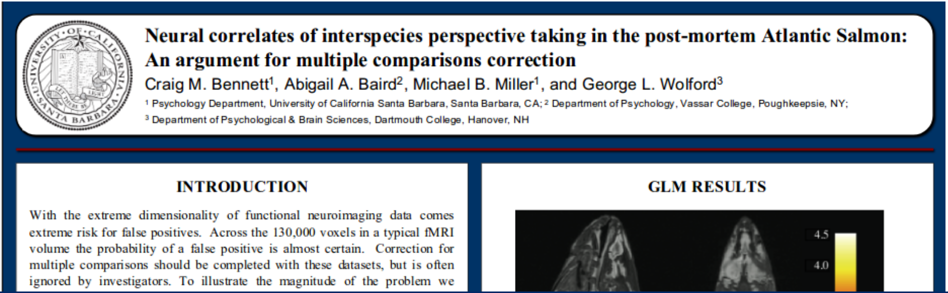
With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we

GLM RESULTS

Data Analysis: Be Careful!

- Dead fish is thinking?!
- fMRI and fNIRS use the same high-level 3-step analysis approach

Preprocessing → **First-level Analysis**



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY;
[3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

INTRODUCTION

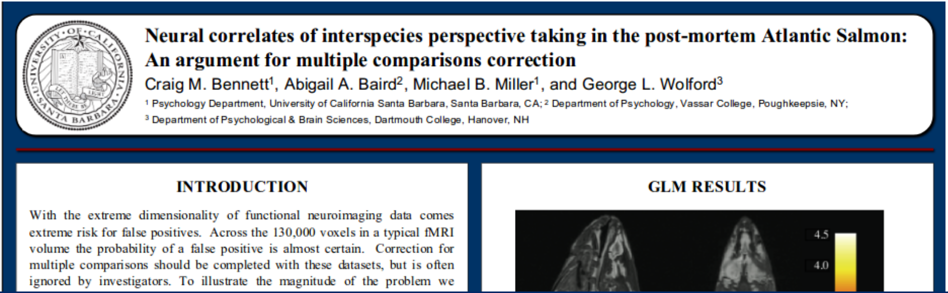
With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we

GLM RESULTS

Data Analysis: Be Careful!

- Dead fish is thinking?!
- fMRI and fNIRS use the same high-level 3-step analysis approach

Preprocessing → **First-level Analysis** → **Contrasts & Group-level Analysis**



Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY; [3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

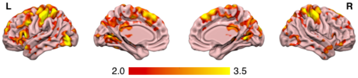
INTRODUCTION

With the extreme dimensionality of functional neuroimaging data comes extreme risk for false positives. Across the 130,000 voxels in a typical fMRI volume the probability of a false positive is almost certain. Correction for multiple comparisons should be completed with these datasets, but is often ignored by investigators. To illustrate the magnitude of the problem we

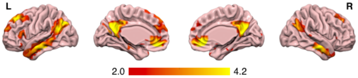
GLM RESULTS

Results: Data Structure Manipulation and Spatial Ability

- **Yes**: data structure manipulations involve spatial ability
  - fMRI: more similarities than differences ($p < 0.001$)
  - fNIRS: activation in the same brain regions ($p < 0.01$)



List/Array vs. Mental Rotation



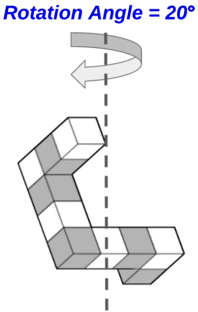
Mental Rotation vs. Tree

Results: The Role of Task Difficulty

- The brain works even harder for more difficult data structure tasks
    - Difficulty measurement

Results: The Role of Task Difficulty

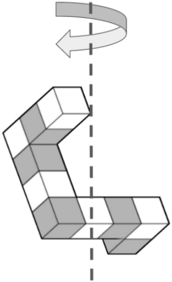
- The brain works even harder for more difficult data structure tasks
  - Difficulty measurement
    - Mental rotations: angle of rotation



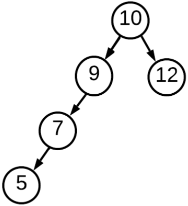
Results: The Role of Task Difficulty

- The brain works even harder for more difficult data structure tasks
  - Difficulty measurement
    - Mental rotations: angle of rotation
    - Data structures: size

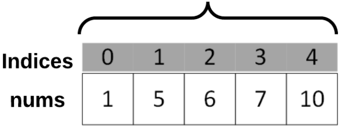
Rotation Angle = 20°



N = 5



N= 5

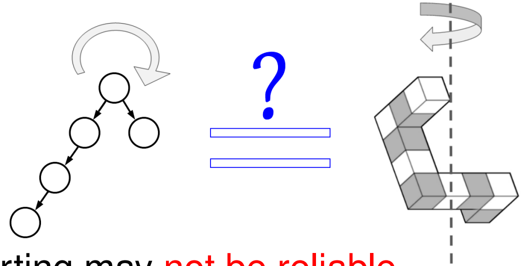


Indices	0	1	2	3	4
nums	1	5	6	7	10

Results: The Role of Task Difficulty

- The brain works even harder for more difficult data structure tasks
  - Difficulty measurement
    - Mental rotations: angle of rotation
    - Data structures: size
  - fMRI: the rate of extra work in your brain is higher for data structure tasks than it is for mental rotation tasks
  - fNIRS: no significant findings for the effect of task difficulty

Results: How Do Self-reporting and Neuroimaging Compare?



- Self-reporting may not be reliable
  - Medical imaging found mental rotation and data structure tasks are very similar
  - 70% of human participants believe there is no connection!

Implications: fMRI vs. fNIRS for Software Engineering

	fMRI	**fNIRS**
Time	~2 hours	~2 hours
Penetration Power	**Strong**	Moderate
Cost	> $20,000 for 36	**~$2000 for 40**
Environment	Restricted	**Free**
Task Accuracy	Lower (85%, $p < 0.01$)	**Higher (92%, $p < 0.01$)**
Effort	**Light**	Heavy
Recruitment	**Easy**	Moderate (hair)

Conclusion

- Large human study: data from 70 participants *

*De-identified data is public: https://web.eecs.umich.edu/~weimerw/fmri.html

Conclusion

- Large human study: data from 70 participants *
    - Data structure manipulations and mental rotations use the same brain regions

*De-identified data is public: https://web.eecs.umich.edu/~weimerw/fmri.html

Conclusion

- Large human study: data from 70 participants *
  - Data structure manipulations and mental rotations use the same brain regions
  - Task difficulty matters for data structures

*De-identified data is public: https://web.eecs.umich.edu/~weimerw/fmri.html

Conclusion

- Large human study: data from 70 participants *
  - Data structure manipulations and mental rotations use the same brain regions
  - Task difficulty matters for data structures
  - Medical imaging can discover more than self-reporting

*De-identified data is public: https://web.eecs.umich.edu/~weimerw/fmri.html

Conclusion

- Large human study: data from 70 participants *
  - Data structure manipulations and mental rotations use the same brain regions
  - Task difficulty matters for data structures
  - Medical imaging can discover more than self-reporting
- This work may inform:
  - Pedagogy and training
  - Technology transfer
  - Programming expertise

*De-identified data is public: https://web.eecs.umich.edu/~weimerw/fmri.html

Bonus Slides

fMRI vs. fNIRS

- **F**unctional **M**agnetic **R**esonance **I**maging
  - Oxygenated and deoxygenated hemoglobin have different magnetic properties that can be detected
  - Lying down in a magnetic tube

- **F**unctional **N**ear-**I**nfra**R**ed **S**pectroscopy
  - Absorption of chromophores (groups of atoms that generate color through the absorption of light) are different between oxygenated and deoxygenated hemoglobin
  - Wearing a specially-designed cap connecting light emitters and detectors

Data Analysis

- fMRI and fNIRS use the same high-level analysis approach
  - Preprocessing
    - Correct systematic sources of noise: VDM for fMRI, autoregressive-whitened robust regression for fNIRS
  - First-level analysis
    - fMRI and fNIRS: GLMs per participant
    - Within individuals
  - Contrasts and group-level analysis
    - False discovery rate (FDR) threshold

Experiment Setup and Data

- 76 Participants: 70 valid
- Experiment design
    1. Set up: background survey, watch a training video
    2. 3 blocks of tasks: 30 stimuli in each block, 2–10 seconds of rest between stimuli
    3. Post-survey: how do you compare these tasks?