Understanding User Cognition: from Everyday Behavior and Spatial Ability to Code Writing and Review

Yu Huang

University of Michigan
Dec 11, 2019
Break Down the Title

● A standard workday of a software developer
Break Down the Title

- A standard workday of a software developer
Break Down the Title

- A standard workday of a software developer
Break Down the Title

A standard workday of a software developer

Problem Introduction and Motivation
Break Down the Title

- A standard workday of a software developer

What could go wrong?
What is currently holding us back?
Break Down the Title

- A standard workday of a software developer

Silicon Valley Has A Mental Health Crisis Too

We need to talk about Silicon Valley's mental health problems

Depression & Anxiety in programming

I recently attended a developer retreat in beautiful Golden, Colorado and ran into something that surprised me. We were doing an ice breaker exercise and...
Break Down the Title

- A standard workday of a software developer

- 62% have mental complaints
- 31% have mental ill-health
- <1% sought for professional help

Leads to impairment in academic functioning and relationship!
Break Down the Title

- How can we be more effective and efficient in programming? What are the cognitive processes of programming? What affects our decisions in programming?
Break Down the Title

- How can we be more effective and efficient in programming? What are the cognitive processes of programming? What affects our decisions in programming?
  - Traditional research solutions: self-reporting
Break Down the Title

- How can we be more effective and efficient in programming? What are the cognitive processes of programming? What affects our decisions in programming?
  - Traditional research solutions: self-reporting
    - Unreliable

Self-Reports in Organizational Problems and Prospects

Philip M. Podsakoff
Dennis W. Organ
Indiana University

Self-reports figure prominently in organizational and research, but there are several problems associated with them. This article identifies six categories of self-reports and describes problems as common method variance, the consistency of social desirability. Statistical and post hoc remedies and dual methods for dealing with artificial bias are presented. Recommendations for future research are also offered.

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

Australian Journal of Advanced Nursing
Volume 25 Issue 4 (June/Aug 2008)
van de Mortel, Thea F

Abstract: Objective: The tendency for questionnaires is called socially desirable answering. Creating false relationships or obscuring scales can be used to detect, minimise questionnaire-based research. The air related studies that used questionnaire...
Break Down the Title

- How can we be more effective and efficient in programming? What are the cognitive processes of programming? What affects our decisions in programming?
  - Traditional research solutions: Unreliable self-reporting
  - Observed potential bias of non-functional factors

Problem Introduction and Motivation
Break Down the Title

- How can we be more effective and efficient in programming? What are the cognitive processes of programming? What affects our decisions in programming?
  - Traditional research solutions: Unreliable self-reporting
  - Observed potential bias of non-functional factors

Lack a foundational understanding
Desired properties for this proposal

- Bring all the concerns together:
Desired properties for this proposal

- Bring all the concerns together:
  - **Objective** measures
    - Not just self-reporting
Desired properties for this proposal

● Bring all the concerns together:
  ● **Objective** measures
    ○ Not just self-reporting
  ● **Foundational understanding** of software activities
    ○ What are the cognitive processes of programming?
Desired properties for this proposal

● Bring all the concerns together:
  ● **Objective** measures
    ○ Not just self-reporting
  ● **Foundational understanding** of software activities
    ○ What are the cognitive processes of programming?
  ● **Higher-level** programming tasks
    ○ Data structures; code writing; code reviews
Desired properties for this proposal

- Bring all the concerns together:
  - **Objective** measures
    - Not just self-reporting
  - **Foundational understanding** of software activities
    - What are the cognitive processes of programming?
  - **Higher-level** programming tasks
    - Data structures; code writing; code reviews
  - **Generalizability** across different user groups
    - How is productivity mitigated by group difference
Insights

• It is now **possible** to conduct studies that acquire **objective data** to understand the underlying **cognitive processes** of certain tasks
  • Mobile crowdsensing (MCS); medical imaging; eye tracking
Insights

- It is now **possible** to conduct studies that acquire **objective data** to understand the underlying **cognitive processes** of certain tasks
  - Mobile crowdsensing (MCS); medical imaging; eye tracking
- We can adapt **scientific approaches and concepts** from **other domains** to assist our investigation and understanding of certain tasks
  - Social anxiety; spatial ability; creative writing
Insights

- It is now **possible** to conduct studies that acquire **objective data** to understand the underlying **cognitive processes** of certain tasks
  - Mobile crowdsensing (MCS); medical imaging; eye tracking
- We can adapt **scientific approaches and concepts** from **other domains** to assist our investigation and understanding of certain tasks
  - Social anxiety; spatial ability; creative writing
- It is now **possible** to study historically-subjective factors by designing rigorous **controlled experiments**
  - Contrast-based experiments
It is possible to meaningfully and objectively measure user cognition to understand the mental status, role of spatial ability, fundamental processes and stereotypical associations in certain software engineering activities by combining mobile crowdsensing (MCS), medical imaging, and eye tracking.
Proposal Overview: Four Components
Proposal Overview: Four Components

Monitoring mental health using mobile crowdsensing
Proposal Overview: Four Components

- Monitoring mental health using mobile crowdsensing
- Understanding the neural representation of data structures
Proposal Overview: Four Components

- Monitoring mental health using mobile crowdsensing
- Understanding the neural representation of data structures
- Comparing prose writing and code writing
Proposal Overview: Four Components

- Monitoring mental health using mobile crowdsensing
- Understanding the neural representations of data structures
- Comparing prose writing and code writing
- Understanding bias in code reviews
Can we monitor humans’ mental health status objectively via their everyday behaviors in a natural setting?
Monitoring Mental Health Using Mobile Crowdsensing

- **Sensus**: Cross-platform, general MCS mobile application for human-subject studies
- **A MCS-based framework**: understanding the relationship between human behaviors and mental health status
Sensus: Cross-Platform, General MCS
Sensus: Cross-Platform, General MCS

1. Target **heterogeneous** mobile infrastructures
Sensus: Cross-Platform, General MCS

1. Target **heterogeneous** mobile infrastructures
2. Support a **wide range** of MCS-based human studies
Sensus: Cross-Platform, General MCS

1. Target **heterogeneous** mobile infrastructures
2. Support a **wide range** of MCS-based human studies
3. **Eliminate** the need for **programming background**

Component 1: Monitoring Mental Health
1. Target heterogeneous mobile infrastructures
2. Support a wide range of MCS-based human studies
3. Eliminate the need for programming background
4. Rely on readily-available mobile devices and cloud storage

Component 1: Monitoring Mental Health
Architecture of Sensus: High-Level Design

- High-level design of Sensus
  - Cloud storage
    - Amazon AWS S3
Architecture of Sensus: High-Level Design

● High-level design of Sensus
  ● Cloud storage
    ○ Amazon AWS S3
  ● Users
    ○ Researchers (study designers)
    ○ Participants

Component 1: Monitoring Mental Health
Architecture of Sensus: High-Level Design

- High-level design of Sensus
  - Cloud storage
    - Amazon AWS S3
  - Users
    - Researchers (study designers)
    - Participants

Component 1: Monitoring Mental Health
Architecture of Sensus: High-Level Design

- High-level design of **Sensus**
  - Cloud storage
    - Amazon AWS S3
  - Users
    - Researchers (study designers)
    - Participants
  - Protocols
    - Sensing plans
      - Probes
      - Surveys
      - Customized scheduling
    - JSON file

Component 1: Monitoring Mental Health
Sensus: An Example Case

● A Sensus protocol example (iOS)

Component 1: Monitoring Mental Health
Sensus: Metrics

- **Sensus** can be used in real-world scalable human-subjects studies
  - Release **Sensus**
  - Conduct real-world studies using **Sensus**
- **Sensus** is easy for researchers without engineering background to use
  - Interview researchers who used **Sensus** but without engineering backgrounds

Component 1: Monitoring Mental Health
**Sensus: Preliminary Results**

- Apple App Store
- Google Play Store: **500+**
- > **200** subjects in research studies

*Sensus development website: [https://predictive-technology-laboratory.github.io/sensus/index.html](https://predictive-technology-laboratory.github.io/sensus/index.html)*


Component 1: Monitoring Mental Health
Sensus: Preliminary Results

- Apple App Store
- Google Play Store: 500+
- > 200 subjects in research studies
- Feedback from the Psychologists (2 studies)
  - Easy to use, intuitive experience


**Sensus: Preliminary Results**

- Apple App Store
- Google Play Store: 500+
- > 200 subjects in research studies
- Feedback from the Psychologists (2 studies)
  - Easy to use, experience is intuitive
  - Does not require extra engineering knowledge as long as you know how to use a smartphone

*Sensus development website: [https://predictive-technology-laboratory.github.io/sensus/index.html](https://predictive-technology-laboratory.github.io/sensus/index.html)*

Sensus: Preliminary Results

- Apple App Store
- Google Play Store: 500+
- >200 subjects in research studies
- Feedback from the Psychologists (2 studies)
  - Easy to use, intuitive experience
  - Does not require extra engineering knowledge as long as you know how to use a smartphone
  - Able to get the data they want and obtain meaningful results

Sensus: Preliminary Results

- Apple App Store
- Google Play Store: 500+
- > 200 subjects in research studies
- Feedback from the Psychologists (2 studies)
  - Easy to use, intuitive experience
  - Does not require extra engineering knowledge as long as you know how to use a smartphone
  - Able to get the data they want and obtain meaningful results
  - A desktop or web-based protocol design tool would be useful

Monitoring Mental Health Using Mobile Crowdsensing

- Recall: Can we monitor humans’ mental health status objectively via their everyday behaviors in a natural setting?
- We already have an MCS mobile application: Sensus
Monitoring Mental Health Using Mobile Crowdsensing

- **Sensus**: Cross-platform, general MCS mobile application for human-subjects studies
- **A MCS-based framework**: understanding the relationship between human behaviors and mental health status
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Fine-grained human behaviors vs. Mental health status
  - Objective measures from *Sensus*
    - **GPS**: mobility patterns with semantics
    - **Accelerometer (3-axis)**: micro-level motions
    - **Smartphone metadata**: call and text logs

Component 1: Monitoring Mental Health
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Fine-grained human behaviors vs. Mental health status
  - Objective measures from **Sensus**
    - **GPS**: mobility patterns with semantics
    - **Accelerometer (3-axis)**: micro-level motions
    - **Smartphone metadata**: call and text logs
  - **Social anxiety levels**: SIAS score (0-80)
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Semantics of locations
  - (42.2930177, -83.718566) => School
  - Point of Interest (POI) information obtained from Foursquare
  - Clustering spatially and temporally
  - Categories of location semantics

Component 1: Monitoring Mental Health
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Semantics of locations
- Micro-level behaviors (behavioral dynamics)
  - Linear dynamic system (LDS)

Motion stimuli caused by social anxiety

Observer system

Control System

Component 1: Monitoring Mental Health
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Semantics of locations
- Micro-level behaviors (behavioral dynamics)
  - Linear dynamic system (LDS)

Smartphone Accelerometer Data

Motion Stimuli

System State

Component 1: Monitoring Mental Health

\[
\begin{align*}
\| u(\bullet) \|_0 & \leq k \\
|u(t)| & \leq 1 \quad \forall t \\
\Sigma_t \| CA^t B \|_1 & \leq \mu \\
x(t + 1) &= Ax(t) + Bu(t) \\
y(t) &= Cx(t) + N
\end{align*}
\]
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Semantics of locations
- Micro-level behaviors (behavioral dynamics)
  - Linear dynamic system (LDS)

Component 1: Monitoring Mental Health

\[
\begin{aligned}
\| u(\bullet) \|_0 &\leq k \\
|u(t)| &\leq 1 \quad \forall t \\
\Sigma_t \| CA^t B \|_1 &\leq \mu \\
x(t + 1) &= Ax(t) + Bu(t) \\
y(t) &= Cx(t) + N
\end{aligned}
\]
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Semantics of locations
- Micro-level behaviors (behavioral dynamics)
  - Linear dynamic system (LDS)

Mathematical formulation:

\[
\sum_t \| CA^t B \|_1 \leq \mu \\
x(t + 1) = Ax(t) + Bu(t) \\
y(t) = Cx(t) + N
\]

\[
\| u(\cdot) \|_{L^\infty} \leq k \\
|u(t)| \leq 1 \quad \forall t
\]
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Semantics of locations
- Micro-level behaviors (behavioral dynamics)
  - Linear dynamic system (LDS)

Component 1: Monitoring Mental Health
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- The architecture of the MCS-based framework
Sliding Window

X-Magnitude
Y-Magnitude
Z-Magnitude
Call
Text

03/14 16:30
03/15 04:30
03/15 16:30
03/16 04:30
03/16 16:30

03/14 16:30
03/15 04:30
03/15 16:30
03/16 04:30
03/16 16:30

CALL_1
CALL_2
CALL_m
TEXT_1
TEXT_2
TEXT_n
WINDOW_1
WINDOW_2
WINDOW_P

Dimension Reduction Through Linear Dynamic Model

Histogram of System Stimulus, u(t)

Distance Matrix (DM)
Sliding Window

Dimension Reduction Through Linear Dynamic Model

Histogram of System Stimulus, \( u(t) \)

Distance Matrix (DM)
A MCS-based Framework: Understanding Behaviors and Mental Health Status

- Feature extraction

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call_Proportion</td>
<td>The proportions of phone calls at different locations</td>
</tr>
<tr>
<td>Text_Proportion</td>
<td>The proportions of text messages at different locations</td>
</tr>
<tr>
<td>FAC₁</td>
<td>The average of the mean values of all distance matrices (DM(i)) belonging to a subject</td>
</tr>
<tr>
<td>FAC₂</td>
<td>The average of the standard deviations of all distance matrices (DM(i)) belonging to a subject</td>
</tr>
<tr>
<td>MC</td>
<td>The metric for a phone call event</td>
</tr>
<tr>
<td>MT</td>
<td>The metric for a text message event</td>
</tr>
</tbody>
</table>

Component 1: Monitoring Mental Health
A MCS-based Framework: Metrics

- In real-world human-subjects studies, we can objectively measure humans’ behaviors in a natural setting.
- From the objectively collected data, we can extract meaningful features.
- We can find features that have a significant correlation with mental health status ($p<0.05$).
A MCS-based Framework: Preliminary Results

- Human study of 52 participants
  - *Sensus*
  - Duration: 14 days
  - SIAS: mean = 35.02, std = 12.10

- Correlations between behavioral dynamics and social anxiety levels under different social contexts

<table>
<thead>
<tr>
<th>Matrix feature</th>
<th>Call (MC)</th>
<th>Text (MT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson r</td>
<td>p-value</td>
</tr>
<tr>
<td>FAC₁</td>
<td>0.2867</td>
<td>0.0457</td>
</tr>
<tr>
<td>FAC₂</td>
<td>0.3041</td>
<td>0.0336</td>
</tr>
</tbody>
</table>

Component 1: Monitoring Mental Health
A MCS-based Framework: Preliminary Results

- Correlations between behavioral dynamics and social anxiety levels under different social contexts

<table>
<thead>
<tr>
<th>Location</th>
<th>Call Proportion</th>
<th>Text Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson r</td>
<td>p-value</td>
</tr>
<tr>
<td>Work</td>
<td>-0.1806</td>
<td>0.2142</td>
</tr>
<tr>
<td>Home</td>
<td>0.3983</td>
<td>0.0045</td>
</tr>
<tr>
<td>Food &amp; leisure</td>
<td>-0.2342</td>
<td>0.1053</td>
</tr>
<tr>
<td>Personal life</td>
<td>0.1234</td>
<td>0.3982</td>
</tr>
<tr>
<td>Transition</td>
<td>-0.0715</td>
<td>0.6141</td>
</tr>
</tbody>
</table>

Recall: Can we monitor humans’ mental health status objectively via their everyday behaviors in a natural setting?

Yes, we can.
Proposal Overview: Four Components

- Monitoring mental health using mobile crowdsensing
- Understanding the neural representations of data structures
- Comparing prose writing and code writing
- Understanding bias in code reviews
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?
Understanding the Neural Representations of Data Structure Manipulations

- 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
Understanding the Neural Representations of Data Structure Manipulations

- **fMRI vs. fNIRS**
  - Measure brain activities by calculating the blood-oxygen level dependent (BOLD) signal
- **Functional Magnetic Resonance Imaging**
  - Magnets
  - Strong penetration power
  - Lying down in a magnetic tube: cannot move
- **Functional Near-Infrared Spectroscopy**
  - Light
  - Weak penetration power
  - Wearing a specially-designed cap: more freedom of movement

Component 2: Neural Representations of Data Structures
Understanding the Neural Representations of Data Structure Manipulations

- Experimental design: 2 tasks
  - Data structure manipulations
    - List/Array operations
    - Tree operations
  - Mental rotations: 3D objects
Understanding the Neural Representations of Data Structure Manipulations

- **Experimental design: 2 tasks**
  - Data structure manipulations
    - List/Array operations
    - Tree operations
  - Mental rotations: 3D objects

![Image of data structure manipulations](image_url)

**Component 2: Neural Representations of Data Structures**

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

<table>
<thead>
<tr>
<th>Indices</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>nums</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

A. 1  B. 2

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

![Image of object rotation](image_url)

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

```
  5
 / \
3   8
```

A. 5, 3, 8  B. 8, 3, 5
Understanding the Neural Representations of Data Structure Manipulations

- Data analysis: we need to be careful
- Spurious correlations due to multiple comparison

Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction
Craig M. Bennett¹, Abigail A. Baird², Michael B. Miller³, and George L. Wulford³

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 100,000 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...
Understand the Neural Representations of Data Structure Manipulations

- Data analysis: we need to be careful
- fMRI and fNIRS use the same high-level 3-step analysis approach
  - False discovery rate correction for multiple comparisons (FDR)
Understanding the Neural Representations of Data Structure Manipulations

- Data analysis: we need to be careful
- fMRI and fNIRS use the same high-level 3-step analysis approach
  - False discovery rate correction for multiple comparisons (FDR)

Preprocessing
Understanding the Neural Representations of Data Structure Manipulations

- Data analysis: we need to be careful
- fMRI and fNIRS use the same high-level 3-step analysis approach
  - False discovery rate correction for multiple comparisons (FDR)

Preprocessing → First-level Analysis

Component 2: Neural Representations of Data Structures
Data analysis: we need to be careful
fMRI and fNIRS use the same high-level 3-step analysis approach
  False discovery rate correction for multiple comparisons (FDR)
Following the best practices in medical imaging, we can find significant relationship between data structure manipulations and spatial ability ($p<0.01$).

We can find significant relationships regarding the difficulty levels of tasks.
Neural Representations of Data Structures: Preliminary Results

- **Experiment setup and data**
  - 76 participants: 70 valid
    - fMRI: 30
    - fNIRS: 40
    - Two hours for each participant: 90 stimuli, qualitative post-survey

De-identified data is public: [https://web.eecs.umich.edu/weimerw/fmri.html](https://web.eecs.umich.edu/weimerw/fmri.html)
Data structure manipulations involve spatial ability

- **fMRI**: more similarities than differences ($p<0.01$)
- **fNIRS**: activation in the same brain regions ($p<0.01$)
Neural Representations of Data Structures: Preliminary Results

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

### Rotation Angle = 20°

![Diagram showing a rotation angle of 20°]

**Indices**

<table>
<thead>
<tr>
<th>N= 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nums</strong></td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

**Component 2: Neural Representations of Data Structures**
The brain works even **harder** for more difficult data structure tasks

- **Difficulty measurement**
  - Mental rotations: angle of rotation
  - Data structure: 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
Neural Representations of Data Structures: Preliminary 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!

Component 2: Neural Representations of Data Structures
Understanding the Neural Representations of Data Structure Manipulations

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

Data structure manipulations and mental rotations (spatial ability) involve very similar brain regions.
Proposal Overview: Four Components

- Monitoring mental health using mobile crowdsensing
- Understanding the neural representations of data structures
- Comparing prose writing and code writing
- Understanding bias in code reviews
Comparing Code Writing and Prose Writing

- 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 Code Writing and Prose Writing

- fMRI: penetration power
- Challenges
  - fMRI-safe bespoke keyboard
    - QWERTY keyboard
    - Allow typing and editing
- Design writing stimuli
  - Prose writing
  - Code writing
Comparing Code Writing and Prose Writing

- fMRI: penetration power
- Challenge: fMRI-safe bespoke keyboard
  - QWERTY keyboard
  - Allow typing and editing
Comparing Code Writing and Prose Writing

● Challenge: Stimuli design
  ● Two categories of tasks for code writing and prose writing
  ● Fill in the blank (FITB)

Prose - FITB

```plaintext
/*Complete the sentence
* such that the sentence
* makes sense*/
Brian was so fond of his dog that their brief ______ left him not just saddened, but in a state of sorrow.
```

Code - FITB

```c
/*Complete the definition of
* the function such that it
* receives an integer parameter
* and returns the absolute
* value of the parameter.*/
int absoluteValue(int num1)
{
    /* YOUR CODE HERE */
    return absoluteValue;
}
```
Comparing Code Writing and Prose Writing

- **Challenge: Stimuli design**
  - Two categories of tasks for **code writing** and **prose writing**
  - Fill in the blank (FITB)
  - Long response (LR)

**Component 3: Comparing Code Writing and Prose Writing**

**Prose - LR**

What would happen if everyone lived in space? (e.g., What type of houses would they live in? What type of clothing would they wear?)

**Code - LR**

Implement a function `is_sorted` that accepts a vector of integer values and returns true if it is non-decreasing, and false otherwise.
Comparing Code Writing and Prose Writing

- Experimental design: 2 categories of tasks for code writing and prose writing
  - Code writing tasks: Turing’s Craft
  - Prose writing tasks: SAT
We can have a bespoke QWERTY keyboard that can safely work in fMRI machine.

We can find a significant relationship between code writing and prose writing ($p<0.01$).

- General relationship
- Relationship between different types of tasks (i.e., FITB and LR)
Code Writing vs. Prose Writing: Preliminary Results

- IRB approved
- Bespoke keyboard
  - Finished deployment and passed safety tests
- Data collection is done
  - 30 participants
    - Two hours for each participant: 52 stimuli
    - For both code writing and prose writing:
      - FITB: 17
      - LR: 9

Component 3: Comparing Code Writing and Prose Writing
Proposal Overview: Four Components

- Monitoring mental health using mobile crowdsensing
- Understanding the neural representations of data structures
- Comparing prose writing and code writing
- Understanding bias in code reviews
Understanding Bias in Code Reviews

- Code reviews
  - The systematic inspection, analysis, evaluation, and revision of code.
  - The latent defect discovery rate of formal code review can be 60%-65%.

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

```cpp
#include <vector>

int majorityElement(vector<int>& num) {
    int cnt = 0;
    int majority = 0;
    for (int i = 0; i < num.size(); i++) {
        if (num[i] == majority)
            cnt++;
        else if (cnt == 0)
            majority = num[i], cnt = 1;
        else
            cnt--;
    }
    return majority;
}
```

Component 4: Bias in Code Reviews
Understanding Bias in Code Reviews

● Code reviews
  ● The systematic inspection, analysis, evaluation, and revision of code.
  ● The latent defect discovery rate of formal code review can be 60%-65%.
Understanding Bias in Code Reviews

- Code reviews
  - The systematic inspection, analysis, evaluation, and revision of code.
  - The latent defect discovery rate of formal code review can be 60%-65%.

- Bias in code reviews
  - Code source
    - Gender
Understanding Bias in Code Reviews

- **Code reviews**
  - The systematic inspection, analysis, evaluation, and revision of code.
  - The latent defect discovery rate of formal code review can be 60%-65%.

- **Bias in code reviews**
  - Code source
    - Gender
    - Automated software repair tools
Understanding Bias in Code Reviews

- How does author information affect software developers’ decision making in code reviews?
- Do software developers have gender bias in code reviews?
- Do software developers have bias against machine-generated code patches?
Understanding Bias in Code Reviews

- **Neural activities in code reviews: fMRI**
- **Visual focus in code reviews: eye tracking**
  - Fixations and saccades
  - Attention over different Area of Interests (AOI)
    - Comment
    - Code changes
    - Author information

```
import java.util.Scanner;
public class eyeTrack2 {
    public static void main(String[] args) {
        Scanner in = new Scanner(System.in);
        System.out.print("First number: ");
        int num1 = in.nextInt();
        System.out.print("Second number: ");
        int num2 = in.nextInt();
        int average = (num1 + num2) / 2;
        System.out.println(average);
    }
}
```
Understanding Bias in Code Reviews

- **Stimuli design**
  - Pull requests from real world open source C and C++ projects (e.g., GitHub)
  - **Relabel** the author information
    - Pictures from **Chicago Face Database**
      - Controlling age, race, attractiveness and facial expressions
    - Avatar picture to represent automated software repair tools

![CFD Chicago Face Database](image)
Understanding Bias in Code Reviews

- **Stimuli design**
  - Pull requests from real world open source projects (C and C++) (e.g., GitHub)
  - **Relabel** the author information
    - Pictures from Chicago Face Database
      - Controlling age, race, attractiveness and facial expressions
      - Avatar picture to represent automated software repair tools
  - We will not tell the participants about the relabeling and the purpose of investigating the author bias in code reviews.
    - Avoid social desirability bias
Understanding Bias in Code Reviews

- Stimuli design
  - Simulating a real-world code review interface

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

Commit message

```cpp
#include <algorithm>

int majorityElement(vector<int> &num) {
    int cnt = 0;
    int majority = 0;

    for (int v : num) {
        if (cnt == 0) {
            majority = v;
            cnt = 1;
        } else if (v == majority) {
            cnt++;
        } else {
            cnt--;
        }
    }

    return majority;
}
```
Understanding Bias in Code Reviews

- Stimuli design
  - Simulating a real-world code review interface

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

Code changes

Component 4: Bias in Code Reviews
Understanding Bias in Code Reviews

- Stimuli design
  - Simulating a real-world code review interface

Component 4: Bias in Code Reviews
Bias in Code Reviews: Metrics

- We are able to involve author deception in the stimuli design (IRB permission)
- We are able to recruit approximately gender-balanced group of participants
- We are able to obtain significant relationship between the brain activities of code reviews with different author information ($p<0.01$)
- We are able to observe significant similarities or differences of the visual focus and strategies for code reviews with different author information ($p<0.01$)
Bias in Code Reviews: Preliminary Results

● **Stimuli design is done**
  ● Two sets of stimuli: 60 stimuli each
    ○ Randomly assign author pictures into three groups
      ● 20 men
      ● 20 women
      ● 20 machine
    ○ Relabel each set with different code-author combinations
      ● Control code quality

● **IRB approved**
● The fMRI lab has a **built-in eye tracker**
● fMRI lab pilot grant to support this study
Ph.D. Timeline

- SENSUS Mobile APP [UbiComp’16]
- Monitoring Social Anxiety [UbiComp’16]
- Clinical Models Using Mobile Sensing [J. MIR’17]
- Behavioral Dynamics [J. Information Fusion’19]
- Spatial Ability in Programming [ICSE’19]
- Code Writing vs. Prose Writing
- Bias in Code Reviews
- Capstone journal article

Ph.D. Timeline
Publications: Supporting this Proposal

1. **Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS.** Yu Huang, Xinyu Liu, Ryan Krueger, Tyler Santander, Xiaosu Hu, Kevin Leach, Westley Weimer. *41st ACM/IEEE International Conference on Software Engineering (ICSE 2019).* 
   *Distinguished Paper Award*


3. **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. *The 1st Workshop on Digital Biomarkers, DigitalBiomarkers ’17, pages 9–14, New York, NY, USA, ACM.*


15. **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.*

Broader Impact

- All the medical imaging and behavioral data will be de-identified and released publicly
- **Sensus** has been released and can be used in a wide range of human-subject studies
- Our research findings can help psychologists monitor mental health status and help computer science educators develop efficient training strategies
- Our studies provide guidelines for future study design and implementation in the community
Proposal Summary: Four Components

- Monitoring mental health using mobile crowdsensing
  - Sensus: Cross-platform, general MCS mobile application for human-subject studies
- Understanding human behaviors and mental health status via MCS
- Understanding the neural representation of data structures
- Comparing prose writing and code writing
- Understanding bias in code reviews