Measuring Neural Efficiency of Program Comprehension

Janet Siegmund

Norman Peitek

Chris Parnin

Sven Apel

Johannes Hofmeister

Christian Kästner

Andrew Begel

Anja Bethmann

André Brechmann

Microsoft Research

Carnegie Mellon University

NC STATE UNIVERSITY

UNIVERSITÄT PASSAU

LIN LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

Bottom-up Comprehension

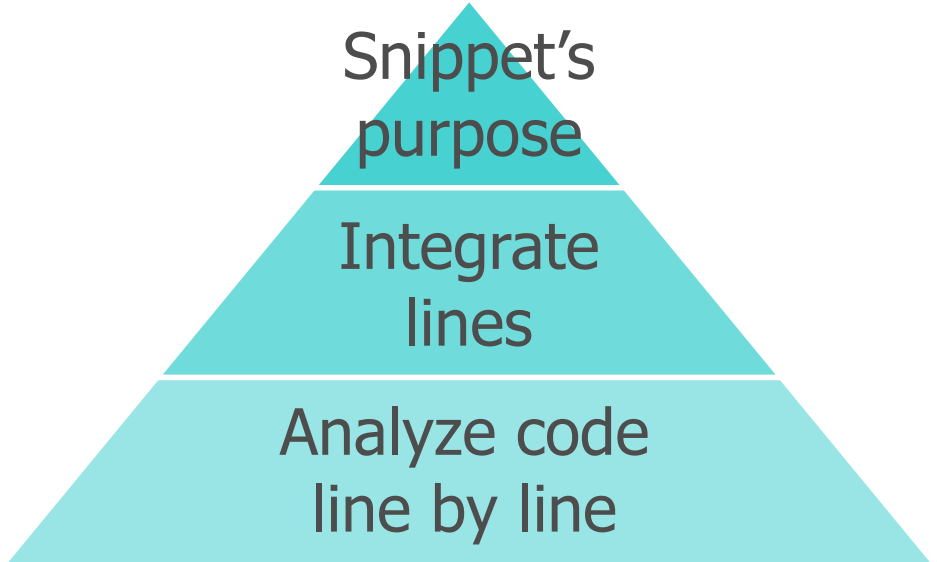

```
public void red() {
    int array[] = {2, 19, 5, 17};
    int result = array[0];
    for (int i = 1; i < array.length; i++)
        if (array[i] > result)
            result = array[i];
    System.out.println(result);
}
```

Not helpful

Variables..

Loop with condition

Output

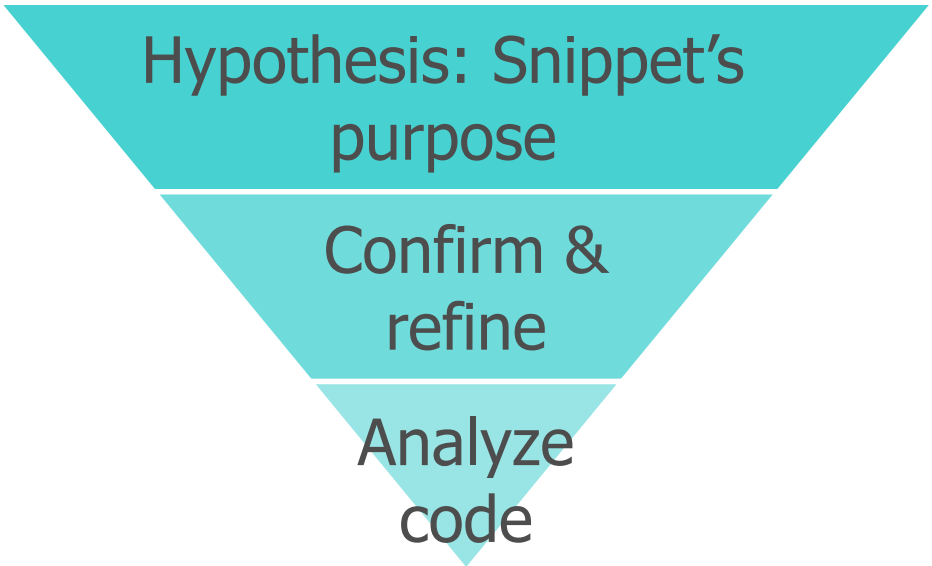


Top-down Comprehension


```
public float arrayAverage(int[] numbers) {
    int counter = 0;
    int sum = 0;

    while (counter < numbers.length) {
        sum = sum + numbers[counter];
        counter = counter + 1;
    }

    float average = sum / (float) counter;
    return average;
}
```



Beacon

Confirmation

Loop

Calculation

Bottom-up Comprehension

Top-down Comprehension

Why does it matter?

Motivation

Bottom-up Comprehension

Top-down Comprehension

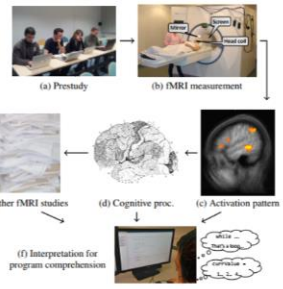
ICSE'14

Understanding Understanding Source Code with Functional Magnetic Resonance Imaging

Janet Siegmund[a,*], Christian Kästner[c], Sven Apel[a], Chris Parnin[g], Anja Bethmann[e], Thomas Leich[d], Gunter Saake[e], and André Brechmann[f]
[a]University of Passau, Germany [c]Carnegie Mellon University, USA
[f]Georgia Institute of Technology, USA [f]Leibniz Inst. for Neurobiology Magdeburg, Germany
[d]Metop Research Institute, Magdeburg, Germany [e]University of Magdeburg, Germany

ABSTRACT

Program comprehension is an important cognitive process that inherently eludes direct measurement. Thus, researchers are struggling with providing suitable programming languages, tools, or coding conventions to support developers in their everyday work. In this paper, we explore whether *functional magnetic resonance imaging (fMRI)*, which is well established in cognitive neuroscience, is feasible to more directly measure program comprehension. In a controlled experiment, we observed 17 participants inside an fMRI scanner while they were comprehending short source-code snippets, which we contrasted with locating syntax errors. We found a clear, distinct activation pattern of five brain regions, which are related to working memory, attention, and language processing—all processes that fit well to our understanding of program comprehension. Our results encourage us and, hopefully, other researchers to use fMRI in future studies to measure program comprehension and, in the long run, answer questions, such as: Can we predict whether someone will be an excellent programmer? How effective are new languages and tools for program understanding? How should we train developers?



1. INTRODUCTION

Today

Measuring Neural Efficiency of Program Comprehension

Janet Siegmund
University of Passau
Passau, Germany

Norman Peitek
Leibniz Institute for Neurobiology
Magdeburg, Germany

Chris Parnin
NC State University
Raleigh, North Carolina, USA

Sven Apel
University of Passau
Passau, Germany

Johannes Hofmeister
University of Passau
Passau, Germany

Christian Kästner
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Andrew Begel
Microsoft Research
Redmond, Washington, USA

Anja Bethmann
Leibniz Institute for Neurobiology
Magdeburg, Germany

André Brechmann
Leibniz Institute for Neurobiology
Magdeburg, Germany

ABSTRACT

Most modern software programs cannot be understood in their entirety by a single programmer. Instead, programmers must rely on a set of cognitive processes that aid in seeking, filtering, and shaping relevant information for a given programming task. Several theories have been proposed to explain these processes, such as "beacons," for locating relevant code, and "plans," for encoding cognitive models. However, these theories are decades old and lack validation with modern cognitive-neuroscience methods. In this paper, we report on a study using functional magnetic resonance imaging (fMRI) with 11 participants who performed program comprehension tasks. We manipulated experimental conditions related to beacons and layout to isolate specific cognitive processes related to bottom-up comprehension and comprehension based on semantic cues. We found evidence of semantic chunking during bottom-up comprehension and lower activation of brain areas during comprehension based on semantic cues, confirming that beacons ease comprehension.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Empirical studies in HCI**;

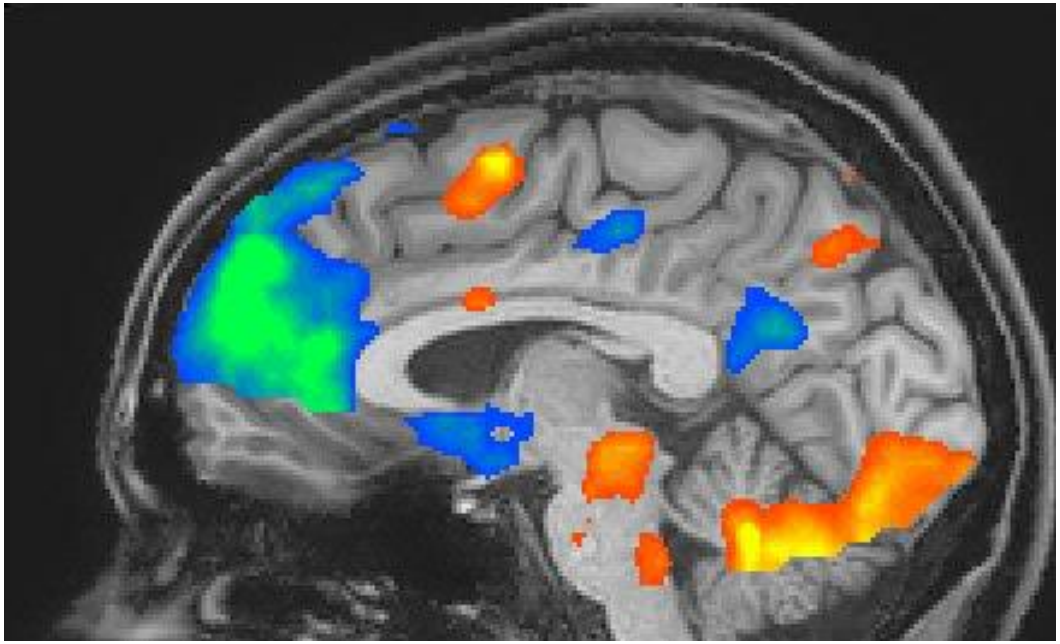
Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering, Paderborn, Germany, September 4–8, 2017 (ESEC/FSE'17), 11 pages.
https://doi.org/10.1145/3106237.3106268

1 INTRODUCTION

During program comprehension, the eyes of programmers glide across a computer screen. In just seconds, they can extract a deep understanding from abstract symbols and text arranged in a source-code file. Expert programmers are especially adept at program comprehension of familiar code—their eyes dance around, finding points of interest, called *beacons* (or semantic cues) that provide hints about a program's purpose, such as method signatures, and common programming idioms. *Top-down comprehension* has been used as an umbrella term to describe cognitive processes related to experience and expectation that guide the understanding of source code [7]. Researchers have also theorized that programmers must use preformed knowledge structures, called *plans*, that represent semantic and syntactical patterns of software [8]. For example, an identifier `bubbleSort` indicates the presence of a sorting algorithm and primes a programmer to expect other elements of the bubble-sort algorithm, such as code related to a swap of array

Neuro-Imaging: fMRI Basics

- Why functional magnetic resonance imaging (fMRI)?
  - **Observe** cognitive processes via **brain activation**

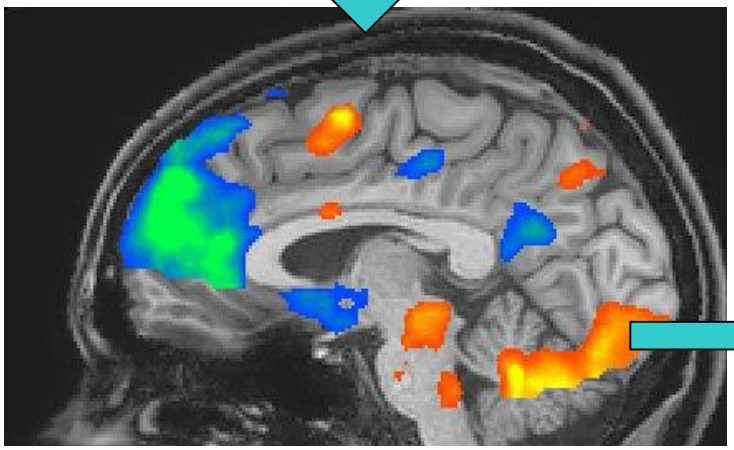


Neuro-Imaging: fMRI Basics

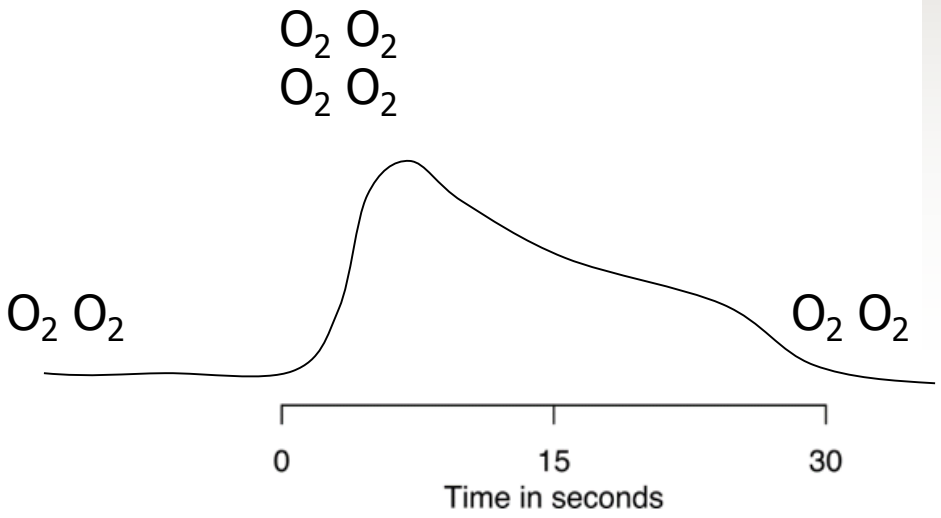

```
public float arrayAverage(int[] numbers) {
    int counter = 0;
    int sum = 0;

    while (counter < numbers.length) {
        sum = sum + numbers[counter];
        counter = counter + 1;
    }

    float average = sum / (float) counter;
    return average;
}
```

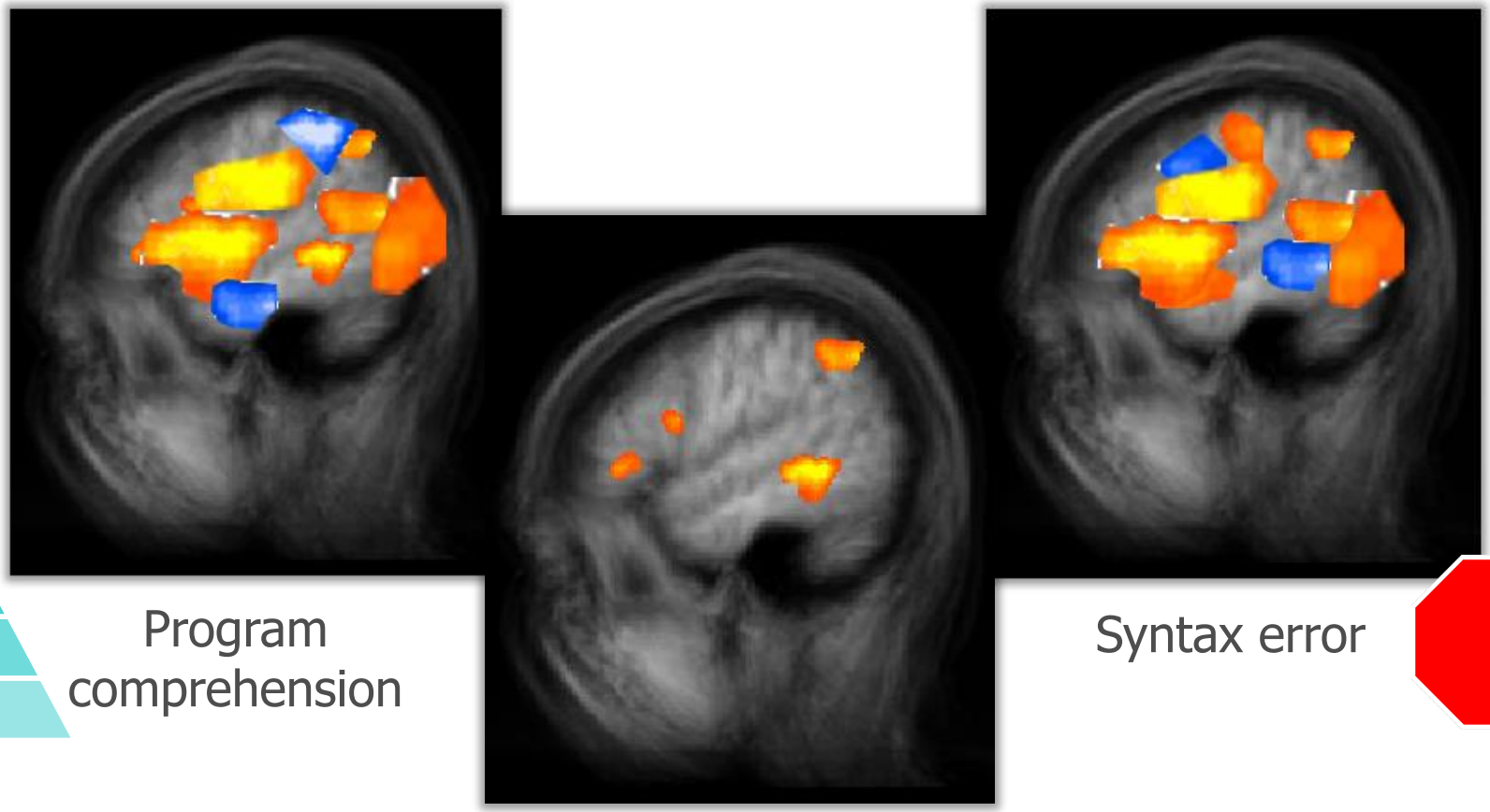


BOLD response



Neuro-Imaging: fMRI Contrasts

- Necessary: **contrasts**
  - What brain activity is actually related to the task?



Program comprehension

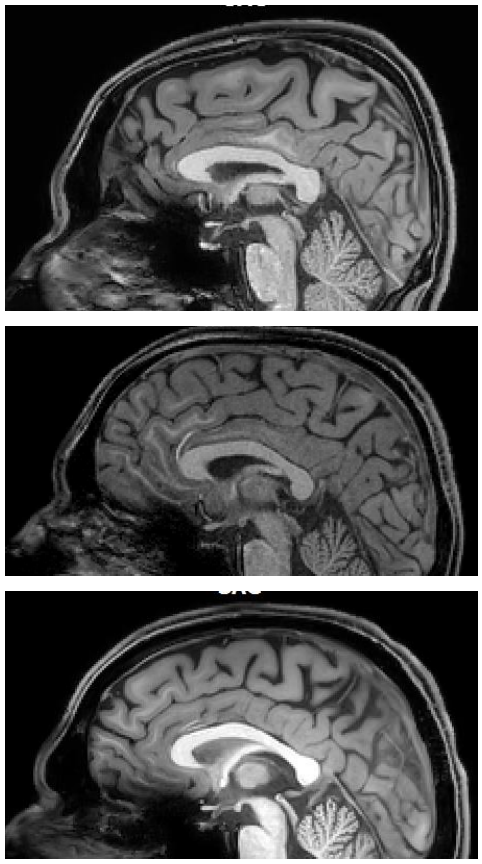
Syntax error

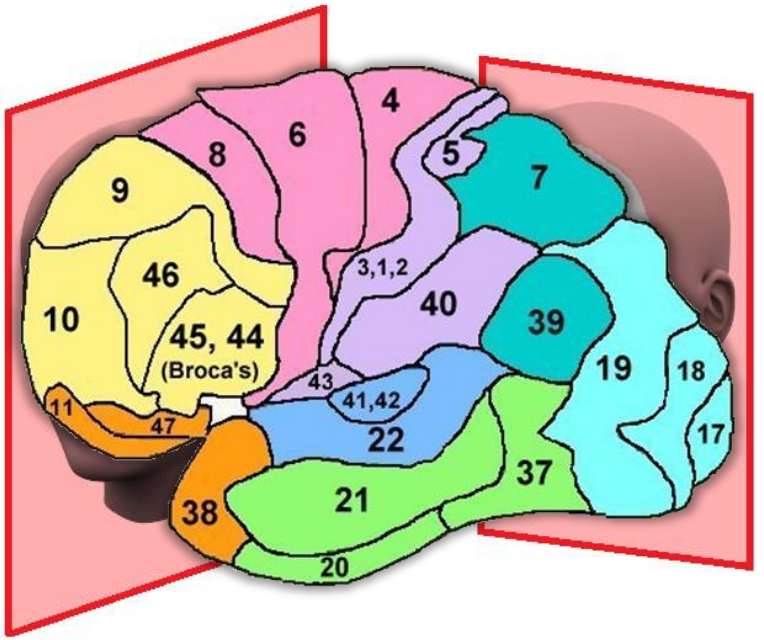
;

Program comprehension – Syntax error

Neuro-Imaging: fMRI Analysis

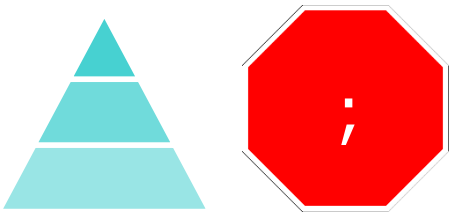
- Activation analysis by mapping into **Talairach space**
  - Brain mapped into Brodmann areas





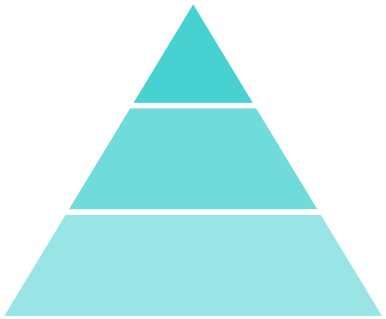
Research Questions

- RQ1: Can we replicate the results of the study by Siegmund and others [ICSE'14]?



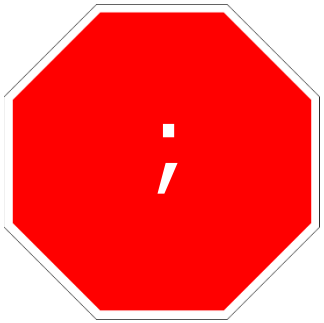
- RQ2: What is the difference between bottom-up program comprehension and top-down comprehension in terms of activation and the brain areas involved?

Experiment Design



Bottom-up comprehension

Top-down comprehension



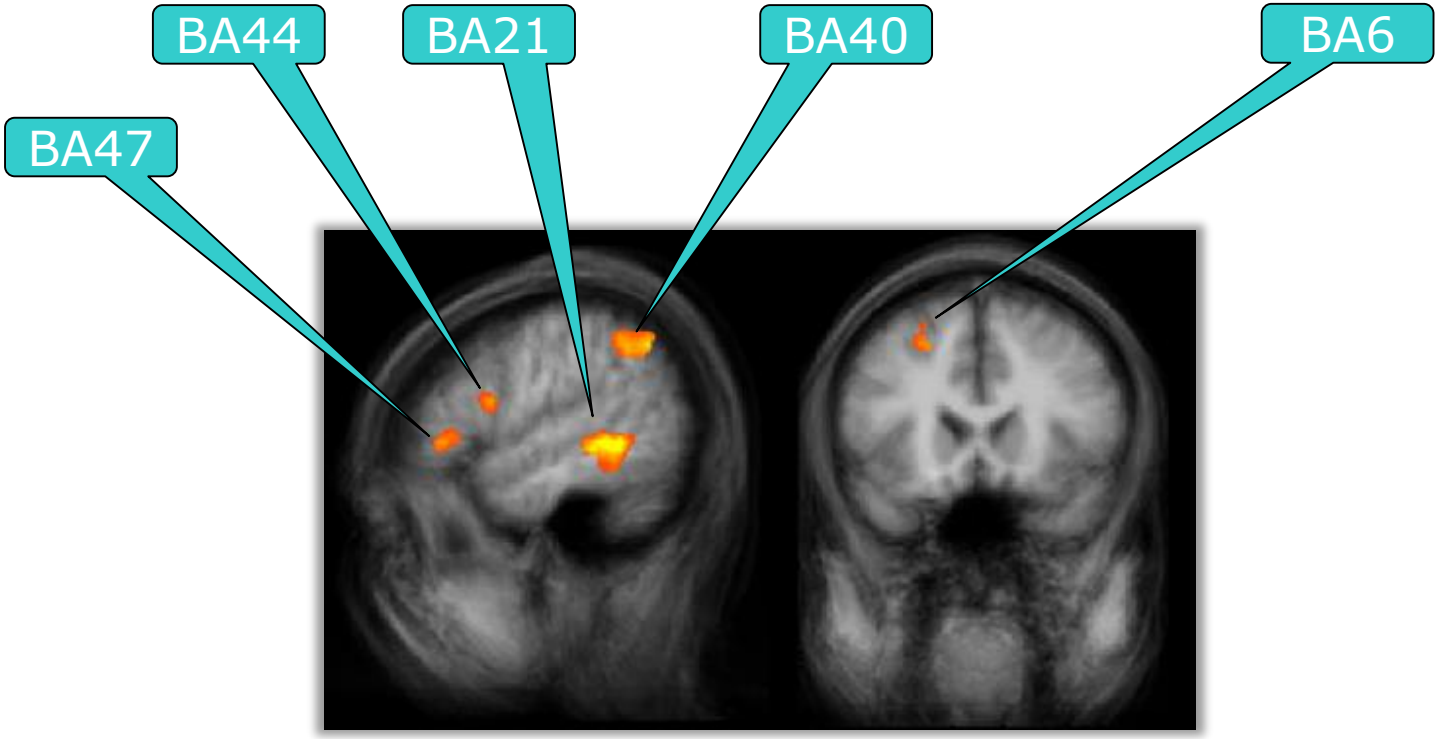
Syntax errors

- Conducted study with 11 participants, 14 sessions

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

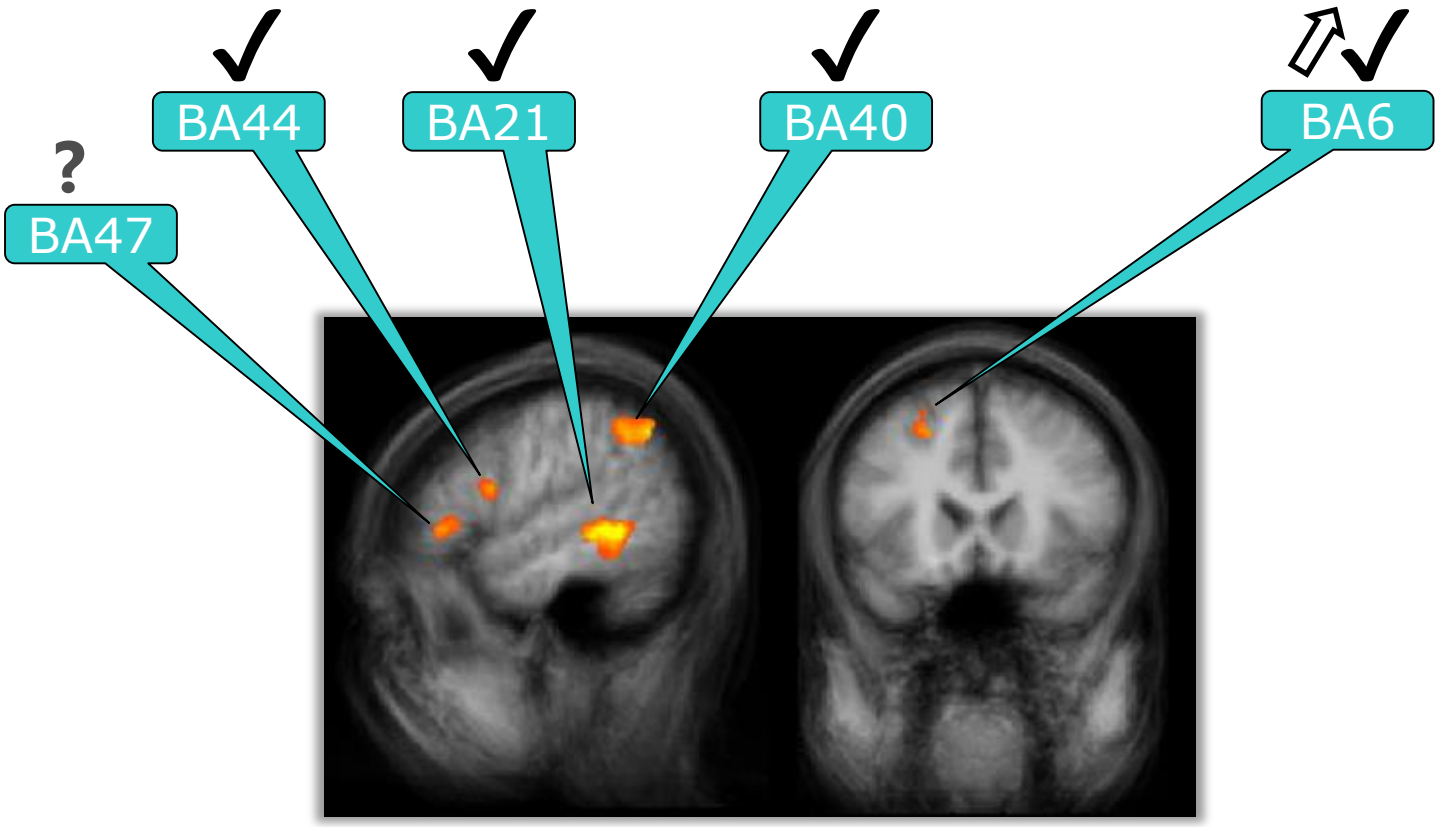
Results & Discussion

- RQ1: Can we replicate the results of the study by Siegmund and others [ICSE'14]?



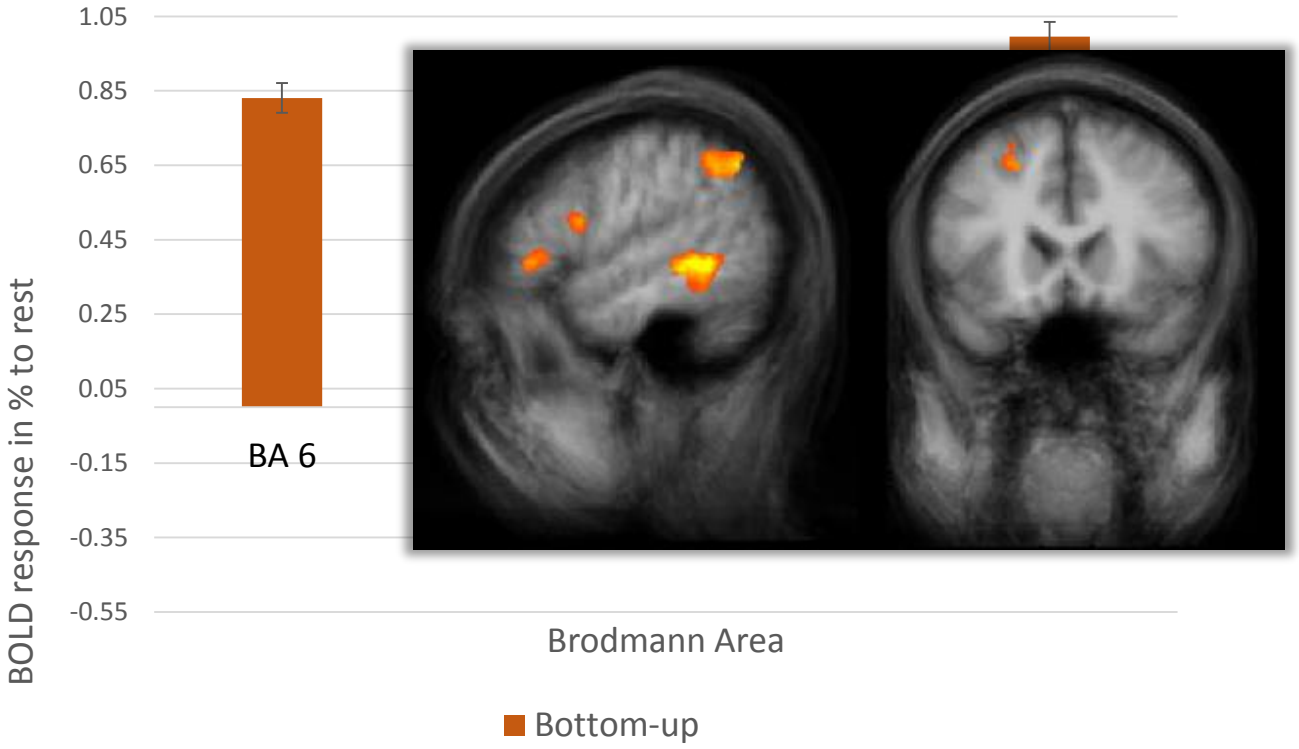
Results & Discussion

- RQ1: Can we replicate the results of the study by Siegmund and others [ICSE'14]?



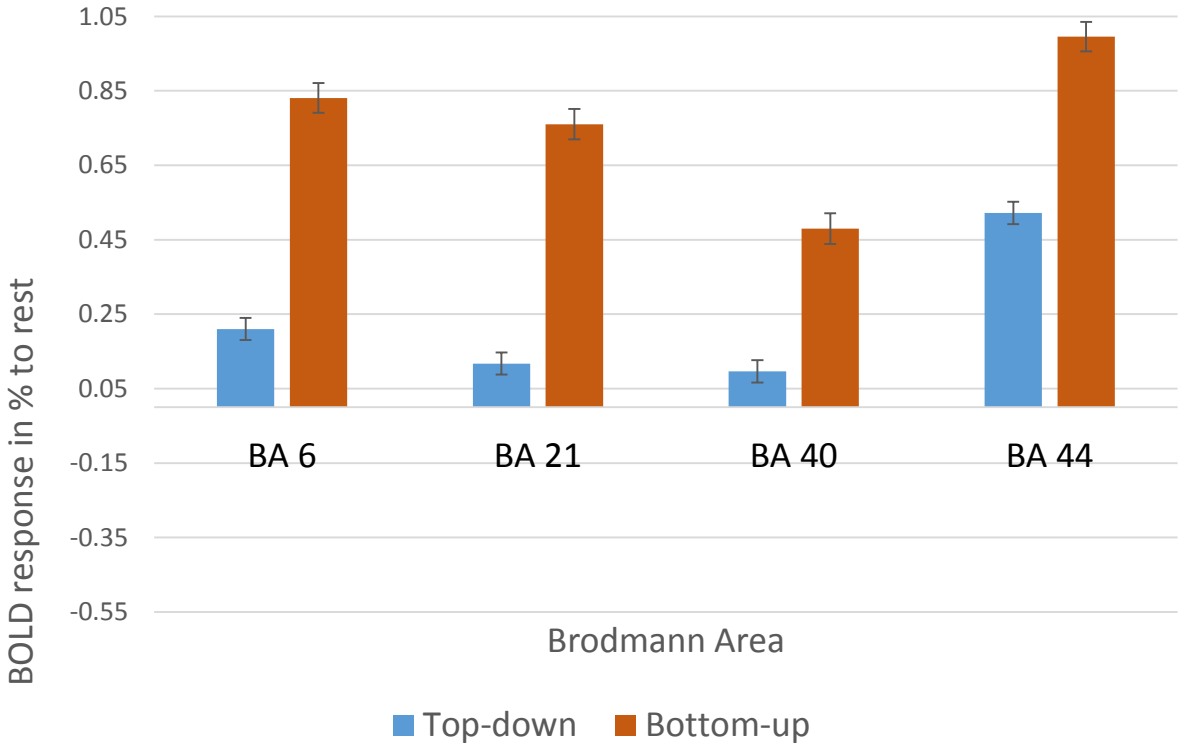
Results & Discussion

- RQ2: What is the difference between bottom-up program comprehension and top-down comprehension in terms of activation and the brain areas involved?



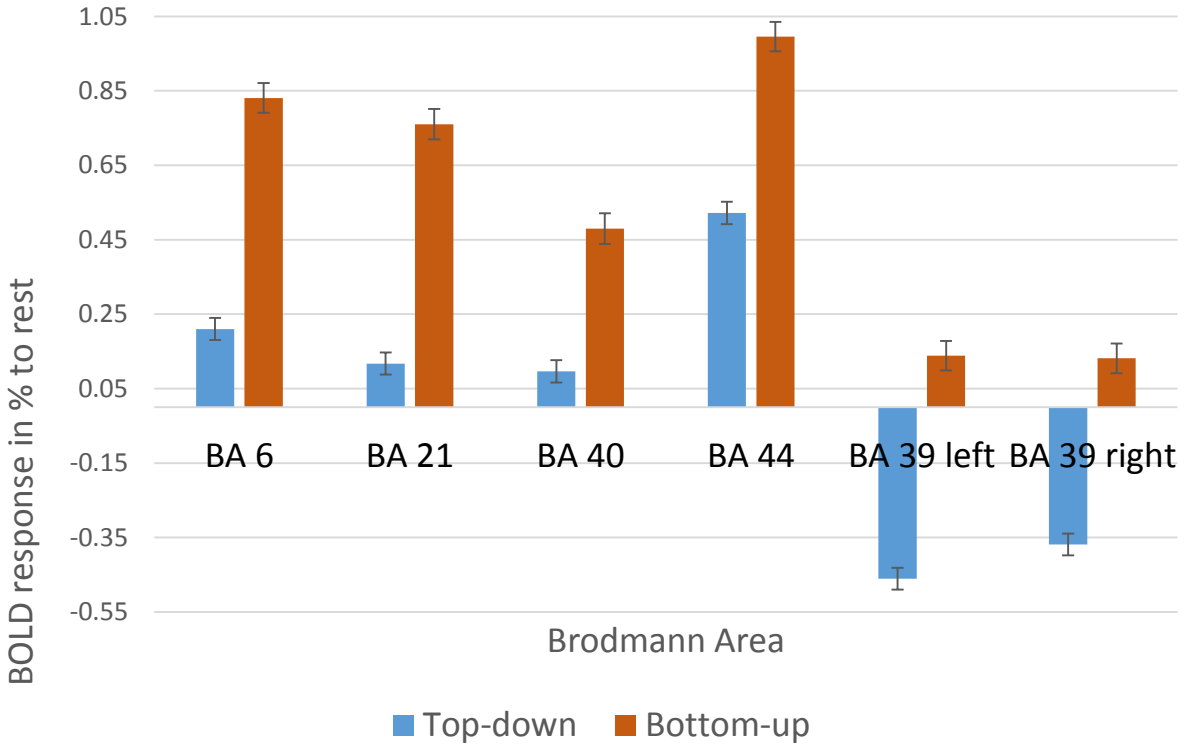
Results & Discussion

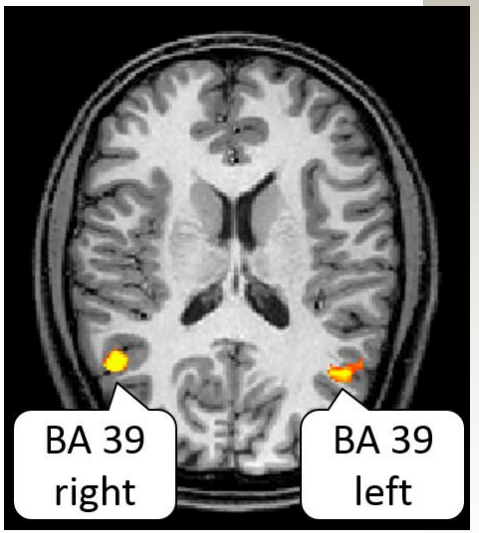
- RQ2: What is the difference between bottom-up program comprehension and top-down comprehension in terms of activation and the brain areas involved?



Results & Discussion

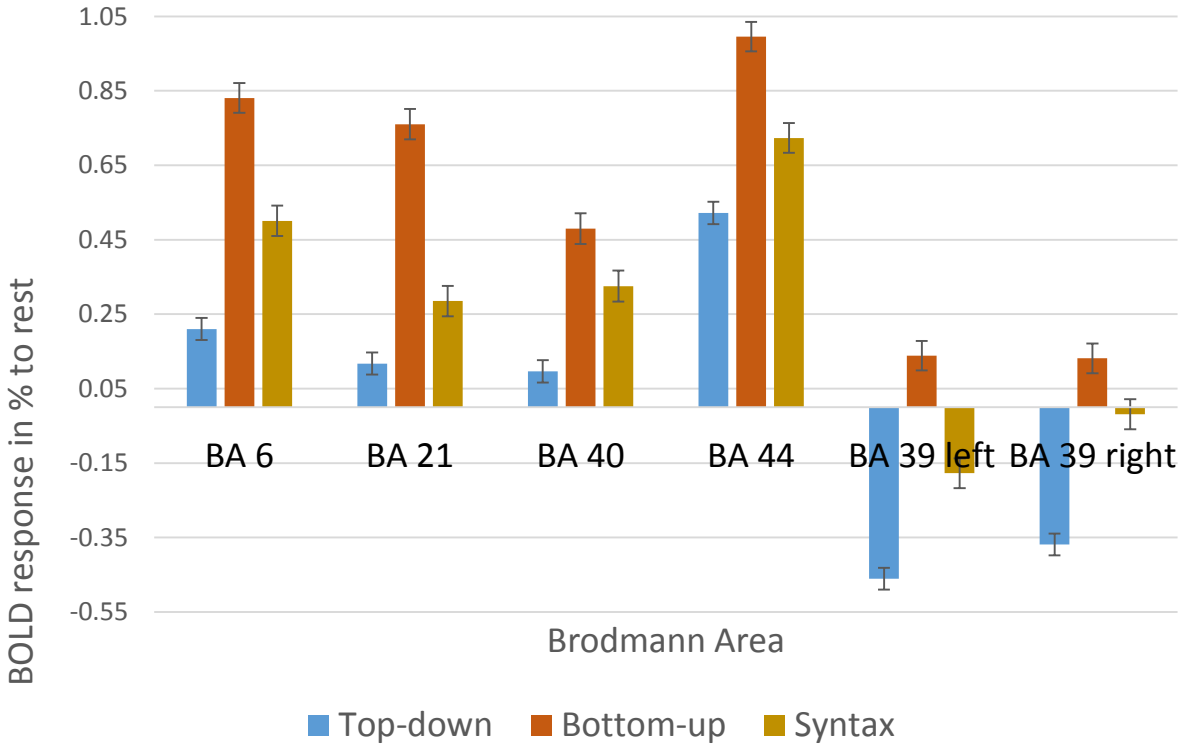
- RQ2: What is the difference between bottom-up program comprehension and top-down comprehension in terms of activation and the brain areas involved?



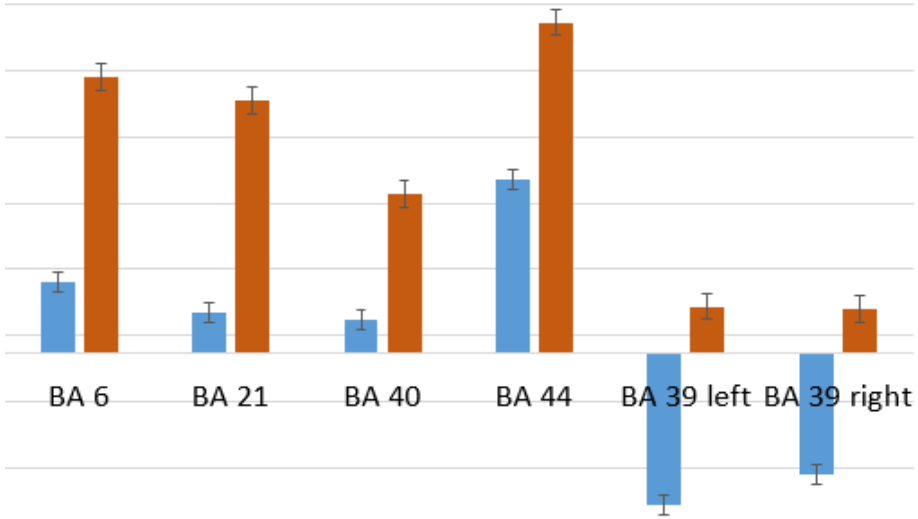


Results & Discussion

- RQ2: What is the difference between bottom-up program comprehension and top-down comprehension in terms of activation and the brain areas involved?



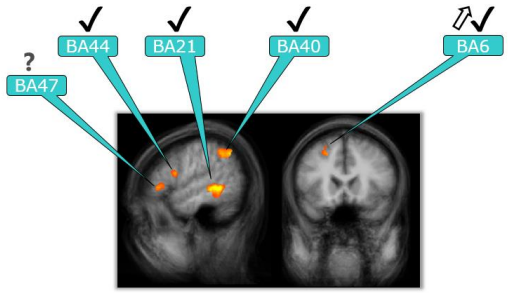
Program Comprehension Theories



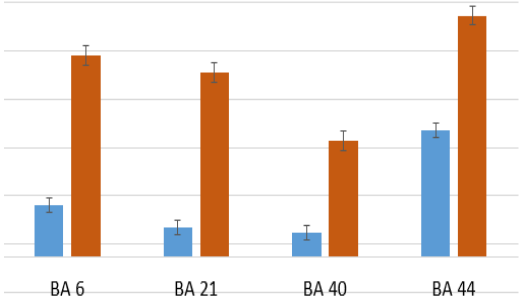
- Semantic Chunking ✓
  - Evidence for semantic chunking during bottom-up comprehension
- Neural Efficiency ✓
  - Increased neural efficiency for top-down comprehension

Conclusions

- **Replication** of bottom-up comprehension study **successful**



- Neuro-cognitive perspective **supports** existing program comprehension theories



- fMRI experiment framework needs further refinement
  - Research field is **wide open**

Backup Slides: Research Questions

- RQ1: Can we replicate the results of the study by Siegmund and others [ICSE14]?
- RQ2: What is the difference between bottom-up program comprehension and comprehension with semantic cues in terms of activation and the brain areas involved?
- RQ3: How do layout and beacons in source code influence program comprehension?

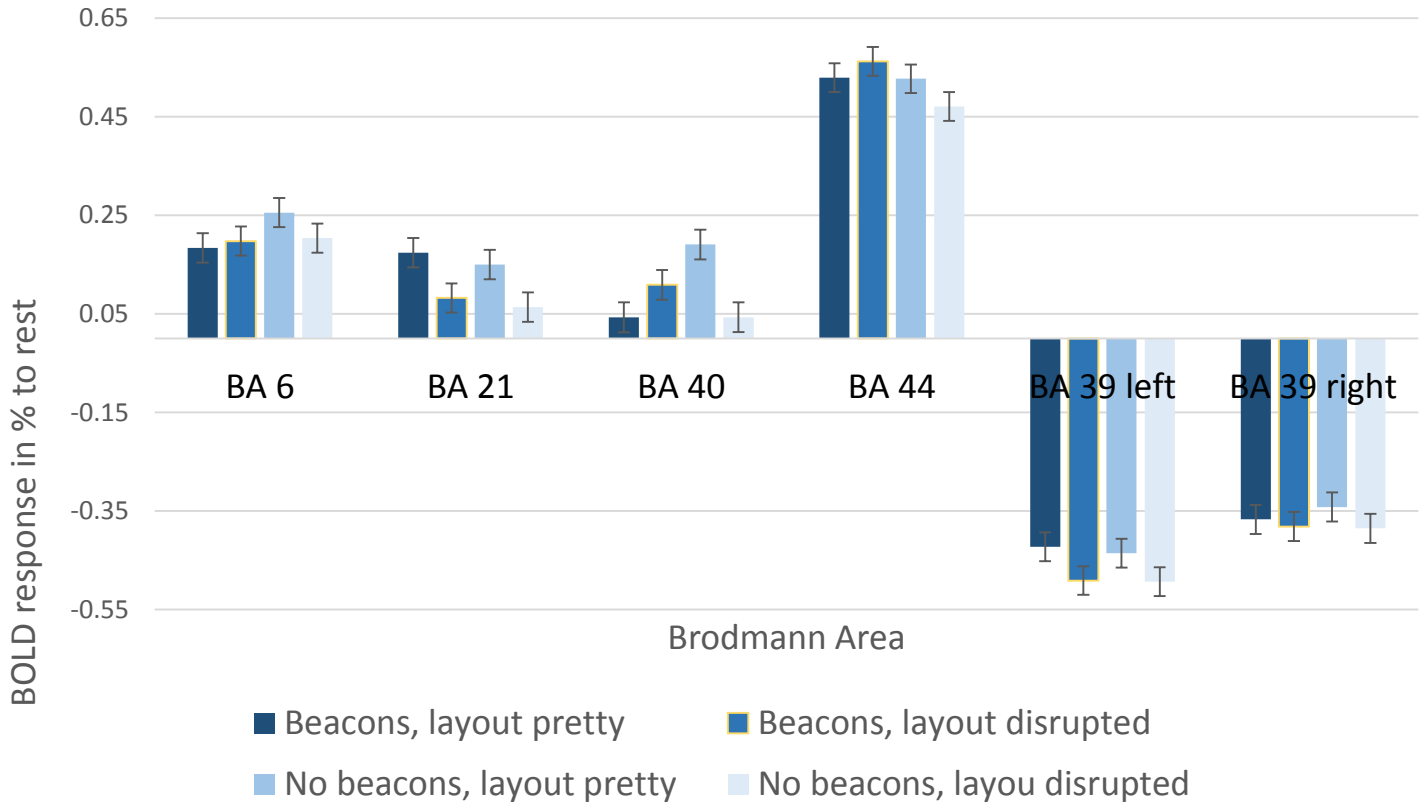
Backup Slides

- Snippets
  - Based on Siegmund and others [ICSE'14]
  - LOC: 8 to 18
  - All material is on our website

Semantic-cues	Bottom-up	Syntax
ArrayAverage	CommonChars	Average
BinaryToDecimal	**CrossSum**	DoubleArray
CrossSum	DoubleArray	**Power**
FirstAboveThreshold	**Factorial**	**ReverseIntArray**
Power	**MaxInArray**	**ReverseWord**
SquareRoot	SumUpToN	**Swap**
ContainsSubstrings		
CountSameCharsAtSamePosition		
CountVowels		
IntertwineTwoWords		
Palindrome		
ReverseWord		

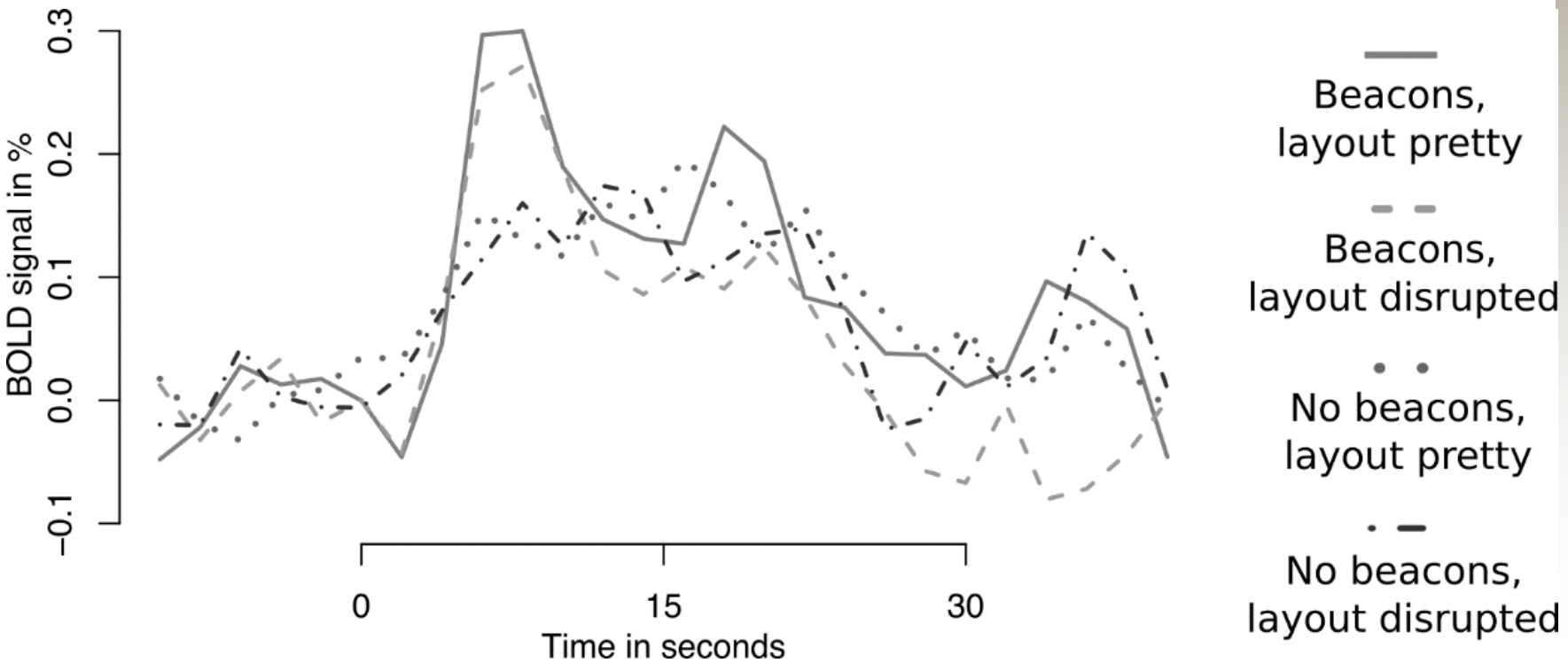
Backup Slides

- RQ3: How do layout and beacons in source code influence program comprehension?



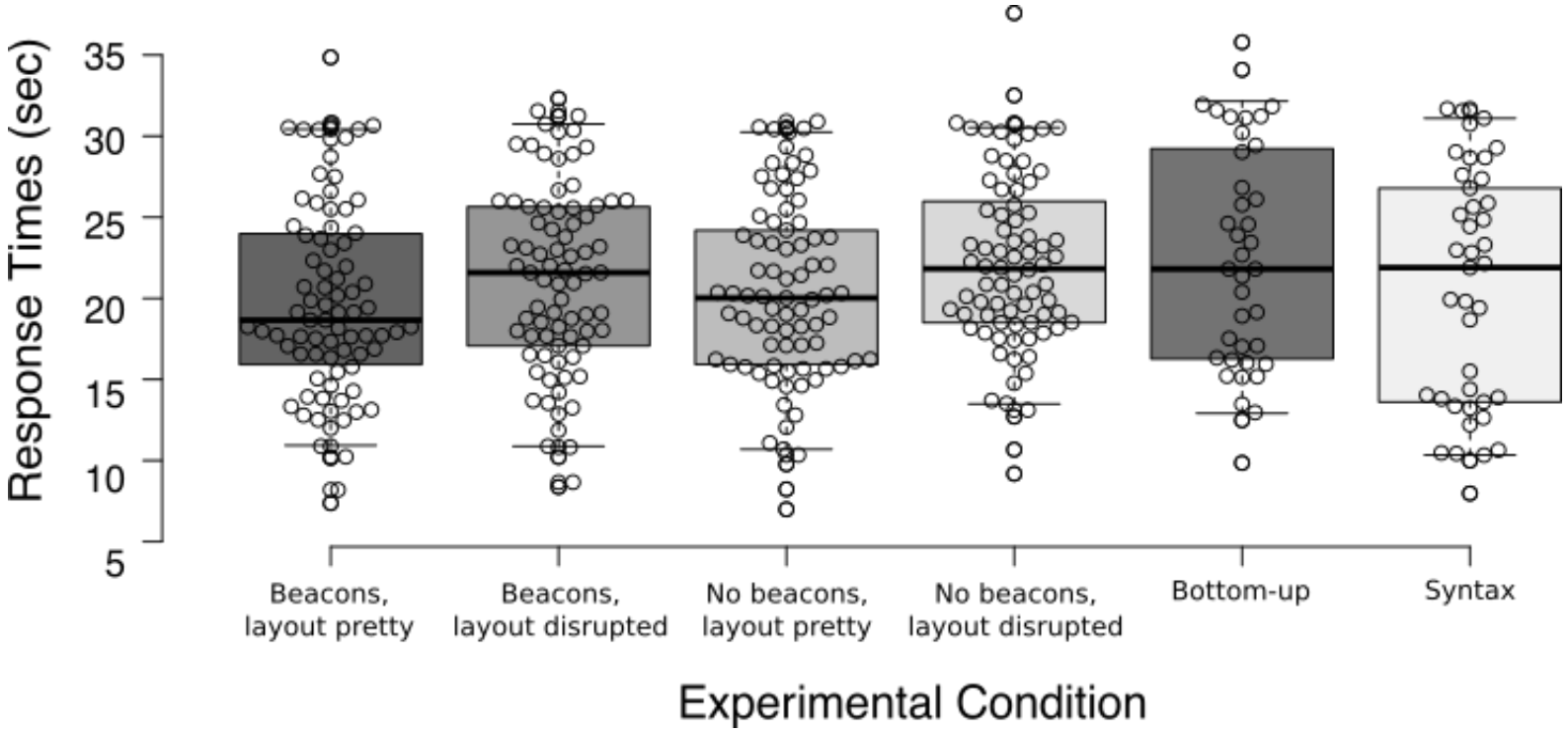
Backup Slides

- RQ3: How do layout and beacons in source code influence program comprehension?



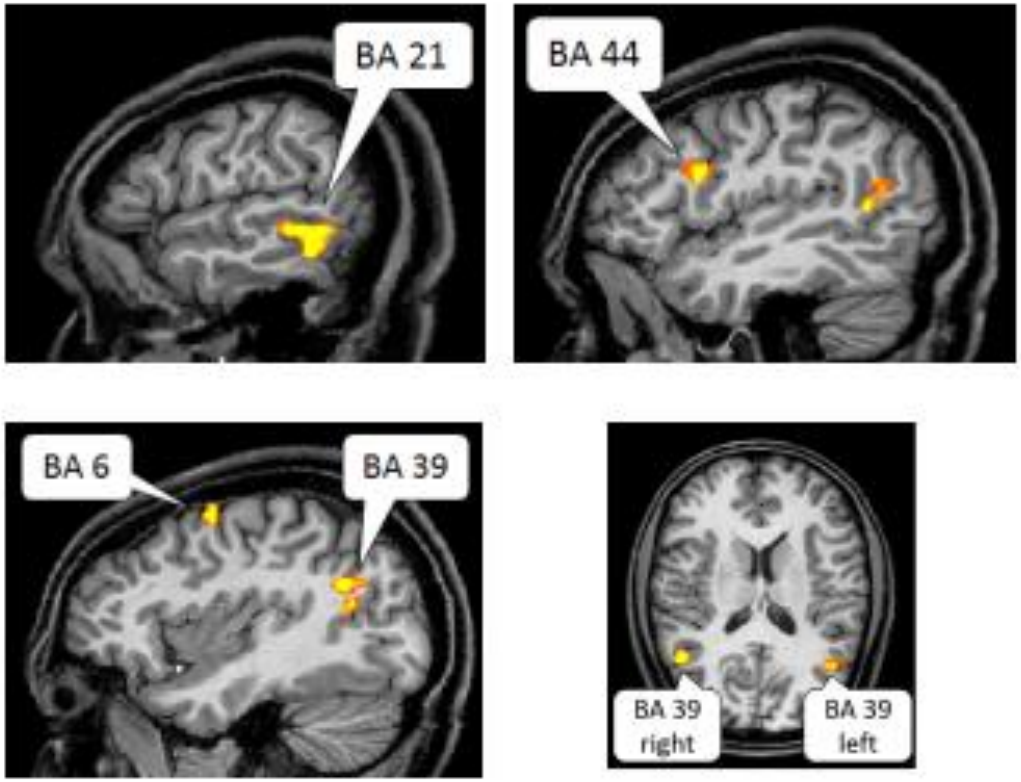
Backup Slides

- Response Times



Backup Slides

- Activation FSE17



Backup Slides

- 11 participants
  - 5 CS students
  - 3 graduate math students
  - 3 professional programmers
- 2f, 9m
- all were right-handed
- Age: 25.3 ± 3.82 years
- Programming experience: 1.7 ± 0.39 years