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

Ph.D. Dissertation Proposal
Yu Huang (yhhy@umich.edu)

Dec 11, 2019

Overview. We propose using objective measures to provide a systematic solution to understand user cognition in programming activities. We focus on understanding the mental status, role of spatial ability, fundamental processes and stereotypical associations in certain software engineering activities by combining mobile crowdsensing, medical imaging, and eye tracking.

Intellectual Merit. This thesis addresses how we can meaningfully and objectively measure cognitive processes in software engineering. We propose to use mobile crowdsensing, medical imaging and eye tracking to answer questions such as: How can we monitor humans' mental status from their natural daily work and life settings? Does the brain treat manipulating data structures like rotating 3D objects? How do difficulty levels of programming tasks affect cognitive load? Is creative coding like creative writing? How do cognitive biases affect the code review process? We improve on previous approaches that used subjective self-reporting, studied activity sub-steps, or restricted attention to homogeneous groups by proposing a solution that uses **objective measures**, provides a **fundamental understanding**, studies **higher-level tasks**, and **generalizes across users** and populations. We thus propose four research efforts:

- **Monitoring Mental Status in Natural Settings.** We propose to design a cross-platform mobile application, Sensus, and a mental health monitoring framework to examine the link between users' behavior and mental health status under different social contexts.
- **Understanding the Relationship between Data Structure Manipulation and Spatial Ability.** We propose to use two medical imaging modalities to compare the brain activities of spatial ability and data structure tasks. We will also compare the trade offs between different medical imaging modalities.
- **Comparing Code Writing and Prose Writing.** We propose to use medical imaging techniques to understand the cognitive processes underlying code writing and to use prose writing as a baseline to ground the results. We will design a bespoke magnet-safe keyboard to allow typing and editing in medical imaging devices.
- **Investigating Bias in Code Reviews.** We propose to combine medical imaging and eye tracking techniques to investigate the effects of stereotypical associations (bias) in code reviews. We will focus on the effect of author identity. We will use code reviews (pull requests) from open source projects with a controlled relabeling author information.

Broader Impact. The potential broader impact of this proposal occurs at different time frames. **Short term.** All the medical imaging and behavioral data will be de-identified and released publicly for replication research. Our mobile application can be used in a wide range of human subject studies. Our findings from the proposed research components can help monitor the mental health status (thus help professionals intervene with potential severe mental issues) and help computer science educators develop more efficient strategies. **Medium term.** PI Huang is involved in the proposal writing, organization, and data collection for two successfully-funded *Rackham Faculty Allies and Student Ally Diversity Grants*. This activity brings in external speakers for student-focused colloquia and meetings to educate students about underrepresented career goals and research options. **Long term.** Human studies with objective measures based on mobile crowdsensing can enrich the research options for researchers with interest in human behaviors in natural settings. While medical imaging research in computer science has the potential to augment unreliable self-reporting, guide pedagogy and technology transfer, and improve productivity, it is still a new topic in computer science research. Our studies can provide guidelines for future study design and implementation in the community.

1 Introduction

A fundamental understanding of computational activities is essential to improving productivity and efficiency in software engineering. With software-related innovations driving a $3.8 trillion global IT market [35, 36] and demand for university computer science courses outstripping the supply of professors [118], the value of developing and maintaining software is increasing rapidly. This importance is already emphasized in the software industry, with Fortune 500 companies, such as Amazon and AT&T, committing massive resources to retrain up to half of their workforces to obtain better productivity and efficiency in programming-intensive areas [19, 38]. Despite this increasing prevalence of software and demand for skilled programmers, we rely on traditional survey instruments rather than an understanding of fundamental cognitive processes in computational activities, when developing methods to support and improve productivity and efficiency. Research has been done in both Psychology [15, 90] and Software Engineering [3, 9, 116] to study the thought processes behind different tasks. However, most of these studies either focused on particular low-level tasks (e.g., identifying syntax errors or measuring blink rates [9, 116]) or relied on observational evidence and self-reporting (e.g., [3, 15, 90]).

While findings relating to cognitive processes have successfully transitioned to guiding behavioral and developmental improvement in domains like mathematics [40] and education [95], we still lack a fundamental understanding of the cognition behind software activities. Such an understanding would help programmers, academics, and industry participants: from understanding productivity and expertise [6, 80, 110] to increasing participation in the modern workforce [18, 86] to guiding pedagogy [8, 109] to augmenting unreliable self-reporting [77, 99]. To obtain such an understanding, we desire a systematic solution that satisfies the following criteria:

- **Objective Measures.** It is critical to measure the relevant factors objectively in computer science tasks. Research in both Psychology and Computer Science has shown that subjective measures may not be adequately trustworthy (e.g., [41, 55, 65, 108]). Objective measures are necessary to understand user cognition in a generalizable and reliable way.
- **Foundational Understanding.** A fundamental understanding of the cognitive processes in programming tasks is essential to help users solve software problems, learn programming, and make decisions in software development, as well as to further improve productivity and efficiency [72, 75].
- **Higher-Level Tasks.** We desire an understanding of higher-level programming tasks. Software development is a complicated process consisting of different components. In the software industry, development usually includes higher-level tasks such as code reviews [7, 50]. The efficiency of such higher-level tasks directly and significantly affects their time and monetary cost [32, 69]. Thus, it is important to understand the cognitive processes for these semantically-rich and industry-related activities.
- **Generalizability Across Users.** We desire an understanding that applies to a wide range of people. Modern software development is conducted in an environment of diverse populations and diversity is important to effectiveness in software engineering [22, 96]. The diversity of the CS workforce has been receiving more attention over time [63, 104]. Thus, we desire a theory that accounts for demographic differences (e.g., gender) across users.

In Psychology, researchers have studied cognitive processes of different tasks, ranging from musical performance [74], to food cravings [94] to prose writing [13, 85, 97]. However, these studies rarely involve computer science tasks. There is also a significant body of work investigating the psychology of programming, ranging from the cognitive prerequisites of programming [90], to programming behaviors [3, 9] to entire theories of the coding process [15]. Unfortunately, these studies usually rely on observational behavior or self-reporting data which may lack foundational evidence. To address the concerns associated with self-reporting, researchers in software engineering have turned to medical imaging to obtain neurological evidence regarding programming activities. Research using medical imaging techniques has examined brain patterns in code comprehension, code readability and bug detection [23, 44, 51, 116]. However, these studies fail to provide an understanding of cognitive processes for higher-level and more industry-related activities, as well as the differences between diversified groups of users. Some studies in software engineering explored individual differences in outcomes (e.g., associated with gender) [53, 68, 121], but they do not illuminate how user demographics actually affect decision making in programming tasks.

We combine several insights to form the basis of a systematic study for understanding user cognition in computer science. **First**, with the increasing emergence of *mobile crowdsensing (MCS)*, *medical imaging* and *eye tracking* (e.g., smartphoness, fMRI, and eye trackers), it is now possible to conduct studies that acquire objective data in a natural setting to provide an understanding of the underlying cognitive processes of certain tasks. For instance, researchers can collect a rich set of data through the embedded sensors in smartphones. Similarly, modern medical imaging techniques allow researchers to investigate the neurological patterns in human brains during different tasks. **Second**, we can adapt *scientific approaches and concepts from other domains* to assist our investigation and understanding for computer science activities. For example, in Psychology, fruitful studies have been done on mental health problems (e.g., social anxiety or depression) that are highly associated with impairment in academic functioning and relationships [89, 131]. These findings shed light on how we can improve productivity and efficiency in computer science. In another example, education researchers have studied the influence of spatial ability and shown that it is a major factor in proficiencies such as mathematics [62, 127] and the natural sciences [122, 133]. They have also designed corresponding interventions to enhance spatial ability [64, 105, 119, 120]. These results inspire us to study the relationship between spatial ability and programming to help with programmers' productivity. **Third**, it is now possible to study historically-subjective factors like cognition by designing *rigorous controlled experiments*. For instance, contrast-based experiments are widely used in medical imaging studies to investigate the brain. Such designs make it possible to focus on brain activities that are only relevant to the actual experimental conditions. These three insights support our systematic solution that satisfies all four of the desired properties we require.

To understand cognition in software engineering activities, we propose to study certain relevant and important behaviors of programmers. Consider a standard workday of a traditional software engineer. She is working on a software project as part of a team. Her mental status, modulated by social interactions with her co-workers, pressure from the work, etc., affects her functionality and productivity [2, 25, 26, 130]. As a software engineer, she must be able to understand the software she is working with [126] and produce high-quality code [81]. She also needs to conduct code reviews to assess her colleagues' patches and decide if they can be merged into the project's code base [50]. We propose a systematic study consisting of four research components that combines the three insights above to investigate these four critical activities in software engineering.

First, we propose to study the mental status of users from their natural daily work and life settings using mobile crowdsensing techniques. We leverage the ubiquity of smartphones to measure and monitor the mental well-being of end users via a specially-designed mobile application. Modern smartphones have a rich set of sensors and logs including GPS, accelerometers, text messages, phone calls and so on, which provide a window into the relationship between behaviors and mental health. We propose designing a mobile application and a mental health monitoring framework to examine the link between users' behavior and mental health status under different social contexts. We choose to focus on social anxiety because it is one of the most common mental health problems [1] and it negatively affects performance due to impaired social functioning [25, 130].

Our second proposed component is to investigate the relationship between data structure manipulation and spatial ability using medical imaging techniques. Studies in Psychology have displayed the importance of, and possible interventions for, enhancement for spatial ability in several domains [14, 64, 119, 127], but it is rarely studied within software development. We hypothesize that spatial ability is highly associated with software tasks on a foundational level. If so, we can leverage training experience for spatial ability in Psychology to improve pedagogy and productivity in computer science. We propose to use medical imaging to compare the brain activities of spatial ability and data structure tasks because (1) data structures are a fundamental element in computer science that affect the performance and cost of many systems and (2) medical imaging techniques can provide us with a neurological basis for the relationship between two tasks. We propose to statistically validate this relationship on humans using two medical imaging modalities.

Our third proposed component investigates the relationship between code writing and prose writing using medical imaging. The goal of this study is to understand the cognitive process of code writing, a crucial activity in software engineering. We propose to use prose writing as a baseline to ground our results. While some studies have explored how software developers read code (e.g., [52,116]), there is no research studying the cognitive processes of creativity in programming such as code writing. One challenge is that normal keyboards cannot be safely placed, or accurately read, near magnetic resonance scanners. We propose to design a bespoke magnet-safe keyboard to allow typing and editing. We will test the brain activity relationship between code writing and prose writing using statistical tests on human participants.

Our fourth proposed component is to study the decision making process in code reviews using medical imaging and eye tracking. The goal of this study is to investigate the effects of stereotypical associations (i.e., bias) in software engineering. There is research showing that software developers do not recognize this potential bias when checking the source of code in code reviews [53] and female developers have lower acceptance rates while their identities are directly recognizable in open source projects [121]. We propose to use open source projects and controlled author information to study the cognitive processes in code reviews. In addition, we will also monitor subjects' eye movement using eye tracking devices. We will compare the brain activities and eye motions across conditions using statistical tests.

The overall thesis statement of the proposed dissertation is:

> *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, medical imaging, and eye tracking.*

1.1 Contribution

The expected main contributions of the proposed dissertation are the following:

- A deployed research mobile application for iOS and Android supporting real-world human-subject studies using mobile crowdsensing.
- A framework for integrating and analyzing users' mobility patterns, micro-behaviors and communication patterns to monitor and understand mental health status.
- A mathematical model of the particular relationship between data structure manipulation and spatial ability based on fundamental medical imaging results.
- A mathematical model of the particular cognitive processes involved in the higher-level software engineering activities of code writing and code review.
- A comparison of cost-benefit and feasibility tradeoffs between different medical imaging techniques for software engineering.

The remainder of this proposal is as follows. In Section 2, we discuss background material and related research efforts. Next, we describe our proposed research and associated technical approaches in Section 3, and in Section 4, we detail our evaluation plan. We present our proposed schedule of research in Section 5. Finally, in Section 6 we present results of preliminary experiments then finish with a discussion of the broader impact in Section 7, and concluding thoughts in Section 8.

2 Background and Related Work

In this section, we discuss the background and related work for our proposed research. First, we introduce studies using mobile crowdsensing to monitor human behavior and mental health status. Second, we discuss medical imaging techniques and the related work in software engineering to help researchers understand the cognitive processes behind programming tasks. Third, we introduce spatial ability, code writing, and code reviews, with a particular focus on stereotypical associations in code reviews.

2.1 Mobile Crowdsensing and Mental Health

Mobile crowdsensing (MCS) refers to a broad range of community sensing paradigms, where individuals with sensing and computing devices collectively share data and extract information to measure and map phenomena of common interest [56]. Human-centric, participatory sensing [20] and mobile crowdsensing [56] have facilitated participant recruitment and data collection for a range of human-subject studies.

Eagle and Pentland *et al.* [46–48] were among the first to use MCS to study human subjects. Pentland's group extended this work to personality [39] and social ties [42], and others have used a similar approach to study mental and physical health relationships [101], connections between mental health and academic performance [128, 129], and human mobility and depression [21]. These and other studies indicate that MCS is a feasible paradigm for recruitment and data collection in human-subject research [45, 102].

Existing MCS systems are predominantly designed for a single purpose, target only one of the major mobile platforms (Android or iOS), and require knowledge of markup and programming languages. Thus, we propose a cross-platform, general-purpose system for MCS-based human-subject studies. The goal of our platform is to lower the barrier to entry into MCS-based research and increase the likelihood that MCS will be applied in a diverse set of fields.

The emergence of MCS research via smartphones has created new opportunities to understand how mental health disorders such as anxiety and depression manifest in real-world settings. Smartphone usage data and sensor data streams (e.g., accelerometers, microphones, and GPS) provide continuous, unobtrusive measurements of variables critical to mental health status (e.g., social interactions, location semantics, and physical activity) [28–30, 66, 78, 79, 128, 135]. The ultimate goal for MCS-based mental health studies is to deliver effective interventions in real time. Though lacking a fine-grained data analysis to support real-time intervention delivery, patient-centered [124] and context-aware mobile sensing systems have the potential to deliver personalized healthcare services [27, 136] at the right time, in the right place, and in the right manner [134]. Researchers have used smartphone sensor data via mobile sensing systems to monitor college students' mental health and educational outcomes [128], detect relevant problematic smartphone usage patterns of users with mental disorders [115], and measure characteristic changes in symptomatic individuals to predict their health status [78, 79].

Unlike the previous research, we propose to provide a more general application for conducting human-subject studies with MCS. We also propose to focus on a more fine-grained analysis of sensor data to understand the behaviors associated with mental health with the goal of delivering effective interventions in real time.

2.2 Medical Imaging in Software Engineering

Medical imaging techniques are used to study brain activity. Over the past 30 years, non-invasive *in vivo* functional neuroimaging techniques have emerged as important tools in understanding cognitive processes.

The most popular technique is *functional magnetic resonance imaging* (fMRI). It is a non-invasive tool that poses significantly less risk and can access a wider range of brain regions than other techniques (e.g., electrocorticography). fMRI relies on the *hemodynamic response*, the metabolic changes in neuronal blood flow to active brain regions, using oxygen consumption as an indirect measurement for brain activity [17].

Siegmund *et al.* introduced the study of software engineering tasks with fMRI, focusing on code comprehension [116]. Their subsequent work presented more potential analyses of fMRI data involving software programming, finding evidence that data-flow-based code complexity metrics (but not control-flow-based metrics) rest on valid assumptions [92]. Floyd *et al.* used fMRI to study code comprehension, code review, and prose review [52]. Newer work has looked at bug detection [23, 44], code comprehension [93, 117] and logic problem solving [112].

We propose to use medical imaging to investigate the cognitive process of high level software engineering tasks, such as code writing and code reviews, as well as exploring the relationship between data structure manipulation and spatial ability.

2.3 Spatial Ability

There are many interpretations of *spatial ability*, including the determination of spatial relationships between objects and the mental manipulation of spatially-presented information. Despite spatial ability's influence in many disciplines [4,14,62,87,122,127,133], it has rarely been studied in software engineering. To the best of our knowledge, only one previous study [3] focused on the relationship between software engineering tasks and spatial ability [3]. This work used interviews but no quantitative relationship has been investigated. Drawing inspiration from previous work, we consider spatial ability in the context of data structures to be the capacity to mentally represent, remember and manipulate spatial relations between elements of data. We propose to use *mental rotation* [34,43,114] tasks to provide a solid neurological basis for spatial ability against which the cognitive processes associated with data structure manipulation can be compared.

2.4 Code Writing and Cognition

There is strong interest from both academia and industry in improving programmers' ability to write code [54, 71, 81, 83]. While the success of a handful of projects (e.g., Scratch [106]) hints at the opportunity present in understanding the mental processes associated with writing code, modern research is limited by our lack of a foundational, neurobiologically-grounded understanding [65].

Researchers since the 1950's have sought to understand the psychology of programmers, but have largely relied on observational data (e.g., [16]). Pea and Kurland, who conducted research in the 1980's on the cognitive prerequisites and effects of programming, emphasized the need to understand programmers' psychology given the rising importance of computer literacy [90, 91].

Despite the massive growth of software engineering since then, Pea and Kurland's call has not been answered with a grounded neurological understanding. By contrary, there is a significant body of research to study the cognitive processes of prose writing (see Berninger and Winn [13] for a survey). Unlike code, prose writing has been studied using medical imaging to establish a more objective understanding. The findings from these medical imaging prose writing studies, such as the brain regions associated with prose writing and the specialization of such regions [12, 85, 97, 98, 107, 113], have in turn successfully informed subsequent research in pedagogy [13, 70, 95]. Inspired by research on prose writing and other domains using medical imaging, we believe similar benefits for code writing may be available.

To the best of our knowledge, there are no previous imaging studies of code writing in general, nor studies comparing code and prose writing in particular. We propose to investigate the relationship between code writing and prose writing using medical imaging.

2.5 Code Review and Bias

Code review is the systematic inspection, analysis, evaluation, and revision of code. It is one of the most essential and common activities in modern software engineering to reduce the cost of maintenance. Modern code review often arises in the form of *pull requests* associated with version control repositories [7]. In code reviews, software developers read a co-worker's patch and decide if the code changes should be merged into the code base. This practice has shown the latent defect discovery rate of formal code code review to be in the 60%–65% range [69].

However, while researchers have found out that diversity is important to productivity in software development, software developers may have unconsciously introduced bias in code reviews that might hurt the efficiency of code reviews. Researchers analyzed pull requests in open source projects on GitHub, and found out that female contributors have a significantly higher acceptance rate when their gender identities are not directly recognizable. In comparison, when the gender identity is directly recognizable, female developers have a significantly lower acceptance rate [121]. Other studies have also shown that in code reviews, software developers unconsciously check the social information of the patch authors [53]. Though such research findings indicate biases in code reviews, there is no studies that provide any objective evidence that these stereotypical associations affect software developers' decision making. Further more, it is unclear how stereotypical associations affect developers' decisions in code reviews.

In recent years, as automated software repair tools are adapted into industry [82], the issue of trusting patches generated by machines has gathered attention in software development. For example, Ryan *et al.*'s studies indicated software developers find machine-generated code repairs less trustworthy than repairs made by a human [111]. In general, we lack a fundamental understanding of such trust and bias differences and thus propose to study them.

3 Proposed Research

Our goal is to provide systematic studies to objectively and meaningfully measure user cognition to understand the mental status, role of spatial ability, fundamental processes and stereotypical association in software engineering activities. We propose the following four research components including mobile crowdsensing, medical imaging and eye tracking to provide this systematic understanding:

1. To design an MCS-based framework for collecting and analyzing users' mobility patterns, micro-behavior and communication patterns to monitor and understand users' mental status (Section 3.1).
2. To develop a mathematical model of the relationship between data structure manipulation and spatial ability using both fMRI and another modality in a human study (Section 3.2).
3. To develop a mathematical model of the cognitive processes in code writing compared with prose writing using fMRI in a human study (Section 3.3).
4. To develop a mathematical model of the effects of stereotypical associations in code reviews using fMRI and eye tracking in a human study involving deception (Section 3.4).

The remainder of this section describes each research effort and proposed approach at a high level. Subsequently, Section 4 details our proposed metrics and experimental setups.

3.1 Monitoring Mental Health Using Mobile Crowdsensing

The goal of this research thrust is to provide a systematic and objective way to measure and understand humans' mental health status. We propose two sub-components to achieve this end. First, we propose Sensus, a mobile crowdsensing (MCS) system (see Section 2.1) for human-subject studies that bridges the gap between human-subject researchers and MCS methods by offering rich features for integrating objective measurements and psychological surveys. Second, we propose a framework to combine smartphone sensor data (e.g., collected from Sensus including accelerometer, GPS) and call and text metadata to study the relationship between human behaviors and their mental well-beings. In particular, in this research thrust, we focus on social anxiety, one of the most common mental health problems among college students [1].

3.1.1 Sensus: Cross-Platform, General MCS for Human-Subject Studies

We propose to extend MCS methods to new disciplines, researchers, and study populations in the design of a cross platform mobile application, Sensus. Our objectives are to (1) target heterogeneous mobile infrastructures while maintaining a uniform user experience, (2) support a wide range of MCS-based studies that leverage sensing and interaction capabilities of modern smartphones, (3) eliminate the need for researchers to understand markup and programming languages when developing their MCS studies, and (4) implement an MCS system such that it relies exclusively on readily-available mobile devices and cloud storage.

Probes for most hardware-based sensors (e.g., altitude, GPS coordinates, 3-axis acceleration) will follow the observer pattern and will be activated upon value changes. Probes will have configurable data storage rates and offer continuous (higher power consumption) and periodic (lower power consumption) sensing modes.

We propose to build Sensus on top of the Xamarin platform.[1] This allows us to use a unified programming interface on top of the native iOS and Android OS.

3.1.2 Understanding Human Behaviors and Mental Health Status via MCS

Using Sensus, our proposed MCS system for human-subject studies, we will be able to distribute our own research protocols to study human behaviors with objective measurements. In this research sub-component, we propose to collaborate with psychologists and use objective measures from smartphone sensors via Sensus to examine socially anxious individuals' fine-grained behaviors around periods in which they engage in some form of social interaction, and how these behaviors differ as a function of location (e.g., at home, at work, or at an unfamiliar location).

We propose to passively sense (1) micro-level motion via accelerometer sensors, and (2) physical locations (e.g., home, restaurant) via GPS sensors before, during, and after phone calls and text messages. We hypothesize that subtle user motions covary with social anxiety level and social context (e.g., hands may shake more when in a public setting where the chance of negative evaluation by others is high). We refer to these motions as the user's *behavioral dynamics*. We propose to examine whether information from these sensors reveals systematic differences in movement patterns associated with social anxiety symptoms. In particular, we propose to explore the relationship between the behavioral dynamics of smartphone use and an individual's social anxiety levels. Additionally, we propose a new analysis method for capturing the behavioral dynamics of smartphone use. We model the smartphone as a linear dynamical system (LDS) [49]. We extract features from the LDS and then use them to understand behavioral differences across social anxiety levels, communication media (text messages and phone calls), and semantic locations.

Although such behavioral dynamic models have not been used to study human's mental health status, scholars have developed similar models to predict subjects' physical health status [59, 73, 103]. We propose to leverage these previous efforts to develop this framework for exploring features and identifying individuals with certain social anxiety levels.

3.2 Understanding the Neural Representations of Data Structures

The second component of our proposed systematic understanding of user cognition is to investigate the relationship between data structure manipulations and spatial ability (see Section 2.3). We propose to use mental rotation tasks to provide a solid neurological basis for spatial ability against which the cognitive processes associated with data structure manipulation can be compared.

We propose to design a controlled experiment involving common data structure tasks as well as spatial ability tasks (i.e., mental rotation tasks from Psychology). We will carefully design the tasks to admit controlling for difficulty measures. Psychologists have developed methods to measure the difficulty levels of mental rotation tasks (e.g., rotation angles [114]). We propose to adapt analogous measures in computer science (e.g., Big-Oh notation) in our experiment design to capture the difficulty of programming tasks.

We also propose to use both functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS). Both fMRI and fNIRS are non-invasive *in vivo* neuroimaging techniques that have enabled new and complex studies of brain function (see Section 2.2). However, they exhibit tradeoffs relevant to studying software engineering which have not been studied yet. We propose to provide some guidelines for future experimental designs to the community.

3.3 Comparing Prose Writing and Code Writing

The third proposed research component is to understand the human cognitive processes underlying code writing. We propose to leverage functional brain imaging to investigate neural representations of code

[1] https://www.xamarin.com

writing (see Section 2.4) in comparison to prose writing.

We will address two main challenges: physics and design. First, normal keyboards cannot be safely placed or accurately read near magnetic resonance scanners. They interfere with the fMRI measurements and the fMRI interferes with keyboard reporting. Second, imaging studies require carefully-controlled experiments, and no high-level design for a code writing contrast has previously been proposed. We propose to design a fMRI-compatible keyboard to allow standard typing and editing for software developers. The keyboard will operate by separating its electromagnetic control logic from its physical chassis, connecting them by specially-shielded cables. We also propose a two-by-two contrast setup: code vs. prose writing and fill-in-the-blank vs. long response (informally, single-word production vs. longer creativity). We propose to use code stimuli from Turing's Craft which are widely used in short programming exercises [5] and prose stimuli from Scholastic Assessment Test (SAT) which is standard English test. Our use of prose writing as a baseline grounds our experiment and clarifies our results.

The contrast between code and prose writing in our experiment will illuminate their differences and similarities at a neurological level. We propose to conduct a human study involving these prose and code writing tasks with fMRI and analyze the data to compare the brain activations involved in these tasks.

3.4 Understanding Bias in Code Reviews

Our fourth proposed component aims to explore the neurobiological processes associated with the critical software engineering activity of code review. In particular, we propose a human study to understand the effects of author bias associated with code change judgments (see Section 2.5).

In Psychology, a *bias* is a tendency, inclination, or prejudice toward or against something or someone[2]. Previous research has suggested a cross-gender bias in the open source software community [68, 121, 125]. Moreover, researchers have also observed potential bias associated with source of code (e.g., human or automated software repair tools) [111]. We aim to understand the role of such author information in code reviews. At the same time, we will also consider the influence of other source code features (e.g., how many lines of code are changed in each review).

We propose to use an eye tracker together with medical imaging to observe participants' behaviors in code reviews. For example, we will analyze where and how long participants direct their attentions in code reviews. Such eye tracking data (e.g., fixation) can provide information of humans' attention and cognitive load during tasks [37, 51]. We propose to use real-world pull requests in open source projects (e.g. GitHub) to check if and how much the author information affects developer's decision making in code reviews. To address the issue of controlling the quality of code, we will relabel the author information (i.e., male, female or machine) of the pull requests as a deception factor. We will also use fMRI to collect brain activation data to explore the nature of this bias (e.g., attentional on code judgments, visual on the author picture).

4 Proposed Experiments and Metrics

In this section we discuss the metrics for evaluating the success of each of our four proposed components, as well as experimental designs for testing our hypotheses.

4.1 Monitoring Mental Health Using Mobile Crowdsensing

We propose both to develop Sensus, a general MCS-based mobile application, and also a framework to address the particular issue of relating social anxiety to behavioral dynamics. Our approach must be capable of conducting human subjects studies that aim at using objective measures in general and must be capable of

[2]https://www.psychologytoday.com/us/basics/bias

calculating the correlations between human's mental well-being status and extracted features in particular. We thus propose a two-stage experiment and evaluation.

For the former, we desire Sensus to support heterogeneous hardware, to support a simple markup language for use by non-CS researchers, and to support off-the-shelf devices and storage. We will assess these goals by cooperating with psychology researchers as they conduct an experiment using Sensus. In particular, we will couple the success or failure of our proof-of-concept feasibility study (for objective factors, such as hardware support) with interviews of the psychologists involved (for subjective factors, such as ease of markup language use).

For the latter, we will use our framework to test hypotheses concerning the relationship between behavioral dynamic and social anxiety. We will conduct a human study and analyze objective sensor data from participants following our proposed framework and then carry out statistical correlation analysis on the features extracted from that data.

4.1.1 Sensus: Cross-Platform, General MCS for Human-Subject Studies

Figure 1 presents the proposed high-level design of Sensus from the perspectives of the researcher (top half) and the study participant (bottom half). These two perspectives coincide in the sharing of sensing plans—called *protocols*—and the storage of study data within Amazon Web Services (AWS) Simple Storage Service (S3). The proposed end-to-end progression is summarized as follows: Researchers use the Sensus mobile app to configure a protocol, which is then disseminated to study participants as an encrypted JSON file. Each study participant receives the protocol file as an email attachment or URL, decrypts the protocol, and loads it into the Sensus mobile app on his or her device. This protocol loading process also involves a participant consent step based upon the consent acquisition strategy researchers choose in their studies. Upon execution the protocol directs the collection and transmission of participant data, which are submitted to AWS S3 for retrieval and analysis by researchers.

To evaluate if Sensus can be applied in human subject studies and provide an easy user interface for researchers without heavy engineering backgrounds, we deploy Sensus in a human subject study conducted with a team of psychologists that focuses on social anxiety in a university student population. We will collaborate to use Sensus to study the effect of real-world social interactions on social anxiety (see Section 4.1.2). The researchers have recruited participants from undergraduate psychology courses. To provide contextual information for social anxiety measures, researchers configure a protocol to poll each participant's location every five minutes using GPS.

The assessment involves correlating social anxiety scores with location data collected by the Sensus protocol. In this proof-of-concept protocol, the location data for each participant are mapped to the discrete venue types (e.g., Transportation, Religious, Work & Study, and Food & Leisure). The total time at each venue is estimated and correlated with participants' social anxiety scores.

To evaluate Sensus, we will use the success or failure of this feasibility study itself as well as interviews with the psychologists (e.g., on the ease of developing Sensus protocols without training in markup languages and mobile development).

4.1.2 Understanding Human Behaviors and Mental Health Status via MCS

We will conduct a human study including two weeks of data collection to asses the ability of Sensus to support a particular investigation in mental health status. We propose to recruit students to study the relationship between behavioral dynamics and social anxiety. Participants' social anxiety levels will be measured using SIAS score [84].

We first collect GPS data, accelerometer data, and communication events (i.e., phone calls and text messages). Next, we pre-process the data by aligning observations temporally. The accelerometer data consist of three dimensions (X, Y, and Z axes), and the communication event data include the time and duration of a phone call or text message. The raw GPS data will be transformed into semantic locations that integrate

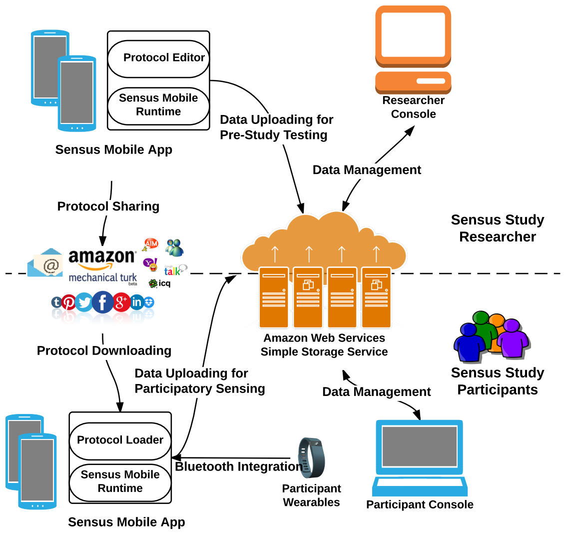

Figure 1: High-level design of Sensus.

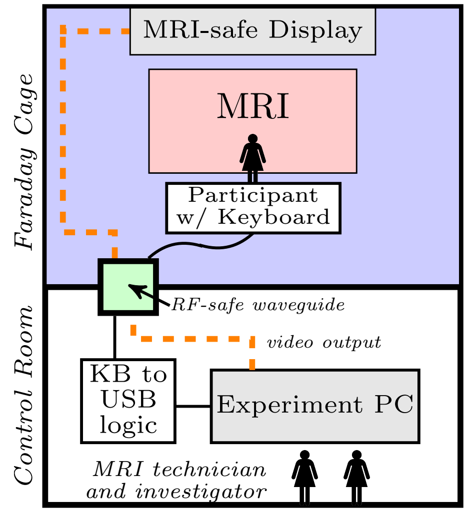

Figure 2: Illustration of fMRI writing setup.

social context. After structuring the data, we will use a sliding window to isolate periods of time before, during, and after each phone call or text message. Next, by modeling the smartphone as a linear dynamic system, the three-dimensional accelerometer data in the sliding windows belonging to a single call or text message can be reduced to a one-dimensional system stimulus. Finally, we will construct a distance matrix using the histogram distribution of the system stimulus and combine this with the GPS semantic labels for further analysis. These distance matrices will allow us to characterize the behavioral dynamics of a participant before, during, and after a phone call or text message. Using the distance matrices, we will further extract features (e.g., statistical data of the matrices, such as mean or max) to examine the relationship between behavioral dynamics and social anxiety levels at different semantic locations and around different communication events. This research activity will be successful if a subset of the features show a significant correlation with the social anxiety levels.

4.2 Understanding the Neural Representations of Data Structures

In our second component, we aim to study the relationship between data structure manipulations and spatial ability at a neurological level. We propose to investigate the following research questions: (1) Do data structure manipulations involve spatial ability? (2) What is the role of task difficulty? (3) Do fMRI and fNIRS agree for software engineering? (4) How do self-reporting and neuroimaging compare?

We propose to conduct a human study in which participants will mentally manipulate lists, arrays, and trees. They will also complete mental rotation tasks involving the ability to determine if two perspective drawings portray the same three-dimensional shapes, as well as questionnaires related to self-reporting.

We will use fMRI to collect high-resolution imaging data following best practices from neuroimaging [58,123]. We will also collect data using the TechEn Inc. CW6 fNIRS system with an above-average number of light detection channels, allowing for a broader view of the brain activities than many published fNIRS studies (cf. [67, 88]). We propose to design a head cap for fNIRS measurement that will deploy at different brain regions which are chosen based on previous neuroimaging studies of program comprehension and mental rotation [33, 116], and consisted of 15 Brodmann[3] areas. For both fMRI and fNIRS experiments, we plan to recruit an above-average number of participants compared with previous studies [88, 116].

[3] The Brodmann anatomical classification system divides the brain into 52 areas, each associated with specific neurological functions [57].

For the analysis for both fMRI and fNIRS data, we propose to follow best practices from psychology, including statistical correction for multiple comparisons and false positives [11]. We will use standard preprocessing techniques to identify and remove artifacts [116], validate model assumptions [52], and standardize locations of brain regions [65] across participants. We will then use general linear models (GLM) to obtain contrast-based estimates of task-related brain activations within voxels (fMRI) or channels (fNIRS) based on the canonical hemodynamic response function. Finally, we propose to perform statistical tests at both individual and group levels to test for significant brain activations, including subsequent correction for false positives. Our goal is to localize significant brain activations from task-related changes in the BOLD response (fMRI) or light absorption (fNIRS). We will analyze the relationships between data structure manipulation and spatial ability in terms of the extent of the brain region overlap between these two types of activities during our contrast-based experiments; to guard against spurious correlations, we will require statistical significance with false discovery rate (FDR) correction ($q < 0.05$).

4.3 Comparing Prose Writing and Code Writing

In our third research activity, we propose to use functional brain imaging to investigate neural representations of code writing in comparison to prose writing. Specifically, we aim to address the following four research questions: (1) Do self reports claim code writing is like prose writing? (2) Does the brain treat code writing like prose writing? (3) What low-level features explain code and prose writing? (4) What high-level features explain code and prose writing?

We will first employ a bespoke QWERTY keyboard that moves all metal and control logic to a separate room. All control logic and metal will be removed from the keyboard, and moving metallic pieces will be replaced with 3D-printed (plastic) equivalents. Briefly, each individual key will be attached to its own shielded wire that extends 30 feet to provide adequate distance from the core of the MRI machine. The wires will be fed through an RF-safe waveguide to the fMRI control room, where a custom-built device reads the state of each key and outputs a standard USB signal. Because no control logic is present near the MRI machine and keystrokes will be processed from the control room, we can eliminate issues caused by electromagnetic interference. This proposed fMRI-safe keyboard organization is illustrated in Figure 2.

Second, we will design two types of tasks: fill-in-the-blank (FITB) and long response (LR). FITB tasks present either a sentence or program containing a blank space, requiring the participant to provide the missing word or code snippet. In LR tasks, participants will write prose or code from scratch to answer an open-ended question or meet a program specification.

We propose to conduct a human study involving these prose and code writing tasks in an fMRI scanner. Participants will complete the two types of writing tasks we introduced above using our bespoke keyboard for the fMRI environment, as well as a questionnaire related to self-reporting.

The data analysis will follow the best practices detailed in Section 4.2. We will collect and analyze the fMRI data with a particular focus on noise mitigation and avoiding false positive correlations. We will restrict attention to those correlations that can be drawn between the brain regions associated with two tasks from our contrast-based experiment that are statistically significant and survive FDR correction ($q < 0.05$).

4.4 Understanding Bias in Code Reviews

In our final research activity, we propose to explore the neurological processes underlying code review and investigate the effects of author bias associated with it. We aim to answer the following research questions: (1) Which pull request aspects affect cognitive load in code reviews? (2) How does code provenance affect developers' decisions in code reviews? (3) Is there any neurological basis for gender-stereotypical associations in code review? (4) Are developers biased regarding computer-generated code patches compared to human-generated patches?

We propose to conduct a human study using fMRI and eye tracking, in which participants will conduct code reviews inside a scanner. This human study will also include a post-experiment survey to gather information

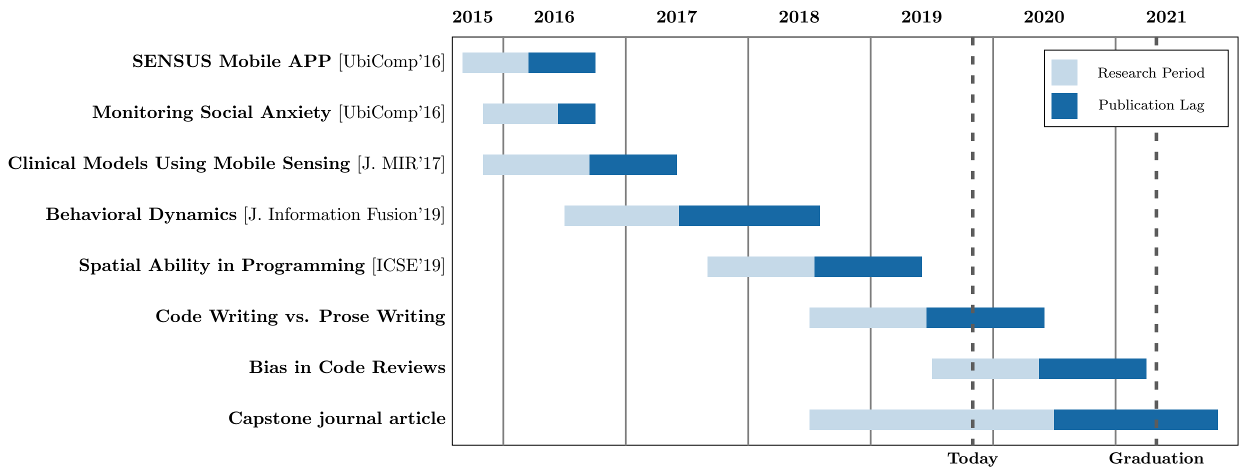

Figure 3: Proposed dissertation work schedule.

of how developers subjectively evaluate their decision making process for code reviews. We will obtain pull requests from real open source projects (e.g., GitHub) and use multiple versions of the stimuli which vary only with respect to the author information. We propose to include female and male authors which are indicated by pictures from the Chicago Face Database, controlling for race, emotional expression and attractiveness [76]. We will also include automated software repair tools as authors of code patches.

The pull request features we will investigate include the number of lines of code presented as context, the number of lines of changes in the code base, and the number of words in pull request comments, and the author provenance of the pull request (a human-generated patch or a computer-generated patch). We will analyze the participants' eye traces to understand how different information in code reviews attracts reviewers' attention. This experiment is designed such that the same code patch will be shown to different participants with different author information. We will check if the author information affects the decision while the quality of code is controlled. We will also check if the author information affects the response time. We will analyze the review judgment, fMRI, eye tracking and post-survey data to check if there is any difference in human developers' answers, brain activity, visual attention, and self-reported subjective thoughts when assessing patches coming from women, men, or computers.

We will recruit a senior student participants and involve approximately the same number of male and female participants for in-group and out-of-group analysis. Similar as described in Section 4.2, we will collect and analyze the fMRI data to study the brain activities in code reviews involving different code features and author information. We will also analyze the eye tracking data in terms of fixation and saccade patterns [37]. We desire an understanding in terms of statistically-significant activations in certain brain regions that are associated with stereotypical associations [10, 24, 61, 100] and/or significant behavioral difference in code reviews tasks with different author information (FDR: $q < 0.05$).

5 Schedule

Figure 3 outlines the research timeline for the proposed dissertation work. We anticipate completing the remaining proposed research efforts in 1.5 years, with an expected graduation in May 2021. Section 6 details preliminary work related to this thesis that has already been conducted or published.

In particular, we have already obtained IRB approval for all human subjects studies proposed here (e.g., HUM00138634 and HUM00161095, including the use of deception related to author provenance to study bias). In addition, all custom equipment (i.e., the fNIRS cap and bespoke fMRI-safe keyboard) has already been constructed and found to be safe for use with medical imaging. The relationships with Psychology researchers are already established (i.e., the Sensus study, fMRI analysis and fNIRS analysis). Finally, we have access to all required equipment, laboratories and hardware (i.e., mobile phones for Sensus, fMRI and fNIRS scanners, and an eye tracker).

We have previously targeted venues such as the *International Conference of Software Engineering* (ICSE), the *International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp), and *Journal of Medical Internet Research* (JMIR). We propose targeting similar venues for the remaining work, as well as other top-tier venues in Software Engineering and Psychology such as the *Foundations of Software Engineering* (FSE), *Transactions on Software Engineering* (TSE) and the *Journal of Brain Imaging and Behavior*.

6 Preliminary Results

In this section, we present preliminary results from ongoing research.

6.1 Monitoring Mental Health Using Mobile Crowd Sensing

We have completed the research described in Section 6.1 to develop a mobile crowdsensing application that supports human subject studies with objective behavioral measures as well as a framework to analyze behavioral dynamic data. The results have been published in peer-reviewed venues [31, 60, 66, 132].

6.1.1 Sensus: Cross-Platform, General MCS for Human-Subject Studies

Sensus has been applied to a feasibility study to collect smartphone sensing data to assess college students' social anxiety, one of the most common disorders in the college student population. In this study, researchers continuously tracked GPS locations of 18 college students, then built an analytic infrastructure to collect the GPS trajectories and finally they analyzed student behaviors (e.g., studying or staying at home) using Point-Of-Interest (POI) data. The collected data provided critical information about how students' social anxiety levels and their mobility patterns are correlated. The primary experiment demonstrated that social anxiety level is positively correlated with the time spent at home or in religious places ($p < 0.05$). We interviewed the researchers to discuss their experience with Sensus. Overall, their feedback was positive and they indicated an interest in using Sensus for future studies.

After that feasibility study, Sensus has been made available to the public. Although the Apple App Store and Google Play Store do not provide precise download information, to the best of our knowledge Sensus has been applied to more than 200 human subjects in several research studies.

6.1.2 Understanding Human Behaviors and Mental Health Status via MCS

We conducted a study of 52 college students and analyzed their accelerometer data using Sensus. We calculated the Pearson correlation coefficient and performed a significance analysis to explore the relationship between human smartphone use and social anxiety status.

Table 1 shows the Pearson correlation and significance analysis between accepted social anxiety measures (SIAS [84]) and the percentages of phone calls ($Call\_Proportion$) and text messages ($Text\_Proportion$) participants made in different locations. We found that the proportions of both phone calls and text messages made at home have a significant positive relation with the social anxiety score (with p-values of 0.0045 and 0.0028). In other words, the higher the social anxiety level is, the more phone calls and text messages are made at home. This may suggest that people with high social anxiety prefer to communicate in the comfort of their homes, perhaps because they are less likely to be exposed to other people or an unfamiliar environment. In this table, the proportion of text messages in personal life locations also shows a significant negative correlation with social anxiety status.

Table 1: Correlation and significance analysis of proportions of phone calls made at different locations vs. SIAS score: Pearson correlation (with p-values) between participants' SIAS scores and the portions of phone calls that were made in specific types of locations. Additionally, the normalized mean ($\bar{x}$) and standard deviation (σ) of participants' portions of phone calls at each location are shown.

	Call_Proportion				Text_Proportion			
Location	Pearson r	p-value	$\bar{x}$	σ	Pearson r	p-value	$\bar{x}$	σ
Work	-0.1806	0.2142	0.0935	0.1074	-0.2511	0.0725	0.1441	0.1040
Home	**0.3983**	**0.0045**	0.3868	0.2484	**0.4059**	**0.0028**	0.3989	0.2128
Food & leisure	-0.2342	0.1053	0.1188	0.1551	-0.0882	0.5340	0.1412	0.1423
Personal life	0.1234	0.3982	0.0138	0.0346	**-0.2917**	**0.0359**	0.0166	0.0228
Transition	-0.0715	0.6141	0.3200	0.1812	-0.0707	0.6045	0.2381	0.1153

Table 2: Correlation and significance Analysis of motion data around phone calls and text messages vs. SIAS score: Pearson correlation (with p-values) between participants' SIAS scores and the features extracted from the smartphone accelerometer data around phone calls and texts. In this table, two features are explored: (1) the average of all distance matrices mean ($\overline{FAC_1}$), and (2) the average of all distance matrices standard deviation ($\overline{FAC_2}$).

	Call (MC)		Text (MT)	
Matrix feature	Pearson r	p-value	Pearson r	p-value
$\overline{FAC_1}$	**0.2867**	**0.0457**	0.1961	0.1634
$\overline{FAC_2}$	**0.3041**	**0.0336**	0.2342	0.0946

Table 2 describes the Pearson correlation between participants' social anxiety scores and the features extracted from their smartphones from the linear dynamic model during phone calls and text messages. For both features, we can conclude that the motion dynamics of phone calls have a significant positive association with anxiety scores (with p-values of 0.0457 and 0.0336). We can infer that people with higher social anxiety levels have more motion variations when they are making phone calls, but there is no significant relationship between text messaging behavioral dynamics and anxiety scores.

6.2 Understanding the Neural Representations of Data Structures

We have conducted a human study involving 76 participants. This population is at least two times larger than previous studies investigating software engineering with medical imaging.

We found that data structure and spatial ability operations are related: both fMRI and fNIRS evidence demonstrate that the two tasks involve activations to the same brain regions ($p < 0.01$). However, the similarity relationship is nuanced: spatial ability operations and tree operations admit a significant contrast and are characterized by differentiated activation magnitudes ($p < 0.001$). Further, some regions relevant to data structures are not accessible to fNIRS: fNIRS lacked the penetrating power to uncover the full evidence reported by fMRI and was unable to distinguish between two distinct tasks. We also found that difficulty matters for data structure tasks: more complicated stimuli result in greater neural activation, and thus an increase in cognitive load. While a neural relationship between spatial ability and data structure manipulation may seem clear in retrospect, it was not obvious to our participants, 70% of whom reported no subjective experience of similarity.

This work was published in ICSE in 2019 [65] where it received a distinguished paper award.

6.3 Comparing Prose Writing and Code Writing

We have completed the hardware construction, safety checks, stimulus design, participant recruitment and scanning for this activity. The formal analysis is ongoing and the work is not yet published.

6.4 Understanding Bias in Code Reviews

We have completed stimulus design and participant recruitment. We also obtained an internal fMRI Pilot Grant from the University of Michigan to support this study. Participant scanning is ongoing.

7 Broader Impact

The potential broader impact of this proposal occurs at different time frames.
Short term. All the medical imaging and behavioral data will be de-identified and released publicly for replication research. Human studies and medical imaging are costly. The data we collect can benefit for future analysis in the community. Our mobile application, SENSUS, can be used in a wide range of human subject studies. Our findings from the proposed research components can help monitor the mental health status (thus help professionals intervene with potential severe mental issues) and help computer science educators develop more efficient strategies.

Medium term. PI Huang is involved in the proposal writing, organization, and data collection for two successfully-funded *Rackham Faculty Allies and Student Ally Diversity Grants*. This activity brings in external speakers for student-focused colloquia and meetings to educate students about underrepresented career goals and research options. In its first year, over 90% of student participants agreed that they learned about different applications of CS and different career options and that the activities helped them better understand how to achieve their career, academic and research goals. PI Huang has helped obtain $30,000 over two years to broaden participation in computing.

Long term. On one hand, human studies with objective measures based on MCS can enrich the research options for researchers with interest in human behaviors in natural settings. On the other hand, medical imaging research in computer science has the potential to augment unreliable self-reporting, guide pedagogy and technology transfer, and improve productivity. While medical imaging is already widely used in other domains, it is still a new topic in computer science research. Our studies can provide guidelines for other researchers for their future study design and implementation. Our findings can also potentially encourage a wider participation in programming community in terms of gender and backgrounds.

8 Conclusion

A fundamental understanding of computational activities is essential to improving productivity and efficiency in software engineering. However, traditionally we have relied only on survey instruments and self-reporting, rather than an understanding of cognitive processes in computational activities, when developing methods to support and improve productivity and efficiency. In this proposal, we present a systematic approach to study the cognitive processes behind multiple important activities in software engineering. We hypothesize that 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 sensing, medical imaging, and eye tracking.

In this thesis, we propose four research components. First, we propose to study the mental status of users in natural settings using mobile crowdsensing. Second, we propose to investigate the relationship between data structure manipulation and spatial ability using medical imaging. Third, we propose to investigate the relationship between code writing and prose writing using medical imaging. Fourth, we propose to study bias in code reviews using medical imaging and eye tracking. This research will improve our fundamental understanding of the cognitive processes involved in critical software engineering activities.

References

[1] *American College Health Association-National College Health Assessment II: Reference Group Executive Summary Spring 2014*. Hanover, MD: American College Health Association (ACHA), 2014.

[2] Silicon Valley Has A Mental Health Crisis Too. https://www.thewealthadvisor.com/article/silicon-valley-has-mental-health-crisis-too, 2019. [Online; accessed 4-September-2019].

[3] Aharoni, D. What you see is what you get the influence of visualization on the perception of data structures. *DOCUMENT RESUME 11*, 4 (2000), 10.

[4] Alias, M., Black, T. R., and Gray, D. E. Effect of instruction on spatial visualization ability in civil engineering students. *International Education Journal 3*, 1 (2002).

[5] Arnow, D., and Weiss, G. Turing's craft. In *https://www.turingscraft.com/* (2019).

[6] Atherton, M., Zhuang, J., Bart, W. M., Hu, X., and He, S. A functional MRI study of high-level cognition. I. the game of chess. *Cognitive Brain Research 16*, 1 (2003), 26–31.

[7] Bacchelli, A., and Bird, C. Expectations, outcomes, and challenges of modern code review. In *Proceedings of the 2013 international conference on software engineering* (2013), IEEE Press, pp. 712–721.

[8] Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., and Grafton, S. T. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences 108*, 18 (2011), 7641–7646.

[9] Bednarik, R., and Tukiainen, M. An eye-tracking methodology for characterizing program comprehension processes. In *Proceedings of the 2006 symposium on Eye tracking research & applications* (2006), ACM, pp. 125–132.

[10] Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., and Sherman, J. W. The quadruple process model approach to examining the neural underpinnings of prejudice. *Neuroimage 43*, 4 (2008), 775–783.

[11] Bennett, C. M., Miller, M., and Wolford, G. L. Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: An argument for proper multiple comparisons correction. *NeuroImage 47* (July 2009).

[12] Berninger, V. W., Richards, T. L., Stock, P. S., Abbott, R. D., Trivedi, P. A., Altemeier, L. E., and Hayes, J. R. fMRI activation related to nature of ideas generated and differences between good and poor writers during idea generation. In *BJEP Monograph Series II, Number 6-Teaching and Learning Writing*, vol. 77. British Psychological Society, 2009, pp. 77–93.

[13] Berninger, V. W., and Winn, W. Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. *Handbook of writing research* (2006), 96–114.

[14] Black, A. A. Spatial ability and earth science conceptual understanding. *Journal of Geoscience Education 53*, 4 (2005), 402–414.

[15] Brooks, R. Towards a theory of the cognitive processes in computer programming. *International Journal of Man-Machine Studies 9*, 6 (1977), 737–751.

[16] Bruner, J. S., Goodnow, J. J., and Austin, G. A. *A study of thinking.* New York: John Wiley & Sons, Inc, 1956.

[17] Buxton, R. B., Uludağ, K., Dubowitz, D. J., and Liu, T. T. Modeling the hemodynamic response to brain activation. *Neuroimage 23* (2004), S220–S233.

[18] Cabeza, R., Daselaar, S. M., Dolcos, F., Prince, S. E., Budde, M., and Nyberg, L. Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval. *Cerebral Cortex 14*, 4 (2004), 364–375.

[19] Caminiti, S. AT&T's $1 billion gambit: Retraining nearly half its workforce for jobs of the future. In *https://www.cnbc.com/2018/03/13/atts-1-billion-gambit-retraining-nearly-half-its-workforce.html* (Mar 2018).

[20] Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Lu, H., Zheng, X., Musolesi, M., Fodor, K., and Ahn, G.-S. The rise of people-centric sensing. *Internet Computing, IEEE 12*, 4 (2008), 12–21.

[21] Canzian, L., and Musolesi, M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), ACM, pp. 1293–1304.

[22] Capretz, L. F., and Ahmed, F. Why do we need personality diversity in software engineering? *ACM SIGSOFT Software Engineering Notes 35*, 2 (2010), 1–11.

[23] Castelhano, J., Duarte, I. C., Ferreira, C., Duraes, J., Madeira, H., and Castelo-Branco, M. The Role of the Insula in Intuitive Expert Bug Detection in Computer Code: An fMRI Study. *Brain Imaging and Behavior* (May 2018).

[24] Cattaneo, Z., Mattavelli, G., Platania, E., and Papagno, C. The role of the prefrontal cortex in controlling gender-stereotypical associations: a tms investigation. *NeuroImage 56*, 3 (2011), 1839–1846.

[25] Chambless, D. L., Fauerbach, J. A., Floyd, F. J., Wilson, K. A., Remen, A. L., and Renneberg, B. Marital interaction of agoraphobic women: A controlled, behavioral observation study. *Journal of Abnormal Psychology 111*, 3 (2002), 502–512.

[26] Charnock, M. Why Is Sillicon Valley So Damn Sad (And Suicidal)? https://medium.com/@MattCharnock/why-is-sillicon-valley-so-damn-sad-and-suicidal-f5772c84b92c, 2018. [Online; accessed 4-September-2019].

[27] Chen, M., Hao, Y., Hwang, K., Wang, L., and Wang, L. Disease prediction by machine learning over big data from healthcare communities. *IEEE Access 5* (2017), 8869–8879.

[28] Chen, M., Qian, Y., Hao, Y., Li, Y., and Song, J. Data-driven computing and caching in 5g networks: Architecture and delay analysis. *IEEE Wireless Communications 25*, 1 (2018), 70–75.

[29] Chen, M., Tian, Y., Fortino, G., Zhang, J., and Humar, I. Cognitive internet of vehicles. *Computer Communications 120* (2018), 58–70.

[30] Chen, M., Zhou, P., and Fortino, G. Emotion communication system. *IEEE Access 5* (2017), 326–337.

[31] Chow, I. P., Fua, K., Huang, Y., Bonelli, W., Xiong, H., Barnes, E. L., and Teachman, A. B. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *J Med Internet Res 19*, 3 (Mar 2017, Impact factor = 4.532), e62.

[32] Cohen, J., Brown, E., DuRette, B., and Teleki, S. *Best kept secrets of peer code review.* Smart Bear Somerville, 2006.

[33] Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A., Bookheimer, S., Rosen, B. R., and Belliveau, J. Changes in cortical activity during mental rotation a mapping study using functional mri. *Brain 119*, 1 (1996), 89–100.

[34] Corballis, M. C. Mental rotation and the right hemisphere. *Brain and language 57*, 1 (1997), 100–121.

[35] Costello, K., and Omale, G. Gartner says global IT spending to grow 1.1 percent in 2019. In *https://www.gartner.com/en/newsroom/press-releases/2019-04-17-gartner-says-global-it-spending-to-grow-1-1-percent-i* (Apr 2019).

[36] Costello, K., and Omale, G. Gartner says global it spending to reach $3.8 trillion in 2019. In *https://www.gartner.com/en/newsroom/press-releases/2019-01-28-gartner-says-global-it-spending-to-reach--3-8-trillio* (Jan 2019).

[37] Crosby, M. E., and Stelovsky, J. How do we read algorithms? a case study. *Computer 23*, 1 (Jan. 1990), 24–35.

[38] Cutter, C. Amazon to retrain a third of its U.S. workforce. In *https://www.wsj.com/articles/amazon-to-retrain-a-third-of-its-u-s-workforce-11562841120* (Jul 2019).

[39] de Montjoye, Y.-A., Quoidbach, J., Robic, F., and Pentland, A. S. Predicting personality using novel mobile phone-based metrics. In *Social computing, behavioral-cultural modeling and prediction*. Springer, 2013, pp. 48–55.

[40] De Smedt, B., Ansari, D., Grabner, R. H., Hannula, M. M., Schneider, M., and Verschaffel, L. Cognitive neuroscience meets mathematics education. *Educational Research Review 5*, 1 (2010), 97–105.

[41] Dell, N., Vaidyanathan, V., Medhi, I., Cutrell, E., and Thies, W. Yours is better!: participant response bias in HCI. In *Human Factors in Computing Systems* (2012), ACM, pp. 1321–1330.

[42] Dong, W., Lepri, B., and Pentland, A. S. Modeling the co-evolution of behaviors and social relationships using mobile phone data. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia* (2011), ACM, pp. 134–143.

[43] Donnon, T., DesCôteaux, J.-G., and Violato, C. Impact of cognitive imaging and sex differences on the development of laparoscopic suturing skills. *Canadian journal of surgery 48*, 5 (2005), 387.

[44] Duraes, J., Madeira, H., Castelhano, J., Duarte, C., and Branco, M. C. WAP: Understanding the Brain at Software Debugging. In *International Symposium on Software Reliability Engineering* (2016), pp. 87–92.

[45] Eagle, N., and Pentland, A. Social serendipity: Mobilizing social software. *Pervasive Computing, IEEE 4*, 2 (2005), 28–34.

[46] Eagle, N., and Pentland, A. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing 10*, 4 (2006), 255–268.

[47] Eagle, N., and Pentland, A. S. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology 63*, 7 (2009), 1057–1066.

[48] Eagle, N., Pentland, A. S., and Lazer, D. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences 106*, 36 (2009), 15274–15278.

[49] Ekkekakis, P. *The measurement of affect, mood, and emotion: A guide for health-behavioral research*. Cambridge University Press, 2013.

[50] Fagan, M. Design and code inspections to reduce errors in program development. In *Software pioneers*. Springer, 2002, pp. 575–607.

[51] Fakhoury, S., Ma, Y., Arnaoudova, V., and Adesope, O. The effect of poor source code lexicon and readability on developers' cognitive load. In *International Conference on Program Comprehension* (2018).

[52] Floyd, B., Santander, T., and Weimer, W. Decoding the representation of code in the brain: An fMRI study of code review and expertise. In *International Conference on Software Engineering* (2017), pp. 175–186.

[53] Ford, D., Behroozi, M., Serebrenik, A., and Parnin, C. Beyond the code itself: how programmers really look at pull requests. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Society* (2019), IEEE Press, pp. 51–60.

[54] Fowler, M. *Refactoring: improving the design of existing code*. Addison-Wesley Professional, 2018.

[55] Fry, Z. P., Landau, B., and Weimer, W. A human study of patch maintainability. In *International Symposium on Software Testing and Analysis* (2012), pp. 177–187.

[56] Ganti, R. K., Ye, F., and Lei, H. Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine 49*, 11 (2011), 32–39.

[57] Garey, L. J. *Brodmann's "Localisation in the Cerebral Cortex"*. World Scientific, 2006.

[58] Glover, G. H. Overview of functional magnetic resonance imaging. *Neurosurgery Clinics 22*, 2 (2011), 133–139.

[59] Gong, J., Asare, P., Qi, Y., and Lach, J. Piecewise linear dynamical model for action clustering from real-world deployments of inertial body sensors. *IEEE Transactions on Affective Computing 7*, 3 (2016), 231–242.

[60] Gong, J., Huang, Y., Chow, P. I., Fua, K., Gerber, M., Teachman, B., and Barnes, L. Understanding behavioral dynamics of social anxiety among college students through smartphone sensors. *Information Fusion: to appear 49* (Sept. 2019), 57–68.

[61] Gozzi, M., Raymont, V., Solomon, J., Koenigs, M., and Grafman, J. Dissociable effects of prefrontal and anterior temporal cortical lesions on stereotypical gender attitudes. *Neuropsychologia 47*, 10 (2009), 2125–2132.

[62] Hegarty, M., and Kozhevnikov, M. Types of visual–spatial representations and mathematical problem solving. *Journal of educational psychology 91*, 4 (1999), 684.

[63] Holloway, L. 4 Big Things Google Is Doing To Improve Diversity In Silicon Valley. https://www.revealnews.org/article/heres-the-clearest-picture-of-silicon-valleys-diversity-yet/, 2015. [Online; accessed 4-September-2019].

[64] Hoyek, N., Collet, C., Rastello, O., Fargier, P., Thiriet, P., and Guillot, A. Enhancement of mental rotation abilities and its effect on anatomy learning. *Teaching and learning in medicine 21*, 3 (2009), 201–206.

[65] Huang, Y., Liu, X., Krueger, R., Santander, T., Hu, X., Leach, K., and Weimer, W. Distilling neural representations of data structure manipulation using fMRI and fNIRS. In *International Conference on Software Engineering* (2019), pp. 396–407.

[66] Huang, Y., Xiong, H., Leach, K., Zhang, Y., Chow, P., Fua, K., Teachman, B. A., and Barnes, L. E. Assessing social anxiety using GPS trajectories and point-of-interest data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2016, acceptance rate = 23.7%), ACM, pp. 898–903.

[67] Ikutani, Y., and Uwano, H. Brain activity measurement during program comprehension with NIRS. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (2014), IEEE, pp. 1–6.

[68] Imtiaz, N., Middleton, J., Chakraborty, J., Robson, N., Bai, G., and Murphy-Hill, E. Investigating the effects of gender bias on github. In *Proceedings of the 41st International Conference on Software Engineering* (2019), IEEE Press, pp. 700–711.

[69] Jones, C. Measuring defect potentials and defect removal efficiency. *CrossTalk The Journal of Defense Software Engineering 21*, 6 (2008), 11–13.

[70] Kellogg, R. T. Training writing skills: A cognitive developmental perspective. *Journal of Writing Research 1*, 1 (2008), 1–26.

[71] Knuth, D. E. *The art of computer programming*, vol. 3. Pearson Education, 1997.

[72] Leff, D. R., Orihuela-Espina, F., Elwell, C. E., Athanasiou, T., Delpy, D. T., Darzi, A. W., and Yang, G.-Z. Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fNIRS) studies. *Neuroimage 54*, 4 (2011), 2922–2936.

[73] Li, K., and Fu, Y. Prediction of human activity by discovering temporal sequence patterns. *IEEE transactions on pattern analysis and machine intelligence 36*, 8 (2014), 1644–1657.

[74] Limb, C. J., and Braun, A. R. Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS one 3*, 2 (2008), e1679.

[75] Lindquist, M. A., Loh, J. M., Atlas, L. Y., and Wager, T. D. Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage 45*, 1 (2009), S187–S198.

[76] Ma, D. S., Correll, J., and Wittenbrink, B. The chicago face database: A free stimulus set of faces and norming data. *Behavior research methods 47*, 4 (2015), 1122–1135.

[77] Mabe, P., and West, S. Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology 67*, 3 (6 1982), 280–296.

[78] Madan, A., Cebrian, M., Lazer, D., and Pentland, A. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (2010), ACM, pp. 291–300.

[79] Madan, A., Cebrian, M., Moturu, S., Farrahi, K., et al. Sensing the" health state" of a community. *IEEE Pervasive Computing 11*, 4 (2012), 36–45.

[80] Maguire, E. A., Woollett, K., and Spiers, H. J. London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus 16*, 12 (2006), 1091–1101.

[81] Maguire, S. *Writing solid code*. Greyden Press, LLC, 2013.

[82] Marginean, A., Bader, J., Chandra, S., Harman, M., Jia, Y., Mao, K., Mols, A., and Scott, A. Sapfix: Automated end-to-end repair at scale. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice* (2019), IEEE Press, pp. 269–278.

[83] Martin, R. C. *Clean code: a handbook of agile software craftsmanship*. Pearson Education, 2009.

[84] Mattick, R. P., and Clarke, J. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety1. *Behaviour Research and Therapy 36*, 4 (1998), 455 – 470.

[85] Menon, V., and Desmond, J. Left superior parietal cortex involvement in writing: integrating fMRI with lesion evidence. *Cognitive brain research 12*, 2 (2001), 337–340.

[86] Milham, M. P., Erickson, K. I., Banich, M. T., Kramer, A. F., Webb, A., Wszalek, T., and Cohen, N. J. Attentional control in the aging brain: Insights from an fMRI study of the Stroop task. *Brain and Cognition 49*, 3 (2002), 277–296.

[87] Moffat, S. D., Hampson, E., and Hatzipantelis, M. Navigation in a “virtual” maze: Sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior 19*, 2 (1998), 73–87.

[88] Nakagawa, T., Kamei, Y., Uwano, H., Monden, A., Matsumoto, K., and German, D. M. Quantifying programmers’ mental workload during program comprehension based on cerebral blood flow measurement: A controlled experiment. In *Companion Proceedings of the 36th International Conference on Software Engineering* (2014), pp. 448–451.

[89] Nordstrom, A. H., Goguen, L. M. S., and Hiester, M. The effect of social anxiety and self-esteem on college adjustment, academics, and retention. *Journal of College Counseling 17*, 1 (2014), 48–63.

[90] Pea, R. D., and Kurland, D. M. On the cognitive prerequisites of learning computer programming. Tech. Rep. 18, Center for Children and Technology, 1983.

[91] Pea, R. D., and Kurland, D. M. On the cognitive effects of learning computer programming. *New ideas in psychology 2*, 2 (1984), 137–168.

[92] Peitek, N., Siegmund, J., Apel, S., Kstner, C., Parnin, C., Bethmann, A., Leich, T., Saake, G., and Brechmann, A. A look into programmers' heads. *IEEE Transactions on Software Engineering* (2018), 1–20.

[93] Peitek, N., Siegmund, J., Parnin, C., Apel, S., Hofmeister, J., and Brechmann, A. Simultaneous Measurement of Program Comprehension with fMRI and Eye Tracking: A Case Study. In *Symposium on Empirical Software Engineering and Measurement* (2018).

[94] Pelchat, M. L., Johnson, A., Chan, R., Valdez, J., and Ragland, J. D. Images of desire: food-craving activation during fMRI. *Neuroimage 23*, 4 (2004), 1486–1493.

[95] Pickering, S. J., and Howard-Jones, P. Educators views on the role of neuroscience in education: Findings from a study of uk and international perspectives. *Mind, Brain, and Education 1*, 3 (2007), 109–113.

[96] Pieterse, V., Kourie, D. G., and Sonnekus, I. P. Software engineering team diversity and performance. In *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries* (2006), South African Institute for Computer Scientists and Information Technologists, pp. 180–186.

[97] Planton, S., Jucla, M., Roux, F.-E., and Démonet, J.-F. The "handwriting brain": a meta-analysis of neuroimaging studies of motor versus orthographic processes. *Cortex 49*, 10 (2013), 2772–2787.

[98] Planton, S., Longcamp, M., Péran, P., Demonet, J.-F., and Jucla, M. How specialized are writing-specific brain regions? an fMRI study of writing, drawing and oral spelling. *Cortex 88* (2017), 66–80.

[99] Podsakoff, P. M., and Organ, D. W. Self-reports in organizational research: Problems and prospects. *Journal of Management 12*, 4 (1986), 531–544.

[100] Quadflieg, S., Turk, D. J., Waiter, G. D., Mitchell, J. P., Jenkins, A. C., and Macrae, C. N. Exploring the neural correlates of social stereotyping. *Journal of Cognitive Neuroscience 21*, 8 (2009), 1560–1570.

[101] Rabbi, M., Ali, S., Choudhury, T., and Berke, E. Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing* (2011), ACM, pp. 385–394.

[102] Raento, M., Oulasvirta, A., and Eagle, N. Smartphones an emerging tool for social scientists. *Sociological methods & research 37*, 3 (2009), 426–454.

[103] Ramos-Garcia, R. I., Muth, E. R., Gowdy, J. N., and Hoover, A. W. Improving the recognition of eating gestures using intergesture sequential dependencies. *IEEE journal of biomedical and health informatics 19*, 3 (2015), 825–831.

[104] Rangarajan, S. Heres the clearest picture of Silicon Valleys diversity yet: Its bad. But some companies are doing less bad. https://www.revealnews.org/article/heres-the-clearest-picture-of-silicon-valleys-diversity-yet/, 2018. [Online; accessed 4-September-2019].

[105] Rauscher, F. H., Shaw, G. L., and Ky, K. N. Listening to mozart enhances spatial-temporal reasoning: towards a neurophysiological basis. *Neuroscience letters 185*, 1 (1995), 44–47.

[106] Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., and Kafai, Y. Scratch: Programming for all. *Commun. ACM 52*, 11 (Nov. 2009), 60–67.

[107] Richards, T. L., Berninger, V. W., Stock, P., Altemeier, L., Trivedi, P., and Maravilla, K. R. Differences between good and poor child writers on fMRI contrasts for writing newly taught and highly practiced letter forms. *Reading and Writing 24*, 5 (2011), 493–516.

[108] Riley, S. Password security: What users know and what they actually do. *Usability News 8*, 1 (2006), 2833–2836.

[109] Ritchey, M., Yonelinas, A. P., and Ranganath, C. Functional connectivity relationships predict similarities in task activation and pattern information during associative memory encoding. *J. Cognitive Neuroscience 26*, 5 (May 2014), 1085–1099.

[110] Ross, J. S., Tkach, J., Ruggieri, P. M., Lieber, M., and Lapresto, E. The mind's eye: Functional MR imaging evaluation of golf motor imagery. *American Journal of Neuroradiology 24*, 6 (2003), 1036–1044.

[111] Ryan, T. J., Alarcon, G. M., Walter, C., Gamble, R., Jessup, S. A., Capiola, A., and Pfahler, M. D. Trust in automated software repair. In *International Conference on Human-Computer Interaction* (2019), Springer, pp. 452–470.

[112] Sato, Y., Masuda, S., Someya, Y., Tsujii, T., and Watanabe, S. An fMRI analysis of the efficacy of Euler diagrams in logical reasoning. In *Visual Languages and Human-Centric Computing* (2015), pp. 143–151.

[113] Shah, C., Erhard, K., Ortheil, H.-J., Kaza, E., Kessler, C., and Lotze, M. Neural correlates of creative writing: an fMRI study. *Human brain mapping 34*, 5 (2013), 1088–1101.

[114] Shepard, R. N., and Metzler, J. Mental rotation of three-dimensional objects. *Science 171*, 3972 (1971), 701–703.

[115] Shin, C., and Dey, A. K. Automatically detecting problematic use of smartphones. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (2013), ACM, pp. 335–344.

[116] Siegmund, J., Kästner, C., Apel, S., Parnin, C., Bethmann, A., Leich, T., Saake, G., and Brechmann, A. Understanding understanding source code with functional magnetic resonance imaging. In *International Conference on Software Engineering* (2014), pp. 378–389.

[117] Siegmund, J., Peitek, N., Parnin, C., Apel, S., Hofmeister, J., Kästner, C., Begel, A., Bethmann, A., and Brechmann, A. Measuring Neural Efficiency of Program Comprehension. In *Foundations of Software Engineering* (2017), pp. 140–150.

[118] Singer, N. The hard part of computer science? getting into class. In *https://www.nytimes.com/2019/01/24/technology/computer-science-courses-college.html* (Jan 2019).

[119] Sorby, S. A. Educational research in developing 3-d spatial skills for engineering students. *International Journal of Science Education 31*, 3 (2009), 459–480.

[120] Sorby, S. A., and Baartmans, B. J. The development and assessment of a course for enhancing the 3-d spatial visualization skills of first year engineering students. *Journal of Engineering Education 89*, 3 (2000), 301–307.

[121] Terrell, J., Kofink, A., Middleton, J., Rainear, C., Murphy-Hill, E., Parnin, C., and Stallings, J. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science 3* (2017), e111.

[122] Tracy, D. M. Toys, spatial ability, and science and mathematics achievement: Are they related? *Sex roles 17*, 3-4 (1987), 115–138.

[123] Triantafyllou, C., Hoge, R., Krueger, G., Wiggins, C., Potthast, A., Wiggins, G., and Wald, L. Comparison of physiological noise at 1.5 t, 3 t and 7 t and optimization of fMRI acquisition parameters. *Neuroimage 26*, 1 (2005), 243–250.

[124] Tsalatsanis, A., Gil-Herrera, E., Yalcin, A., Djulbegovic, B., and Barnes, L. Designing patient-centric applications for chronic disease management. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Aug 2011), pp. 3146–3149.

[125] Vasilescu, B., Posnett, D., Ray, B., van den Brand, M. G., Serebrenik, A., Devanbu, P., and Filkov, V. Gender and tenure diversity in github teams. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (2015), ACM, pp. 3789–3798.

[126] Von Mayrhauser, A., and Vans, A. M. Program comprehension during software maintenance and evolution. *Computer 28*, 8 (1995), 44–55.

[127] Wai, J., Lubinski, D., and Benbow, C. P. Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology 101*, 4 (2009), 817.

[128] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2014), ACM, pp. 3–14.

[129] Wang, R., Harari, G., Hao, P., Zhou, X., and Campbell, A. T. Smartgpa: how smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), ACM, pp. 295–306.

[130] Wenzel, A., Graff-Dolezal, J., Macho, M., and Brendle, J. R. Communication and social skills in socially anxious and nonanxious individuals in the context of romantic relationships. *Behaviour Research and Therapy 43*, 4 (2005), 505 – 519.

[131] Wood, J. Effect of anxiety reduction on children's school performance and social adjustment. *Developmental psychology 42*, 2 (2006), 345.

[132] Xiong, H., Huang, Y., Barnes, L. E., and Gerber, M. S. Sensus: a cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2016, acceptance rate = 23.7%), ACM, pp. 415–426.

[133] Yang, E.-M., Andre, T., Greenbowe, T. J., and Tibell, L. Spatial ability and the impact of visualization/animation on learning electrochemistry. *International Journal of Science Education 25*, 3 (2003), 329–349.

[134] Zhang, D., Yu, Z., and Chin, C.-Y. Context-aware infrastructure for personalized healthcare. *Studies in health technology and informatics 117* (2005), 154 – 163.

[135] Zhao, S., Ramos, J., Tao, J., Jiang, Z., Li, S., Wu, Z., Pan, G., and Dey, A. K. Discovering different kinds of smartphone users through their application usage behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2016), ACM, pp. 498–509.

[136] Zheng, J., Shen, Y., Zhang, Z., Wu, T., Zhang, G., and Lu, H. Emerging wearable medical devices towards personalized healthcare. In *Proceedings of the 8th International Conference on Body Area Networks* (ICST, Brussels, Belgium, Belgium, 2013), BodyNets '13, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 427–431.