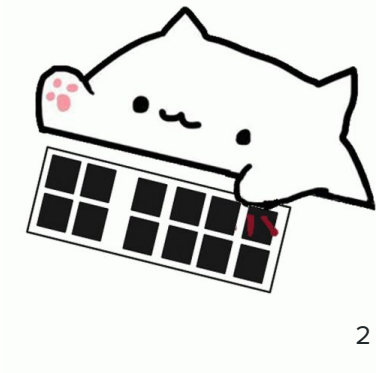
Causal Relationships and Programming Outcomes: A *Transcranial Magnetic Stimulation* (TMS) Experiment

Hammad Ahmad, Madeline Endres, **Kaia Newman**,
Priscila Santiesteban, Emma Shedden, and Westley Weimer

Talk Outline

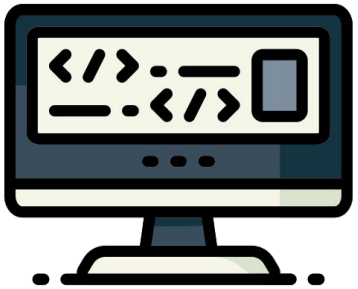
1. Research Motivation and Overview
2. Experimental Design and Data Collection
3. Analysis and Results
4. Discussion and Summary



Understanding the cognitive basis of software engineering is important

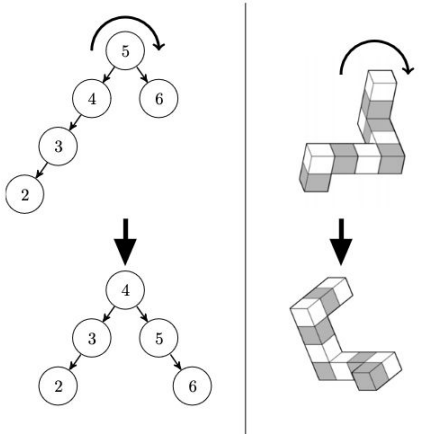
- Can inform **pedagogy** and **training**
- **Make novices more like experts** – faster!

Medical imaging increasingly common for this in SE research



Prior research links spatial reasoning to programming tasks

Two different neuroimaging techniques saw activity in spatial reasoning-related regions of the brain in a study by Huang *et al.* [1]



fMRI machine

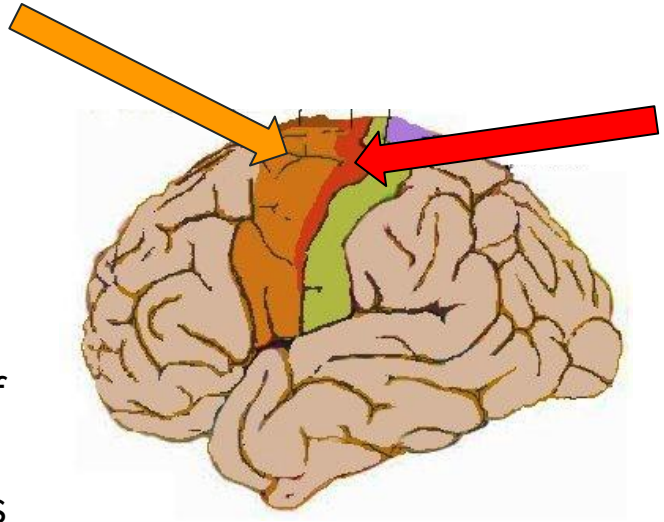
[“95% of [brain activity] was statistically indistinguishable between Mental and Tree tasks.”]

[1] Huang *et al.* Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS. (2019)

Prior research links spatial reasoning to programming tasks

***Supplementary motor area* (SMA)** potentially relevant for:

- Code comprehension [1]
- Data structure manipulation [2]
- Comprehension of code with higher complexity metrics [3]



***Primary motor cortex* (M1)** potentially relevant for:

- Code writing (as opposed to prose writing) [4]

[1] Siegmund *et al.* Understanding understanding source code with functional magnetic resonance imaging. (2014). [2] Huang *et al.* Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS. (2019). [3] Peitek *et al.* Program Comprehension and Code Complexity Metrics: An fMRI Study. (2021). [4] Krueger *et al.* Neurological Divide: An fMRI Study of Prose and Code Writing. (2020)

But causality is not definitively determined!

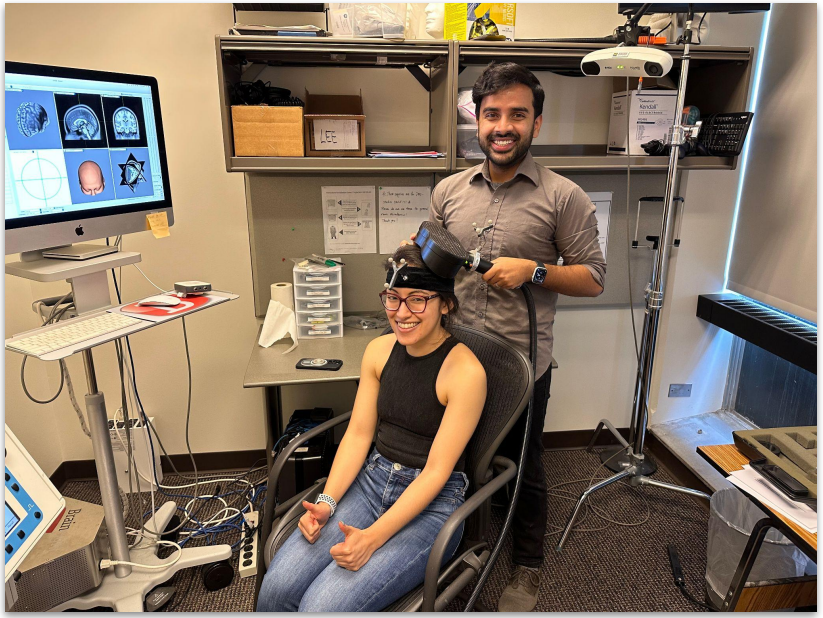
Unfortunately, prior **SE neuroimaging studies only show correlation**, not causation.

To investigate a causal relationship, we want an experimental intervention that is:

- Noninvasive
- Safe
- Time-efficient
- Well-understood and previously validated

Enter: *Transcranial Magnetic Stimulation* (TMS)

TMS uses magnets to induce an electrical current in a targeted region, which can stimulate localized activity



TMS:

- Takes only minutes to localize and administer
- Is used in a variety of scientific and medical cases – 1000+ papers published each year use TMS
- Is safe and well-tolerated

TMS has been used to uncover causality and has been shown to improve outcomes in other fields

Mental math: TMS resulted in a **30% accuracy increase** in memorization and addition of numbers [1].

Language: TMS resulted in **faster verb processing speed** for manual actions [2].

[1] J. Gill *et al*. It's the thought that counts: Examining the task-dependent effects of transcranial direct current stimulation on executive function. 2015. [2] R. M. Willems *et al*. A functional role for the motor system in language understanding: Evidence from theta-burst transcranial magnetic stimulation. 2011.

Our study, from a bird’s-eye view

We present the **first study to investigate causal relationships between previously-implicated brain regions and programming tasks** using TMS

Main RQs:

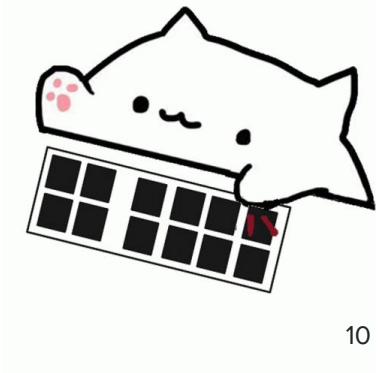
1. Can we **replicate previous results** finding TMS can impact spatial ability?

2. Is there a **direct, causal relationship** between TMS of spatially-related brain regions (M1, SMA) and programming task completion time or outcomes?

3. Can TMS **affect programming tasks at all?**

Talk Outline

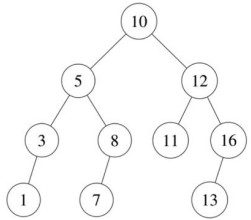
1. Research Motivation and Overview
2. Experimental Design and Data Collection
3. Analysis and Results
4. Discussion and Summary



Stimuli used

Data Structure Manipulation

Consider the AVL tree below. After inserting 14 into the tree (and performing rotations to keep the tree balanced as necessary), which of the following will be produced by a pre-order traversal of the resulting tree?



A: 10, 5, 3, 1, 8, 7, 12, 11, 14, 13, 16

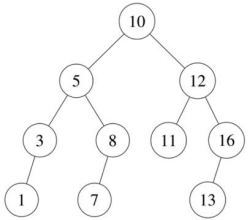
B: 10, 5, 3, 1, 8, 7, 12, 11, 16, 13, 14

Stimuli used

Data Structure Manipulation

Code Comprehension

Consider the AVL tree below. After inserting 14 into the tree (and performing rotations to keep the tree balanced as necessary), which of the following will be produced by a pre-order traversal of the resulting tree?



A: 10, 5, 3, 1, 8, 7, 12, 11, 14, 13, 16

B: 10, 5, 3, 1, 8, 7, 12, 11, 16, 13, 14

What is true about the following code for reversing a vector? Assume that n = *v*.size().

```
vector<int> v{1, 2, 3, 4};
stack<int> s;

for (size_t i = 0; i < v.size(); ++i)
    s.push(v[i]);

for (size_t i = 0; i < v.size(); ++i) {
    v[i] = s.top();
    s.pop();
}  // for ..i
```

A: The memory complexity is O(1).

B: The memory complexity is O(n).

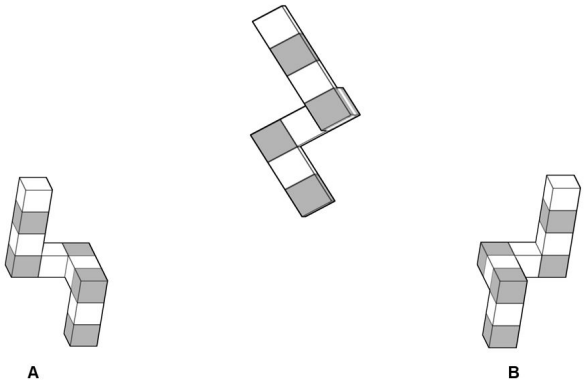
Stimuli used

Data Structure Manipulation

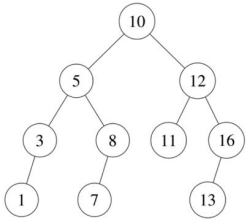
Code Comprehension

Spatial Rotation

Please type the corresponding letter of the bottom objects to give your answer.



Consider the AVL tree below. After inserting 14 into the tree (and performing rotations to keep the tree balanced as necessary), which of the following will be produced by a pre-order traversal of the resulting tree?



A: 10, 5, 3, 1, 8, 7, 12, 11, 14, 13, 16

B: 10, 5, 3, 1, 8, 7, 12, 11, 16, 13, 14

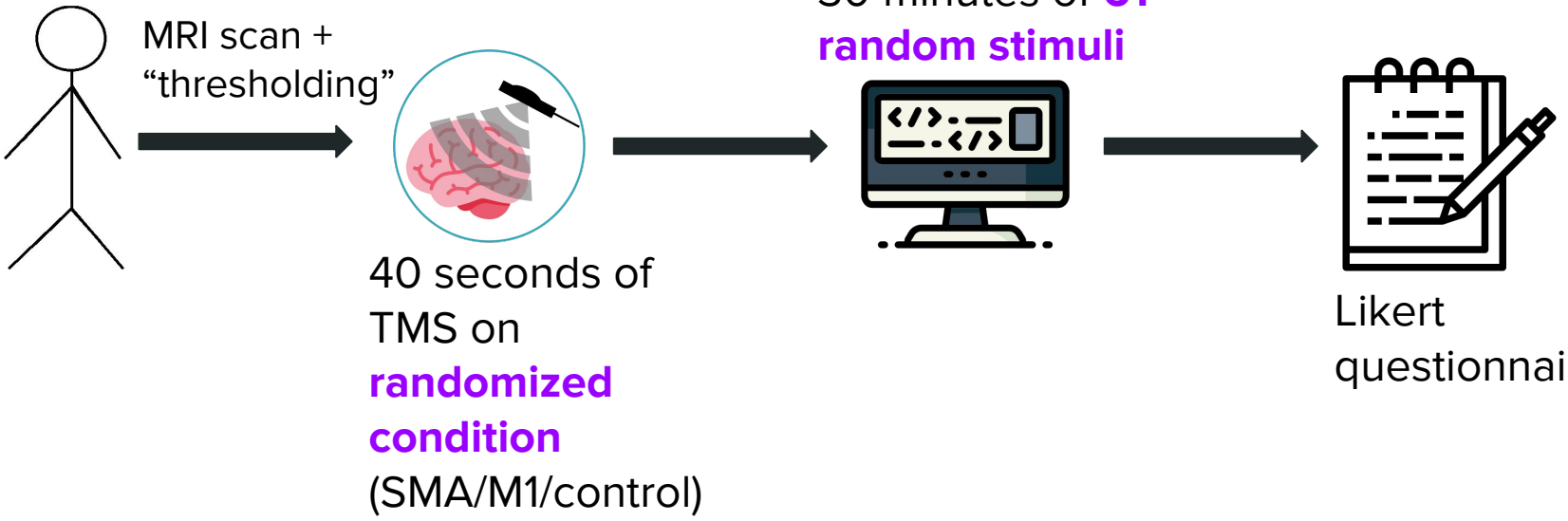
What is true about the following code for reversing a vector? Assume that n = *v*.size().

```
vector<int> v{1, 2, 3, 4};
stack<int> s;

for (size_t i = 0; i < v.size(); ++i)
    s.push(v[i]);

for (size_t i = 0; i < v.size(); ++i) {
    v[i] = s.top();
    s.pop();
}  // for ..i
```

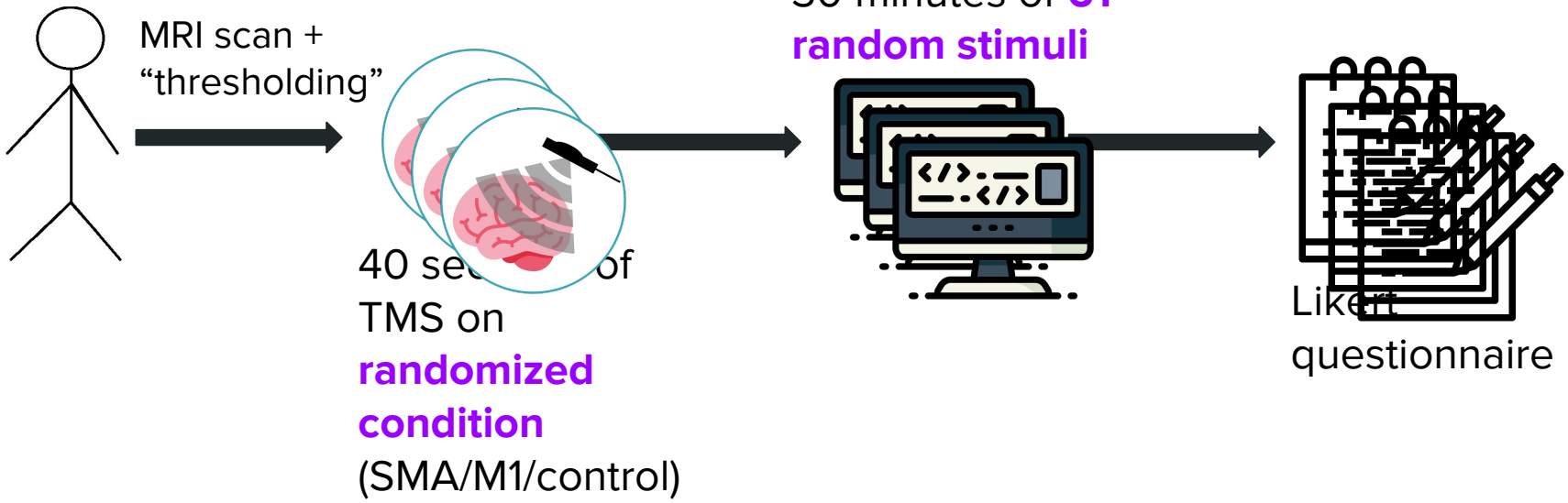
A: The memory complexity is O(1).

B: The memory complexity is O(n).

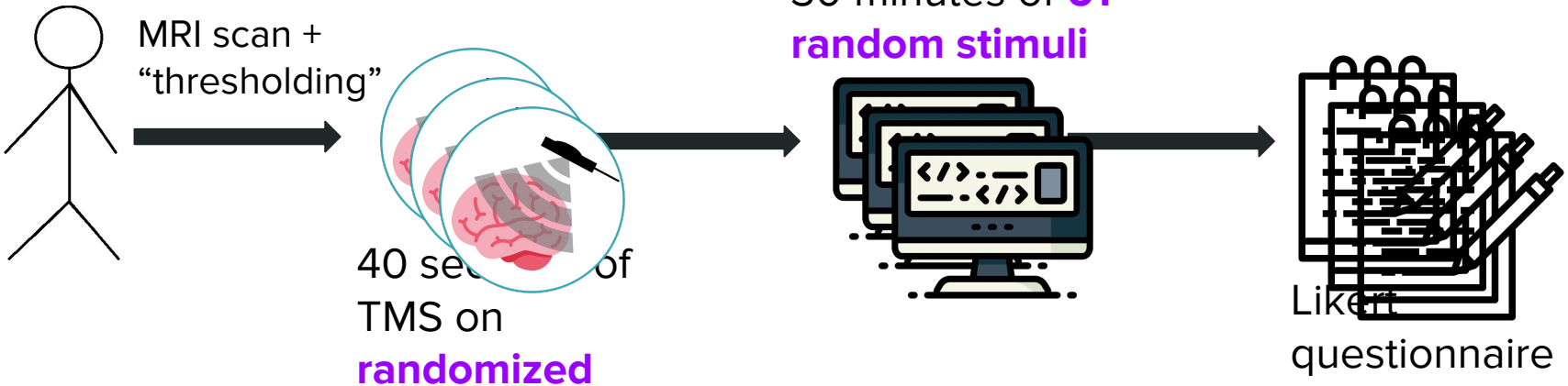
We did an experiment on 16 participants for **3 sessions of TMS each**, recording timing and correctness on stimuli



We did an experiment on 16 participants for **3 sessions of TMS each**, recording timing and correctness on stimuli



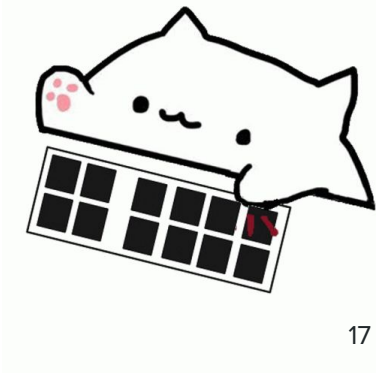
We did an experiment on 16 participants for **3 sessions of TMS each**, recording timing and correctness on stimuli



*We **pre-registered** our primary experimental design and hypothesis to mitigate for p-hacking!

Talk Outline

1. Research Motivation and Overview
2. Experimental Design and Data Collection
3. Analysis and Results
4. Discussion and Summary



We used mixed-effects modeling to analyze timing and correctness data

- **Mixed-effects modeling** was used to account for both systematic (direct) and heterogeneous effects
  - TMS has been shown to have a heterogeneous treatment effect that varies between people in dozens of psychology studies
- Relevant analyses were conducted blind to **reduce researcher bias**
- BH adjustment used to **control for multiple comparisons** when calculating p-values

Up next: the answers to our questions!

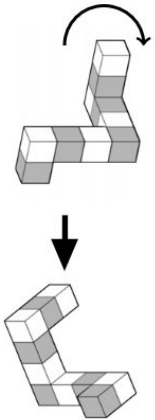
RQ1, *Replication and Correct Application*: Can we **replicate previous results** finding TMS can impact spatial ability?

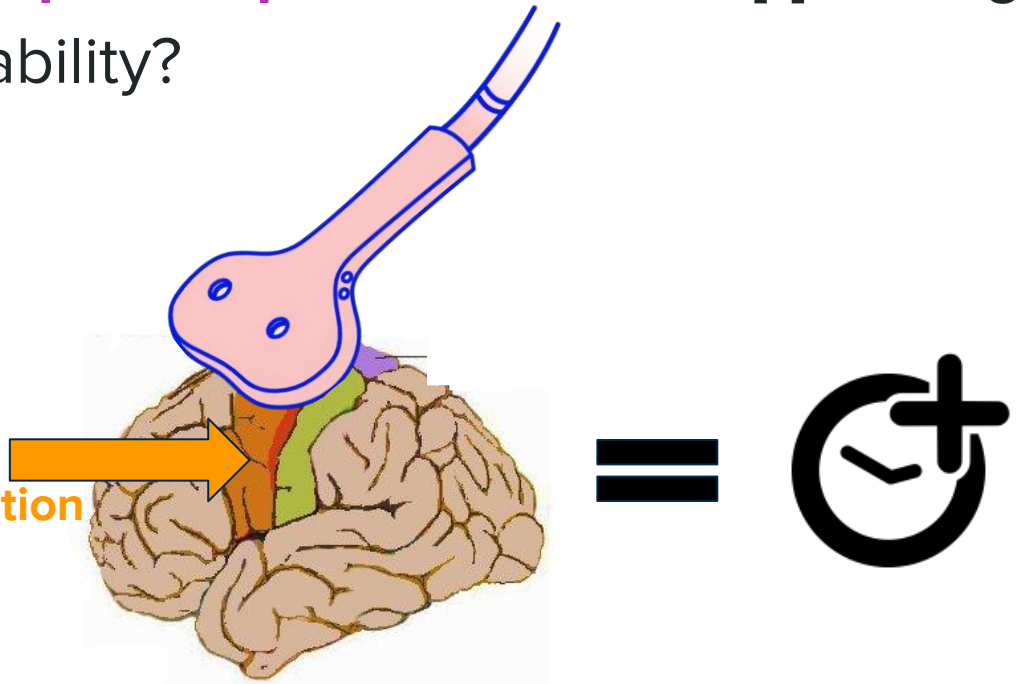
RQ2, *Direct Relationship*: Is there a **direct, causal relationship** between TMS of spatially-related brain regions and programming task completion time or outcomes?

RQ3, *Affecting Programming Outcomes*: Can TMS **affect programming tasks at all?**

RQ1, *Replication and Correct Application*

Question: Can we **replicate previous results** [1] finding TMS can impact spatial ability?





[1] Cona *et al.* TMS of supplementary motor area (SMA) facilitates mental rotation performance: Evidence for sequence processing in SMA. (2017)

RQ1, *Replication and Correct Application*

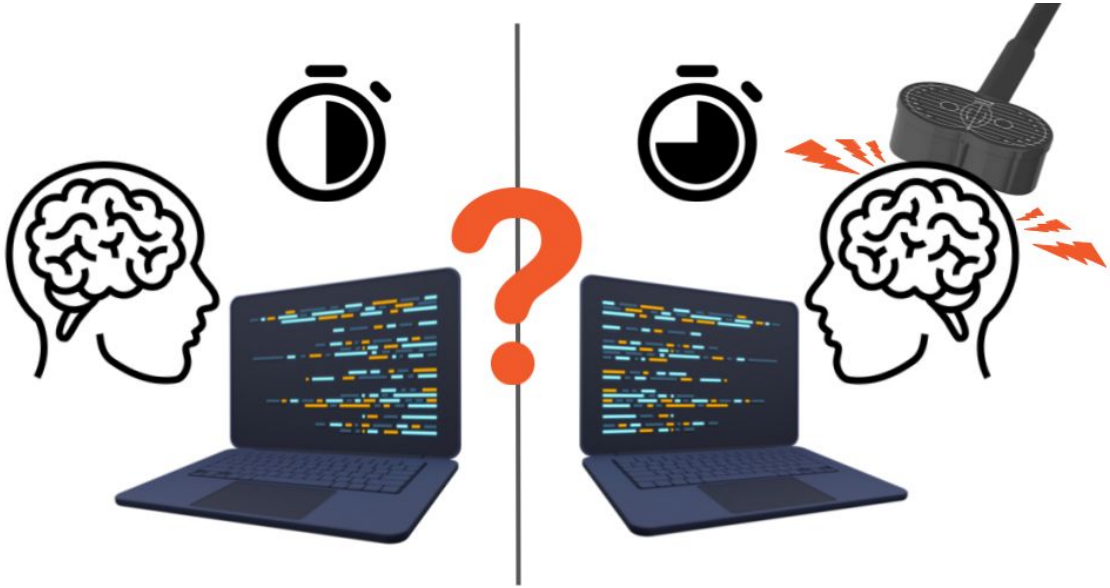
Question: Can we **replicate previous results** [1] finding TMS can impact spatial ability?

Yes! ($p \leq .02$, 15.3% increase in raw response time)

[1] Cona *et al.* TMS of supplementary motor area (SMA) facilitates mental rotation performance: Evidence for sequence processing in SMA. (2017)

RQ2, *Direct Relationship*

Question: Is there a **direct, causal relationship** between TMS of spatially-related brain regions and programming task completion time or outcomes?



RQ2, *Direct Relationship*

Question: Is there a **direct, causal relationship** between TMS of spatially-related brain regions and programming task completion time or outcomes?

We found no significant, direct relationships between experimental condition and behavioral programming outcomes in this experiment.

Result: Contrary to multiple previous correlative studies, we **do not** find evidence for one!

RQ3, *Affecting Programming Outcomes*

Question: Can TMS **affect programming tasks at all?**

- Our model **predicts variance in outcomes** according to various factors
  - Common to TMS studies

RQ3, *Affecting Programming Outcomes*

Question: C[illegible]TMS[illegible]**at all?**

- Our[illegible]ding to

Spoiler: YES! TMS **can affect** programming outcomes!

RQ3, *Affecting Programming Outcomes*

Question: Can TMS **affect programming tasks at all?**

Which factors matter more for time taken?

Factor	**Effect Size (Normalized)**
“How hard is the question?”	`1.00`
“Participant expertise”	`0.18`

RQ3, *Affecting Programming Outcomes*

Question: Can TMS **affect programming tasks at all?**

Which factors matter more for time taken?

Factor	Effect Size (Normalized)
“How hard is the question?”	`1.00`
“Participant expertise”	`0.18`
“TMS”	`0.05`

The effect of TMS was one-twentieth the effect of question difficulty, i.e., **not zero.** (Our 95% confidence interval excludes zero).

RQ3, *Affecting Programming Outcomes*

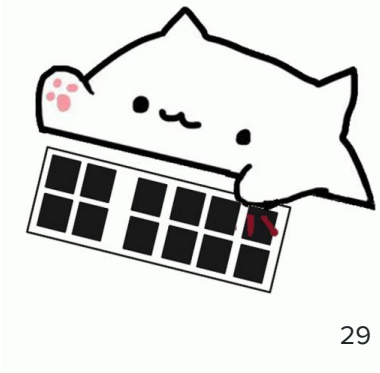
Question: Can TMS **affect programming tasks at all?**

We found that the participant by brain region stimulated random effect significantly **accounted for 2.2% of the variance in the response time** when controlling for other plausible effects.

Result: Yes, TMS **can affect** actual programming outcomes!

Talk Outline

1. Research Motivation and Overview
2. Experimental Design and Data Collection
3. Analysis and Results
4. Discussion and Summary



Why are we excited?

TMS has been used to **uncover causality** from suspected correlations in other psychology studies and has been used for **real, positive interventions.**

2.2% variance is a small, **indirect effect**, but for the future, TMS:

- Can be done alongside other approaches
- Doesn’t require shared language
- Only takes five minutes

This is the **first TMS study for programming;** we are currently investigating more!

Future research could involve varying the TMS protocol or brain regions stimulated

Could try:

- Language-related brain regions
- Working memory-related brain regions
- Using different TMS protocols

In summary:

We **replicate previous results in psychology**, gaining confidence we apply TMS correctly.

We find no evidence for a causal relationship between spatial reasoning-related regions and programming outcomes, suggesting that interpreting **cognition in software engineering may be more complex.**

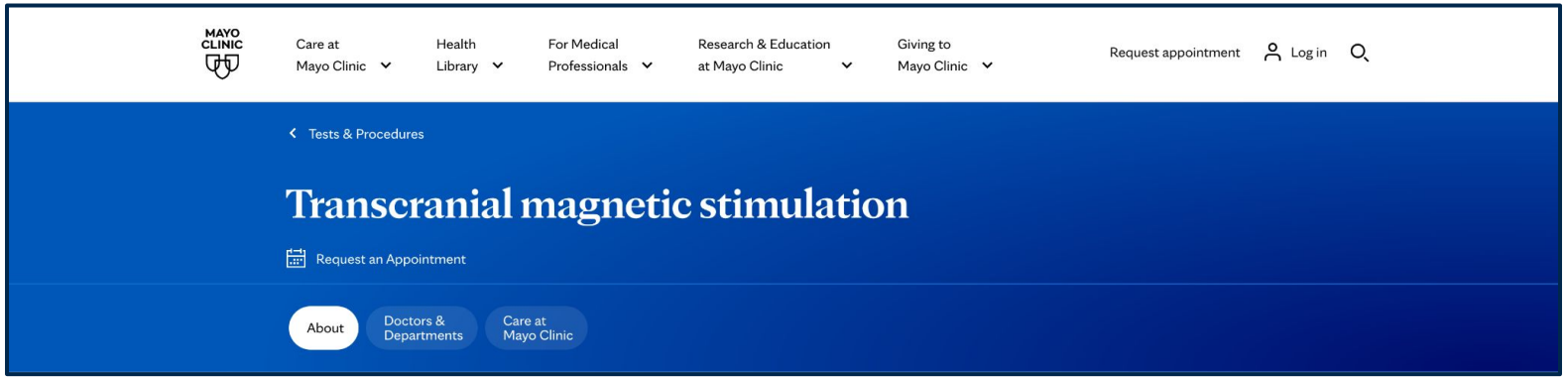
We have evidence that **TMS can affect programming outcomes**, implying the use of it as a tool to **investigate causality** and **directly affect outcomes** in other spaces extends to software engineering.

Hammad Ahmad, Madeline Endres, **Kaia Newman: kaian@cmu.edu,** Priscila Santiesteban, Emma Shedden, and Westley Weimer
Replication package at: **https://github.com/hammad-a/ICSE24_TMS**

“Is *Transcranial Magnetic Stimulation* (TMS) safe?”

Thresholding: before our experiment, we assessed a tolerable level of stimulation **per-person**, and then applied 80% of that to other brain regions.

This was an IRB-approved protocol with professional oversight. TMS is FDA-approved and used by the Mayo Clinic as a treatment for depression.



Pictures of the TMS machine