Program Comprehension Research with Neuroscience Methods

Norman Peitek
Leibniz Institute for Neurobiology Magdeburg

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

- 2009-2014: Master’s Degree in Business Information Systems
- 2014-2016: Software Consultant
- Since 2016: PhD Student

ULTRA TENDENCY
Pioneering the Future

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

Morning:

- Part I: Program Comprehension Research – The Past
- Part II: Neuroscience Basics (for SE)
- Part III: Program Comprehension Studies with fMRI

Afternoon:

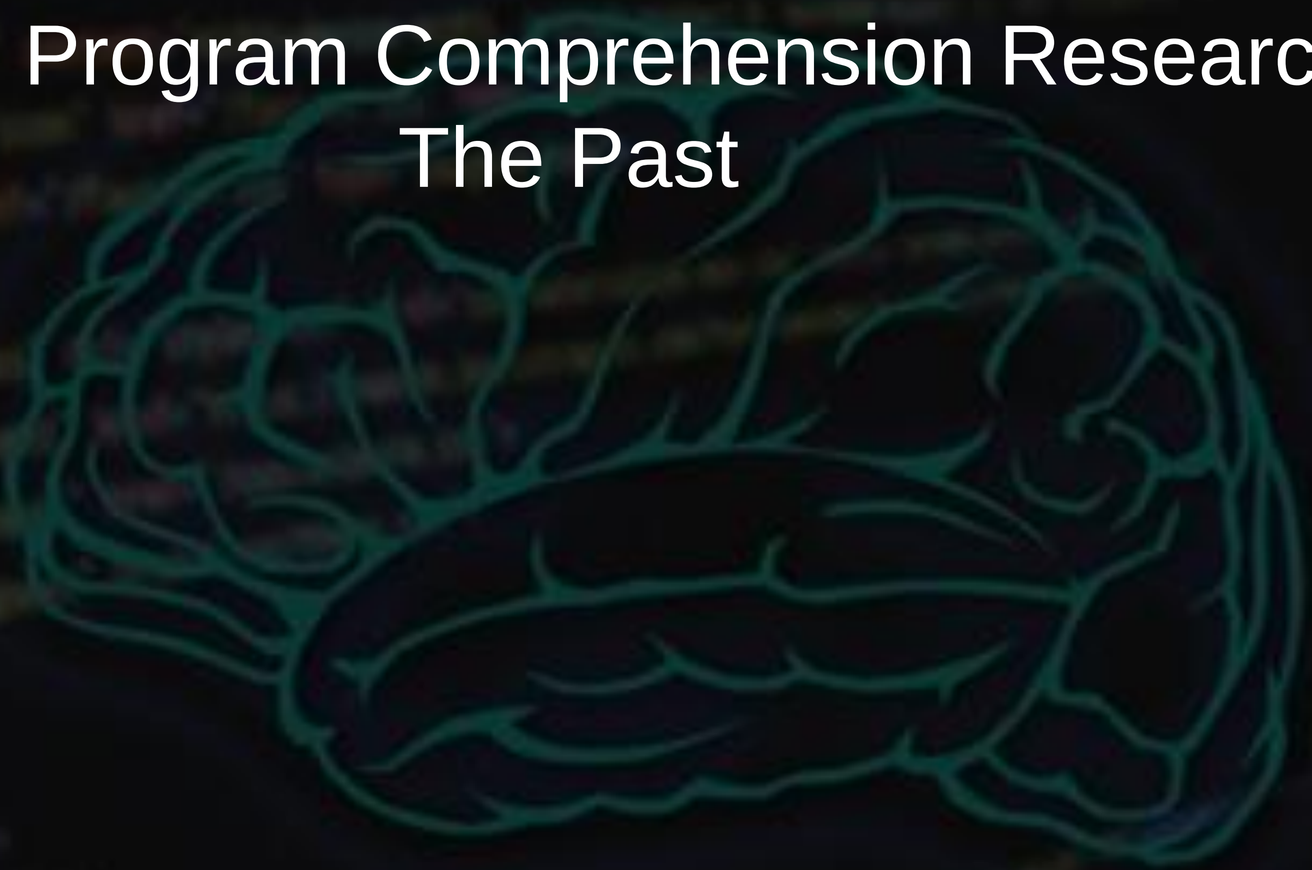
- Part IV: Designing an Experiment

LIN LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

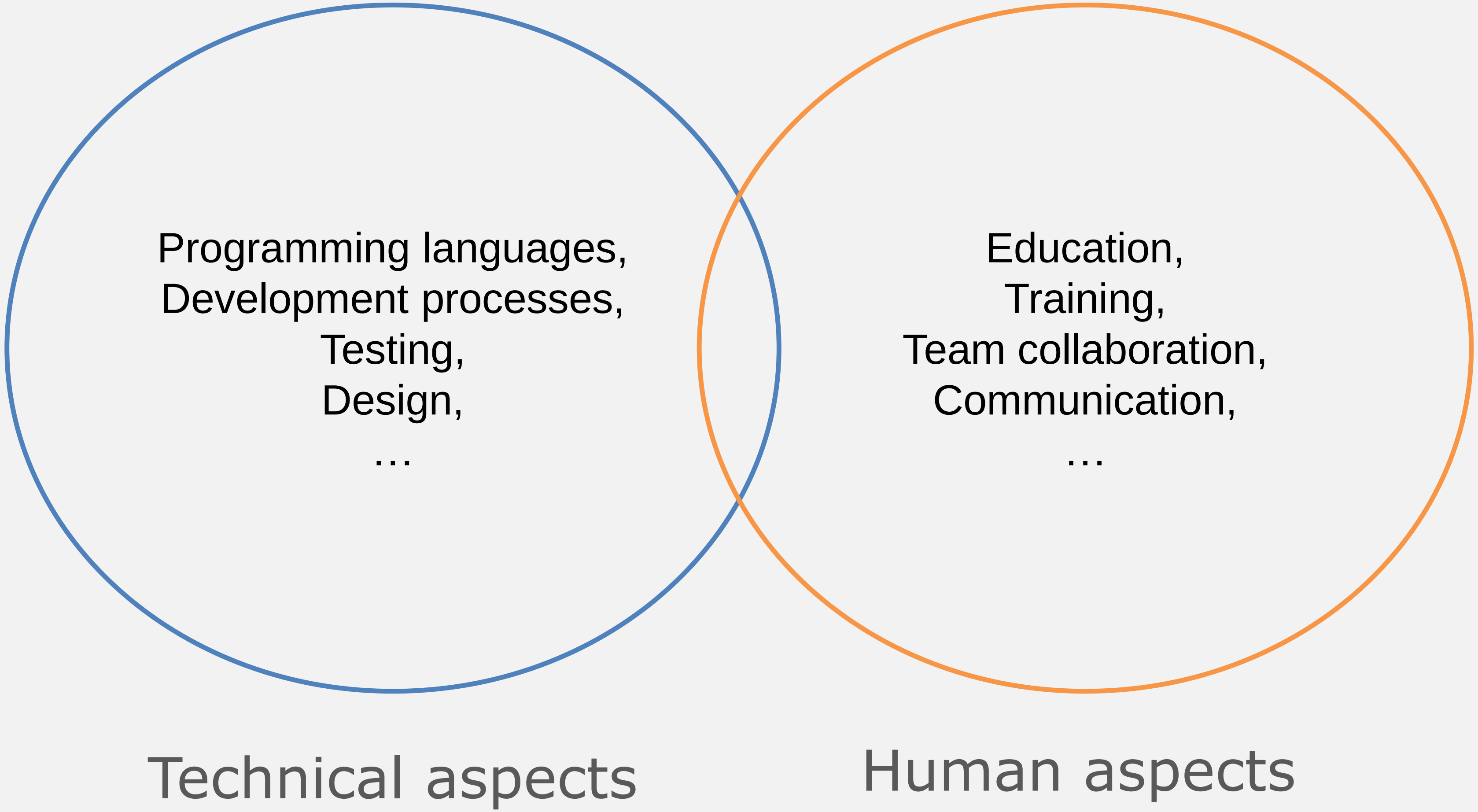
TECHNISCHE UNIVERSITÄT CHEMNITZ

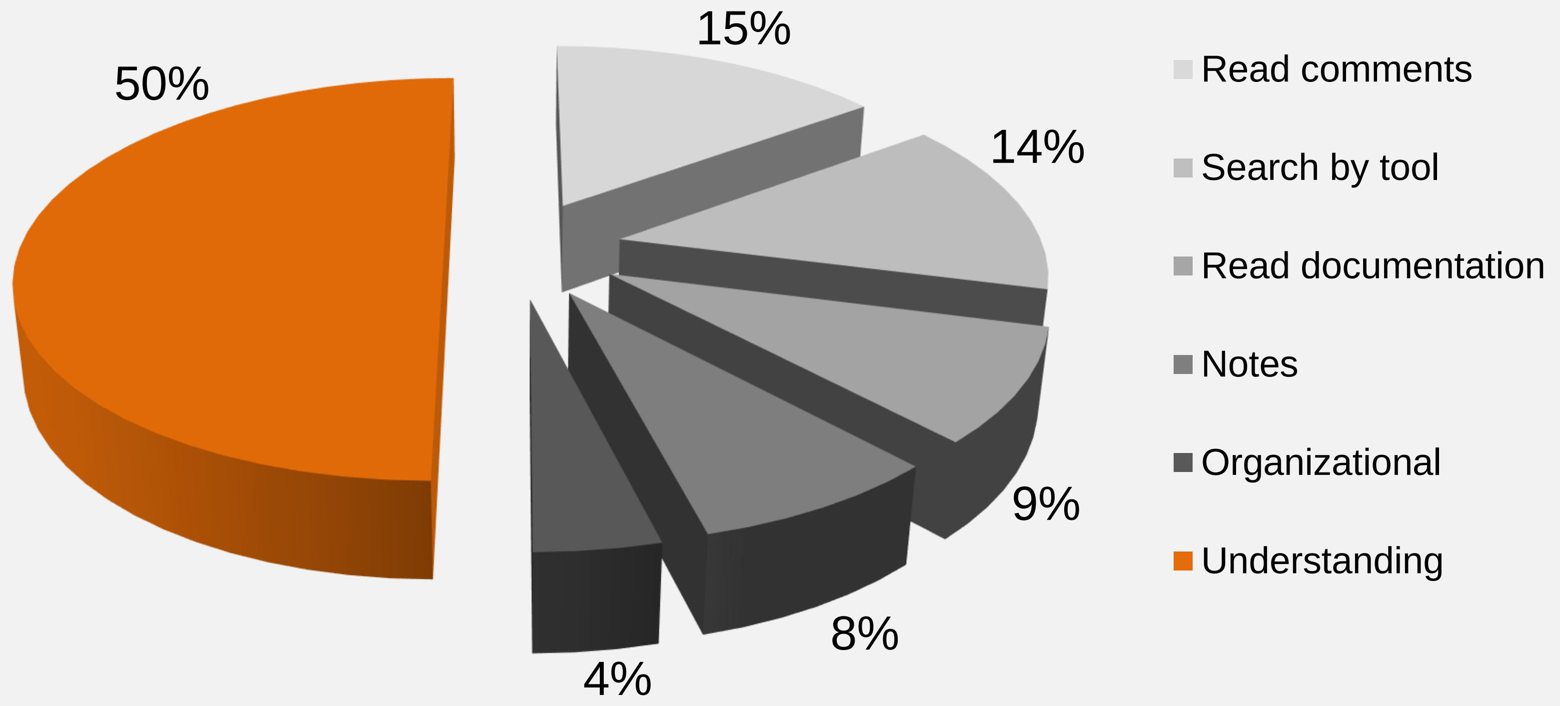
- Understand potentials, limitations, and challenges of empirical research with neuroscientific methods
- Gain familiarity with neuroscience
- Obtain insights into a current SE research area
- Differentiate between typical neuroscience and SE experiments

Part I: Program Comprehension Research
The Past



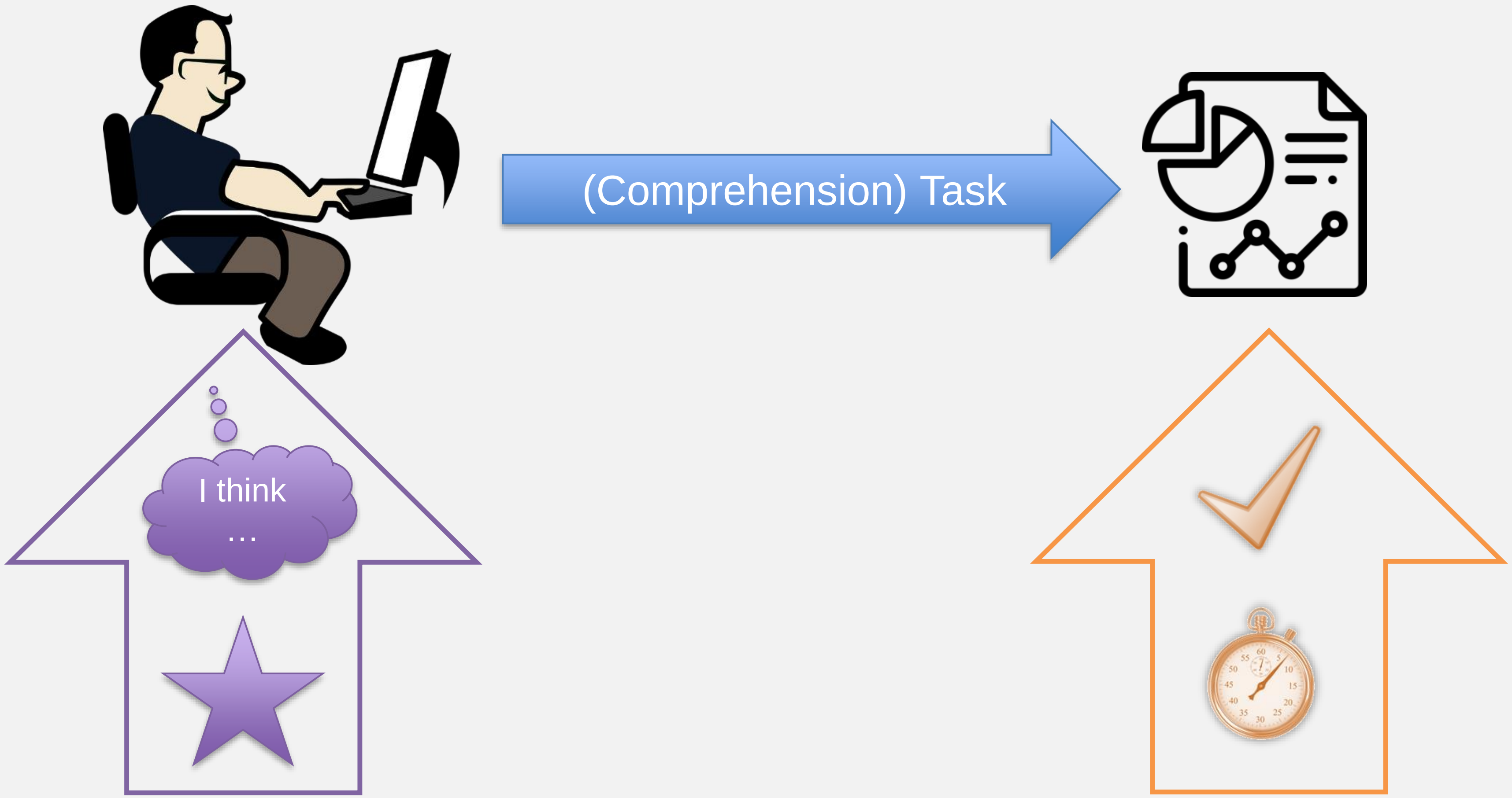
- Software engineering




15%
50%
14%
9%
8%
4%
Read comments
Search by tool
Read documentation
Notes
Organizational
Understanding

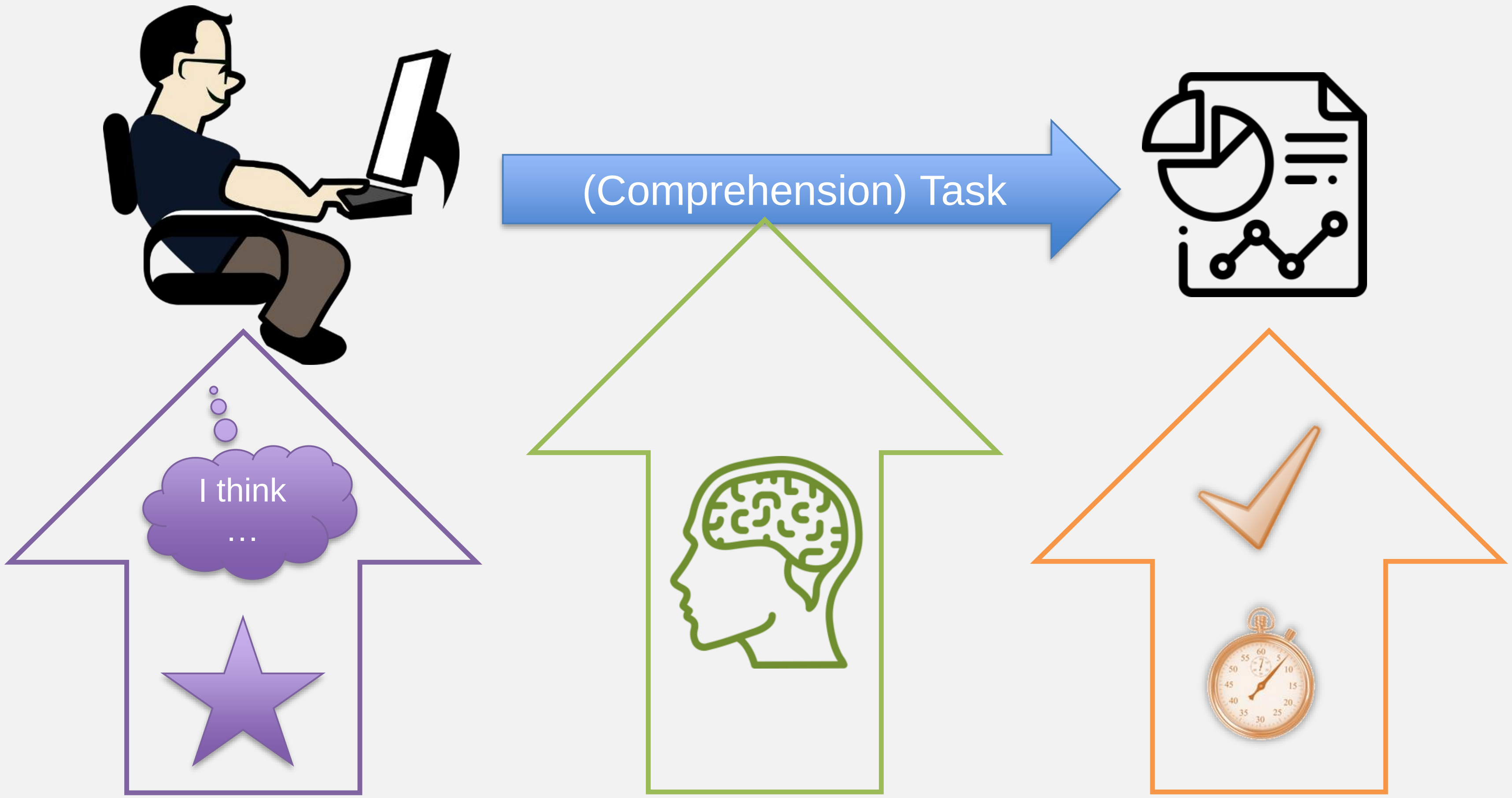
[Tiarks11]

> Ice age of research on programmers



Measure
subjective rating

Measure
task performance

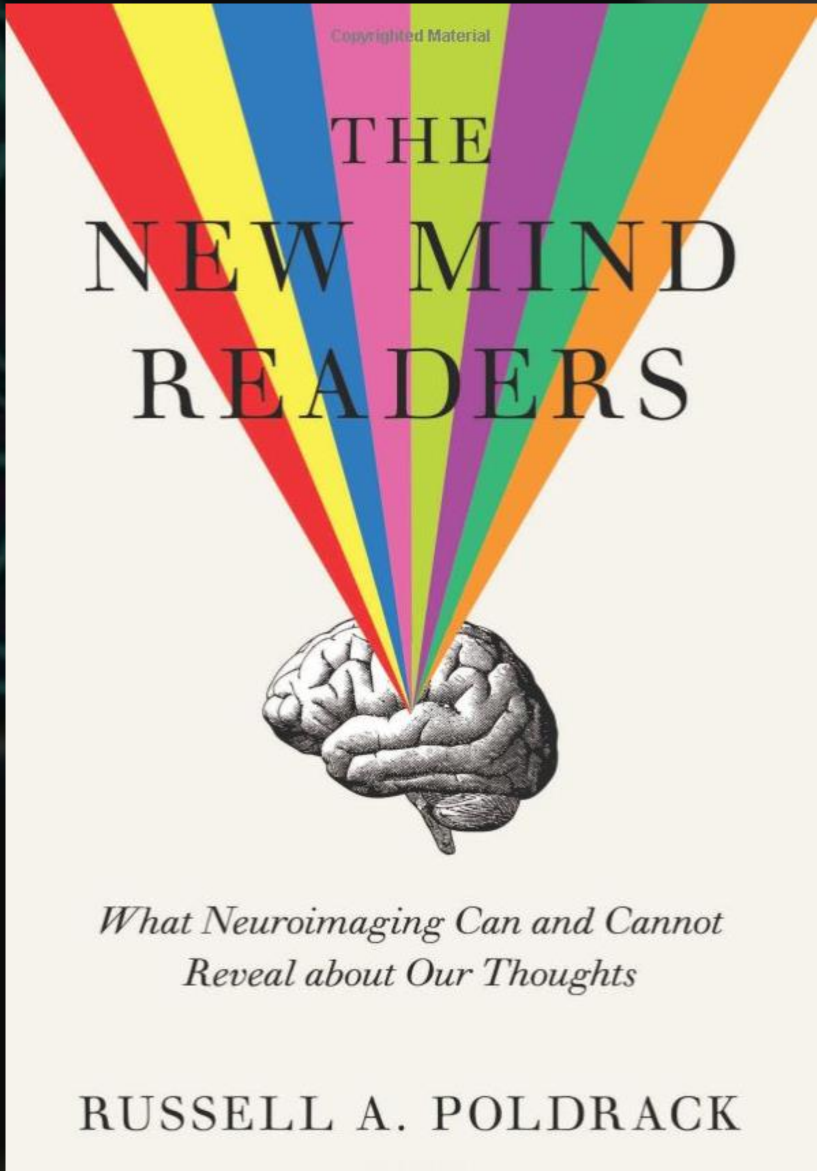


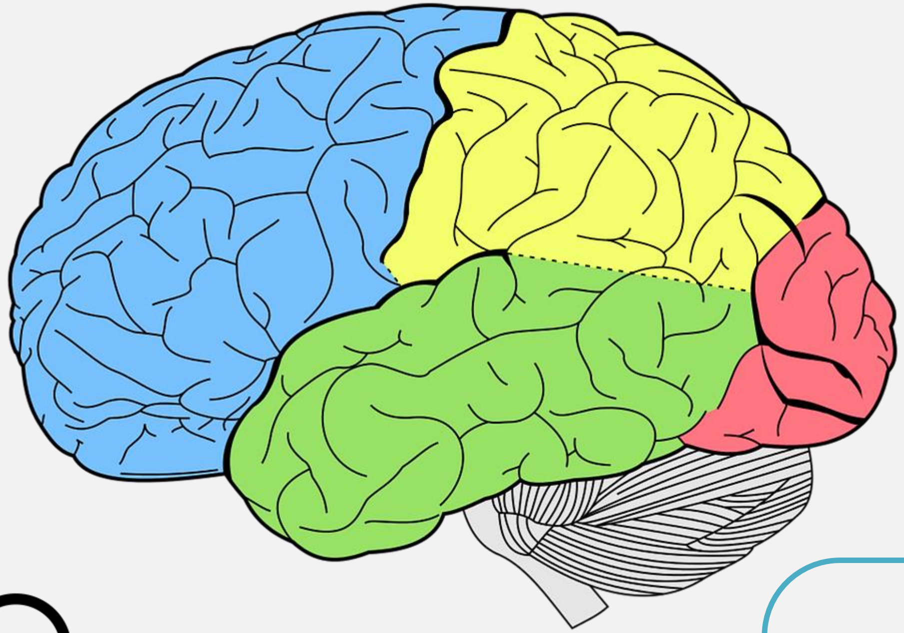
Measure
Subjective rating

Measure with
Neuroscience?

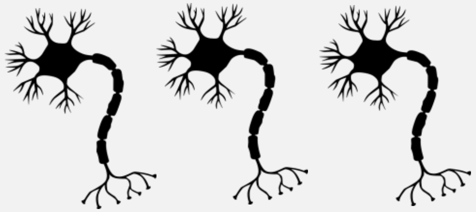
Measure
Task performance

Part II: Neuroscience Basics (for SE)





Micro imaging

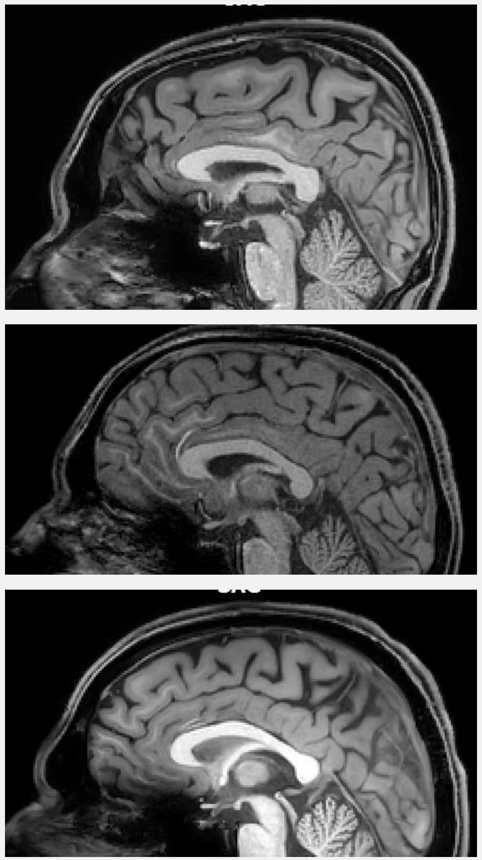


Macro imaging

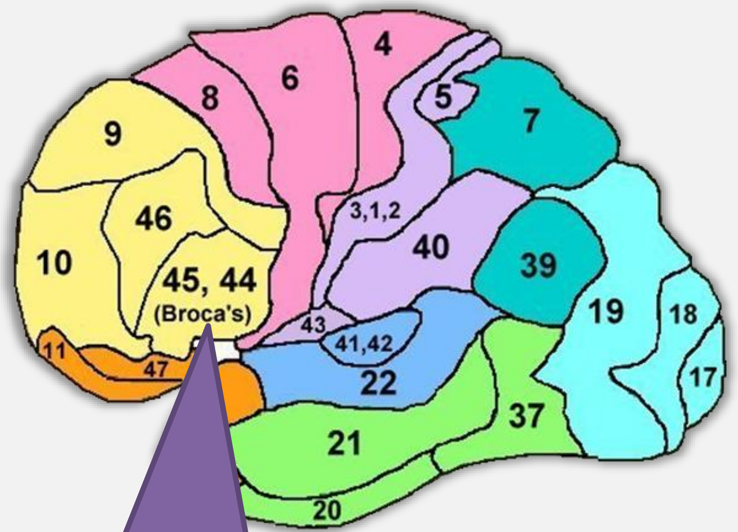
Overall cognitive load

Brain area activation

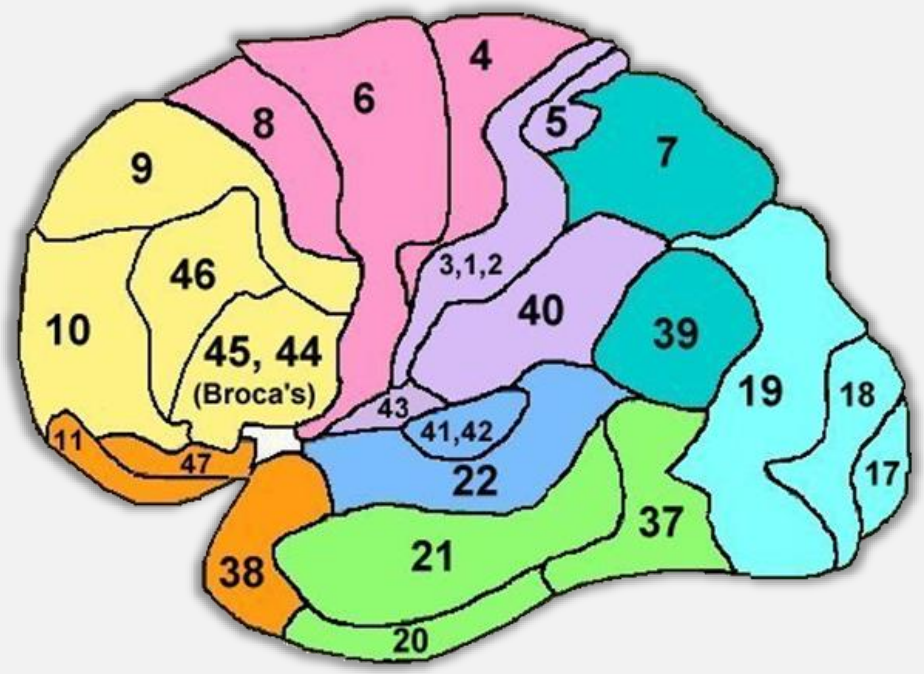
We all have different brains!

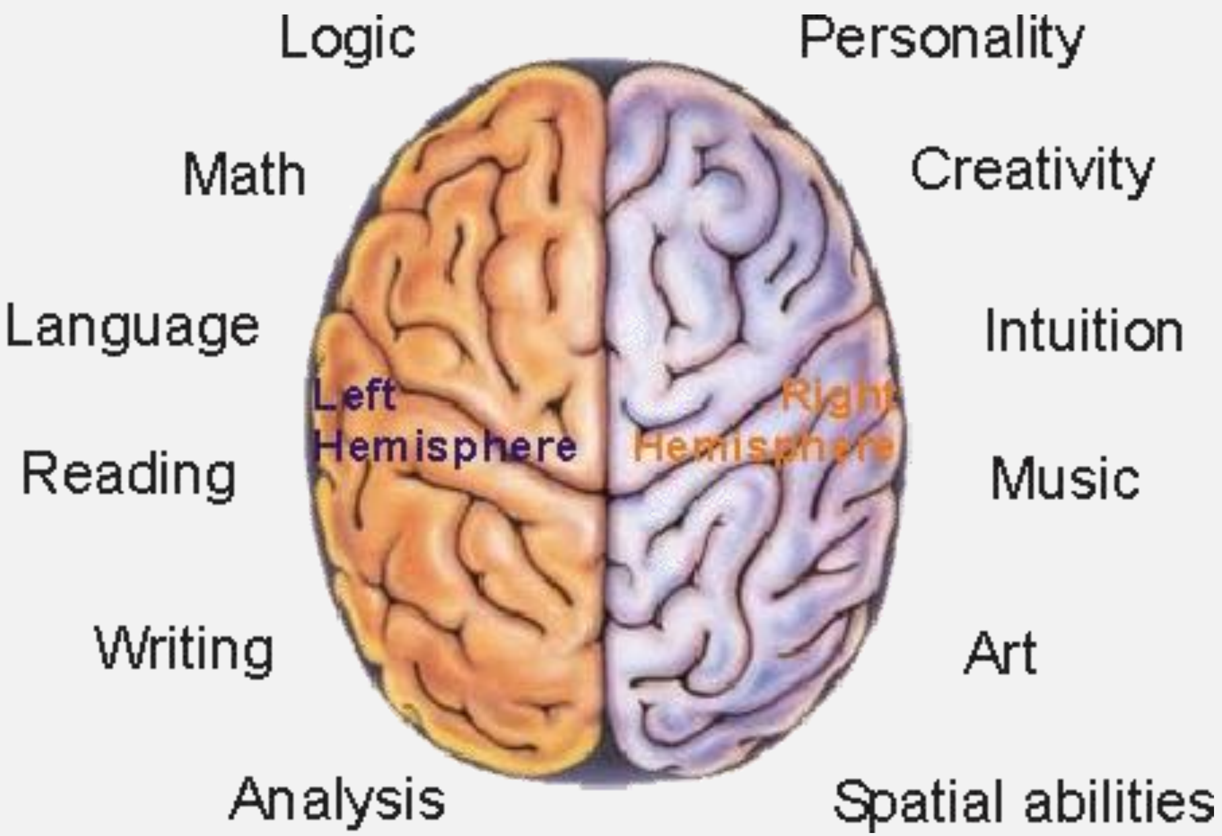


Standardization: Talairach space



Brodmann areas (BA)


4
6
8
5
7
9
46
3,1,2
40
10
39
45, 44
(Broca's)
19
18
43
41,42
11
47
22
17
37
21
38
20


Logic
Personality
Math
Creativity
Language
Intuition
Left Hemisphere
Right Hemisphere
Reading
Music
Writing
Art
Analysis
Spatial abilities

LIN LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

n=2

Position	0	1	2	3	4	5	6
Number							

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

n=2

Position	0	1	2	3	4	5	6
Number	4						

n=2

Position	0	1	2	3	4	5	6
Number		1					

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

n=2

Position	0	1	2	3	4	5	6
Number			9				

n=2

Position	0	1	2	3	4	5	6
Number				8			

LIN LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

n=2

Position	0	1	2	3	4	5	6
Number					0		

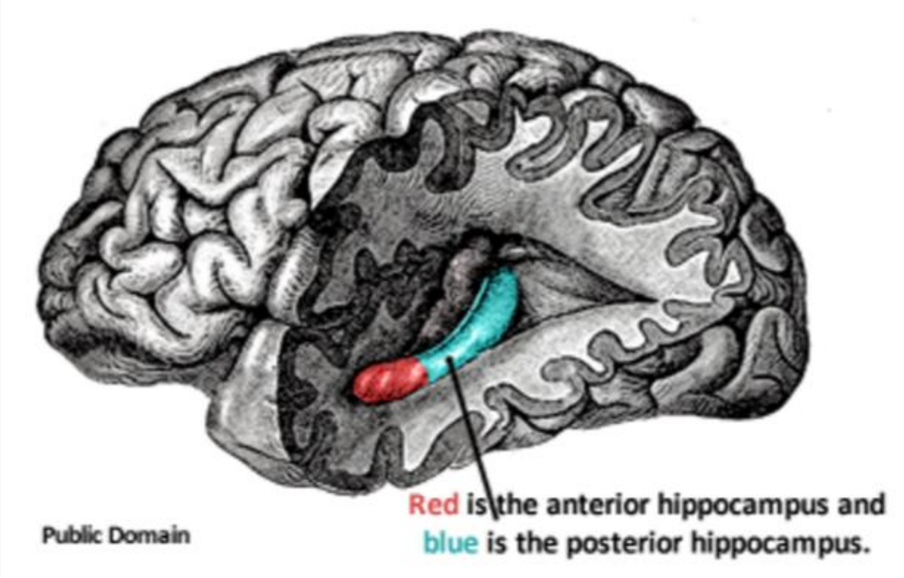
n=3

Position	0	1	2	3	4	5	6
Number					0		

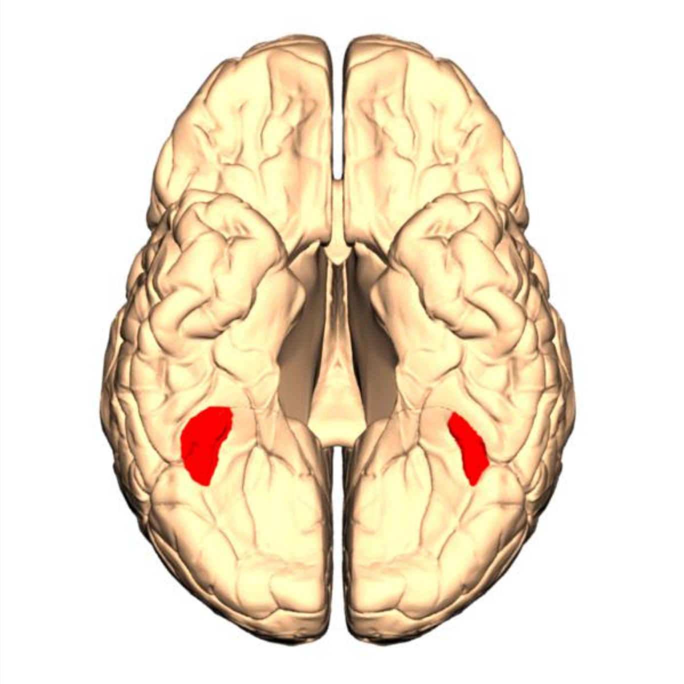
n=4

Position	0	1	2	3	4	5	6
Number					0		

New *software* or new *hardware*?



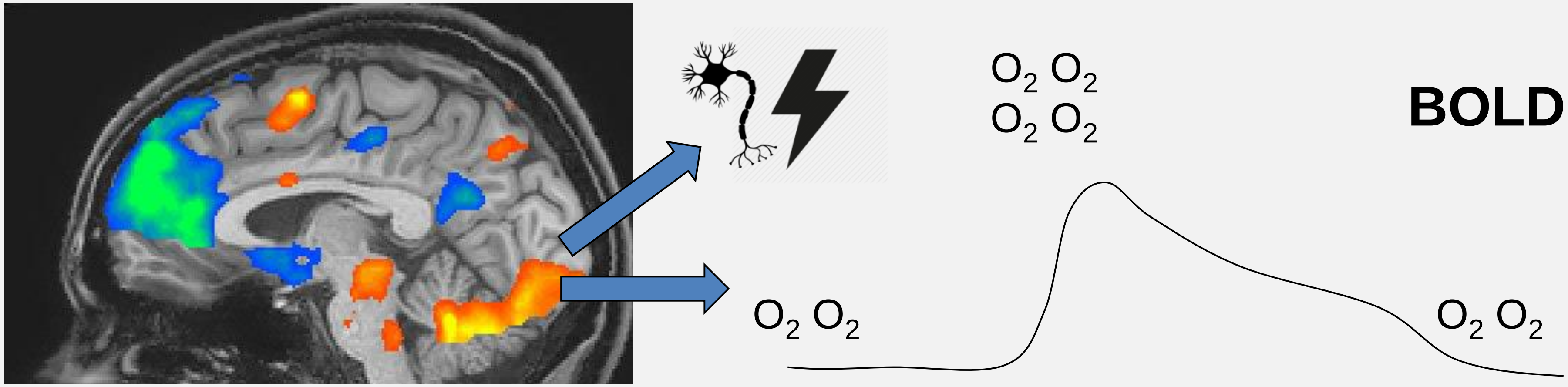
Our brain can physically adapt to our activities.



Our brain can learn to efficiently process information.

Fusiform face area

Position	0	1	2	3	4	5	6
Number			9				



BOLD effect

(functional) magnetic resonance imaging

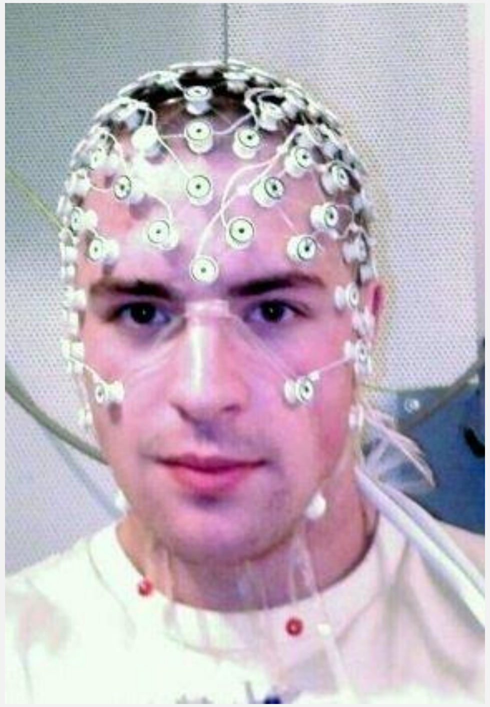
fMRI

functional near-infrared spectroscopy

fNIRS

Electroencephalogram

EEG



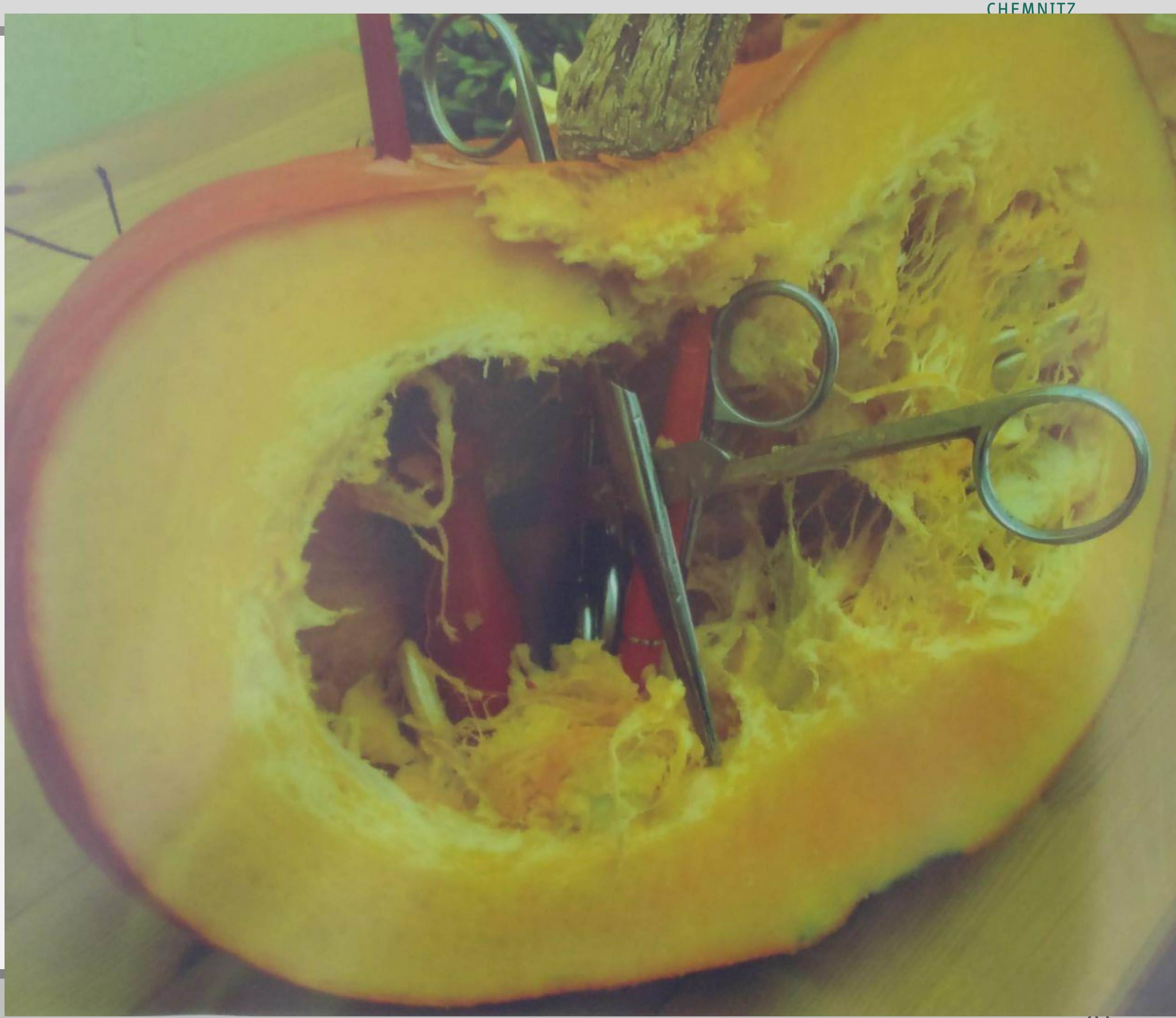
	fMRI	fNIRS	EEG
Based on/Measures	BOLD	BOLD	Electrical activity
Temporal resolution	Slow (1-2 sec), delayed	Slow (1-2 sec), delayed	Very fast (<100ms)
Spatial resolution	Good (full brain, 3D)	Weak	Very weak
Costs	~200 Euro/hour	~75 Euro/hour	~35 Euro/hour
Limitations	Many	Some	Some

→ Chose modality depending on the research question!

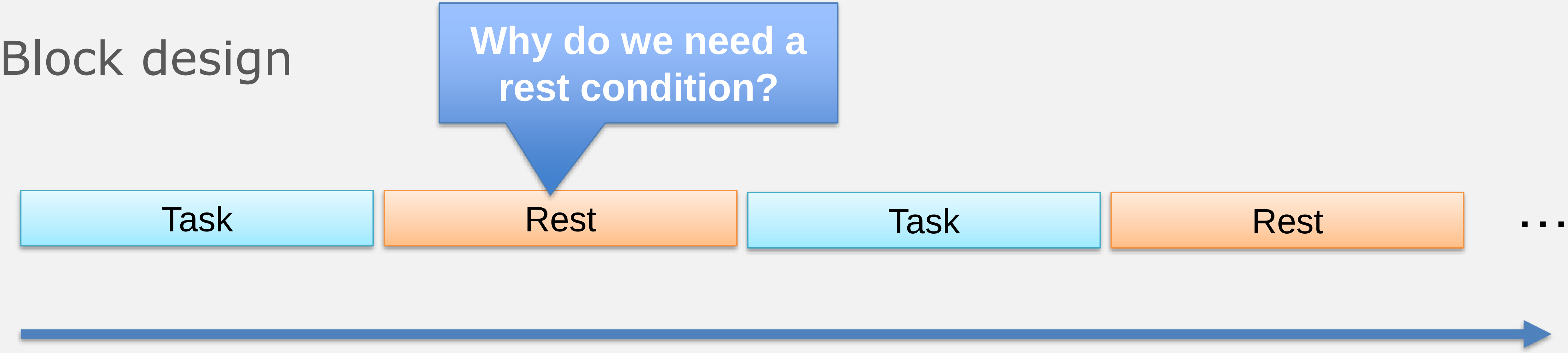
- Technical Limitations
  - Extremely strong magnetic field (1 to 7 Tesla)
    - For reference, magnetic field of the earth is 0.00005 Tesla
  - No paramagnetic/electrical devices
  - Small screen
- Many participant limitations
  - Claustrophobia
  - But also: no tattoos

- Technical Limitations
  - Extremely strong magnetic field (1 to 7 Tesla)
    - For reference, magnetic field of the earth is 0.00005 Tesla
  - No paramagnetic/electrical devices
  - Small screen

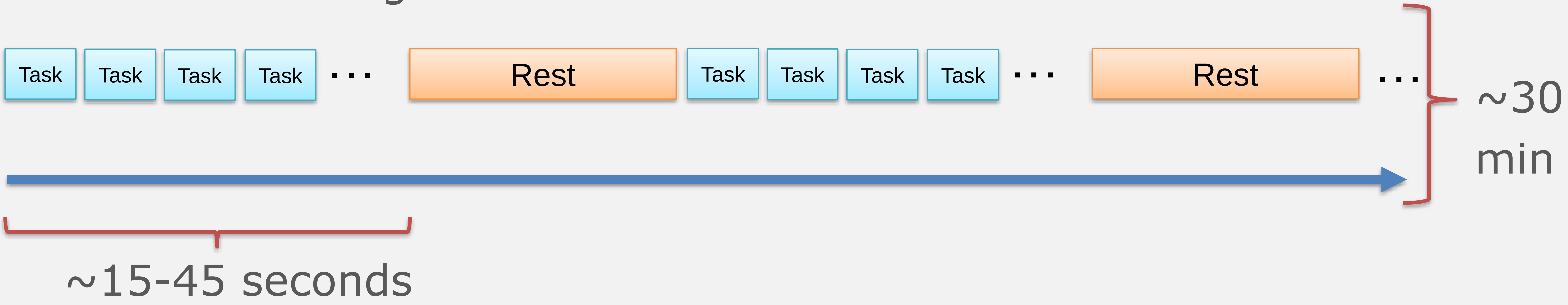
- Many participant limitations
  - Claustrophobia
  - But also: no tattoos



- Block design

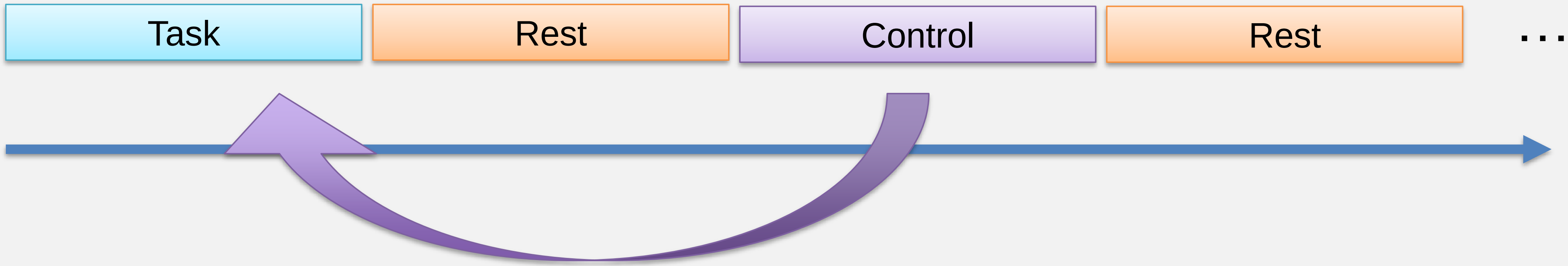


- Event-related design



LIN LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG
TECHNISCHE UNIVERSITÄT CHEMNITZ

- Experiment design with “control” condition

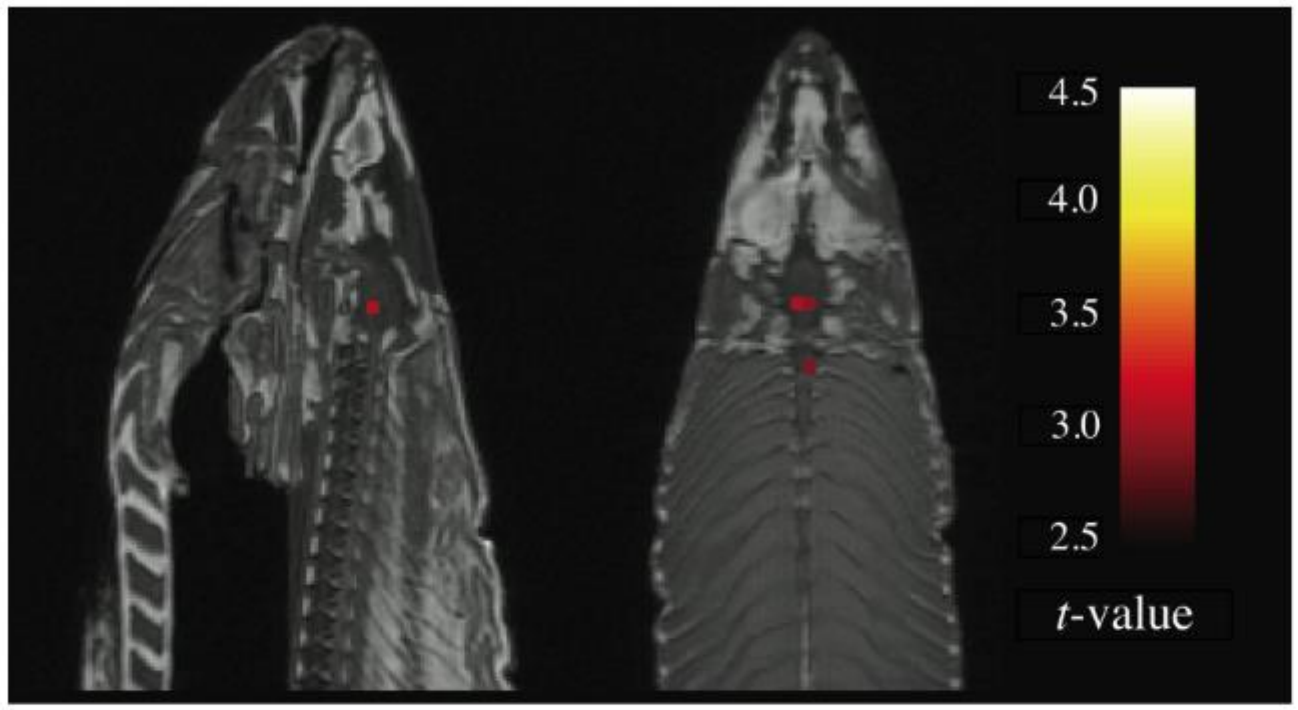


Similar task without cognitive process to be studied!

	Conventional Experiment	**fMRI Experiment**
Observed variable	Time to complete task	+ brain activation
Data points per task	1	~100’000/sec
Data	Single point	Continuous time series
Data points in 30min experiment	30	~180’000’000
Typical statistical test	T-test	GLM
Typical threshold	$p < 0.05$	$p < 0.05$, *FDR corrected*

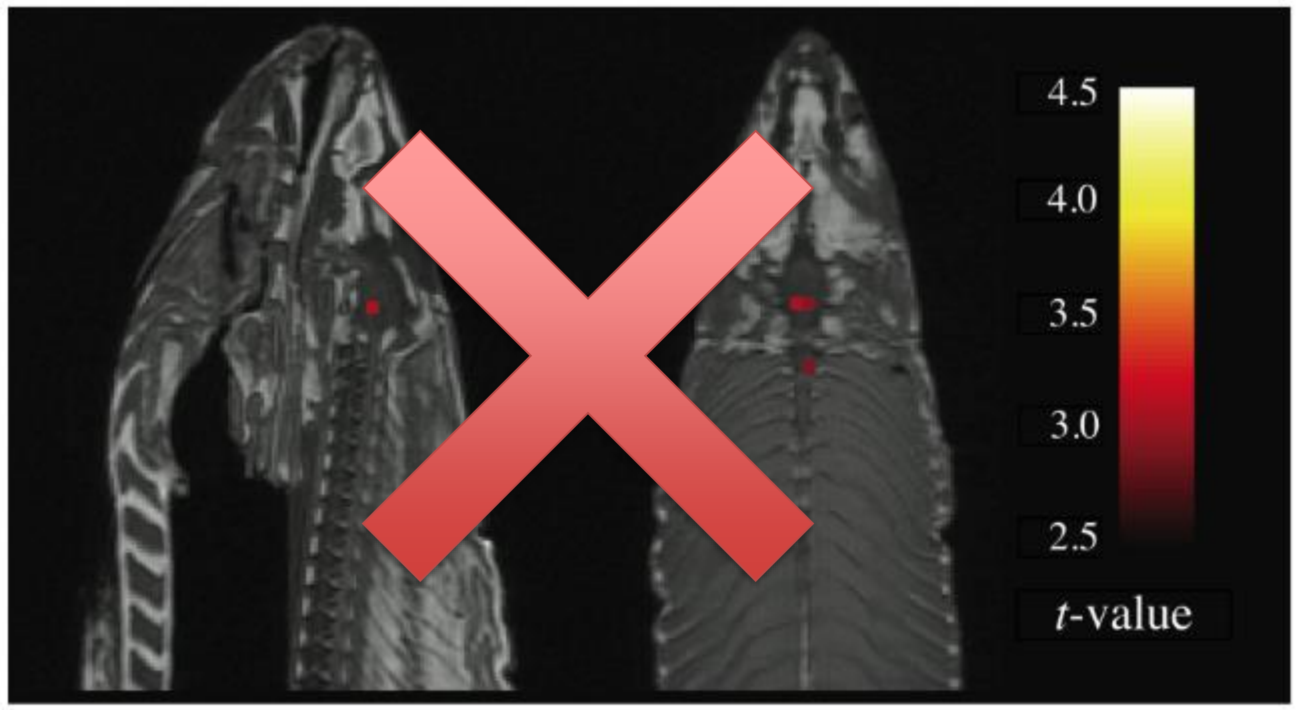
Bennet et al.

"Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon"



Bennet et al.

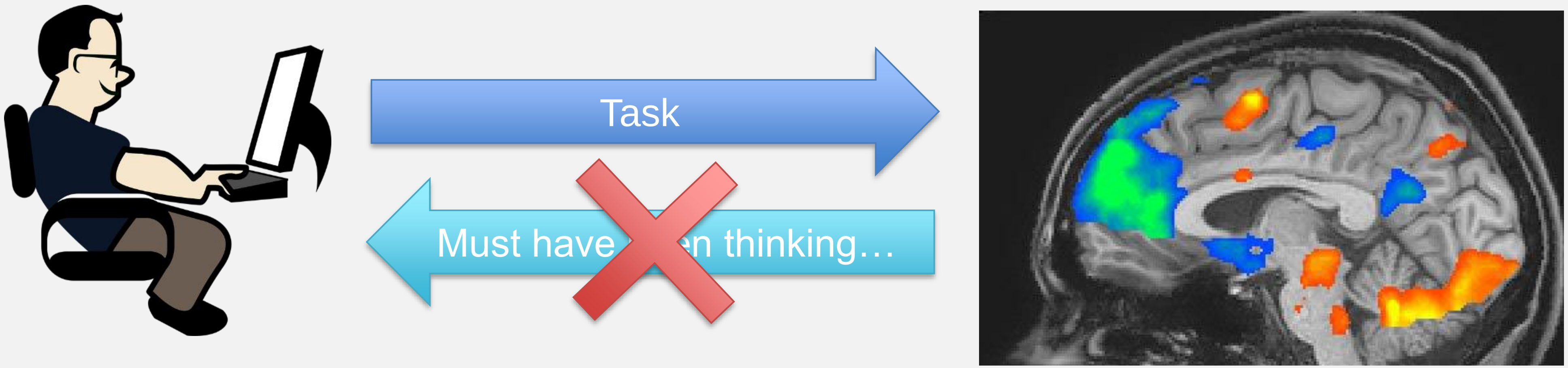
“Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: **An argument for multiple comparisons correction**”



When the analysis was controlled for false discovery rate (FDR), there was no active voxels left, even at $p < 0.25$.

Should we use super conservative thresholds: $p < 0.001$?

> Watch out for reverse inference



Part III: Program Comprehension Studies with fMRI

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

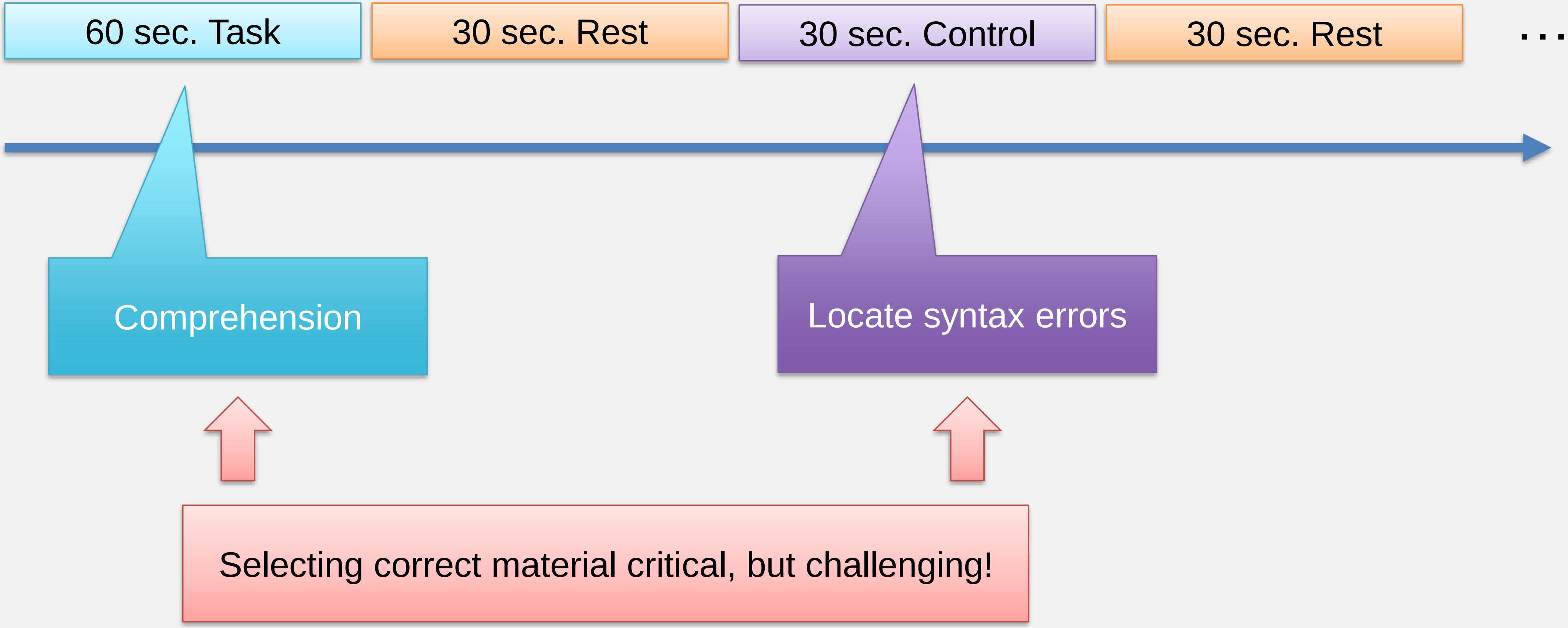
TECHNISCHE UNIVERSITÄT CHEMNITZ

```java
public void method_name() {
    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);
}
```

RQ: Which brain regions are activated during program comprehension?

- Controlled experiment
- Independent variable
  - Task: Program comprehension, control condition, rest
- Dependent variable
  - Brain activation
  - Behavior

- Block design

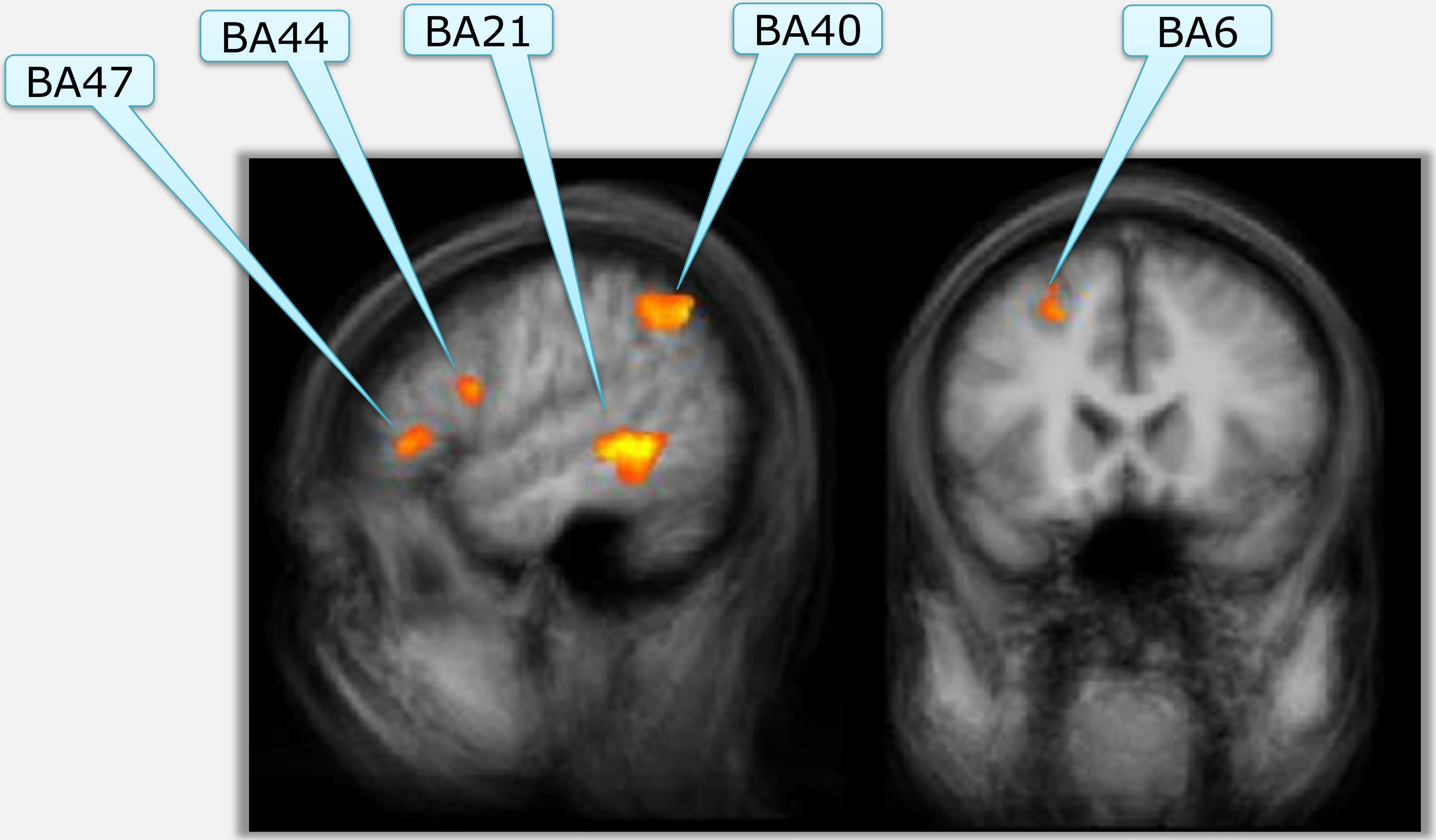


LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

```
public void method_name()
    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);
}
```

- Contrast between Comprehension and Syntax errors
- Network of brain areas
- No "classic" logic brain area
- All left lateralized



LIN LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

```
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;
}
```

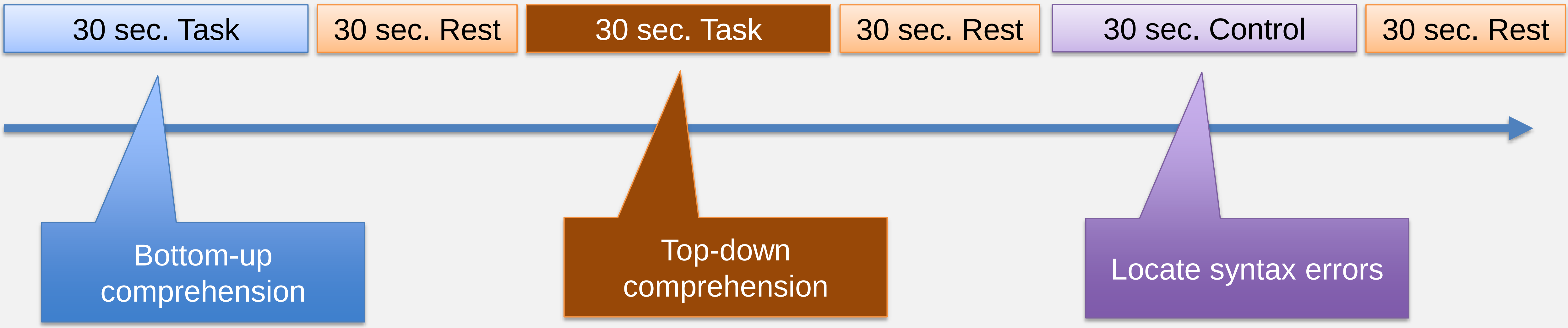
RQ: What is the difference between bottom-up program comprehension and top-down comprehension in terms of activation and the brain areas involved?

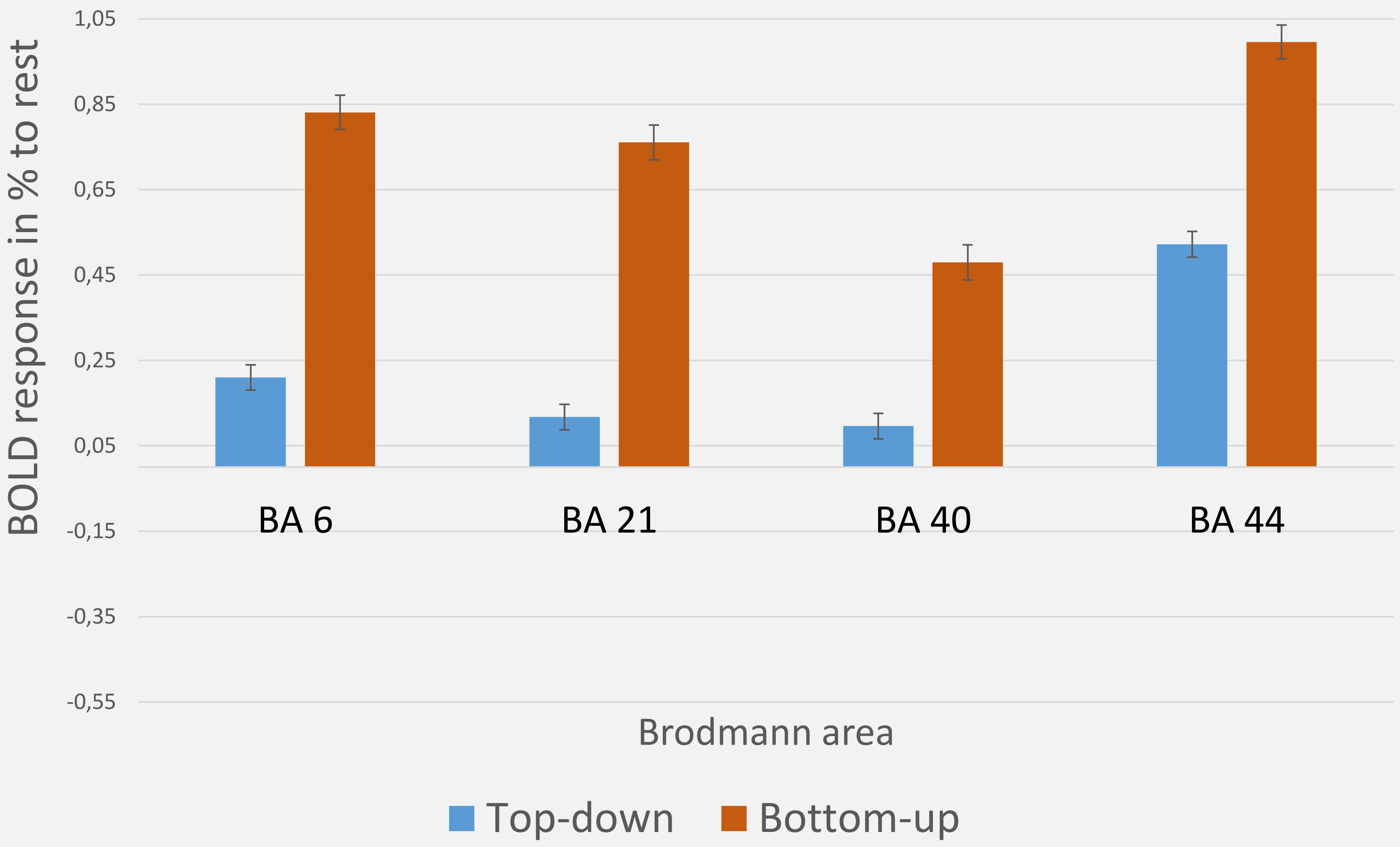
- Controlled experiment
- Independent variable
  - Task: Program comprehension, control condition, rest
- Dependent variable
  - Brain activation
  - Behavior

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

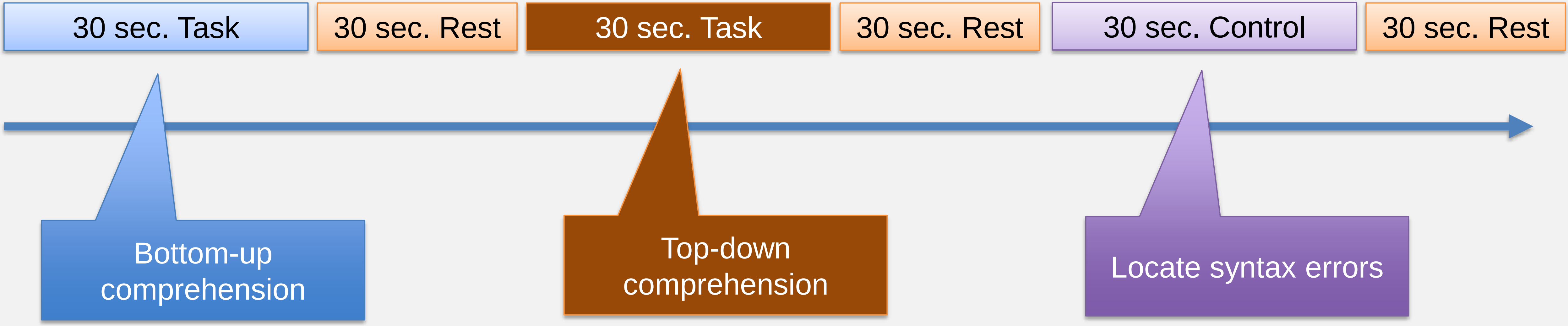
- Controlled experiment
- Independent variable
  - Task: Program comprehension, control condition, rest
  - Comprehension strategy: top-down or bottom-up
- Dependent variable
  - Brain activation
  - Behavior

> Top-down vs. bottom-up comprehension

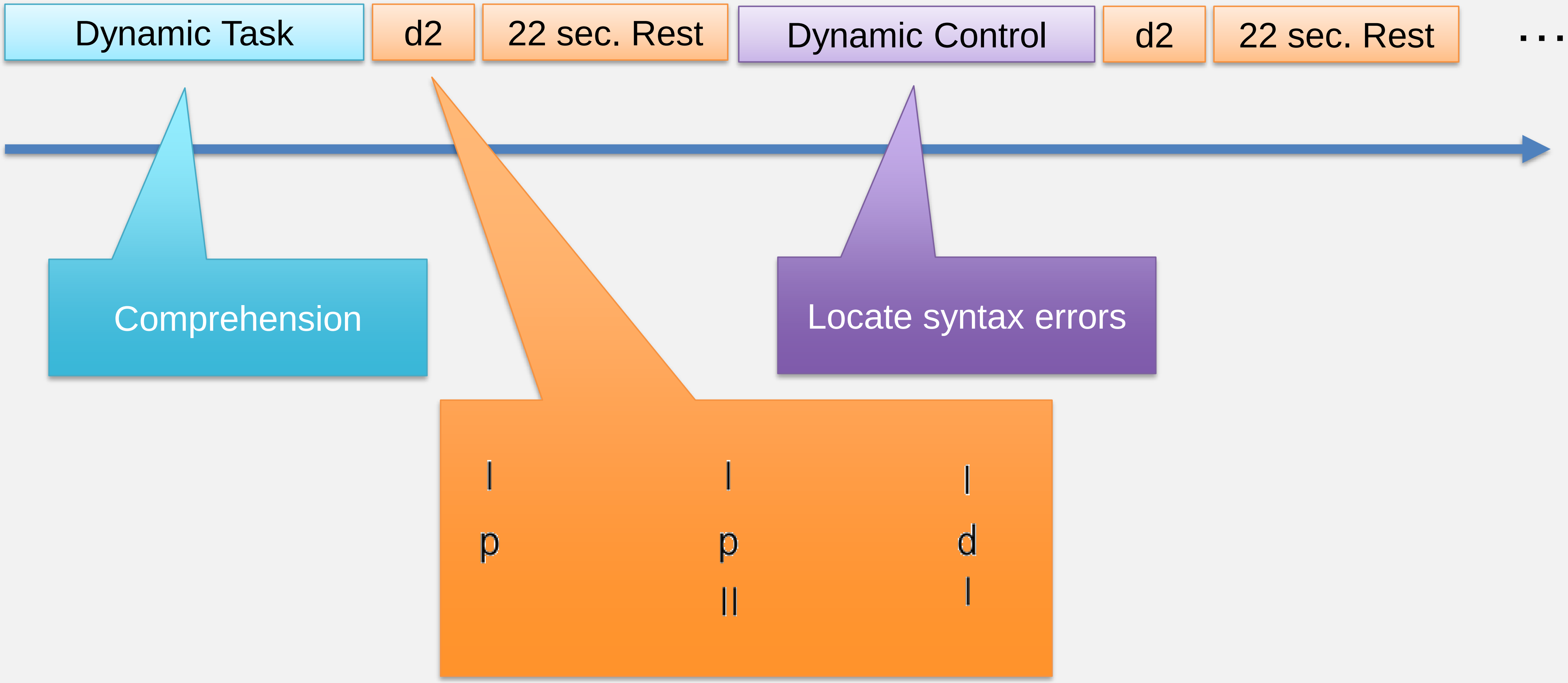



BOLD response in % to rest
1,05
0,85
0,65
0,45
0,25
0,05
-0,15
-0,35
-0,55
BA 6
BA 21
BA 40
BA 44
Brodmann area
Top-down
Bottom-up

> Problems with experiment design

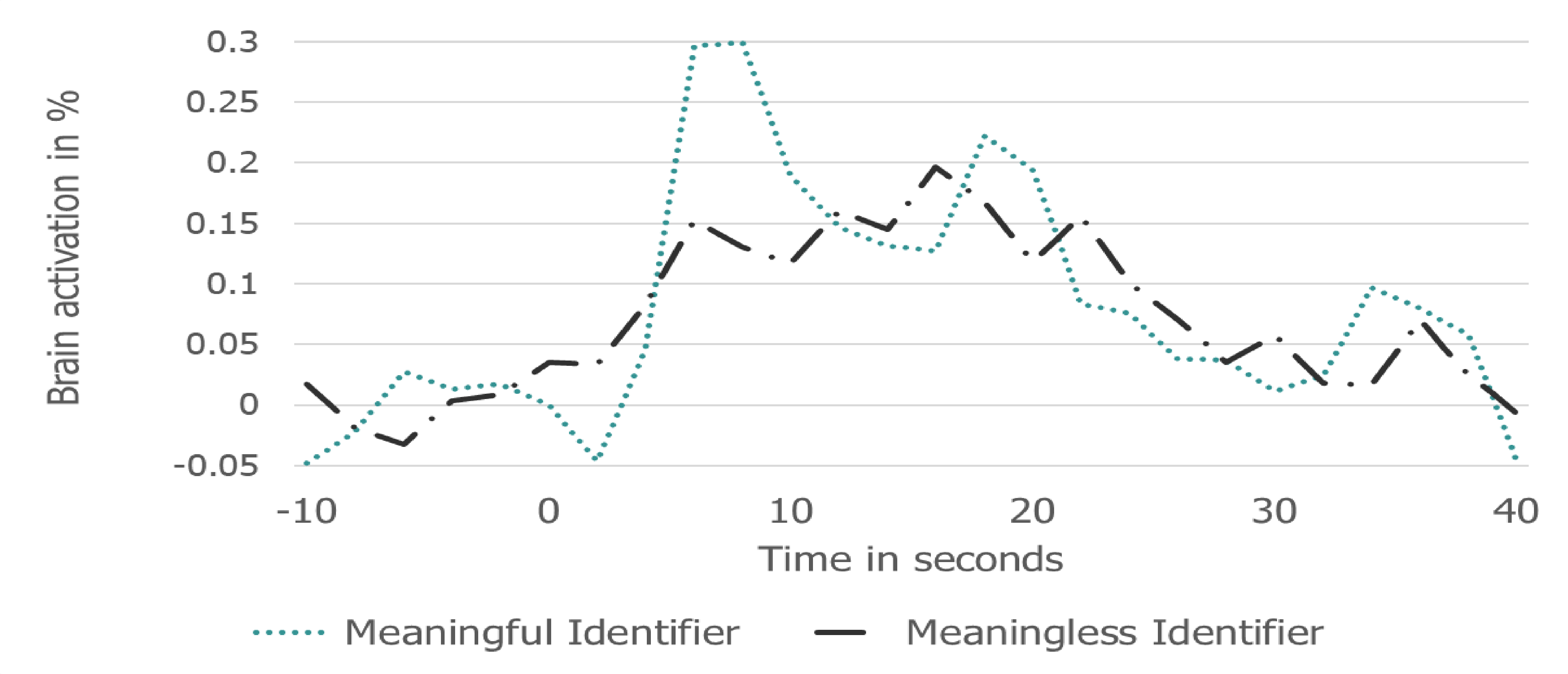


> Refined experiment design



LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

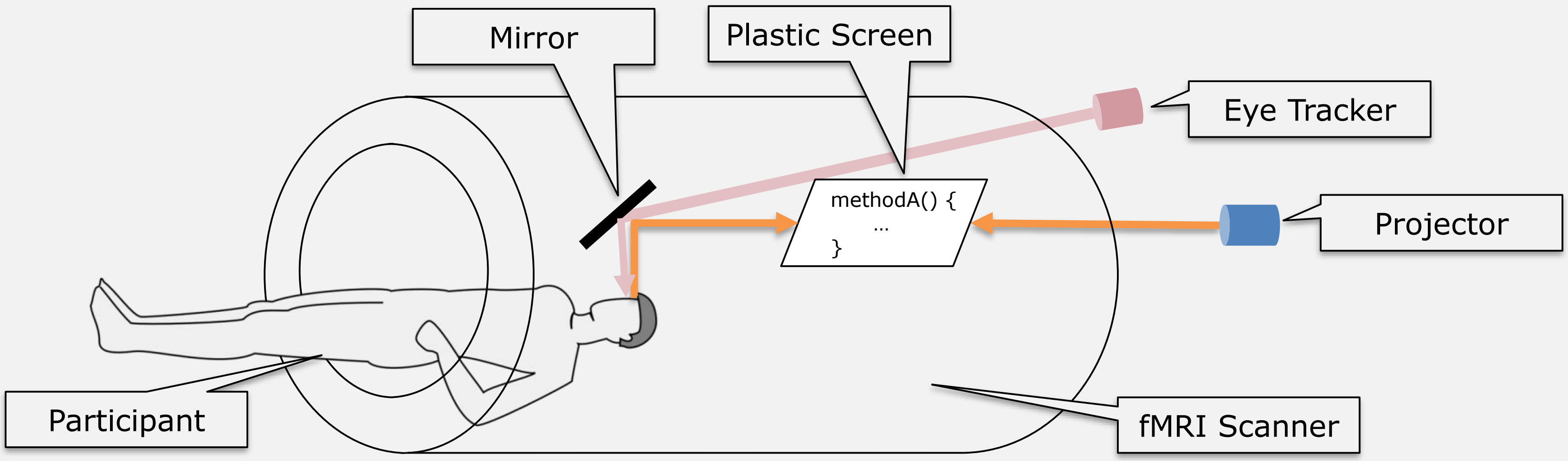


Program Comprehension

Overview | Strategy | Comprehension | Computation

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

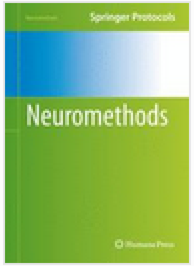


Eye tracking in the fMRI scanner is **challenging**

- Collected for only 10 out of 22 participants a complete data set
- Reduced spatial accuracy
- Negligible drift

- Eye tracking

Springer Link



pp 1-15 | Cite as

A Practical Guide to Functional Magnetic Resonance Imaging with Simultaneous Eye Tracking for Cognitive Neuroimaging Research

Authors | Authors and affiliations

Michael Hanke, Sebastiaan Mathôt, Eduard Ort, Norman Peitek, Jörg Stadler, Adina Wagner

Protocol

First Online: 02 October 2019

LIN LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

	2014	2017	2018	2019
Topic	Comprehension	Top-down and bottom-up comprehension	Simultaneous eye tracking	[redacted]
Published at	ICSE	FSE	ESEM	*Under review*
Participants	17 students	14 students	22 students	20 students

→ Finding a homogenous group in SE is much harder than classic neuroscience experiments

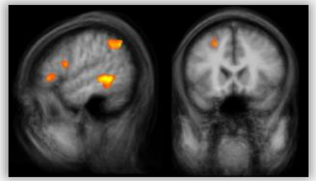
LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

- How do activated brain areas and activation strength differ between expert and novice programmers?
- What are the neural representations of if-then-else statements, loops, and recursion?
- Which programming language or paradigm is best to start with when teaching students programming given the cognitive processes involved in program comprehension?
- **Your own ideas...**

Possibility for a thesis or PhD!

- Neuroscience methods promise to understand programmers’ brains



- Several studies provided insights
  - Experiment design is still being refined

[Hanke19] Hanke, M., Mathôt, S., Ort, E., Peitek, K., Stadler, J., & Wagner, A.. A practical guide to functional magnetic resonance imaging with simultaneous eye tracking for cognitive neuroimaging research. *Neuromethods.* 2019

[Peitek18] Norman Peitek, Janet Siegmund, Chris Parnin, Sven Apel, Johannes Hofmeister, and Andre Brechmann. *"Simultaneous Measurement of Program Comprehension with fMRI and Eye Tracking: A Case Study."* In: Proc. Int'l Symposium Empirical Software Engineering and Measurement (ESEM). ACM, 2018

[Siegmund14] Janet Siegmund, Christian Kaestner, Sven Apel, Chris Parnin, Anja Bethmann, Thomas Leich, Gunter Saake, and Andre Brechmann. *"Understanding Understanding Source Code with Functional Magnetic Resonance Imaging."* In: Proc. Int'l Conf. Software Engineering (ICSE). ACM, 2014, pp. 378–389

[Siegmund17] Janet Siegmund, Norman Peitek, Chris Parnin, Sven Apel, Johannes Hofmeister, Christian Kaestner, Andrew Begel, Anja Bethmann, and Andre Brechmann. "Measuring Neural Efficiency of Program Comprehension." In: Proc. Europ. Software Engineering Conf./Foundations of Software Engineering (ESEC/FSE). Paderborn, Germany: ACM, 2017, pp. 140–150.

[Tiarks11] Rebecca Tiarks. What Programmers Really Do: An Observational Study. 2011

Designing a Program Comprehension fMRI Experiment

Norman Peitek
Leibniz Institute for Neurobiology Magdeburg

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

- Software complexity metrics
  - Measure “complexity” of source code based on some of their properties
  - Widespread in industry as proxy for program comprehension
    - → Based on the metric values, decisions for checking in code or refactoring code could be made
  - Examples: LOC, McCabe cyclomatic complexity, Halstead, DepDegree, …
    - But there are 100+ metrics

- A handful of existing studies with conventional methods suggest that some complexity metrics are only valid to a limited degree
  - Can we shed light on this issue with fMRI?

- Task: Create an experiment design for an fMRI study to understand how complexity metrics correlate to program comprehension!

- Consider
  - Hypotheses
  - Independent and dependent variables
  - Participants (-selection)
  - Task and material
  - Experiment design
  - …

- Task: Create an experiment design for an fMRI study to understand how brain activation differs between novices and expert programmers!

- Consider
  - Hypotheses
  - Independent and dependent variables
  - Participants (-selection)
  - Task and material
  - Experiment design
  - …

- Examples:
  - Source code size measured in lines of code does not affect the strength of brain activation in Broca‘s area
  - Data-flow of source code measured in DepDegree increases brain activation strength in Brodmann area 6

- Independent Variables
  - Task (program comprehension, control task, rest)
  - Code complexity
    - Based on one metric or multiple ones?
    - Watch out for correlation between metrics

- Dependent Variables
  - Brain activation
  - Behavior
  - Eye tracking
  - …

- Who should we invite to our experiment?
  - Students or professional programmers?
    - Or both? Does it matter?
- Do we need multiple participant groups?
  - → Between-subject and within-subject design
- How do we ensure homogenous group(s)?
- How do we establish that they’re sufficiently qualified?

- Which code snippets are we using for the tasks?
  - How do we investigate the effect of complexity metrics?
  - How many snippets do we need?
  - How do we develop suitable snippets?

- Calculate the experiment time! Can we fit enough snippets in to expect statistically significant results?
- Can we display the snippets in the scanner (30 lines max)?
- Are the snippets similar enough and can they be solved by an average participant (frustration potential)?

- What is the task for the participants?
  - How do we ensure they are fulfilling the task?
  - Is the task doable in the scanner?

- Which control condition do we use?
  - Must allow for a suitable fMRI contrast

LEIBNIZ INSTITUTE FOR NEUROBIOLOGY MAGDEBURG

TECHNISCHE UNIVERSITÄT CHEMNITZ

- Novice and expert programmers differ substantially in their knowledge, productivity, and code quality.
  - Conventional studies indicate that they approach understanding code differently (e.g., top-down vs. bottom-up comprehension)
  - Can we quantify this difference with fMRI?

> Conclusion

- Designing a neuroscience experiment is challenging