Understanding User Cognition:
From Spatial Ability to Code Writing and Review

Yu Huang
http://www-personal.umich.edu/~yhhy

June 17, 2021

COLLEGE OF ENGINEERING
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

We want to improve productivity and reduce cost in software development and maintenance.

The Human Aspect Matters

- Early study of industrial developers found order-of-magnitude individual variations

Metric	Poorest	Best	Ratio
Debugging Hours Algebra	170	6	28:1
Debugging Hours Maze	26	1	26:1
CPU Seconds Algebra	3075	370	8:1
CPU Seconds Maze	541	50	11:1
Code Writing Hours Algebra	111	7	16:1
Code Writing Hours Maze	50	2	25:1
Program Size Algebra	6137	1050	6:1
Program Size Maze	3287	651	5:1
Run Time Algebra	7.9	1.6	5:1
Run Time Maze	8.0	0.6	13:1

H. Sackman, W. J. Erikson and E. E. Grant. *Exploratory Experimental Studies Comparing Online and Offline Programming Performance.* Communications of the ACM, 1968.

The Human Aspect Matters

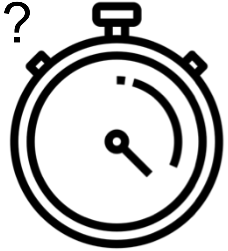
- Early study of industrial developers found ***order-of-magnitude*** individual variations

Metric	Poorest	Best	Ratio
Debugging Hours Algebra	170	6	**28:1**
Debugging Hours Maze	26	1	**26:1**
CPU Seconds Algebra	3075	370	**8:1**
CPU Seconds Maze	541	50	**11:1**
Code Writing Hours Algebra	111	7	**16:1**
Code Writing Hours Maze	50	2	**25:1**
Program Size Algebra	6137	1050	**6:1**
Program Size Maze	3287	651	**5:1**
Run Time Algebra	7.9	1.6	**5:1**
Run Time Maze	8.0	0.6	**13:1**

H. Sackman, W. J. Erikson and E. E. Grant. *Exploratory Experimental Studies Comparing Online and Offline Programming Performance.* Communications of the ACM, 1968.

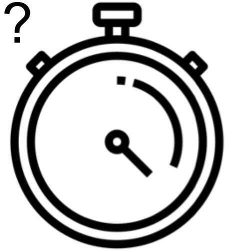
The Human Aspect Matters

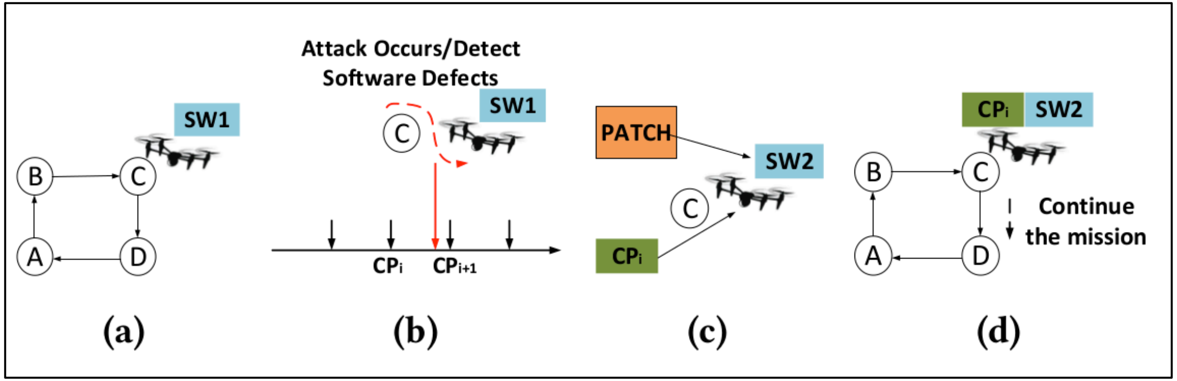
- How to ***measure*** cognitive processes?
  - Conduct behavioral experiments
  - "***Stopwatch***" and "***Scoresheet***"
    - *Time* and *accuracy*



The Human Aspect Matters

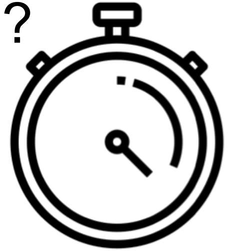
- How to ***measure*** cognitive processes?
  - Conduct behavioral experiments
  - "***Stopwatch***" and "***Scoresheet***"
    - *Time and accuracy*





The Human Aspect Matters

- How to ***measure*** cognitive processes?
  - Conduct behavioral experiments
  - "***Stopwatch***" and "***Scoresheet***"
    - *Time* and *accuracy*
    - ***What*** but not ***why***
      - Generalization, recommendation, transformation
    - **Overlook** what is actually going on
      - Miss information
    - **Limited** research findings



The Human Aspect Matters

- How to ***measure*** cognitive processes?
  - Conduct behavioral experiments
  - "***Stopwatch***" and "***Scoresheet***"
  - ***Self-reporting***

The Human Aspect Matters

- How to ***measure*** cognitive processes?
  - Conduct behavioral experiments
  - "***Stopwatch***" and "***Scoresheet***"
  - ***Self-reporting***
    - **Unreliable**

International Journal for Quality in Health Care 1999; Volume 11, Number 3: pp. 187–192

Evidence of self-report bias in assessing adherence to guidelines

HAN LOMAS[2] AND DENNIS ROSS-DEGNAN[1]

d the

Self-Reports in Organizational R Problems and Prospect

Philip M. Podsakoff
Dennis W. Organ
Indiana University

Self-reports figure prominently in organizational and m research, but there are several problems associated wit This article identifies six categories of self-reports and dis problems as common method variance, the consistency mo cial desirability. Statistical and post hoc remedies and s dural methods for dealing with artifactual bias are presented and evaluated. Recommendations for future research are also offered.

Faking It: Social Desirability Response Bias in Self-report Research

Australian Journal of Advanced Nur
Volume 25 Issue 4 (June/Aug 2008)

van de Mortel, Thea F[1]

Abstract: Objective: The tendency for questionnaires is called socially desirab creating false relationships or obscuring scales can be used to detect, minimise, questionnaire- based research. The ai related studies that used questionnaire

Journal of Business and Psychology, Vol. 17, No. 2, Winter 2002 (©2002)

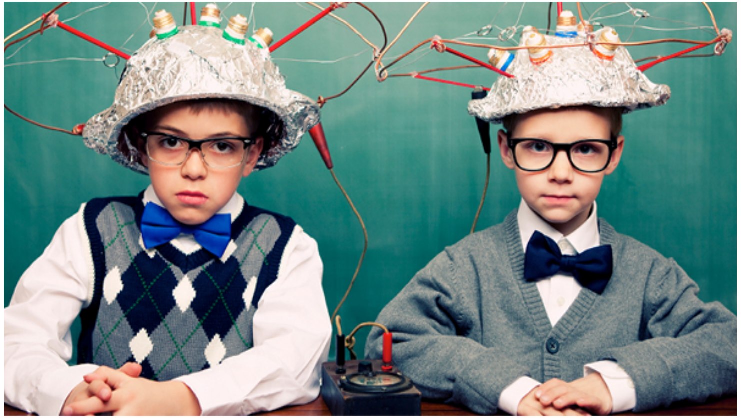
UNDERSTANDING SELF-REPORT BIAS IN ORGANIZATIONAL BEHAVIOR RESEARCH

Stewart I. Donaldson
Claremont Graduate University
Elisa J. Grant-Vallone
California State University, San Marcos

The Human Aspect Matters

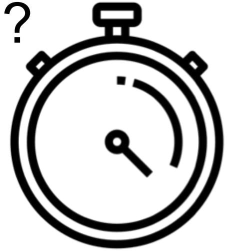
- How to ***measure*** cognitive processes?
  - Conduct behavioral experiments
  - “***Stopwatch***” and “***Scoresheet***”
  - ***Self-reporting***
    - **Unreliable**

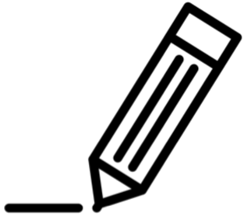
“Can we read your mind?”



The Human Aspect Matters

- How to ***measure*** cognitive processes?
  - Conduct behavioral experiments
  - "*Stopwatch*" and "*Scoresheet*"
  - Self-reporting
    - **Unreliable**





"Can we read your mind?"

Medical Imaging & Eye Tracking

Thesis Statement

It is possible to meaningfully and objectively measure user cognition to understand the role of spatial ability, fundamental processes and stereotypical associations in certain software engineering activities by combining medical imaging and eye tracking.

Challenges

- Many techniques: EEG, PET, fMRI, fNIRS, …
  - **Non-invasive**
  - **fMRI** and **fNIRS**
    - Sampling the brain rapidly with high spatial resolution

Challenges

- Many techniques: EEG, PET, fMRI, fNIRS, …
  - **Non-invasive**
  - **fMRI** and **fNIRS**
    - Sampling the brain rapidly with high spatial resolution
    - > 37,000 publications collected in PubMed in 2020 but relatively new to CS: less than two dozens in SE

Challenges

- Many techniques: EEG, PET, fMRI, fNIRS, …
  - **Non-invasive**
  - **fMRI** and **fNIRS**
    - Sampling the brain rapidly with high spatial resolution
    - > 1500 publications in Psychology but relatively new to CS: less than a dozen in SE
    - Streetlight effects?

Challenges

- Many techniques: EEG, PET, fMRI, fNIRS, ...
  - **Non-invasive**
  - **fMRI** and **fNIRS**
    - Sampling the brain rapidly with high spatial resolution
    - > 1500 publications in Psychology but relatively new to CS: less than a dozen in SE
    - Streetlight effects?
      - Environment

Challenges

- Many techniques: EEG, PET, fMRI, fNIRS, …
  - **Non-invasive**
  - **fMRI** and **fNIRS**
    - Sampling the brain rapidly with high spatial resolution
    - > 1500 publications in Psychology but relatively new to CS: less than a dozen in SE
    - Streetlight effects?
      - Environment
      - Contrast-based tasks

Challenges

- Many techniques: EEG, PET, fMRI, fNIRS, ...
  - **Non-invasive**
  - **fMRI** and **fNIRS**
    - Sampling the brain rapidly with high spatial resolution
    - > 1500 publications in Psychology but relatively new to CS: less than a dozen in SE
    - Streetlight effects?
      - Environment
      - Contrast-based tasks
      - Experimental design

Outline

- Introduction
- Investigating cognition in software engineering

  Understanding the neural representations of data structures

  Comparing prose writing and code writing

  Understanding bias in code reviews

- Career Plan
- Summary

Hypothesis Time!

Is balancing AVL trees like playing arcade claw machines?

Hypothesis Time!

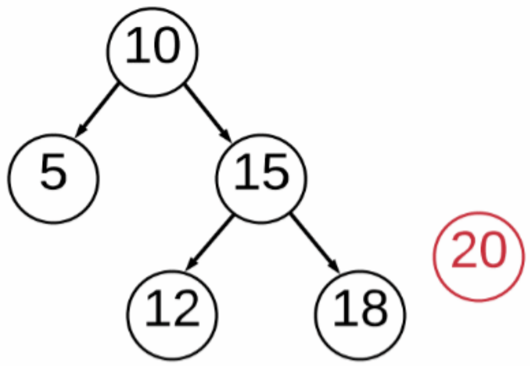
Is balancing AVL trees like playing arcade claw machines?

Yes!
Very similar!

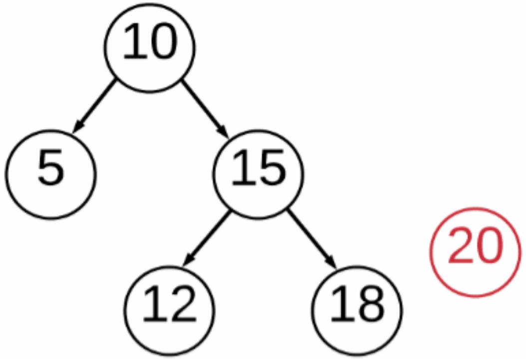
No!
Very Different!

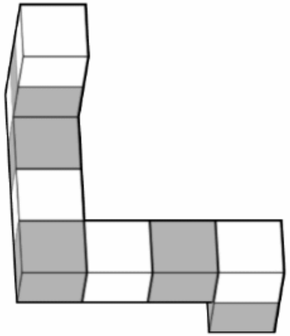
How do human brains represent data structures? Is it more like text or more like 3D objects?

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



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

About Medical Imaging

- **fMRI** and **fNIRS**
- **BOLD** signals
- Contrast design
- Rigorous data analysis: false positives

fMRI vs. fNIRS

Measure brain activities by calculating the **b**lood-**o**xygen **l**evel **d**ependent **(BOLD)** signal

- **F**unctional **M**agnetic **R**esonance **I**maging
- **F**unctional **N**ear-**I**nfra**R**ed **S**pectroscopy

fMRI vs. **fNIRS**

Measure brain activities by calculating the **b**lood-**o**xygen **l**evel **d**ependent **(BOLD)** signal

- **F**unctional **M**agnetic **R**esonance **I**maging
  - ***Magnets***
  - **Strong** penetration power
  - Lying down in a magnetic tube
- **F**unctional **N**ear-**I**nfra**R**ed **S**pectroscopy
  - ***Light***
  - **Weak** penetration power
  - Wearing a specially-designed cap

fMRI vs. fNIRS

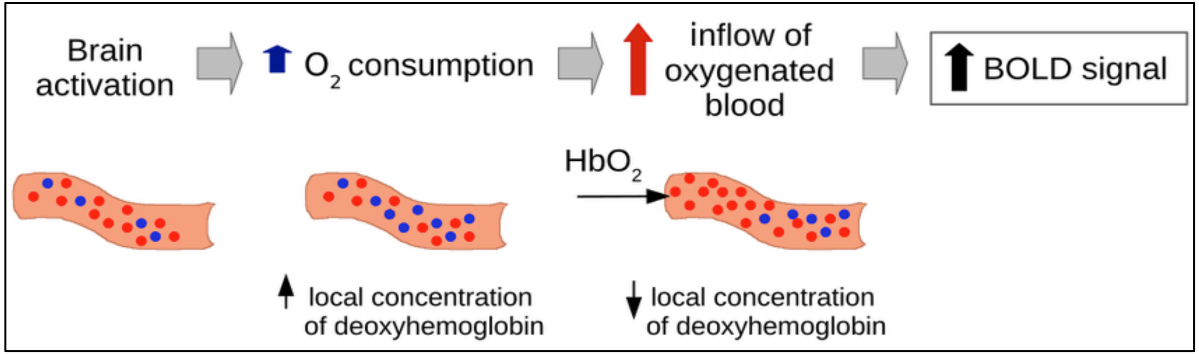
Measure brain activities by calculating the **b**lood-**o**xygen **l**evel **d**ependent **(BOLD)** signal

- **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

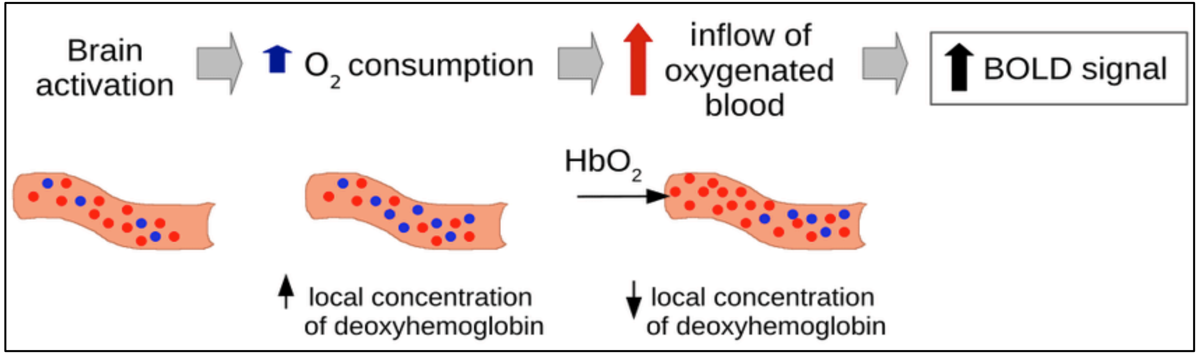
What is *BOLD* signal?

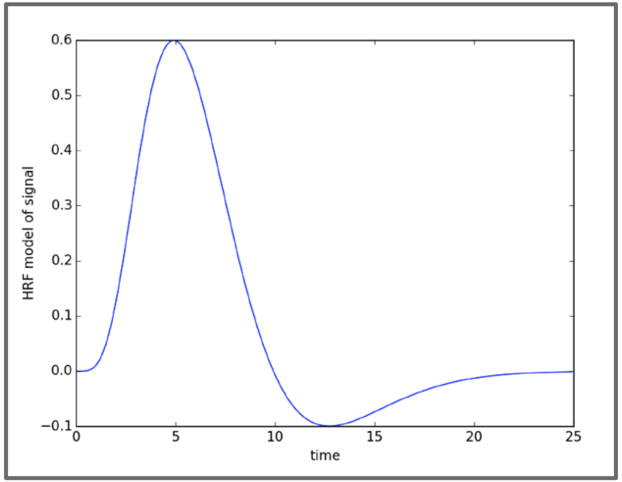
- **B**lood-**O**xygen **L**evel **D**ependent **(BOLD)** signal
- Blood flow and oxygen consumption as a **proxy** for brain activity



What is *BOLD* signal?

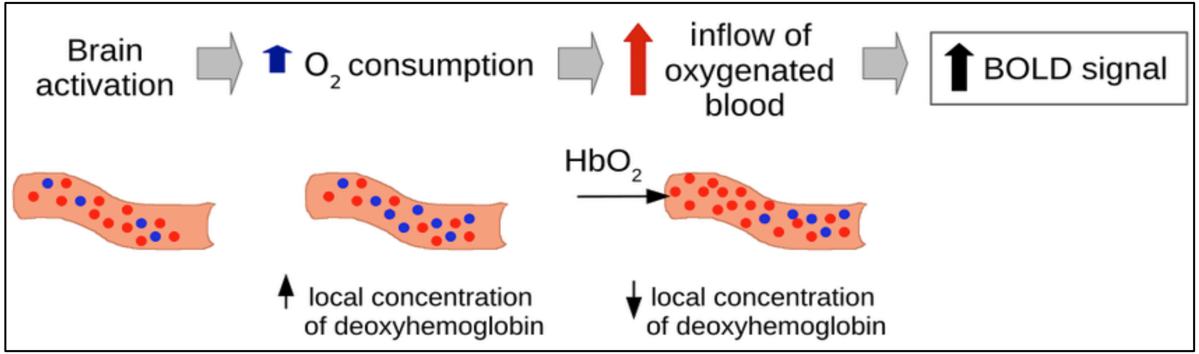
- **B**lood-**O**xygen **L**evel **D**ependent **(BOLD)** signal
- Blood flow and oxygen consumption as a **proxy** for brain activity
- Activation model: hemodynamic response function (HRF)

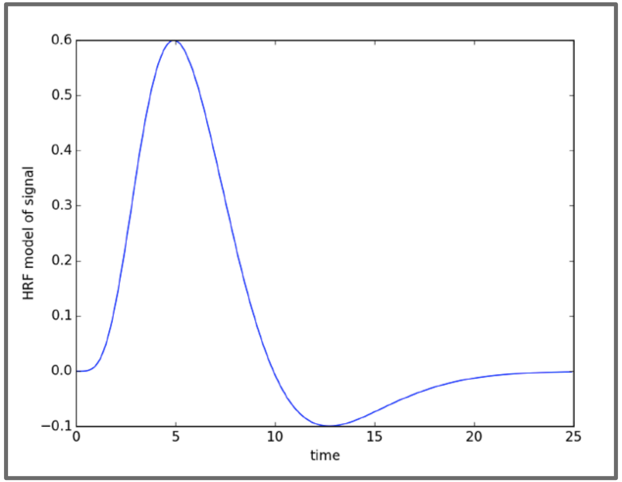




What is *BOLD* signal?

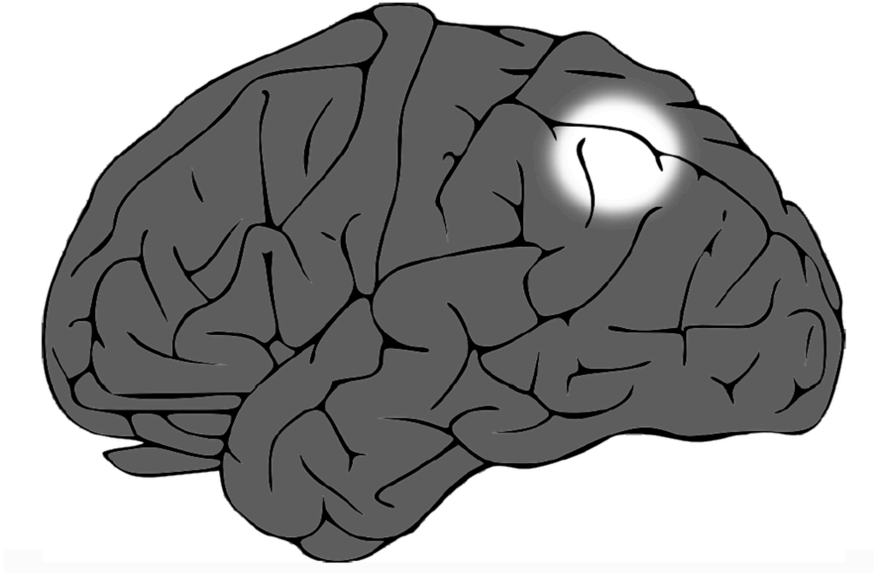
- **B**lood-**O**xygen **L**evel **D**ependent **(BOLD)** signal
- Blood flow and oxygen consumption as a **proxy** for brain activity
- Activation model: hemodynamic response function (HRF)
- Stimulus, HRF, design matrix, noise
  - Comprehensive quantitative model of BOLD signals
    - General Linear Model (GLM)





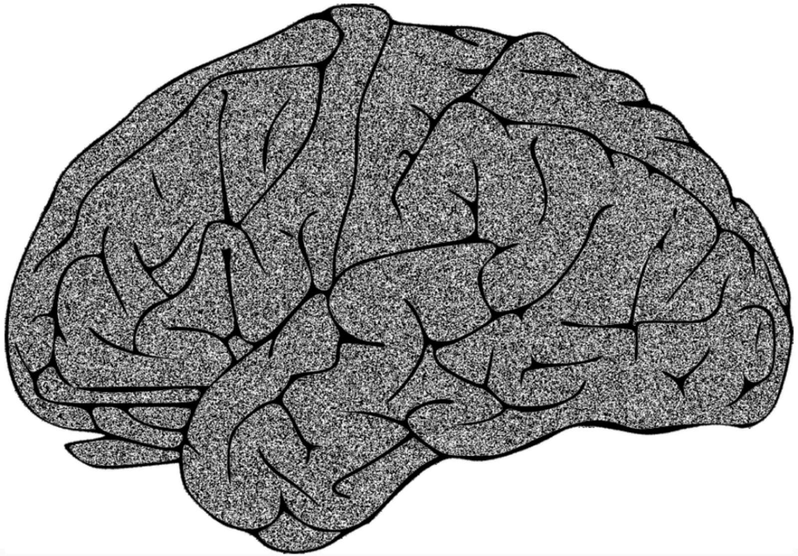
But it's not so easy

- **Brain activation does not work like this:**



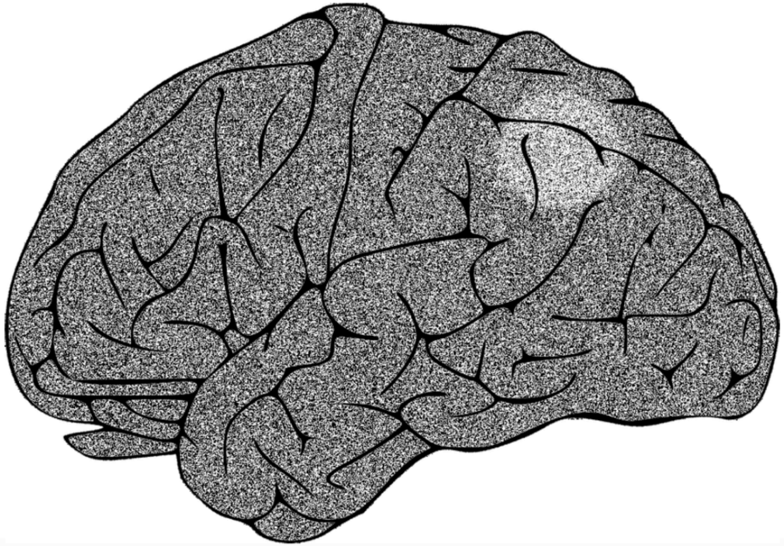
But it’s not so easy

- The brain signals are **noisy**



But it’s not so easy

- The brain signals are **noisy**
- Signal changes are **small**



Think in Terms of Contrasts!

Think in Terms of Contrasts!

- Controlled experimental design
  - Task A = "**balancing trees** + nervous + …"
  - Task B = "**rotating 3D objects** + nervous + ..."
  - Contrast **A** > **B**: brain activations that vary between the tasks

Data Analysis

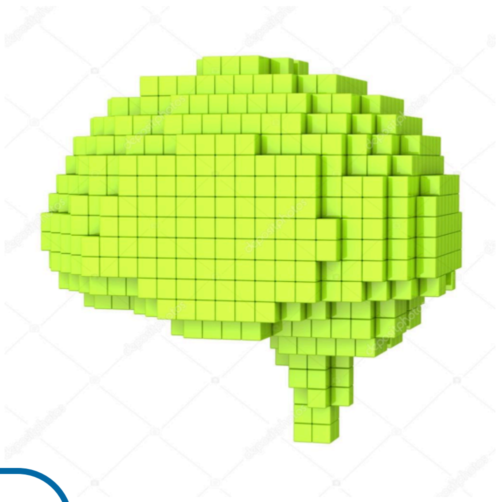
Preprocessing → First-level Analysis → Contrast & Group-level Analysis

Data Analysis

Preprocessing

Slice-time Correction → Voxel Displacement Maps (VDMs), Motion Correction, Co-registration & Normalization (MNI152) → Spatial Smoothing

Data Analysis



Preprocessing

Slice-time Correction → Voxel Displacement Maps (VDMs), Motion Correction, Co-registration & Normalization (MNI152) → Spatial Smoothing

Data Analysis

Preprocessing → **First-level Analysis**

- Within individuals
- General Linear Models
  - Robust weighted least squares (rWLS)
  - For each experiment condition

Data Analysis

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

- Pairwise contrast
  - Mean differences between conditions
  - Group-level random effects analyses
- Second-level GLM
  - Assess average activity across all subjects

Data Analysis

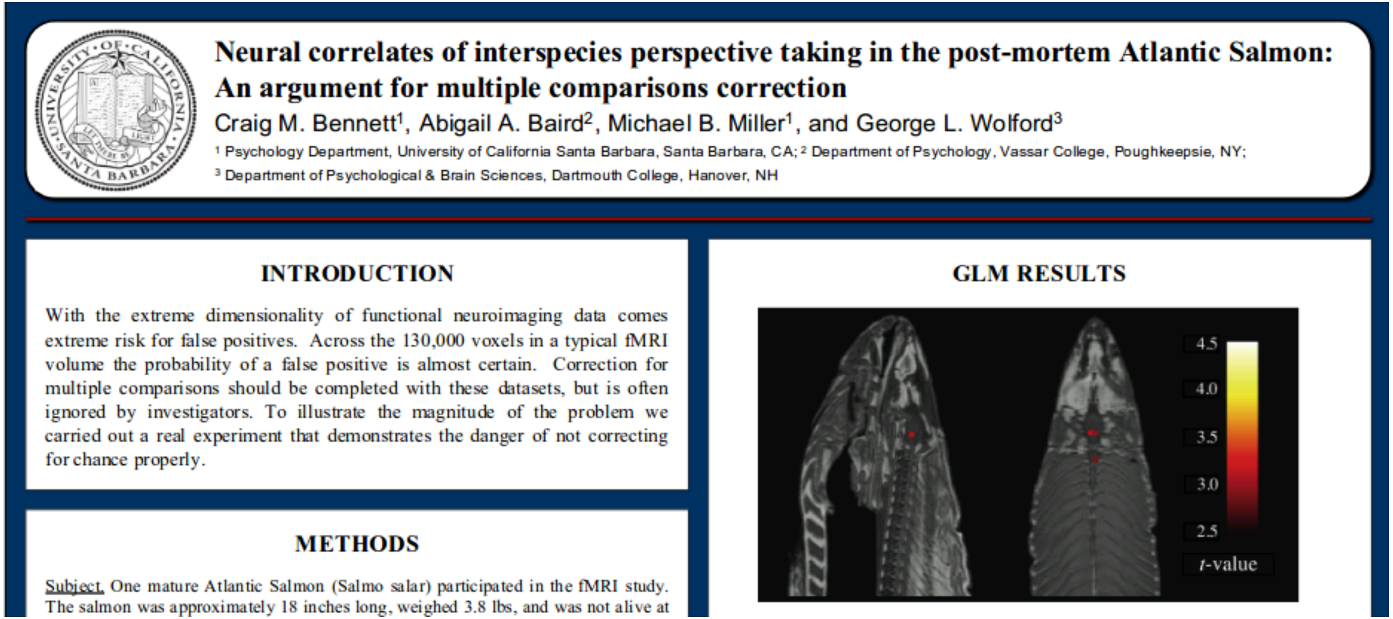
- Final results
  - A statistical parametric map of t-values (t-map) describing clusters of significant activity for a given task-related comparison

Data Analysis

- We need to be ***careful***
  - 153,000 voxels or more
  - Spurious correlations due to multiple comparison: false positives

Data Analysis

- We need to be ***careful***
  - 153,000 voxels or more
  - Spurious correlations due to multiple comparison: false positives



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

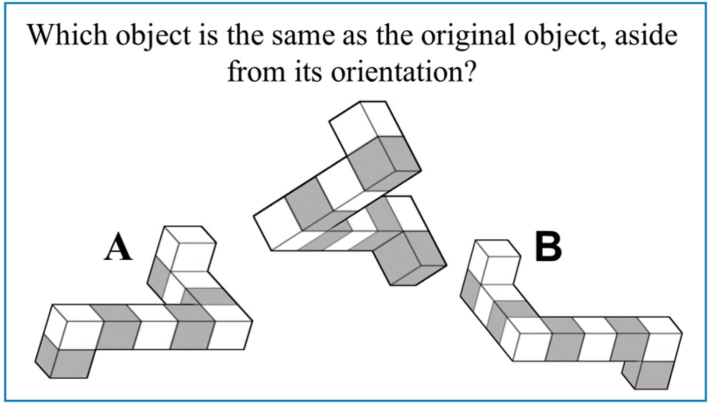
Data Analysis

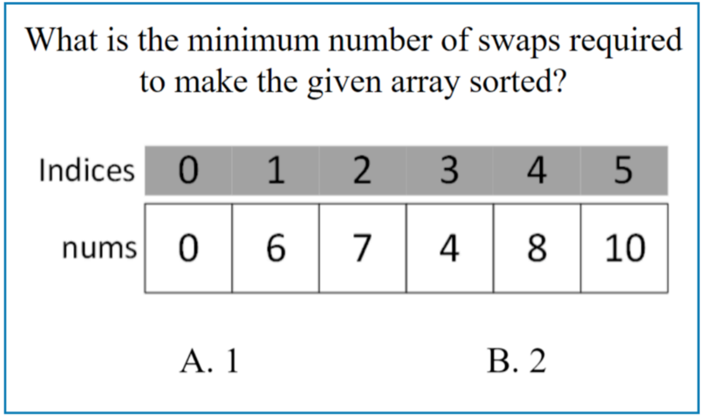
Preprocessing → **First-level Analysis** → **Contrast & Group-level Analysis**

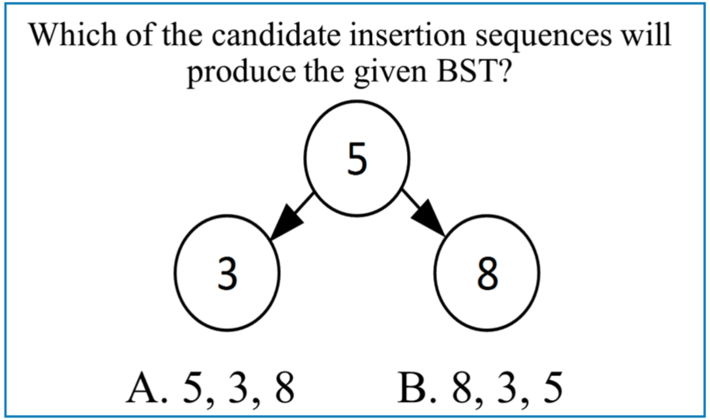
- False discovery rate (FDR) correction ($q<0.05$)

Experimental Design

- Two types of tasks
  - Data structure manipulations
  - Mental rotations: 3D objects

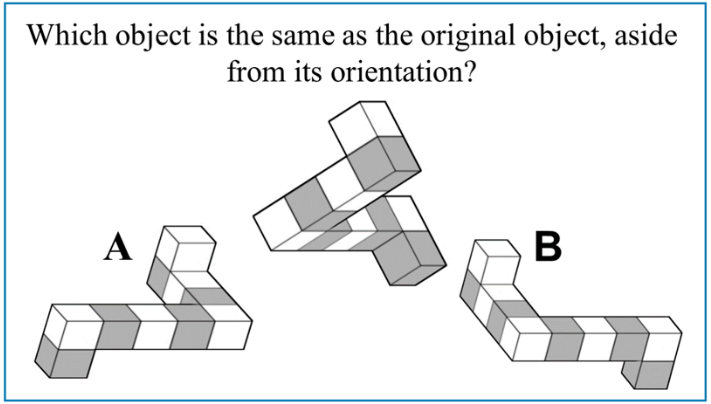


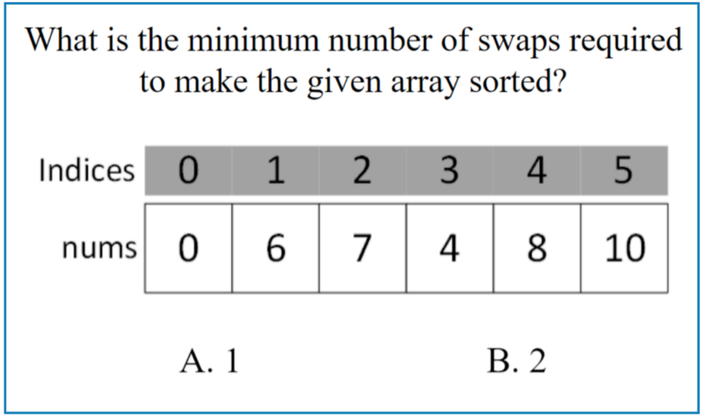


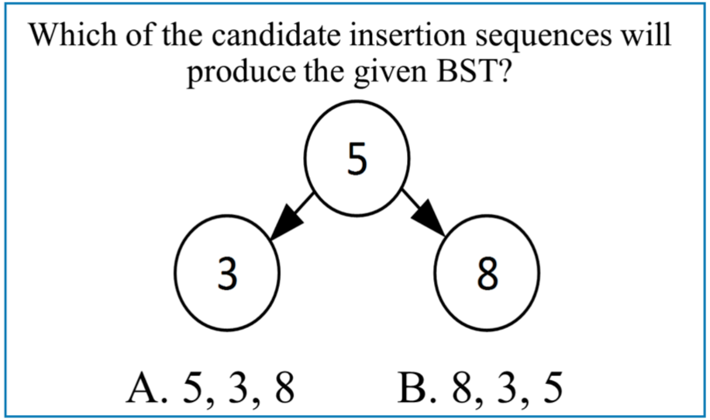


Experimental Design

- Two types of tasks
  - Data structure manipulations
  - Mental rotations: 3D objects
- **fMRI** and **fNIRS**: ***1st time in SE***

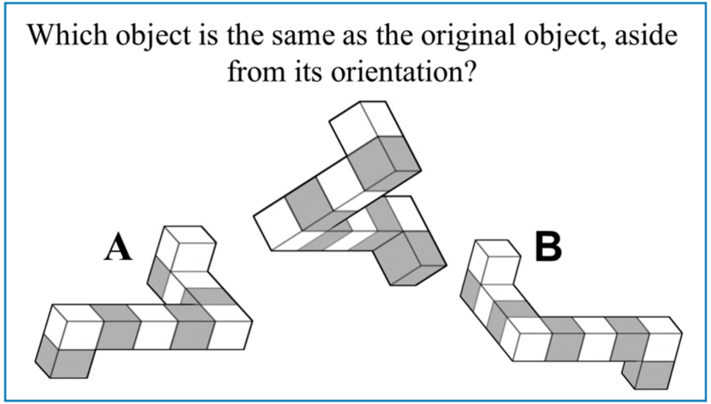


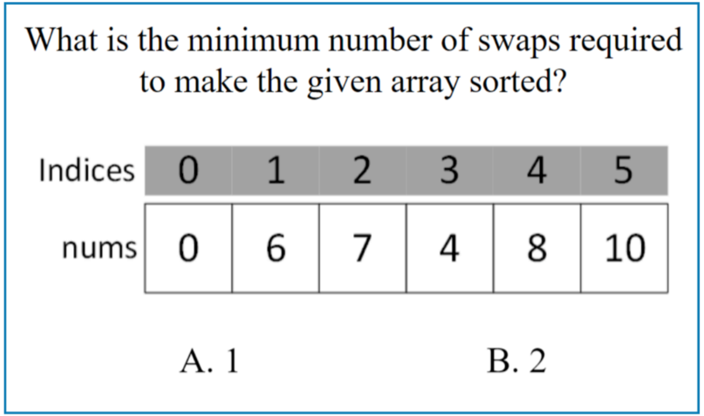


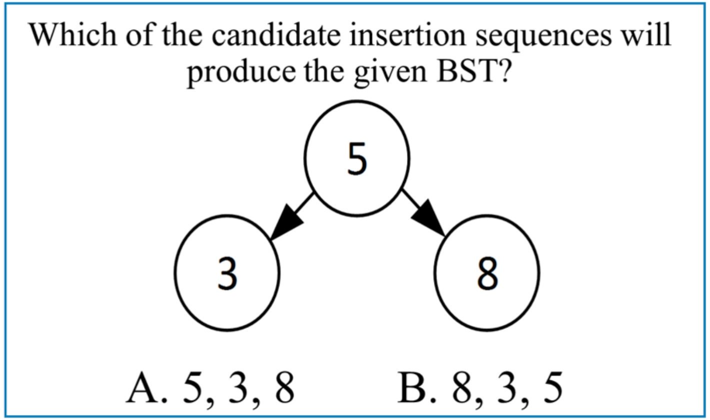


Experimental Design

- Two types of tasks
  - Data structure manipulations
  - Mental rotations: 3D objects
- **fMRI** and **fNIRS**: ***1st time in SE***
- ***Largest*** **in SE: 76 participants**

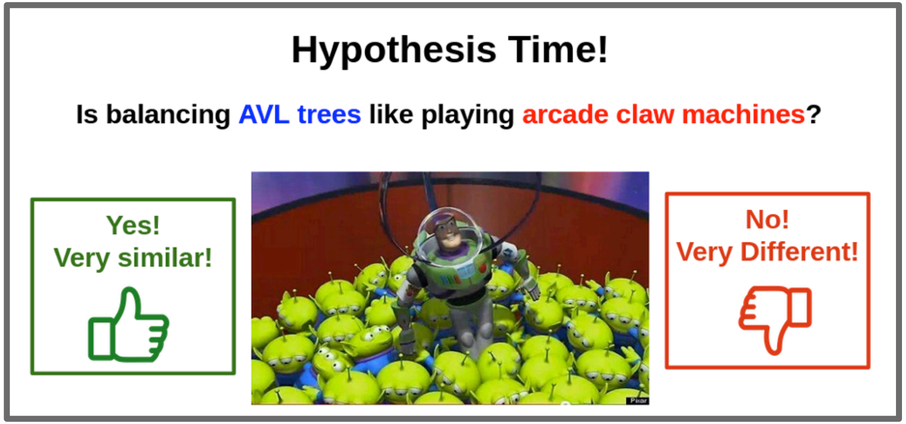


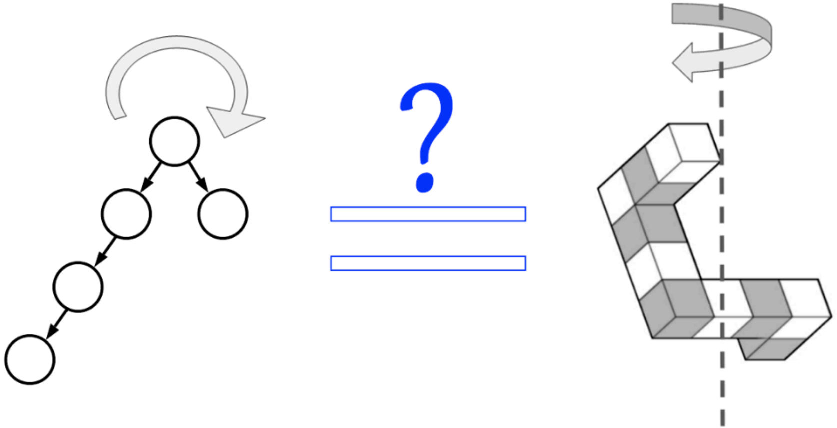




Data Structures vs. Spatial Ability

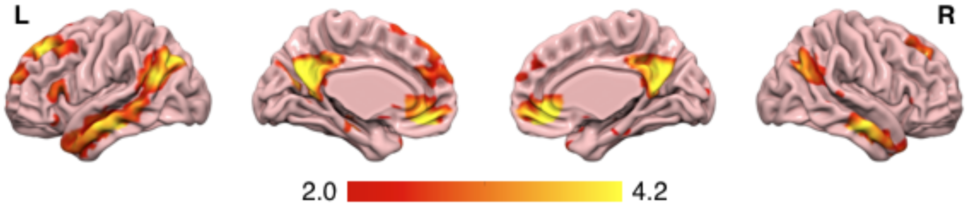
- 70% of human participants believe there is no connection
- What is your answer?





Data Structures vs. Spatial Ability

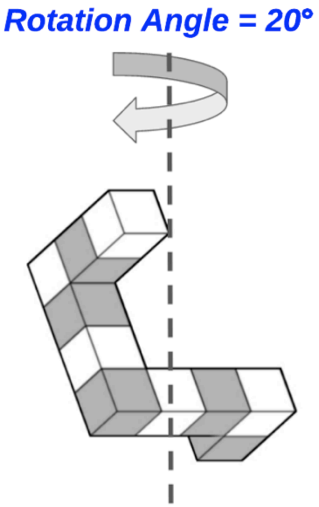
- Data structure manipulations use the same parts of the brain as rotating 3D objects in the real world (spatial ability)
  - fMRI: more similarities than differences ($p<0.01$)
  - fNIRS: activation in the same brain regions ($p<0.01$)

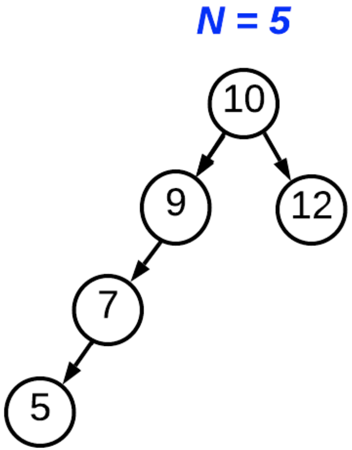


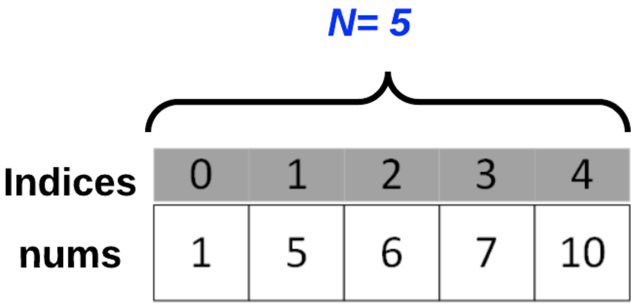
Mental Rotation > Tree

The Role of Task Difficulty

- The brain works even harder for more difficult data structure tasks
  - Difficulty measurement
    - Mental rotations: angle of rotation
    - Data structure: size
- fNIRS: no significant findings for the effect of task difficulty







Summary: Data Structures vs. Spatial Ability

- Large human study: 76 participants
- fMRI vs. fNIRS
- 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

Summary: Data Structures vs. Spatial Ability

- This work may inform:
  - Pedagogy and training
  - Technology transfer
  - Programming expertise

Summary: Data Structures vs. Spatial Ability

- This work may inform:
  - Pedagogy and training
  - Technology transfer
  - Programming expertise
- The findings have been used to direct a longitudinal study
  - Improve CS students' performance on programming using spatial training

Summary: Data Structures vs. Spatial Ability

- This work may inform:
  - Pedagogy and training
  - Technology transfer
  - Programming expertise

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] and Westley Weimer[1]

[1]{yhhy, xinyuliu, ryankrue, xiaosuhu, kjleach, weimerw}@umich.edu, University of Michigan
[2]t.santander@psych.ucsb.edu, Univeristy of California, Santa Barbara

Outline

- Introduction
- Investigating cognition in software engineering
  - **Understanding the neural representations of data structures**
  - **Comparing prose writing and code writing**
  - **Understanding bias in code reviews**
- Career Plan
- Summary

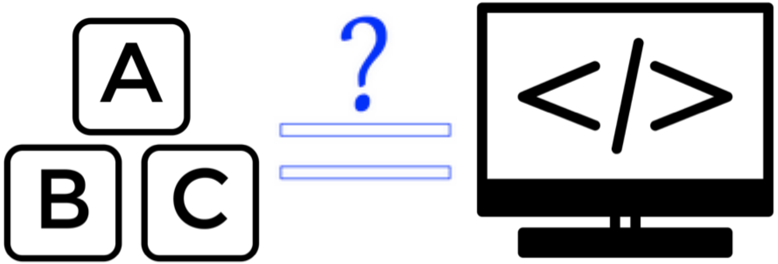
How do we tell truths that might hurt?

Besides a mathematical inclination, an exceptionally good mastery of one's native tongue is the most vital asset of a competent programmer.

Hypothesis Time!

Is writing code like writing English?

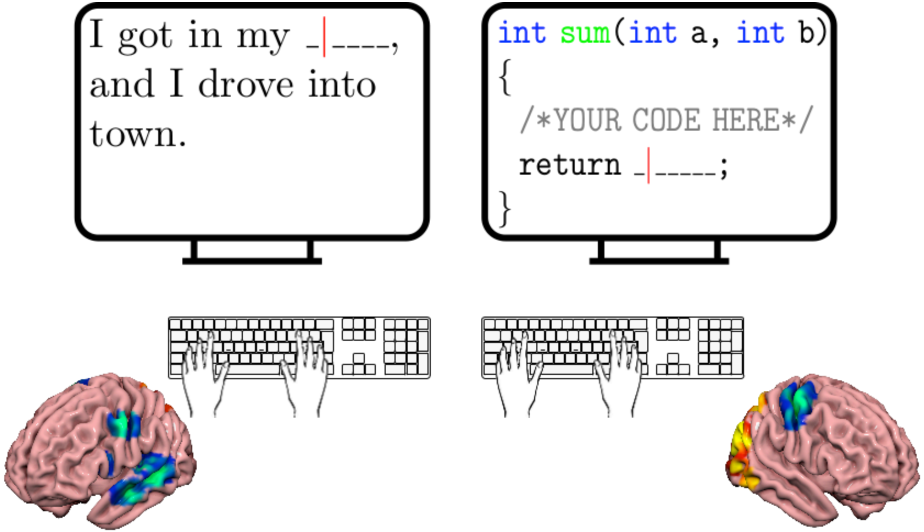
Yes!
Very similar!



No!
Very Different!

Comparing prose writing and code writing *(ICSE'20)*

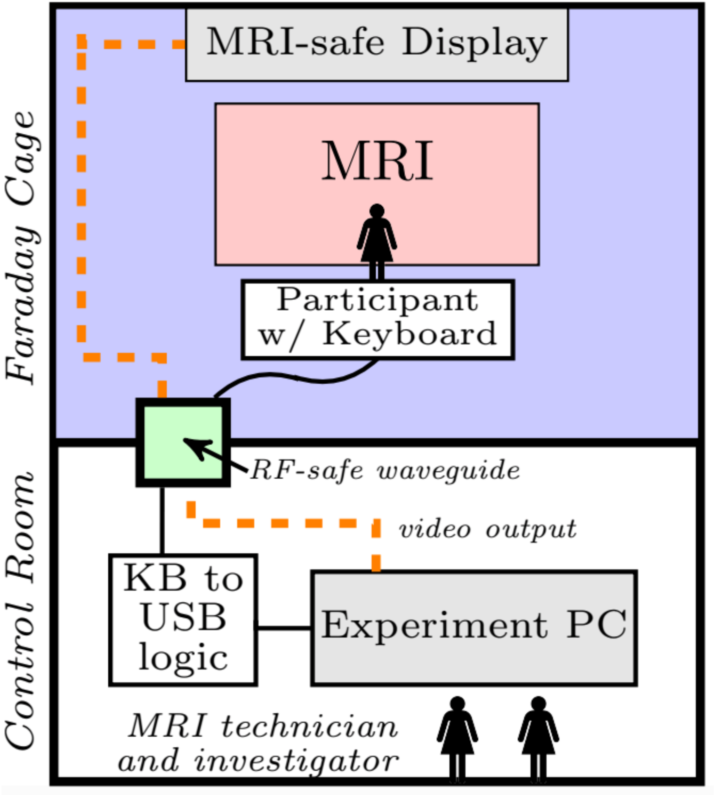
Are code writing and prose writing similar neural activities? Do I have to be good at English writing to become a good software developer?



Comparing prose writing and code writing *(ICSE'20)*

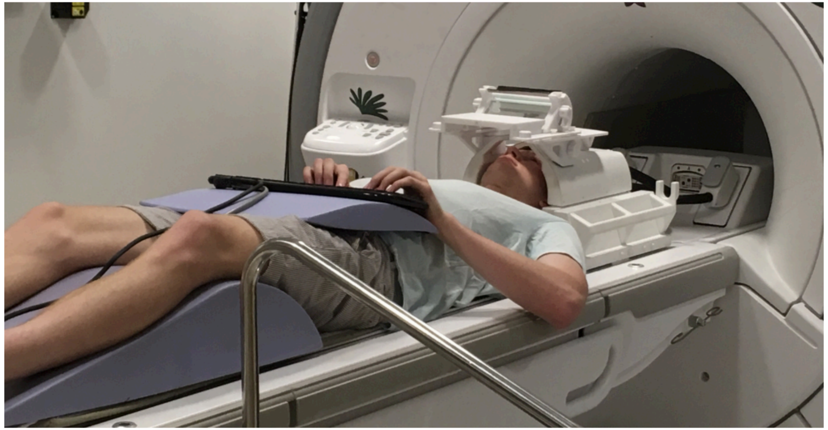
Challenge: Typing in fMRI

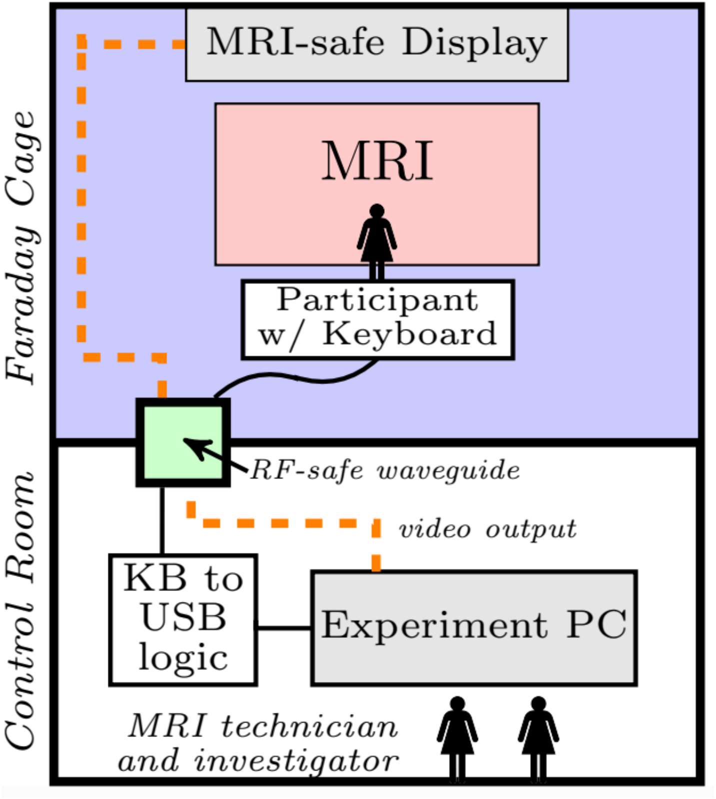
- fMRI-safe bespoke keyboard
  - QWERTY keyboard
  - Allow typing and editing



Challenge: Typing in fMRI

- fMRI-safe bespoke keyboard
  - QWERTY keyboard
  - Allow typing and editing





Experimental Design

- Two-by-two contrast task design: 30 participants

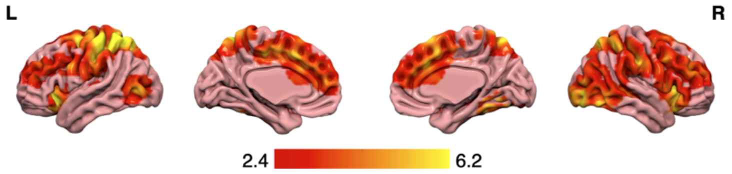
FITB: Fill in the Blank

LR: Long Response

Comparing prose writing and code writing *(ICSE'20)*

Summary: Prose Writing vs. Coding Writing

- Code writing and prose writing are very ***distinct*** neural activities! ($2.4<t<6.2$)
  - Code writing: top-down control, memory, planning, spatial ability
  - Prose writing: language-related regions

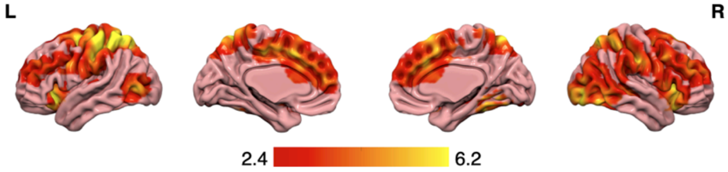


Code > Prose

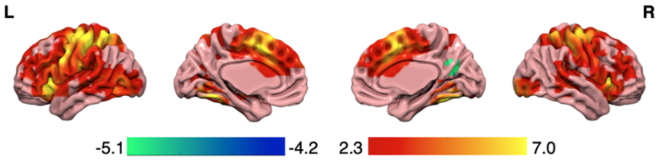
Comparing prose writing and code writing *(ICSE'20)*

Summary: Prose Writing vs. Coding Writing

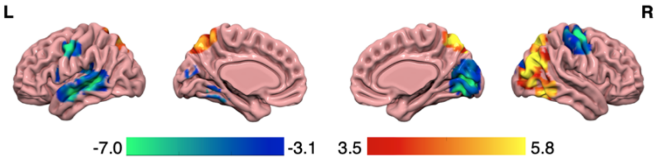
- Code writing and prose writing are very ***distinct*** neural activities! $(2.4<t<6.2)$
  - Code writing: top-down control, memory, planning, spatial ability
  - Prose writing: language-related regions



Code > Prose



Code FITB > Prose FITB

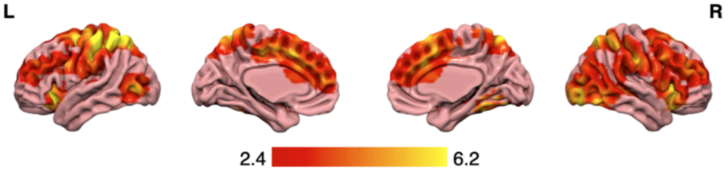


Code LR > Prose LR

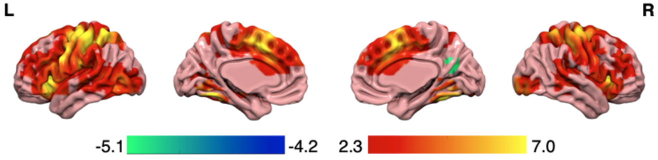
Comparing prose writing and code writing *(ICSE'20)*

Summary: Prose Writing vs. Coding Writing

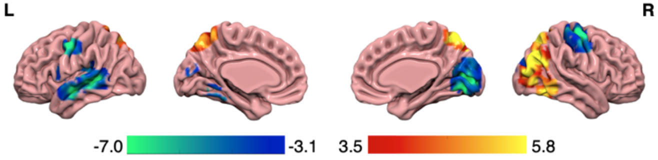
- Code writing and prose writing are very ***distinct*** neural activities! (2.4<t<6.2)
  - Code writing: top-down control, memory, planning, spatial ability
  - Prose writing: language-related regions
- Implications
  - Training and pedagogy
  - Broadening participation
  - Writing proofs?



Code > Prose



Code FITB > Prose FITB



Code LR > Prose LR

Outline

- Introduction
- Investigating cognition in software engineering
  - Understanding the neural representations of data structures
  - Comparing prose writing and code writing
  - Understanding bias in code reviews
- Career plan
- Summary

Code Review

- **Code review is *critical* for software development**
  - ***Systematic*** inspection**,** analysis, evaluation, and revision of code.

Code Review

- **Code review is *critical* for software development**
  - ***Systematic*** inspection**,** analysis, evaluation, and revision of code.

Delete the equal mark in case the array is like {x,x,x...(n),y,y,y,y...(n+1)}

```
2 algorithms/cpp/majorityElement/majorityElement.cpp

@@ -32,7 +32,7 @@ int majorityElement(vector<int> &num) {
                cnt++;
            }else{
                majority == num[i] ? cnt++ : cnt --;
-               if (cnt >= num.size()/2) return majority;
+               if (cnt > num.size()/2) return majority;
            }
        }
        return majority;
```

Code Review

- **Code review is *critical* for software development**
  - ***Systematic*** inspection**,** analysis, evaluation, and revision of code.

Delete the equal mark in case the array is like {x,x,x...(n),y,y,y,y...(n+1)}

Code changes

```diff
2 algorithms/cpp/majorityElement/majorityElement.cpp
@@ -32,7 +32,7 @@ int majorityElement(vector<int> &num) {
                 cnt++;
             }else{
                 majority == num[i] ? cnt++ : cnt --;
-                if (cnt >= num.size()/2) return majority;
+                if (cnt > num.size()/2) return majority;
             }
         }
         return majority;
```

Code Review

- **Code review is *critical* for software development**
  - ***Systematic*** inspection**,** analysis, evaluation, and revision of code.

Commit message

Delete the equal mark in case the array is like {x,x,x...(n),y,y,y,y...(n+1)}

Code changes

```
2 algorithms/cpp/majorityElement/majorityElement.cpp

@@ -32,7 +32,7 @@ int majorityElement(vector<int> &num) {
                     cnt++;
                 }else{
                     majority == num[i] ? cnt++ : cnt --;
-                    if (cnt >= num.size()/2) return majority;
+                    if (cnt > num.size()/2) return majority;
                 }
             }
         return majority;
```

Understanding bias in code reviews *(FSE’20)*

Code Review

- **Code review is *critical* for software development**
  - ***Systematic*** inspection**,** analysis, evaluation, and revision of code.

Code Review

- **Code review is *critical* for software development**
  - ***Systematic*** inspection, analysis, evaluation, and revision of code.

Beyond the Code Itself:
How Programmers *Really* Look at Pull Requests

Denae Ford, Mahnaz Behroozi
North Carolina State University
Raleigh, NC, USA
{dford3, mbehroo}@ncsu.edu

Alexander Serebrenik
Eindhoven University of Technology
Eindhoven, The Netherlands
a.serebrenik@tue.nl

Chris Parnin
North Carolina State University
Raleigh, NC, USA
cjparnin@ncsu.edu

Code Review

- **Code review is *critical* for software development**
  - ***Systematic*** inspection, analysis, evaluation, and revision of code.

Beyond the Code Itself:
How Programmers *Really* Look at Pull Requests

Denae Ford, Mahnaz Behroozi
North Carolina State University
Raleigh, NC, USA
{dford3, mbehroo}@ncsu.edu

Alexander Serebrenik
Eindhoven University of Technology
Eindhoven, The Netherlands
a.serebrenik@tue.nl

Chris Parnin
North Carolina State University
Raleigh, NC, USA
cjparnin@ncsu.edu

Code Review

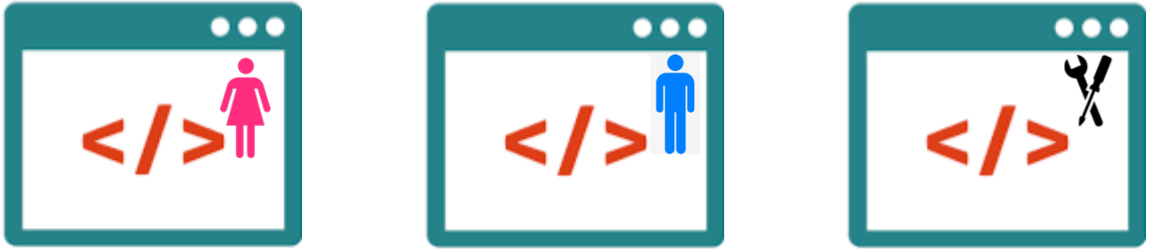
- **Code review is *critical* for software development**
  - ***Systematic*** inspection**,** analysis, evaluation, and revision of code.

Beyond the Code Itself:
How Programmers *Really* Look at Pull Requests

Denae Ford, Mahnaz Behroozi
North Carolina State University
Raleigh, NC, USA
{dford3, mbehroo}@ncsu.edu

Alexander Serebrenik
Eindhoven University of Technology
Eindhoven, The Netherlands
a.serebrenik@tue.nl

Chris Parnin
North Carolina State University
Raleigh, NC, USA
cjparnin@ncsu.edu



Understanding bias in code reviews *(FSE'20)*

Hypothesis Time!

Do women and men review code in the same way?

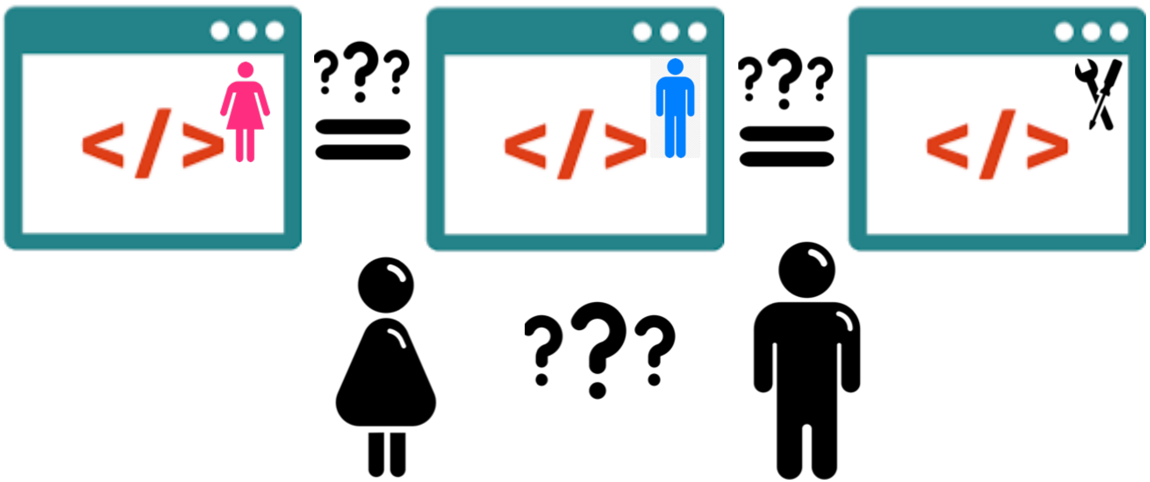
Yes!
Very similar!

No!
Very Different!

Is there bias on gender and identities in code review?

How do we characterize the bias?

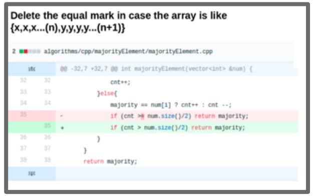
- Systematically
- Objectively
- Rigorously



Experimental Design

- **60 C/C++ pull requests from GitHub**

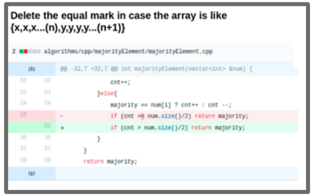
60 Pull Requests



Experimental Design

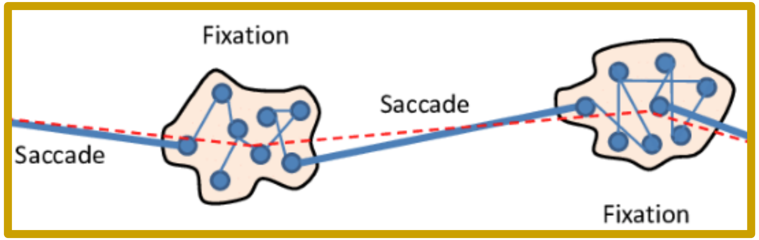
- **60 C/C++ pull requests from GitHub**

60 Pull Requests

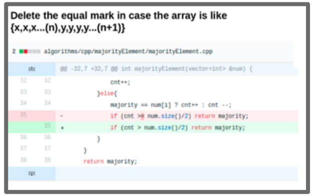


Experimental Design

- **60 C/C++ pull requests from GitHub**



60 Pull Requests



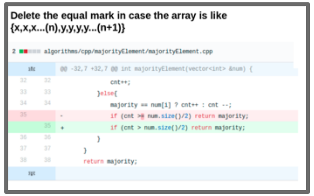
Delete the equal mark in case the array is like {x,x,x...(n),y,y,y,y...(n+1)}

```
2 algorithms/cpp/majorityElement/majorityElement.cpp
@@ -32,7 +32,7 @@ int majorityElement(vector<int> &num) {
                cnt++;
            }else{
                majority == num[i] ? cnt++ : cnt --;
-               if (cnt >= num.size()/2) return majority;
+               if (cnt > num.size()/2) return majority;
            }
        }
        return majority;
```

Experimental Design

- **60 C/C++ pull requests from GitHub**
- **Author images: Relabel the author information**
  - Chicago Face Database

60 Pull Requests ➕


Delete the equal mark in case the array is like {x,x,x...(n),y,y,y,y...(n+1)}

```
2 algorithms/cpp/majorityElement/majorityElement.cpp
@@ -32,7 +32,7 @@ int majorityElement(vector<int> &num) {
32 32              cnt++;
33 33          }else{
34 34              majority == num[i] ? cnt++ : cnt --;
35    -             if (cnt >= num.size()/2) return majority;
   35 +             if (cnt > num.size()/2) return majority;
36 36          }
37 37      }
38 38      return majority;
```

20 Women

20 Men

20 Machine

Experimental Design

- 60 C/C++ pull requests from GitHub
- Author images: Relabel the author information
- Construction of code review stimuli

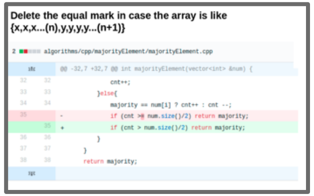
60 Pull Requests

+

20 Women

20 Men

20 Machine



60 Stimuli: V1

60 Stimuli: V2

Experimental Design

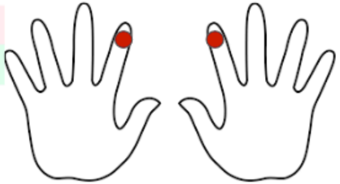
- 60 C/C++ pull requests from GitHub
- Author images: Relabel the author information
- Const

60 Pu

Delete the equal mark in case the array is like {x,x,x...(n),y,y,y,y...(n+1)}

```
2 algorithms/cpp/majorityElement/majorityElement.cpp
@@ -32,7 +32,7 @@ int majorityElement(vector<int> &num) {
                cnt++;
            }else{
                majority == num[i] ? cnt++ : cnt --;
-               if (cnt >= num.size()/2) return majority;
+               if (cnt > num.size()/2) return majority;
            }
        }
        return majority;
```

Owner:



Accept Reject

li: V1

li: V2

Experimental Design

Recruitment → Deception → Code Reviews → Post Survey → Debriefing

- Avoid social desirability bias
- 37 Participants
- Post-survey questions

Demographic	Number of Participants		
	Total	Version I	Version II
Men	21	11	10
Women	16	7	9
Undergraduate	26	11	15
Graduate	11	7	4

Summary: Biases in Code Review

- We find **universal biases** in how all participants treat code reviews as a function of the **reviewers' gender** and **apparent author**:

Summary: Biases in Code Review

- We find **universal biases** in how all participants treat code reviews as a function of the **reviewers' gender** and **apparent author**:
  - **Behavioral difference**
    i. All participants spend ***less time*** evaluating the Pull Requests of ***women*** ($p<0.01$)
    ii. All participants are ***less likely to accept*** the Pull Requests of ***machines*** ($p<0.05$)
    iii. Women reviewers spent ***less time*** on all Pull Requests ($p<0.0001$)

Summary: Biases in Code Review

- We find **universal biases** in how all participants treat code reviews as a function of the **reviewers' gender** and **apparent author**:
  - **Behavioral difference**
    1. All participants spend ***less time*** evaluating the Pull Requests of ***women*** ($p<0.01$)
    2. All participants are ***less likely to accept*** the Pull Requests of ***machines*** ($p<0.05$)
    3. ***Women reviewers*** spent ***less time*** on all Pull Requests ($p<0.0001$)
  - **Visual difference**
    1. Men and women reviewers employ ***different high-level problem-solving strategies*** ($p<0.001$)
    2. **Men fixated more frequently** ($p<0.001$), while **women** spent significantly **more time** analyzing **Pull Requests messages** and **author pictures.**

Summary: Biases in Code Review

- We find **universal biases** in how all participants treat code reviews as a function of the **reviewers' gender** and **apparent author:**
  - **Neurological difference**
    i. It is possible to ***distinguish*** women and men conducting code review

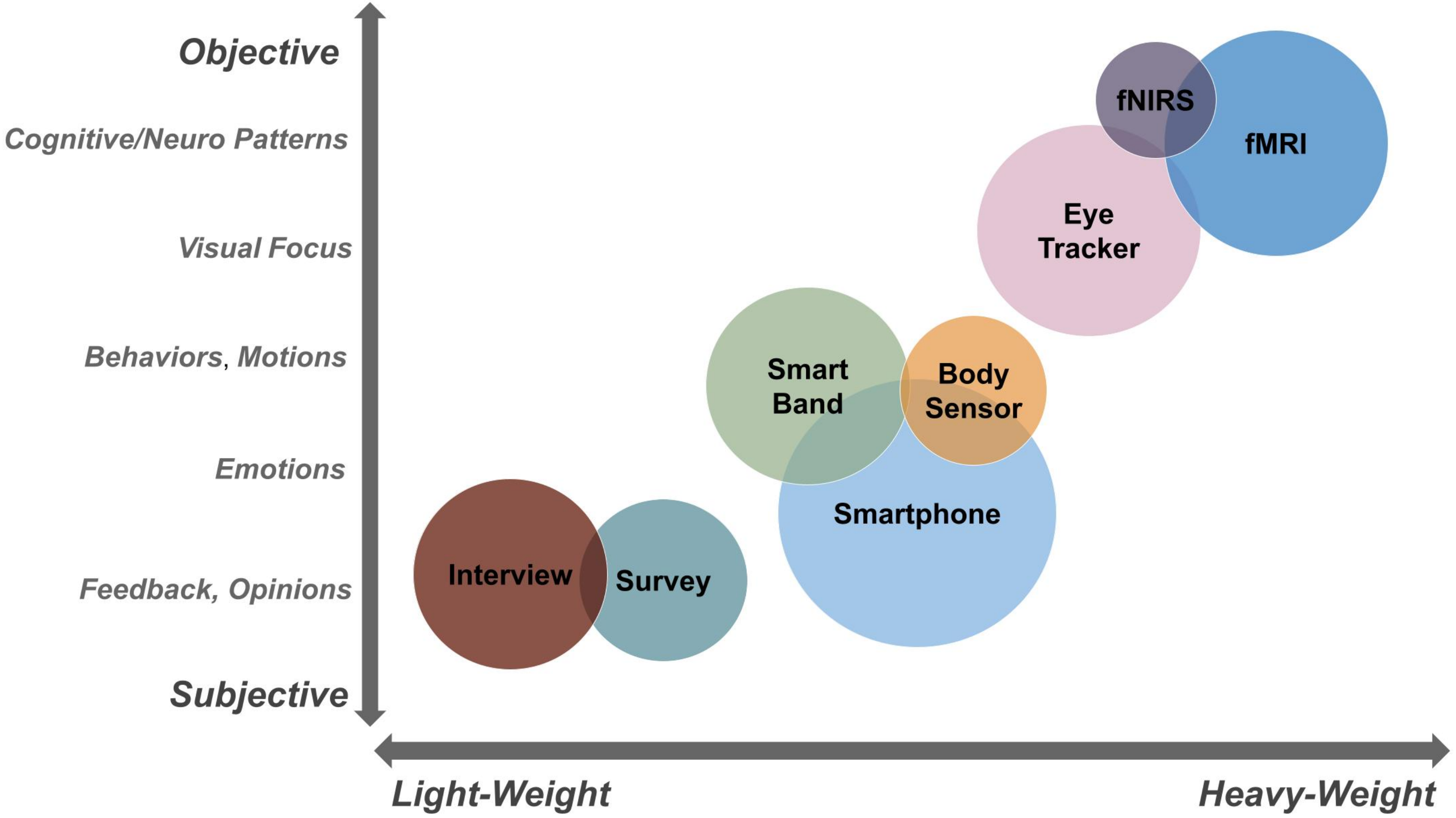
Summary: Biases in Code Review

- We find **universal biases** in how all participants treat code reviews as a function of the **reviewers' gender** and **apparent author:**
    - **Neurological difference**
        i. It is possible to ***distinguish*** women and men conducting code review
- Participants' **self-reported perception** in code review **do not align** with the objective observations.
    - Do not realize the existence of difference on gender
    - Bias against machines exists

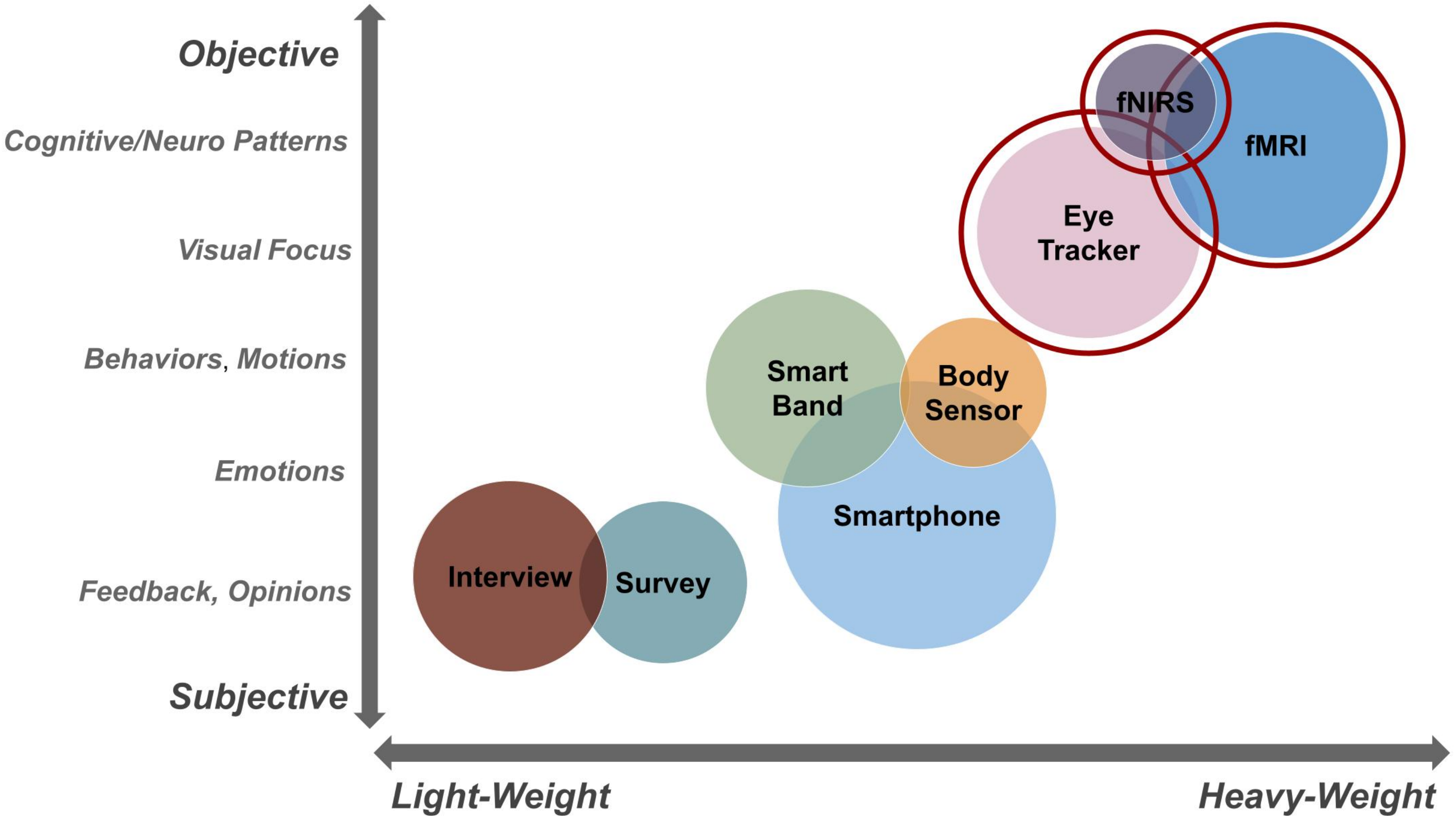
Summary: Biases in Code Review

- We find **universal biases** in how all participants treat code reviews as a function of the **reviewers' gender** and **apparent author:**
  - **Neurological difference**
    i. It is possible to ***distinguish*** women and men conducting code review
- Participants' **self-reported perception** in code review **do not align** with the objective observations.
  - Do not realize the existence of difference on gender
  - Bias against machines exists
- Implications
  - How should we design code review environment based on the **differences**?
  - Should we avoid showing **authors' profiles**?
  - Is there any effective training to **mitigate the biases**?

Thesis Scope: Metrics for Human Factors in SE



Thesis Scope: Metrics for Human Factors in SE



Publications: Supporting this Thesis

1. **Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS**. ***(SIGSOFT Distinguished Paper Award)***
   Yu Huang, Xinyu Liu, Ryan Krueger, Tyler Santander, Xiaosu Hu, Kevin Leach, Westley Weimer.
   41st ACM/IEEE International Conference on Software Engineering (ICSE 2019).
2. **Neurological Divide: An fMRI Study of Prose and Code Writing.**
   Ryan Krueger, **Yu Huang**, Xinyu Liu, Tyler Santander, Westley Weimer, Kevin Leach.
   42nd ACM/IEEE International Conference on Software Engineering (ICSE 2020).
3. **Biases and Differences in Code Review using Medical Imaging and Eye-Tracking: Genders, Humans, and Machines.**
   Yu Huang, Kevin Leach, Zohreh Sharafi, Tyler Santander, Westley Weimer.
   ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2020)
4. **Towards an Objective Measure of Developers' Cognitive Activities.**
   Zohreh Sharafi, **Yu Huang**, Kevin Leach, Westley Weimer.
   ACM Transactions on Software Engineering and Methodology, Volume 30, Issue 3 (TOSEM 2021)

Publications: Others

1. **Connecting the Dots: Rethinking the Relationship between Code and Prose Writing with Functional Connectivity.** Zachary Karas, Andrew Jahn, Westley Weimer, **Yu Huang**. *ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2021). To Appear.*
2. **Applying Automated Program Repair to Dataflow Programming Languages. Yu Huang**, Hammad Ahmad, Stephanie Forrest, Westley Weimer. *10th International Workshop on Genetic Improvement (GI 2021 @ ICSE 2021). To Appear.*
3. **Leaving My Fingerprints: Motivations and Challenges of Contributing to OSS for Social Good. Yu Huang,** Denae Ford, Thomas Zimmermann. *43rd ACM/IEEE International Conference on Software Engineering (ICSE 2021). To Appear.*
4. **Trustworthiness Perceptions in Code Review: An Eye-tracking Study.** Ian Bertram, Jack Hong, **Yu Huang**, Westley Weimer, Zohreh Sharafi. *Empirical Software Engineering and Measurement, Emerging Results and Vision Papers (ESEM 2020).*
5. **A Human Study of Comprehension and Code Summarization.** Sean Stapleton, Yashmeet Gambhir, Alexander LeClair, Zachary Eberhart, Westley Weimer, Kevin Leach, **Yu Huang.** *28th IEEE/ACM International Conference on Program Comprehension (ICPC 2020).*
6. **Understanding Behavioral Dynamics of Social Anxiety Among College Students Through Smartphone Sensors.** Jiaqi Gong, **Yu Huang**, Philip I Chow, Karl Fua, Matthew Gerber, Bethany Teachman, Laura Barnes. *Information Fusion, 49:57–68, September 2019.*
7. **Physiological Changes Over the Course of Cognitive Bias Modification for Social Anxiety**. Mehdi Boukhechba, Jiaqi Gong, Kamran Kowsari, Mawulolo K Ameko, Karl Fua, Philip I Chow, **Yu Huang**, Bethany A Teachman, and Laura E Barnes. *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on, pages 422–425.*

Publications: Others

8. **I Did OK, But Did I Like It? Using Ecological Momentary Assessment to Examine Perceptions of Social Interactions Associated with Severity of Social Anxiety and Depression.** Emily C Geyer, Karl C Fua, Katharine E Daniel, Philip I Chow, Wes Bonelli, **Yu Huang**, Laura E Barnes, and Bethany A Teachman. *Behavior therapy, 49(6):866–880, 2018*
9. **Discovery of Behavioral Markers of Social Anxiety From Smartphone Sensor Data. Yu Huang**, Jiaqi Gong, Mark Rucker, Philip Chow, Karl Fua, Matthew S. Gerber, Bethany Teachman, and Laura E. Barnes. *1st Workshop on Digital Biomarkers, DigitalBiomarkers '17, pages 9–14, New York, NY, USA, ACM.*
10. **Using Mobile Sensing to Test Clinical Models of Depression, Social Anxiety, State Affect, and Social Isolation Among College Students.** Philip I. Chow, Karl Fua, **Yu Huang**, Wesley Bonelli, Haoyi Xiong, Laura E. Barnes, and Bethany Teachman. *Journal of Med Internet Res, 19(3):e62, Mar 2017.*
11. **Monitoring Social Anxiety From Mobility and Communication Patterns.** Mehdi Boukhechba, **Yu Huang**, Philip Chow, Karl Fua, Bethany A. Teachman, and Laura E.Barnes. *The ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, UbiComp '17, pages 749–753.*
12. **Daehr: A Discriminant Analysis Framework for Electronic Health Record Data and an Application to Early Detection of Mental Health Disorders.** Haoyi Xiong, Jinghe Zhang, **Yu Huang**, Kevin Leach, and Laura E. Barnes. *ACM Trans. Intell. Syst. Technol., 8(3):47:1–47:21, February 2017.*
13. **Assessing Social Anxiety Using GPS Trajectories and Point-of-Interest Data. Yu Huang**, Haoyi Xiong, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua, Bethany A Teachman, and Laura E Barnes. *2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16), pages 898–903.*

Publications: Others

14. **Sensus: a Cross-Platform, General-Purpose System for Mobile Crowdsensing in Human-Subject Studies.** Haoyi Xiong, **Yu Huang**, Laura E Barnes, and Matthew S Gerber. *In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp* '16, pages 415–426.
15. **Demons: an Integrated Framework for Examining Associations Between Physiology and Selfreported affect Tied to Depressive Symptoms.** Philip Chow, Wesley Bonelli, **Yu Huang**, Karl Fua, Bethany A Teachman, and Laura E Barnes. *In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, pages 1139–1143.*
16. **A Design and Theoretical Analysis of a 145 mV to 1.2 V Single-Ended Level Converter Circuit for Ultra-Low Power Low Voltage ICs.** **Yu Huang**, Aatmesh Shrivastava, Laura E Barnes, and Benton H Calhoun. *Journal of Low Power Electronics and Applications, 6(3):11, 2016*.
17. **M-SEQ: Early Detection of Anxiety and Depression via Temporal Orders of Diagnoses in Electronic Health Data.** Jinghe Zhang, Haoyi Xiong, **Yu Huang**, Hao Wu, Kevin Leach, and Laura Barnes. *In Proceedings of the 2015 IEEE International Conference on Big Data (BigData 2015), September 2015*.
18. **A 145 mV to 1.2 V Single Ended Level Converter Circuit for Ultra-Low Power Low Voltage ICs.** **Yu Huang**, Aatmesh Shrivastava, and Benton H Calhoun. *In SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2015 IEEE, pages 1–3*.
19. **Optimizing Energy Efficient Low Swing Interconnect for Sub-Threshold FPGAs.** He Qi, Oluseyi Ayorinde, **Yu Huang**, and Benton Calhoun. *In Field Programmable Logic and Applications (FPL), 2015 25th International Conference on, pages 1–4. IEEE, 2015*.

Publications: Others

20. **Using Island-Style Bi-directional Intra-CLB Routing in Low-Power FPGAs.** Oluseyi Ayorinde, He Qi, **Yu Huang**, and Benton H Calhoun. *In Field Programmable Logic and Applications (FPL), 2015 25th International Conference on, pages 1–7. IEEE, 2015*.

Career Plan

- Next stop: Assistant Professor (tenure-track) in Vanderbilt

Career Plan

- Next stop: Assistant Professor (tenure-track) in Vanderbilt
- Research plan

Career Plan

- Next stop: Assistant Professor (tenure-track) in Vanderbilt
- Research plan
  - Cognitive modeling for SE
    - Non-functional factors in software development process
    - Development of expertise in CS
    - Integation of cognitve models in automated program repair

Career Plan

- Next stop: Assistant Professor (tenure-track) in Vanderbilt
- Research plan
    - Cognitive modeling for SE
        - Non-functional factors in software development process
        - Development of expertise in CS
        - Integation of cognitve models in automated program repair
    - Sustaining participation in/via Open Source Software for Social Good
    - Supporting HW/SW co-design

Acknowledgement

Acknowledgement

My advisor: Prof. Westley Weimer

Acknowledgement

My advisor: Prof. Westley Weimer

Acknowledgement

My committee members

Prof. Westley Weimer

Prof. Stephanie Forrest

Prof. Mark Guzdial

Prof. Ioulia Kovelman

Dr. Kevin Leach
@UM

Dr. Zohreh Sharafi
@UM

Dr. Tyler Santander
@UC Santa Barbara

Dr. Xiaosu Hu
@UM

Dr. Andrew Jahn
@UM

Dr. Denae Ford
@MSR

Dr. Tom Zimmermann
@MSR

Hammad Ahmad
@UM

Dr. Stephanie Forrest
@ASU

Dr. Kevin Angstadt
@St. Lawrence

Xinyu Liu
@UM

Ryan Krueger
@UM

Ian Bertram
@UM

Nick McKay
@UM

Jack Hong
@UM

Mike Flanagan
@UM

Sean Stapleton
@UM

Zach Karas
@UM

Heather Lukas
@UVa

Yuyan Zhang
@Uva

Wesley Bonelli
@Uva

Sarai Alvarez
@UVa

Acknowledgement

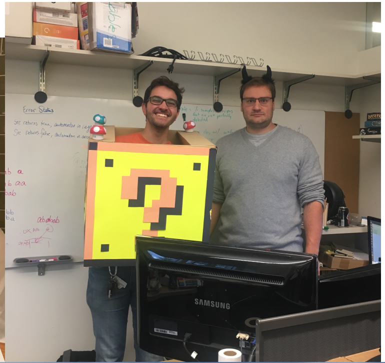
WRG: Kevin Leach, Zohreh Sharafi, Kevin Angstadt, Madeline Endres, Colton Holoday, Yirui Liu, Hammad Ahmad, Annie Li, Fee Christoph, Yiannos Demetriou…

And all my friends!

Acknowledgement

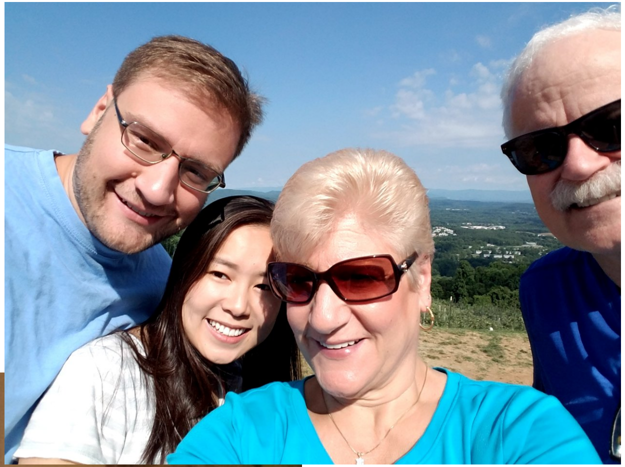
WRG: Kevin Leach, Zohreh Sharafi, Kevin Angstadt, Madeline Endres, Colton Holoday, Yirui Liu, Hammad Ahmad, Annie Li, Fee Christoph, Yiannos Demetriou…

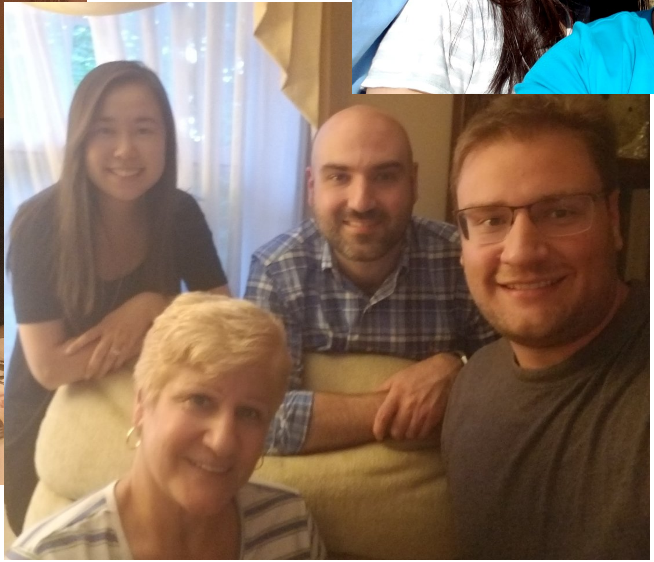
And all my friends!



Acknowledgement

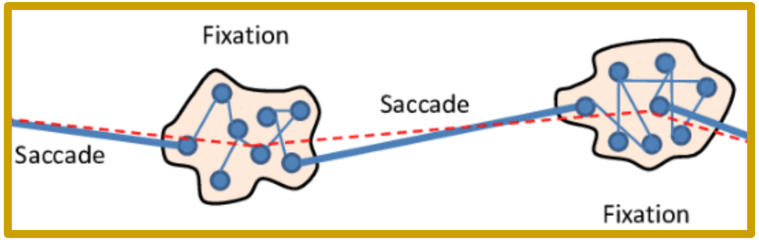
- My parents: Zhonglin Huang and Qian Xu
- My husband: Kevin Leach
- My parents-in-law: Richard and Linda Leach
- My brother-in-law: Eric Leach





Summary

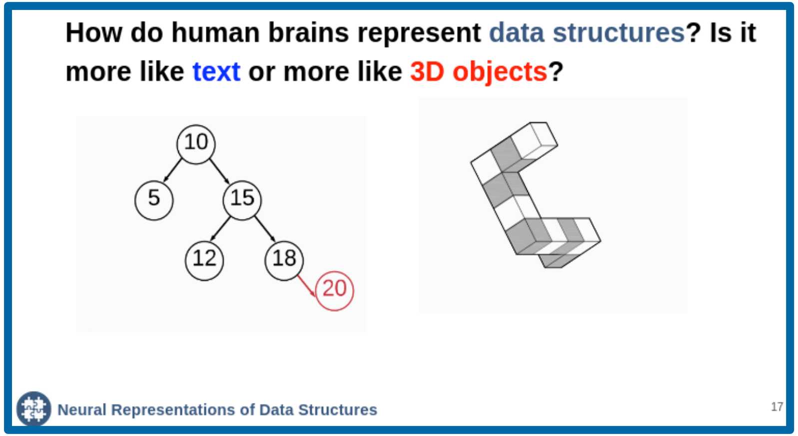
- Measure cognitive processes in software activities



De-identified data is public: http://www-personal.umich.edu/~yhhy/

Summary

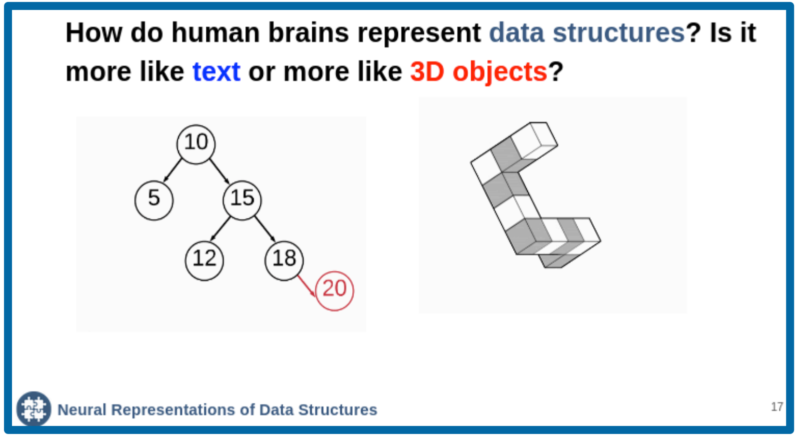
- Measure cognitive processes in software activities
- Novel concept, problems, and approaches

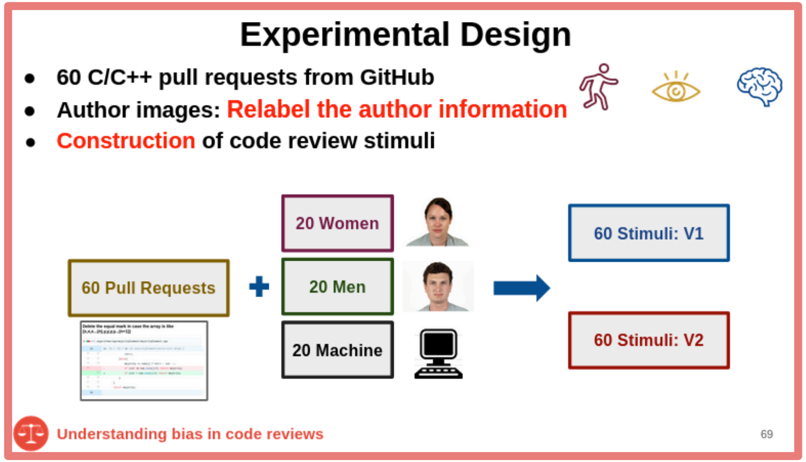


De-identified data is public: http://www-personal.umich.edu/~yhhy/

Summary

- Measure cognitive processes in software activities
- Novel concept, problems, and approaches
- Rigorous controlled experiments

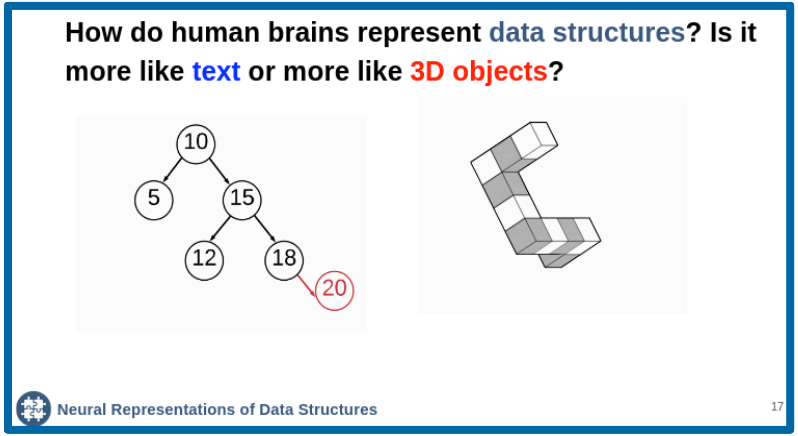


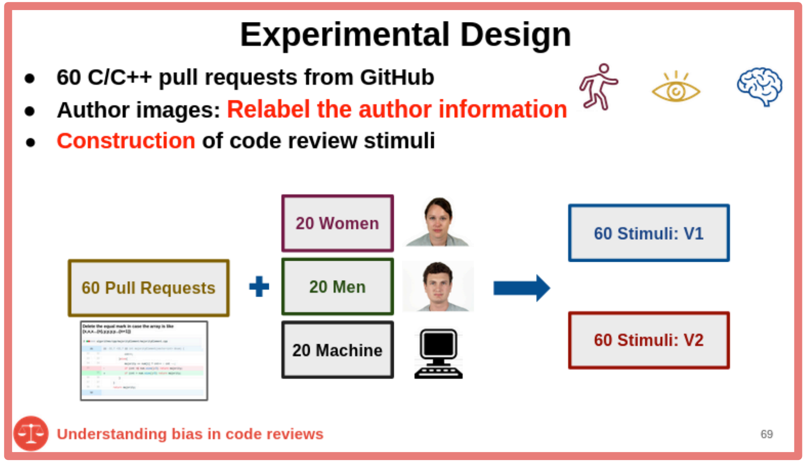


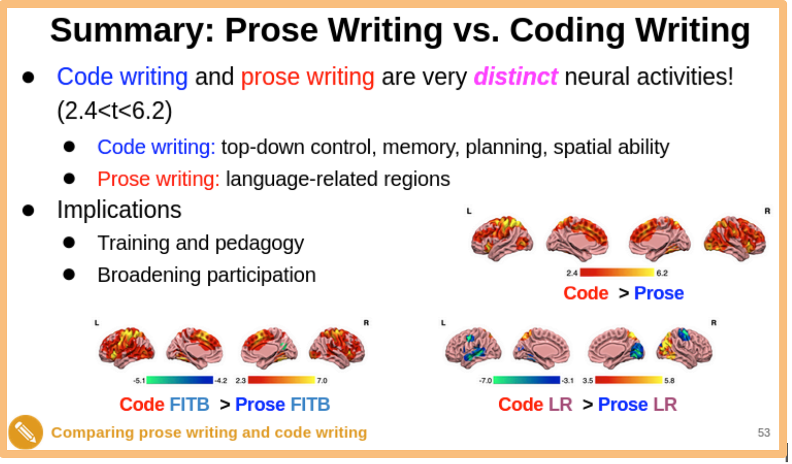
De-identified data is public: http://www-personal.umich.edu/~yhhy/

Summary

- Measure cognitive processes in software activities
- Novel concepts, problems, and approaches
- Rigorous controlled experiments
- Potentials for broad impact







De-identified data is public: http://www-personal.umich.edu/~yhhy/