Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS

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

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

***Abstract*—Data structures permeate many aspects of software engineering, but their associated human cognitive processes are not thoroughly understood. We leverage medical imaging and insights from the psychological notion of spatial ability to decode the neural representations of several fundamental data structures and their manipulations. In a human study involving 76 participants, we examine list, array, tree, and mental rotation tasks using both functional near-infrared spectroscopy (fNIRS) and functional magnetic resonance imaging (fMRI).**

We find a nuanced relationship: data structure and spatial operations use the same focal regions of the brain but to different degrees. They are related but distinct neural tasks. In addition, more difficult computer science problems induce higher cognitive load than do problems of pure spatial reasoning. Finally, while fNIRS is less expensive and more permissive, there are some computing-relevant brain regions that only fMRI can reach.

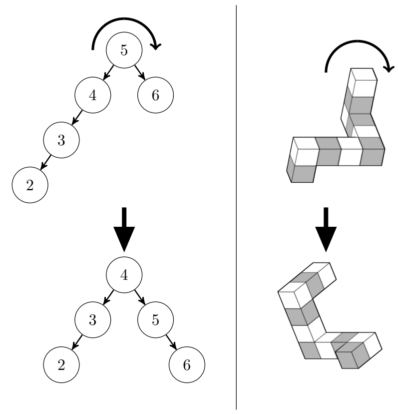


Fig. 1: A representation of the investigated relationship between data structures and spatial ability. On the left, an unbalanced binary tree is rotated about node 1 to produce the tree on the bottom left. On the right, a three-dimensional object is rotated in space as shown in the bottom right. We investigate how the brain represents these two activities using medical imaging techniques.

I. Introduction

Data structures are a fundamental element in computer science that affect the performance and cost of many systems [5], [29], [84], [90]. Data structure choice and usage influence many aspects of software engineering, including maintainability [60], fault tolerance [8], reliability [81], and scalability [70]. Despite the importance of data structures in software development, we have a limited understanding of the subjective cognitive processes underlying their employment. Understanding these processes is important to augment unreliable self-reporting [32], [45], [46] and inform pedagogy, technology transfer, and programming expertise (Section II). In this paper, we present the first investigation that uses medical imaging to decode the neural representations of several classes of data structures and their manipulation.

We leverage two key insights to study the neurological bases associated with data structures. First, we investigate the relationship between data structures and spatial ability. *Spatial ability* is often measured via *mental rotation* tasks like illustrated in Figure 1 [18], [25], [75]. Second, we use two medical imaging techniques, functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS), to provide objective measurements of active brain function and establish a grounded understanding of mental processes associated with data structure manipulation. By comparing these neuroimaging modalities, we develop best practices for imaging investigations of software engineering.

Psychology research has shown spatial ability to be a major factor in proficiencies such as mathematics [38], [89], natural sciences [85], [92], engineering [6], meteorology [11], and map navigation [52]. 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 a wide range of disciplines, it has rarely been studied within software engineering. To the best of our knowledge, only one previous study (conducted by Aharoni [4]) focused on the relationship between software engineering tasks and spatial ability [4]. This previous work relied on interviews with students to understand their thought processes and suggested that programmers use visual representations to reduce the level of abstraction of data structures. However, 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 conducted a human study in which 76 participants mentally manipulated lists, arrays, and trees. Participants also completed mental rotation tasks involving the ability to determine if two perspective drawings portray the same three-dimensional shapes. In our study, we 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 carry out our study using fMRI and fNIRS, two non-invasive *in vivo* neuroimaging techniques that have enabled new and complex studies of brain function. By indirectly measuring changes in oxygen consumption, these two imaging modalities can be used to isolate the brain regions recruited for specific tasks. However, they exhibit tradeoffs relevant to studying software engineering. fMRI allows for sampling across the whole head and offers excellent spatial resolution while requiring a restrictive experimental environment that limits its range of use cases. By contrast, fNIRS is significantly less expensive and admits more freedom in the experimental environment and participant motion. However, fNIRS provides inferior spatial resolution and weaker penetration power below the scalp. Both fMRI and fNIRS have been widely used in psychological and clinical research to develop a deeper understanding of brain functions such as sensory, verbal, and motor processing [2], [33], [45], [59], [67], [80]. fMRI and fNIRS allow us to investigate the physical substrates underlying data structure manipulation and its relationship with spatial ability.

We note that the use of medical imaging in software engineering is still exploratory; since 2014, only nine publications have studied its associated cognitive processes with either fMRI or fNIRS alone [15], [26], [28], [30], [41], [54], [65], [76], [77]. Given the tradeoffs between these two neuroimaging techniques, the community has not settled on the best option for studying software engineering tasks. Our work is the first to use both fMRI and fNIRS to study software engineering, validating fNIRS as an effective neuroimaging technique comparable to fMRI for future research in this field.

The contributions of this paper are as follows:

1) We report on a human study involving 76 participants and two medical imaging techniques, the largest such study we are aware of for software engineering.
2) We find that **data structure and spatial operations are related but distinct neural tasks**: they use the same focal regions of the brain but to different degrees.
3) We demonstrate that **problem difficulty matters at a neural level** in computer science, with more complex stimuli inducing a relatively higher cognitive load in data structure tasks than in mental rotation.
4) We find that fMRI and fNIRS measurements broadly agree for the claims in this study. However, fNIRS cannot distinguish some activities as clearly as can fMRI. On the other hand, fMRI may influence participant accuracy. **Care is needed when using medical imaging** for software engineering.
5) We present evidence from a qualitative investigation showing that imaging can find connections that subjective self-perceptions may overlook.

This paper contributes to a fundamental understanding of cognitive processes in software engineering. To the best of our knowledge, this is the first paper to (1) study data structures with neuroimaging, (2) study the relationship between data structure manipulation and spatial ability, and (3) compare fMRI and fNIRS in the context of software engineering.

II. Background And Motivation

We summarize results and techniques related to medical imaging and psychology for a computer science audience. Section II-A overviews the mechanism and research use of fMRI and fNIRS, including their relative advantages and disadvantages for our experiment. Section II-B summarizes the study of mental rotation in psychology, supporting our experimental use of it as a neurological basis for spatial ability.

A. Medical Imaging

Functional neuroimaging 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. We discuss the most popular techniques: fMRI and its counterpart, fNIRS. First, as non-invasive tools, these two imaging modalities pose significantly less risk and can access a wider range of brain regions than previous invasive techniques (e.g., electrocorticography). Second, fMRI and fNIRS provide a wider field of view and higher spatial resolution than other functional neuroimaging techniques (e.g., EEG, MEG), enabling characterization of a brain region's contribution to a specific task. Third, fMRI and fNIRS do not use ionizing radiation or radioactive elements common in many other neuroimaging modalities (e.g., CT, PET). Instead, both techniques rely on the *hemodynamic response*, the metabolic changes (e.g., oxygen, glucose) in neuronal blood flow to active brain regions, using oxygen consumption as an indirect measurement for brain activity [14].

As a result, fMRI and fNIRS have experienced a dramatic rise in popularity in research. In 2010 alone, fMRI was used in more than 1500 published articles [78]. Among other examples, fMRI has been used to study face recognition, decision making, resting, and vegetative states [47], [53], [57], [78], [79], [88]. Similarly, the use of fNIRS has grown significantly [12]. The applications of fNIRS span many fields such as behavioral development, psychiatric conditions, and brain injury [12], [27], [48], [56].

However, fMRI and fNIRS also share limitations. One limitation is *hemodynamic lag*: the onset of changes in neuronal blood flow peaks several seconds after the onset of stimuli [1], [39]. Similarly, the hemodynamic response saturates over time [46], resulting in weaker signals for sustained tasks. Both characteristics enforce experimental restrictions such as limited task duration (commonly 30 seconds) and require robust mathematical analysis [10], [74]. In this study, we follow best practices in neuroscience and psychology to analyze hemodynamic response signals (Section IV).

1) How fMRI Works: fMRI provides indirect measurements of brain activity by calculating the blood-oxygen level dependent (BOLD) signal, defined as the ratio of oxygenated to deoxygenated hemoglobin [58]. fMRI captures BOLD signals via the application and removal of a series of magnetic fields. As task-related brain activity is mapped onto an anatomical scan of the participant's brain in the associated mathematical analysis, participants must lie still in the fMRI bore throughout the experiment with minimal head movement.

2) How fNIRS Works: fNIRS also measures the hemodynamic response to determine active brain regions. It relies on differences in the absorption of chromophores, groups of atoms that generate color through the absorption of light, between oxygenated and deoxygenated hemoglobin. Light is emitted and detected through devices placed at specific locations on a scalp cap worn by the participant. fNIRS admits relative freedom of motion and has few environmental restrictions. For example, participants can sit in front of a computer and perform in a more realistic software development setting.

3) Comparison of fMRI and fNIRS: fMRI provides excellent spatial resolution and penetrating power. It is a precise neuroimaging modality that captures activations across the whole brain. In contrast, fNIRS provides inferior spatial resolution and depth due to inconsistent photon paths and the limited penetration of near-infrared light. As a result, fNIRS also provides a noisier signal, requiring careful considerations in experiment and analysis design. Likewise, fNIRS requires deciding in advance on the placement of light emitter-detector devices. The number of regions fNIRS can measure simultaneously is limited by physical space on the scalp.

However, fNIRS is gaining traction as a neuroimaging technique due to its portability, ease of administration, ecological validity, and lower cost. In contrast, the high cost, restrictive environment, and high sensitivity to participant motion of fMRI limit its practicality for a broad spectrum of use cases. In this paper, we present recommendations for the use of fMRI and fNIRS to study software engineering.

4) Motivating Functional Neuroimaging: We outline the importance of using medial imaging to understand the mental processes associated with data structure manipulations (discussed further in previous work [30], [61], [76]):

Unreliable self-reporting. Previous software engineering [32] and psychology studies [45], [46] demonstrate that humans' self-reporting are often unreliable. Medical imaging can give accurate, objective explanations of subjective processes.

Pedagogy and training. We can take advantage of visual representations to help students learn programming [4]. Additionally, studies have found different patterns of activation across ages for other tasks. Knowing whether this occurs in software engineering could guide workforce retraining.

Technology transfer. Better models of human judgments of software tools may help improve tool design.

Programming expertise. Previous studies have shown how brain structures change with expertise for other tasks [50] and suggested how fMRI could be used to investigate expertise [63]. Imaging may help us understand the reported productivity gap between experienced and novice programmers.

We hope medical imaging studies can help software engineering researchers to gain a better understanding and improvement of certain tasks.

B. Mental Rotation of Objects

Mental rotation is defined as the capacity to quickly and accurately rotate two- or three-dimensional figures in imagination [25]. Mental rotation tasks generally involve comparing two three-dimensional objects rotated about an axis (Figure 1 right), and are a standard paradigm for testing spatial ability [18]. Neuroimaging suggests that mental rotation involves the right parietal lobe, a region believed to be responsible for spatial ability [17], [21], [37]. In our experiments we use mental rotation as a validated test case for spatial ability.

Mental rotation tasks have a natural notion of difficulty: angle of rotation. Shepard and Metzler found that the time required to solve mental rotation tasks is a linearly-increasing function of the angular difference between the orientations of the two objects [75]. Gogos et al. studied the difficulty of mental rotation using medical imaging. They used fMRI to examine differential activations in regions within the parietal lobe, identifying rises in the BOLD signal with increased angles of rotation [36]. Such findings support the use of mental rotation as a meaningful comparison for the investigation of difficulty in this neuroimaging study.

III. Experimental Setup And Method

We present our study protocol to decode the neurological bases of data structures and their relationship with spatial ability and difficulty. Materials (e.g., all stimuli and de-identified data) are available at the project's website.[1]

A. Overview

In this human study, participants completed three blocks of tasks while being scanned by either fMRI or fNIRS. Stimuli consisted of data structure (i.e., list, array, tree) and mental rotation tasks with varying levels of difficulty. This setup permits the controlled investigation of the relationship between data structures and spatial ability through the lens of difficulty and the choice of medical imaging modality.

B. Recruitment

We recruited 76 students from the University of Michigan for this study. Email solicitations were made to a graduate student list as well as brief presentations in four upper-level undergraduate CS classes. Monetary compensation was offered. After standard filtering (see Section III-C), the final pool contained measurements from 30 fMRI participants and 40 fNIRS participants. Prior to each experiment, participants were screened for the requisite computing background. Table I summarizes the demographic information for all participants. The protocol was approved by our Institutional Review Board.

C. Data Collection

Each participant completed the experiment in a single session. Upon arriving, they provided informed consent and completed a background questionnaire. After watching a training video, participants were prepared for scanning and began the task activities. Participants completed three task blocks of 30 stimuli each (90 stimuli in total). All stimuli were presented for up to 30s and required an A or B response. A fixation cross, a mark used to center participants' gaze, was

[1] http://web.eecs.umich.edu/~weimerw/fmri.html

TABLE I: Demographic data of eligible participants

Demographic Variables		# fMRI	# fNIRS
Sex	Male	16	30
	Female	14	10
Degree Pursuing	Undergraduate	23	31
	Graduate	7	9

shown before each stimulus for 2s–10s. Both fMRI and fNIRS experiments used the same set of 90 stimuli.

Stimuli were subdivided into three categories: (1) lists and arrays (collectively referred to as "sequences"), (2) trees, and (3) mental rotation. Each task block consisted of 10 stimuli from each category. The stimuli order was chosen randomly per participant. Participants were directed to respond as quickly and accurately as possible. After the scanning, participants completed a post survey to provide verbal explanations of their choices and actions.

Our experimental task protocol was designed to accommodate both fMRI and fNIRS. For the fMRI experiments, participants lay in an fMRI machine (see Section II-A1) holding MR-compatible buttons and remained in the machine for the entire scan (see Figure 2). In contrast, fNIRS participants sat in a chair wearing an fNIRS device (see Section II-A2) using a standard keyboard and monitor (see Figure 3). Participants were asked to remain still, but were permitted five minute breaks between each task block. As mentioned in III-B, data from 6 individuals were removed due to difficulties presented when collecting fMRI data (e.g., discomfort in the machine, incomplete dataset, or excessive head motion). In the fNIRS analyses, data from all 40 individuals could be used.[2]

We now provide technical details suitable for conducting or replicating similar research. Section III-D continues with a discussion of the stimuli used in our experiment.

1) fMRI Acquisition: In this experiment, we used fMRI to collect high-resolution imaging data following best practices from neuroimaging [35], [86]. All imaging procedures were conducted on a 3T General Electric MR750 with a 32-channel head coil at the University of Michigan Functional MRI Laboratory. High-resolution anatomical images were acquired with a T_1-weighted spoiled gradient recall (SPGR) sequence (TR = 2300.80 ms, TE = 24 ms, TI = 975 ms, FA = 8°; 208 slices, 1 mm thickness). Prior to the functional scans, we obtained estimates of the static magnetic field using spin-echo fieldmap sequences (TR = 7400 ms, TE = 80 ms; 2.4 mm slice thickness). Functional MRI data were then acquired during both a resting state and during three task-related runs. All scans employed a $T_2{}^*$-weighted multiband echo planar imaging sequence (TR = 800 ms, TE = 30 ms, FA = 52°; acceleration factor: 6), with whole-brain coverage over 60 slices (2.4 mm^3 isotropic voxels).

2) fNIRS Acquisition: In this experiment, we collected 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. [41], [54]). This system contains two laser diodes at 690 nm and 830 nm with fiber optic cables to transmit light between the instrument and a sensor probe on the participant's head. We designed three head caps to accommodate different head sizes (head circumference: 58 cm, 60 cm, 62 cm) based on the international 10–20 system [42], [83], [91] (see Figure 3a). To fit the fNIRS cap to each participant [42], we aligned the cap center with the 10–20 point fPZ (above the bridge of the nose, see [83]). The cap included 16 light emitters and 32 detectors, spaced 3 cm apart, yielding 61 data collection channels[3] deployed at different regions. Regions were chosen based on previous neuroimaging studies of program comprehension and mental rotation [17], [76], and consisted of 15 Brodmann[4] areas. Signals were sampled at 50 Hz and then resampled to 2 Hz for analysis.

Fig. 2: fMRI machine used in our experiment. The participant lies flat in the center of the bore.

(a) fNIRS cap (b) fNIRS environment

Fig. 3: The fNIRS cap on the head of a participant providing coverage of Brodmann areas 6–9, 17–19, 21, 39, 40, 41, 44–47 is shown on the left. On the right, a participant is shown completing the tasks in the fNIRS experimental environment.

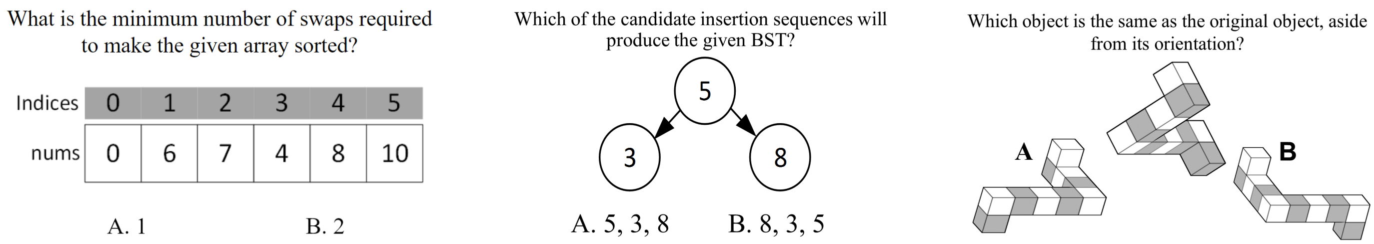
D. Materials and Design

As described in Section III-C, participants were presented with three categories of stimuli: (1) sequences, (2) trees, and (3) mental rotation. Each stimulus from the first and second categories included a starting data structure, an operation to perform, and two answer choices (Figure 4). Answers were either numerical values to describe the outcome of an operation or candidate data structures resulting from an operation. A sequence appeared as either a linked list or an array. For simplicity of modeling, we defined the *difficulty* of a sequence

[2] Although no fNIRS data were removed due to noise, fNIRS does rely on differences in the absorption of near-infrared light, which can be obstructed depending on properties of a participant's hair such as color and thickness.

[3] In theory, each emitter-detector pair could form a channel. In practice, our fNIRS hardware throughput limited us to 61 channels.

[4] The Brodmann anatomical classification system divides the brain into 52 areas, each associated with specific neurological functions [34].



(a) Sequence (List or Array) (b) Tree (c) Mental Rotation

Fig. 4: Example task stimuli, reduced for presentation space. Sequence and Tree stimuli examples shown are simplified for clarity.

or tree task to be the total number of elements present — the N in Big-Oh notation.

The sequence tasks include merge, insert, and swap operations. The tree tasks include binary search tree (BST) rotation, insertion, and traversal operations. In mental rotation tasks, participants were shown a three-dimensional object and two candidate objects, then chose the candidate that could result from a rigid rotation of the original (Figure 4c). We adapted the Mental Rotation Stimulus Library established by Peters and Battista [66] with rotational angle difficulty (see Section II-B). Figure 4 shows simplified examples. Stimuli are available at the project's website.

In the fMRI experiment, the stimuli were presented as images on a screen in the back of the scanner. Participants viewed stimuli via a mirror mounted atop the head coil.[5] Conversely, in the fNIRS experiment, the stimuli were presented as images on a computer monitor next to the fNIRS device (Figure 3b).

IV. Approach

In this section, we present details on the mathematical analyses applied to fMRI and fNIRS data. Our goal is to localize brain activations from task-related changes in the BOLD response (fMRI) or light absorption (fNIRS). Such analyses pose complicated statistical challenges, involving the interpretation of *hemodynamic* responses across anatomically and functionally diverse participants, which themselves are indirect metabolic proxies for underlying *neuronal* (i.e., molecular/cellular) responses. We used standard preprocessing techniques to identify and remove artifacts, validate model assumptions, and standardize locations of brain regions across participants. We then used general linear models to obtain estimates of task-related brain activations within voxels (fMRI) or channels (fNIRS) based on the canonical hemodynamic response function. Finally, we performed statistical tests at both individual and group levels to test for significant brain activations, including subsequent correction for false positives.

Notation. We use the neuroimaging notation $\mathsf{A} > \mathsf{B}$ to refer to the *contrast* (or difference) between two task conditions. For example, Sequence $>$ Tree refers to the comparison of brain activations during sequence vs. tree manipulation. Contrasts are *directional* tests: the aforementioned Sequence $>$ Tree contrast will specifically attempt to identify regions in which average sequence task activity is *greater* than tree manipulation. Critically, this does *not* imply that the inverse contrast (Tree $>$ Sequence) will reveal regions in which tree activity is significantly greater than sequence activity, as differences in the opposite direction may be too small to be statistically meaningful (particularly with the conservative thresholds we use to guard against false positives).

A. fMRI Analysis Approach

Preprocessing. A critical first step in the analysis of fMRI data is *preprocessing*, which serves to correct systematic sources of noise and transform individual brains into a standard space for cross-participant comparison. We employed a number of standard preprocessing procedures using the Statistical Parametric Mapping 12 (SPM12, Wellcome Trust Centre for Neuroimaging, London) software in Matlab. First, we computed *voxel displacement maps* (VDMs) using images from the fieldmap sequence. We then realigned the functional scans after accounting for head motion over time; the VDMs were used to "unwarp" geometric distortions from motion. Next, the anatomical scans were segmented, skull-stripped, and spatially coregistered to the functional data. All images were then transformed into a standard space according to the Montreal Neurological Institute (MNI152) template [51]. Finally, we computed a brain mask using gray and white matter segments of the anatomical scans — this was applied in subsequent statistical analyses to prevent identification of false positive signals within ventricles or outside of brainspace.

First-level analysis. Functional MRI analyses are *multi-level*. First-level models are specified on individual participant data — the results are then combined in a group-level model to assess average task-related changes in brain activity. We specified two first-level general linear models (GLMs) per participant. Briefly, these analyses require us to *predict* the BOLD response to each condition — voxels whose timeseries align with the predicted response are "task-sensitive". In each GLM, we specified regressors for Sequence, Tree, and Mental stimuli across all runs. The duration of each event was curtailed to participant response times. These were convolved with the canonical hemodynamic response function (HRF) and high-pass filtered ($\sigma = 128$ s) to remove low-frequency noise. In one model, we additionally specified a *parametric modulator* for each condition to determine whether the magnitude of the BOLD response scaled linearly with trial difficulty. All models were fit using robust weighted least squares (rWLS) [24],

[5] A helmet-like casing that surrounds the head and is essential for capturing high-quality images [16].

which first obtains estimates of the error variance at each timepoint and reweights the images by a factor of 1/variance to reduce the influence of noisy scans (e.g., due to head motion). This procedure homogenizes the residual timeseries and obtains optimal parameter estimates for each condition.

Contrasts and group-level analysis. Following first-level model estimation, we computed pairwise contrasts to determine mean differences in activity between conditions. These were estimated on a within-participant basis (i.e., on first-level models). We applied a 5 mm^3 full-width at half maximum (FWHM) Gaussian smoothing kernel to each contrast map and carried them upward into group-level *random effects* analyses. A GLM in this context allows us to assess average activity across *all* participants, accounting for inter-individual variance to make some population-level inference. The end result is a *statistical parametric map* of t-values describing clusters of significant activity for a given task-related comparison. Importantly, all models and tests described here were done *voxelwise* — that is, a GLM was specified and estimated for each of nearly 73,000 voxels in brainspace. We therefore applied a *false discovery rate* (FDR) threshold at $q < .05$ to control for false positives as a result of multiple comparisons.

B. fNIRS Analysis Approach

Preprocessing. The raw fNIRS data are light signals transmitted through the channels between emitters and adjacent detectors on the fNIRS cap. The light signals were converted to a measure of the optical density[6] change over time that results from hemodynamic responses.

First-level analysis. Statistical analyses for fNIRS follow the same general principles as fMRI. We specified within-subject, first-level GLMs to model fNIRS optical density measurements in all the channels that were statistically related to the timing of the hemodynamic responses (as determined by convolving timeseries of stimulus events with the canonical HRF). In fNIRS, systemic physiology and motion-induced artifacts are major sources of noise and false positives. We therefore fit our models using autoregressive-whitened robust regression [9], which controls for such confounds and affords optimal parameter estimation. Then, we applied t-tests to the regression coefficients describing the task-related brain activations modeled for every participant. We additionally separated tasks into three difficulty levels and constructed GLMs to analyze the effect of task difficulty on neural activity.

Contrasts and group-level analysis. As with the fMRI analysis, we computed pairwise contrasts to determine mean differences in activity between conditions, estimated on a within-participant basis. Next, we conducted a group-level analysis to summarize the first-level regression coefficients. A mixed effects model was used to examine the average group-level response, with individual participants treated as random effects. Finally, we applied an FDR threshold at $q < .05$ to control for false positives from multiple comparisons.

[6]The degree to which a refractive medium retards transmitted rays of light.

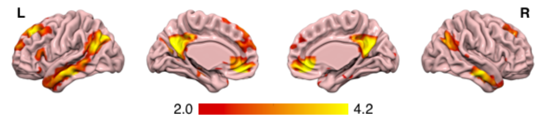


Fig. 5: Significant clusters of activity for Mental > Tree, independent of task difficulty. "Hotter" colors indicate regions showing a larger magnitude difference between the two tasks (i.e., more activity during mental rotation relative to tree manipulation).

V. Results and Analysis

We present quantitative and qualitative analyses to address the following research questions:

RQ1 Do data structure manipulations involve spatial ability?
RQ2 What is the role of task difficulty?
RQ3 Do fMRI and fNIRS agree for software engineering?
RQ4 How do self-reporting and neuroimaging compare?

For simplicity of presentation, we use Code to refer to sequence (array and list) and tree tasks collectively.

A. RQ1 (Data Structures & Spatial Ability) — fMRI

We began with a broad examination of mental rotation vs. code tasks, independent of task difficulty: this would allow us to determine whether there were reliable differences between mental rotation and the two data structure tasks on average. A group-level test of Code > Mental yielded no significant activations after FDR thresholding (i.e., no regions showed consistently stronger activations across both tree and sequence tasks relative to mental rotation). However, Mental > Code revealed robust increases in activation (FDR-corrected) of several regions commonly associated with the brain's "default mode network" (DMN) [13]. Most notably, we observed bilateral recruitment of wide swaths of posterior cingulate cortex (PCC; BA 31) and medial prefrontal cortex (mPFC; BA 8), including subgenual anterior cingulate cortex (sgACC; BA 32). On the lateral face, there emerged a large cluster of activity in the left angular gyrus (AG) / temporoparietal junction (TPJ) (BA 39, 21–22), with additional clusters extending rostrally along the superior temporal sulcus (pSTS) and middle temporal gyrus (MTG) to the temporal pole (BA 21, 38). These anterior temporal cortex clusters were also largely bilateral. The DMN is heavily implicated in various types of *mental simulation*, as required by the tasks performed here.

Given that mental rotation reliably activated DMN regions more than the two code tasks, we applied more focal contrasts to determine whether there were specific differences between Mental > Tree and Mental > Sequence. This revealed that the Mental > Code effect was primarily driven by Mental > Tree (Figure 5). While Mental > Sequence yielded significant differential activations in midline DMN regions such as the PCC and mPFC, these clusters had relatively minimal spatial extent. Patterns of activity related to Mental > Tree, however, were nearly identical to those observed in the comprehensive Mental > Code contrast (Pearson's $r = 0.97, p < .001$). As with the omnibus Code > Mental contrast above, the inverse contrasts (Tree > Mental and Sequence > Mental) also had no voxels survive FDR thresholding.

fMRI results suggest that there are more similarities than differences during mental rotation vs. software engineering tasks. A number of DMN regions involved in mental simulation were recruited more heavily during mental rotation; nevertheless, 95% of voxels were statistically indistinguishable between Mental and Tree tasks.

B. RQ1 (Data Structures & Spatial Ability) — fNIRS

Table II summarizes the fNIRS results. We first examined brain activations comparing each task to a rest condition. The columns Sequence, Mental and Tree show the Brodmann Areas that are significantly activated during the task categories ($p < 0.01$ and $q < 0.05$). The t-values range from 8 (much stronger activation) to -8 (much weaker). We observe that the three categories of tasks all involve significant activations in exactly the same brain regions: BA 6–9, 17–19, 39 and 46.

In the frontal lobe, the premotor cortex and supplementary motor cortex (BA 6), and the frontal eye field (BA 8) showed activation. In the parietal lobe, the part which is associated with visuomotor coordination presented activation (BA 7) and part of Wernicke's area showed activation (BA 39). We also observed strong activation in the primary, secondary and associative visual cortex (BA 17–19). Finally, regions of the dorsolateral prefrontal cortex (BA 9, 46) showed activations for all tasks. All the brain areas listed in the table passed FDR correction ($q < 0.05$).

Having established a broad similarity in how the three tasks each differ from a rest state, we narrowed the investigation by examining how the tasks differ from each other. In Table II, the column Sequence > Mental shows the brain activation results when comparing sequence tasks and mental rotation tasks. Areas related to vision (BA 17–19), Wernicke's area (BA 39) and the prefrontal cortex (BA 46) showed very different patterns of activation between the data structure task and mental rotation. In addition, areas related to language processing (BA 41, 44–45, and 47, which include Broca's Area) strongly distinguish the two. As we observe here, an area (e.g., BA 41) may not significantly distinguish Sequence from a rest state or Mental from a rest state, but may significantly distinguish them *from each other*.

However, the Mental > Tree and Sequence > Tree distinctions are far less compelling. In a comparison, t-values near to either 8 or -8 are relevant. While Sequence > Mental features three areas that reach a magnitude of 5 or more, the other two contrasts never reach a magnitude of 5 and involve fewer regions and channels. In an fNIRS analysis [40], [82], contrasts of that strength result in a conclusion that Mental and Tree, as well as Sequence and Tree, are similar tasks.

fNIRS results demonstrate that mental rotation and data structure tasks involve activations to the same brain regions. However, while Sequence > Mental may be a compelling contrast, the fNIRS evidence does not support the claim that the other tasks are distinct.

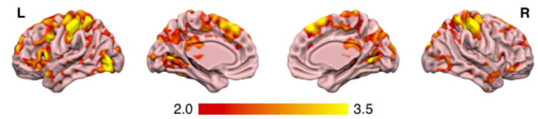


Fig. 6: Significant clusters of activity for Sequence > Mental, accounting for task difficulty. "Hotter" colors indicate regions showing a larger magnitude difference between the two tasks (i.e., more activity during difficult sequence manipulation trials relative to difficult mental rotation trials).

C. RQ2 (Task Difficulty) — fMRI

When we considered the difficulty of each task, we found a significant effect in Sequence > Mental (Figure 6). Larger sequence tasks elicited stronger activations across a wide extent of the brain (FDR-corrected). With the exception of PCC, there was little to no overlap with DMN regions (as seen in the contrasts in Section V-A). Rather, the largest clusters included bilateral postcentral gyrus (BA 40), left inferior frontal gyrus (IFG; BA 44–45), bilateral dorsomedial PFC (dmPFC; BA 6, 8), bilateral anterior insula (BA 13), and bilateral ventral precuneus extending into visual association cortex (BA 18). The heavy recruitment of frontoparietal regions (particularly in the left hemisphere) suggests an increase in cognitive load [22] scaling with the total size of the stimuli.

That is, we found that the brain works measurably "harder" (i.e., there is a larger magnitude BOLD response) for more difficult problems. Because the relationship between mental rotation difficulty and the BOLD signal is so well-established in psychology and cognitive neuroscience [36], [75], it is particularly compelling that we observe a significantly larger effect (in terms of cognitive load and top-down control rising with more complex stimuli) for sequence data structures in software engineering than for mental rotation.

A similar analysis with our fNIRS data revealed no significant findings for the effect of task difficulty on neural activity. This is likely due to fNIRS lacking the penetrative depth and spatial resolution of fMRI.

The brain works measurably harder for more difficult software engineering problems (in terms of cognitive load). Moreover, the regions activated suggest a greater need for effortful, top-down cognitive control when completing challenging sequence manipulation tasks.

D. RQ3 (fMRI and fNIRS Agreement)

Our fMRI and fNIRS measurements and analyses both support the claim that mental rotation and data structure tasks differentially recruit a number of brain regions. However, while fMRI evidence supports a very robust Mental > Tree contrast, the fNIRS evidence is insufficient to support that same claim. This is sensible when we consider the regions yielding the largest differences in fMRI: they largely correspond to structures (e.g., the medial prefrontal cortex and posterior cingulate) that fNIRS cannot measure. Very informally, the parts of the brain that distinguish mental rotation from tree

TABLE II: Summary of fNIRS results. Each column corresponds to a particular task. Each row corresponds to a particular Brodmann Area used during that task along with the range of t-values measured by all fNIRS channels on that BA. Positive t-values indicate stronger activation while negative t-values indicate weaker activation. We report all t-values with $p < 0.01$: all reported results are significant.

Sequence		Mental		Tree		Sequence > Mental		Mental > Tree		Sequence > Tree	
BA	t-value range	BA	t-value range	BA	t-value range	BA	t-value range	BA	t-value range	BA	t-value range
6	2.5 – 5.0	6	2.8 – 4.3	6	3.8 – 4.6					6	2.7 – 2.7
7	4.7 – 5.5	7	5.9 – 6.4	7	5.1 – 7.2						
8	2.6 – 5.1	8	2.9 – 5.5	8	2.5 – 5.6						
9	2.6 – 5.1	9	5.5 – 5.5	9	2.7 – 5.3						
17	3.1 – 4.9	17	3.2 – 6.2	17	2.6 – 5.3	17	-2.4 – -2.4				
18	3.8 – 5.2	18	5.3 – 6.9	18	4.2 – 5.3	18	-2.4 – -2.4	18	2.6 – 2.6		
19	4.0 – 6.6	19	5.3 – 9.1	19	4.2 – 7.3	19	-4.3 – -3.2	19	2.4 – 4.3		
39	3.7 – 7.1	39	4.1 – 7.9	39	4.4 – 7.9	39	-3.3 – -3.3	39	2.4 – 2.4		
						41	-2.3 – -2.3				
						44	-3.3 – -2.6	44	2.6 – 3.4		
						45	-5.0 – -2.4	45	3.5 – 3.5		
46	3.8 – 4.1	46	3.5 – 4.6	46	4.7 – 5.6	46	-5.9 – -2.4	46	2.7 – 4.3	46	-2.6 – -2.6
						47	-5.9 – -5.0	47	3.4 – 4.3		

manipulations are too far "inside the skull" for fNIRS to see: its near-infrared light cannot penetrate deeply beyond regions near the cortical surface.

However, while fMRI is more spatially-resolved, its restrictive and alien environment can also be more daunting for participants. We compared participant performance (i.e., whether or not they gave the correct answer and how long it took) for fMRI and fNIRS; such information was available for 30 fMRI and 40 fNIRS participants. Recall that the questions were identical and the participants were drawn from the same pool. The average accuracy of fNIRS participants, 92%, was significantly higher than the 85% accuracy of fMRI participants ($t = 4.50$, $p < 0.01$) with no significant difference in response time ($t = 0.70$, $p = 0.25$). This could be a very relevant concern for medical imaging studies of productivity, expertise, accuracy or similar software engineering issues.

> fMRI and fNIRS agreed that many areas similarly activate during data structure and mental rotation tasks. However, there were also differences between the tasks that fNIRS was not able to observe. In addition, the fMRI environment had a significant effect on participant accuracy.

E. RQ4 (Self-Reporting & Neuroimaging) — Qualitative

We also conducted a qualitative analysis of survey data focusing on the correlation between explanations provided by participant and neuroimaging data. Data was available for 72 of our 76 participants. At a high level, we find that self-reporting often subtly contrasts with analyses from fMRI and fNIRS data. Complete (de-identified) survey information is available with our other experimental materials and measurements; for space we focus here on a single indicative question.

Participants were asked to compare and contrast a mental rotation task with an BST rotation task. Of the 72 responses, 70% reported *no similarity* between the two tasks. The following quote is indicative: "I don't think those two kinds of tasks were similar. Tree rotation was an idea acquired from CS classroom [sic], but mental rotation was an action more natural to me and easier to perform." However, this subjective experience does not align with measured observation that the same brain regions are recruited to solve both tasks. Even if mental rotation and tree rotation feel subjectively different, changes to brain regions and brain region connectivity have been shown to correlate with learning rates and expertise [50], [73]. It may be, for example, that exercises related to spatial ability can help improve student performance on certain data structure tasks (e.g., because mastering one changes a brain region recruited by the other). Speculatively, this is one example of a research avenue that is encouraged by medical imaging data but entirely hidden if only self-reporting data is used.

These findings reinforce a considerable body of work on unreliable self-reporting (both in psychology [49], [68] and in computer science, including fields such as security [69], human-computer interaction [23], and software maintenance [32]). As previous studies have relied on self-reporting to study mental processes associated with data structures [3], [4], this evidence informs future research of the importance of neuroimaging (or similar techniques) when studying the cognitive processes underlying software engineering tasks.

> While medical imaging data found a nuanced relationship between mental rotation and data structure tasks, including the involvement of the same brain regions, subjective self-report only rarely mentioned any connections.

VI. Threats to Validity

In this section, we describe threats to internal and external validity in our experiment.

One potential threat to internal validity concerns whether or not our data structure tasks measure what they claim to be measuring (i.e., data structure manipulation). The thought processes each participant used when answering may not be identical: indeed, there is significant inter-participant variance in the neural representation of this task. In addition, the particular data structures and tasks we chose are not representative of all of software engineering (e.g., we did not consider skip lists, tries, heaps, maps, etc.). While we mitigate this somewhat by considering fundamental structures (linear sequential structures and branching trees), it is important not to generalize our results far beyond what was directly measured.

Our use of mental rotation tasks as a baseline for spatial ability is one potential threat to external validity, as mental rotation and data structure manipulations differ in their *rigidity*. Rigid transformations are those where distances between every pair of points on an object is preserved [7]. When studying the relationship between data structure manipulation and spatial ability, operations such as insertion, tree rotation, and merging may be more amenable to comparison with non-rigid transformations. However, we believe that mental rotation serves as a useful baseline (Section II-B) for relating data structures to spatial ability. Mental rotation is a paradigm case of spatial ability that has been classified by difficulty both with and without medical imaging [17], [21], [37].

Due to the inherent limitations of fMRI and fNIRS (see Section II-A), we explicitly used stimuli that took no longer than 30 seconds to finish. Thus, our results may not generalize to real-world software engineering tasks. We mitigate this threat slightly by choosing stimuli from college-level courses, which commonly focus on associated fundamental skills. However, this emphasis on tasks that are much shorter than many of those performed by practicing developers is a limitation of the current use of imaging techniques in software engineering [15], [26], [28], [30], [41], [54], [65], [76], [77].

Next, the models used in our data analysis may threaten external validity. To the best of our knowledge, we followed best practices in neuroscience and psychology for our analyses.

Finally, we only recruited undergraduate and graduate students. Thus, our results may only generalize to those with university-level programming experience and education.

VII. Related Work

In this section, we discuss previous work related to computer science and neuroscience, as well as studies from other domains that have used both fMRI and fNIRS. Additionally, we briefly discuss previous research on the cognitive processes underlying data structures and their manipulation.

A. Computer Science and Neuroscience

Siegmund et al. introduced the study of software engineering tasks with fMRI, focusing on code comprehension [76]. Their analyses identified five brain regions with distinct activation patterns, all of which are relevant to working memory, attention, and language processing. 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 [64]. Floyd et al. used fMRI to study code comprehension, code review, and prose review [30]. In addition to identifying brain regions associated with verbal processing, their paper focused on expertise and classification. Newer work has explored the relationship between bug detection and brain activities [15], [26], code comprehension together with other techniques such as eye tracking [65] and the effects of beacons (semantic cues) on code comprehension [77]. Sato et al. studied logic problem solving and diagrams with fMRI [71]. Our study applies Siegmund et al.'s innovative use of neuroimaging, and adopts these previously-identified brain regions as an established basis for verbal processing in software engineering. Unlike previous work, we examine data structures and spatial ability.

Similar to fMRI, fNIRS has been used to study the relationship between program comprehension and brain activity. NIRS researchers found an increase in cerebral blood flow when analyzing obfuscated code and code that requires variable memorization [41], [54]. Subsequent work studied the effect of code readability on cognitive load [28], [77]. The use of fNIRS in these studies provides additional evidence for its application in understanding the cognitive processes in software engineering. However, the relatively small size of these studies ($<$11 participants each) reinforces the need to validate fNIRS as a viable technique in this field. Using 70 participants with two modalities, our study supports the feasibility of using fNIRS to study software engineering.

Besides fMRI and fNIRS, researchers have tried other medical imaging tools to study software engineering. Crk et al. used electroencephalography (EEG) to find that the brain's electrical activity can indicate both prior programming experience and self-reported experience levels [19]. Lee et al. used EEG in a similar setting [44] to Floyd et al.'s work [30]. Parnin used electromyography (EMG) to explore the roles of subvocalization for different programming tasks [62]. Researchers have linked programming tasks and cognitive load [31], [43] using EEG, EMG, and eye tracking.

Beyond neuroimaging, Parnin examined theories of how programmers work and the design of programming environments from a cognitive neuroscience perspective. Parnin proposed a model focused on how a programmer manages task memory, specifically during multi-tasking and interruptions [61]. Similar to our study, Parnin's work investigates computer science from a neuroscience perspective.

Of the previous studies combining neuroimaging or cognitive neuroscience with software engineering, none have investigated the effect of data structures on brain activity or explicitly investigated the relationship between data structures and spatial ability. In addition, no previous study has compared fMRI to fNIRS in the domain of software engineering.

B. Combining fMRI and fNIRS in Psychology

The trade-offs presented by fMRI and fNIRS (see Section II-A) have motivated researchers to conduct combined fMRI and fNIRS studies.

Cui et al. were the first to compare the two modalities through a suite of cognitive tasks [20]. Their findings align with ours: similar conclusions can be drawn but with inferior resolution in the NIRS data. Liu et al. measured fMRI and fNIRS signals from the prefrontal cortex at the same time during cognitive tasks. Their results showed fNIRS-measured cortical signals can be used to infer deep-brain fMRI-measured signal but with lower prediction power. Others comparing fNIRS and fMRI report similar findings [55], [72], [87]: fMRI and fNIRS provide highly-correlated neural responses and fNIRS can be an appropriate substitute for fMRI. However, fNIRS studies must be carefully designed when involving activities in regions more distal from the scalp. Beyond raw signal-level correlations, we found that the resulting models of the two modalities do not draw identical conclusions on low-level explorations in software engineering (see Section V-D), a relevant concern for future imaging research in this area.

C. Data Structures

There are two main papers that have considered data structures at a cognitive level. In a qualitative study involving nine computer science majors, Aharoni investigated student thought processes when dealing with data structures [3], [4]. Through semi-structured interviews, Aharoni found a phenomenon termed the *perception of a data structure as static or dynamic*. This work formally studied the notion that data structures could be subject to mental manipulation. In a second study applying new analyses to the same data, Aharoni found evidence that visual representations influenced students' perceptions of the overall properties of data structures, suggesting that programmers use visual representations to reduce levels of abstraction. While we draw inspiration from Aharoni's investigation of data structure mental manipulations, we use medical imaging to conduct a quantitative analysis that does not focus on self-reporting or qualitative experiments.

VIII. Costs, fMRI, fNIRS, and Research

Medical imaging studies, while still quite rare, are becoming more common in the software engineering literature [15], [26], [28], [30], [41], [54], [65], [76], [77]. fMRI and fNIRS present tradeoffs between cost, fidelity, experimental convenience, and experimental verisimilitude. In this section, we discuss their tangible and intangible costs, including those associated with participant recruitment, equipment cost and time.

As discussed in Section II-A, fMRI poses higher monetary costs than fNIRS. In our study, the cost of fMRI was $575/hour (including the equipment and fMRI technician, etc.). By contrast, in our institute, the use of fNIRS equipment was free. In both cases, participants required 30 minutes for preparation, up to 75 minutes of scanning, and two researchers present.

In addition, each approach comes with recruitment restrictions. For example, fMRI typically requires corrected-to-normal vision (because of the mirror/projection setup) and is not approved for certain populations (e.g., pregnant women or metalworkers). In some cases, participants may not be able to finish a fMRI scanning due to claustrophobia. On the other hand, fNIRS may place significant practical restrictions on the use of participants with dark, thick hair. In practice, we found the fNIRS restrictions to be less onerous (resulting in 0 unusable applicants compared to 4 for fMRI).

Software engineering researchers must carefully weigh the costs and benefits. At a high level, both approaches provide broadly similar evidence. fNIRS requires the researcher to identify relevant brain areas in advance for cap construction (Section II-A3) and cannot penetrate some areas relevant to software engineering (Section V-D). While fMRI is regarded as the gold standard for imaging accuracy, it cost roughly $20,000 more to acquire the fMRI data than the fNIRS data for this experiment, and the environmental constraints of fMRI may influence participant accuracy (Section V-D). As a broad generalization, researchers investigating a computer science topic for the first time may favor fMRI; once the relevant brain areas have been identified, if those regions are accessible via fNIRS, a more cost effective and ecologically-valid study can be conducted via fNIRS. If the proposed study requires more freedom of motion or a quiet environment, involves more than one participant (e.g., pair programming, face-to-face communication, etc.), or uses metal equipment (e.g., a tablet or cellphone), fMRI is not an option without extra effort.

IX. Summary

We investigated the neural representations of fundamental data structures and their manipulations. We hypothesized that data structures are related to spatial ability. Our two key insights were the use of multiple medical imaging approaches and the use of the mental rotation paradigm to serve as a baseline for measuring spatial ability.

Our study involved 76 participants, at least two times larger than previous studies investigating software engineering with medical imaging and is the first to investigate the **neural representations of data structures**.

We found that **data structure and spatial ability operations are related**: both fMRI and fNIRS evidence demonstrates that they involve activations to the same brain regions (e.g., Section V-A and Section V-B, $p < 0.01$).

However, **the similarity relationship is nuanced**: spatial ability operations and tree operations admit a significant contrast and are characterized by differentiated activation magnitudes (e.g., Section V-A, $p < 0.001$).

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 (Section V-D) 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 (Section V-C).

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 (Section V-E).

Since our **direct comparison of fMRI and fNIRS** is unique in software engineering, we elaborate on both measurement and performance issues (Section V-D), as well as monetary, protocol and recruitment issues (Section VIII).

Data structures are critical to many aspects of software engineering, but no previous work has quantitatively investigated their neurological underpinnings. Using medical imaging to understand cognitive processes in software engineering is still very new: this study is exploratory rather than definitive. Indeed, from our perspective it may raise more questions than it answers, as related to implications. We hope that our concrete analysis of a lower-cost medical imaging alternative, and our direct analysis of a computing activity beyond the realm of code comprehension, will encourage more researchers to investigate the cognitive aspects of software engineering.

REFERENCES

[1] R. Aamand, T. Dalsgaard, Y.-C. Lynn Ho, A. Moller, A. Roepstorff, and T. Lund. A NO way to BOLD?: Dietary nitrate alters the hemodynamic response to visual stimulation. *NeuroImage*, 83, 07 2013.

[2] B. P. Acevedo, E. N. Aron, A. Aron, M.-D. Sangster, N. Collins, and L. L. Brown. The highly sensitive brain: an fMRI study of sensory processing sensitivity and response to others' emotions. *Brain and behavior*, 4(4):580–594, 2014.

[3] D. Aharoni. Cogito, ergo sum! cognitive processes of students dealing with data structures. *ACM SIGCSE Bulletin*, 32(1):26–30, 2000.

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

[5] M. A.-M. Al-Shandawely. *Impacts of Data Structures and Algorithms on Multi-core Efficiency*. Skolan för datavetenskap och kommunikation, Kungliga Tekniska högskolan, 2010.

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

[7] K. Atit, T. F. Shipley, and B. Tikoff. Twisting space: are rigid and non-rigid mental transformations separate spatial skills? *Cognitive processing*, 14(2):163–173, 2013.

[8] Y. Aumann and M. A. Bender. Fault tolerant data structures. In *Foundations of Computer Science*, pages 580–589. IEEE, 1996.

[9] J. W. Barker, A. Aarabi, and T. J. Huppert. Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical optics express*, 4(8):1366–1379, 2013.

[10] C. M. Bennett, M. Miller, and G. Wolford. Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: an argument for multiple comparisons correction. *Neuroimage*, 47(Suppl 1):S125, 2009.

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

[12] D. A. Boas, C. E. Elwell, M. Ferrari, and G. Taga. Twenty years of functional near-infrared spectroscopy: introduction for the special issue, 2014.

[13] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter. The brain's default network. *Annals of the New York Academy of Sciences*, 1124(1):1–38, 2008.

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

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

[16] Center for Diagnostic Imaging. I'm getting an MRI, so what's a coil? *https://www.mycdi.com/viewpoints/im_getting_an_mri_so_whats_a_coil_103*, Jan 2016.

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

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

[19] I. Crk, T. Kluthe, and A. Stefik. Understanding programming expertise: an empirical study of phasic brain wave changes. *Transactions on Computer-Human Interaction*, 23(1):2, 2016.

[20] X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage*, 54(4):2808–2821, 2011.

[21] J. C. Culham and N. G. Kanwisher. Neuroimaging of cognitive functions in human parietal cortex. *Current opinion in neurobiology*, 11(2):157–163, 2001.

[22] D. T. D, L. Chang, E. Caparelli, and T. Ernst. Different activation patterns for working memory load and visual attention load. *Brain research*, 1132, 2007.

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

[24] J. Diedrichsen and R. Shadmehr. Detecting and adjusting for artifacts in fMRI time series data. *NeuroImage*, 27(3):624 – 634, 2005.

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

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

[27] A.-C. Ehlis, S. Schneider, T. Dresler, and A. J. Fallgatter. Application of functional near-infrared spectroscopy in psychiatry. *Neuroimage*, 85:478–488, 2014.

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

[29] N. Faria, R. Silva, and J. L. Sobral. Impact of data structure layout on performance. In *Parallel, Distributed and Network-Based Processing*, pages 116–120, 2013.

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

[31] T. Fritz, A. Begel, S. C. Müller, S. Yigit-Elliott, and M. Züger. Using psycho-physiological measures to assess task difficulty in software development. In *International Conference on Software Engineering*, pages 402–413, 2014.

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

[33] W. D. Gaillard, B. C. Sachs, J. R. Whitnah, Z. Ahmad, L. M. Balsamo, J. R. Petrella, S. H. Braniecki, C. M. McKinney, K. Hunter, B. Xu, et al. Developmental aspects of language processing: fMRI of verbal fluency in children and adults. *Human brain mapping*, 18(3):176–185, 2003.

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

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

[36] A. Gogos, M. Gavrilescu, S. Davison, K. Searle, J. Adams, S. L. Rossell, R. Bell, S. R. Davis, and G. F. Egan. Greater superior than inferior parietal lobule activation with increasing rotation angle during mental rotation: an fMRI study. *Neuropsychologia*, 48(2):529–535, 2010.

[37] I. M. Harris, G. F. Egan, C. Sonkkila, H. J. Tochon-Danguy, G. Paxinos, and J. D. Watson. Selective right parietal lobe activation during mental rotation: a parametric pet study. *Brain*, 123(1):65–73, 2000.

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

[39] R. N. Henson, C. J. Price, M. D. Rugg, R. Turner, and K. J. Friston. Detecting latency differences in event-related bold responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage*, 15(1):83–97, 2002.

[40] T. J. Huppert. Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics*, 3:010401, 03 2016.

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

[42] G. H. Klem, H. O. Lüders, H. Jasper, C. Elger, et al. The ten-twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol*, 52(3):3–6, 1999.

[43] S. Lee, D. Hooshyar, H. Ji, K. Nam, and H. Lim. Mining biometric data to predict programmer expertise and task difficulty. *Cluster Computing*, pages 1–11, 2017.

[44] S. Lee, A. Matteson, D. Hooshyar, S. Kim, J. Jung, G. Nam, and H. Lim. Comparing programming language comprehension between novice and expert programmers using EEG analysis. In *International Conference on Bioinformatics and Bioengineering*, pages 350–355, Oct 2016.

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

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

[47] J. Liu, A. Harris, and N. Kanwisher. Perception of face parts and face configurations: an fMRI study. *Journal of cognitive neuroscience*, 22(1):203–211, 2010.

[48] S. Lloyd-Fox, A. Blasi, and C. Elwell. Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews*, 34(3):269–284, 2010.

[49] P. A. Mabe and S. G. West. Validity of self-evaluation of ability: A review and meta-analysis. *Journal of applied Psychology*, 67(3):280, 1982.

[50] E. Maguire, D. Gadian, I. Johnsrude, C. Good, J. Ashburner, R. Frackowiak, and C. Frith. Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, 97(8):4398–4403, 2000.

[51] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

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

[53] M. M. Monti, A. Vanhaudenhuyse, M. R. Coleman, M. Boly, J. D. Pickard, L. Tshibanda, A. M. Owen, and S. Laureys. Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine*, 362(7):579–589, 2010.

[54] T. Nakagawa, Y. Kamei, H. Uwano, A. Monden, K. Matsumoto, and D. M. German. 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*, pages 448–451, 2014.

[55] J. A. Noah, Y. Ono, Y. Nomoto, S. Shimada, A. Tachibana, X. Zhang, S. Bronner, and J. Hirsch. fMRI validation of fNIRS measurements during a naturalistic task. *Journal of visualized experiments*, (100), 2015.

[56] H. Obrig. NIRS in clinical neurology — a ‘promising’ tool? *Neuroimage*, 85:535–546, 2014.

[57] J. P. O’Doherty, A. Hampton, and H. Kim. Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of sciences*, 1104(1):35–53, 2007.

[58] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

[59] M. Okamoto, M. Matsunami, H. Dan, T. Kohata, K. Kohyama, and I. Dan. Prefrontal activity during taste encoding: an fNIRS study. *Neuroimage*, 31(2):796–806, 2006.

[60] P. Oman and J. Hagemeister. Metrics for assessing a software system’s maintainability. In *Software Maintenance, 1992. Proceerdings., Conference on*, pages 337–344. IEEE, 1992.

[61] C. Parnin. A cognitive neuroscience perspective on memory for programming tasks. *Programming Interest Group*, page 27, 2010.

[62] C. Parnin. Subvocalization-toward hearing the inner thoughts of developers. In *International Conference on Program Comprehension*, pages 197–200. IEEE, 2011.

[63] C. Parnin, J. Siegmund, and N. Peitek. On the nature of programmer expertise. In *Psychology of Programming Interest Group Workshop*, 2017.

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

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

[66] M. Peters and C. Battista. Applications of mental rotation figures of the shepard and metzler type and description of a mental rotation stimulus library. *Brain and cognition*, 66(3):260–264, 2008.

[67] L.-A. Petitto, M. S. Berens, I. Kovelman, M. H. Dubins, K. Jasinska, and M. Shalinsky. The “perceptual wedge hypothesis” as the basis for bilingual babies phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and language*, 121(2):130–143, 2012.

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

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

[70] H. Samet. *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1990.

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

[72] V. Scarapicchia, C. Brown, C. Mayo, and J. R. Gawryluk. Functional magnetic resonance imaging and functional near-infrared spectroscopy: insights from combined recording studies. *Frontiers in human neuroscience*, 11:419, 2017.

[73] J. Scholz, M. Klein, T. Behrens, and H. Johansen-Berg. Training induces changes in white matter architecture. *Nature neuroscience*, 12(11):1370–1371, 2009.

[74] Scicurious. IgNobel prize in neuroscience: The dead salmon study. *Scientific American Blog Network*, Sep 2012.

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

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

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

[78] K. Smith. fMRI 2.0. *Nature*, 484(7392):24, 2012.

[79] S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, et al. Correspondence of the brain’s functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, 106(31):13040–13045, 2009.

[80] C. J. Stoodley, E. M. Valera, and J. D. Schmahmann. Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *Neuroimage*, 59(2):1560–1570, 2012.

[81] R. E. Strom and S. Yemini. Typestate: A programming language concept for enhancing software reliability. *IEEE Transactions on Software Engineering*, SE-12(1):157–171, 1986.

[82] S. Tak and J. C. Ye. Statistical analysis of fNIRS data: A comprehensive review. *NeuroImage*, 85:72–91, 2014. Celebrating 20 Years of Functional Near Infrared Spectroscopy (fNIRS).

[83] T. C. Technologies. 10/20 system positioning. *https://www.trans-cranial.com/local/manuals/10_20_pos_man_v1_0_pdf.pdf*, 2012.

[84] V. Tkachenko. How three fundamental data structures impact storage and retrieval. *https://dzone.com/articles/how-three-fundamental-data-structures-impact-stora*, Mar 2016.

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

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

[87] A. Turner, S. B. Erdogan, and A. Akin. Translating fMRI to fNIRS. In *International Conference on Neural Engineering*, pages 493–496. IEEE, 2013.

[88] M. P. Van Den Heuvel and H. E. H. Pol. Exploring the brain network: a review on resting-state fMRI functional connectivity. *European neuropsychopharmacology*, 20(8):519–534, 2010.

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

[90] R. Williams. A survey of data structures for computer graphics systems. In *Data Structures, Computer Graphics, and Pattern Recognition*, pages 105–152. Elsevier, 1977.

[91] X. Xiao, H. Zhu, W.-J. Liu, X.-T. Yu, L. Duan, Z. Li, and C.-Z. Zhu. Semi-automatic 10/20 identification method for MRI-free probe placement in transcranial brain mapping techniques. *Frontiers in Neuroscience*, 11(4), 2017.

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