



# Studying Human Planning with the “Think-Aloud” Method

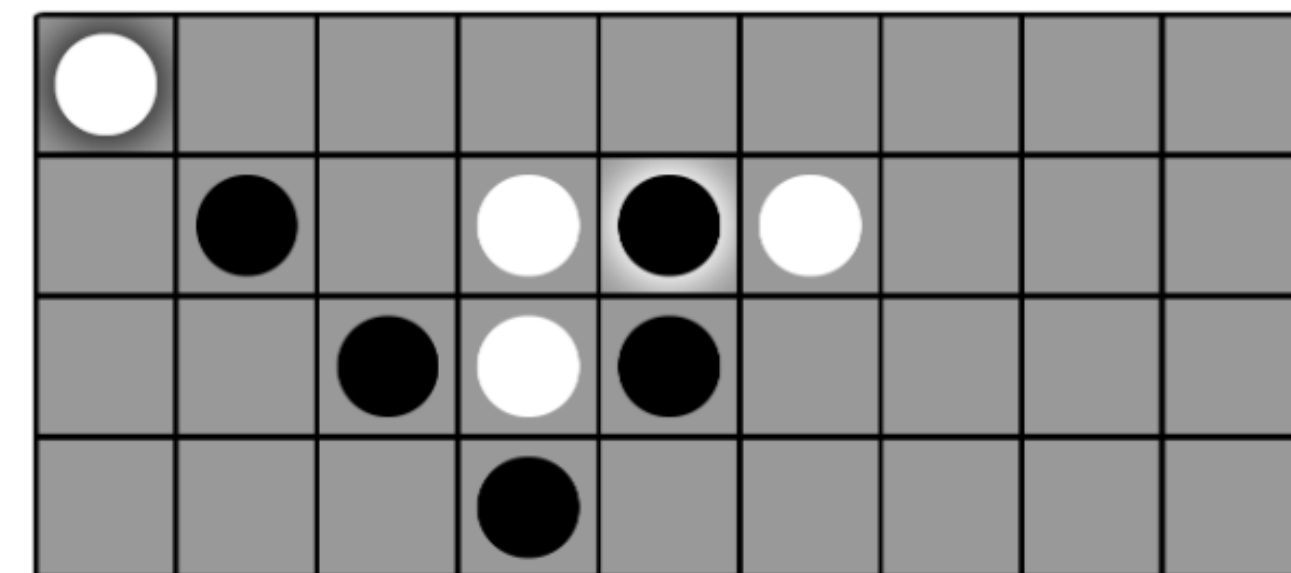
Shucheng Li<sup>1,3</sup>, Daisy Lin<sup>1</sup>, Wei Ji Ma<sup>1,2</sup>

Center for Neural Science<sup>1</sup> and Department of Psychology<sup>2</sup>, New York University; New York University Shanghai<sup>3</sup>



## Introduction

- Planning is a complex cognitive process that requires mental stimulation for future scenarios.
- In the Wei Ji Ma Lab, we study the mechanisms of complex planning through a novel game called 4-in-a-row.



4-in-a-row in a  $4 \times 9$  board: Put 4 pieces of your side in a row while keeping your opponent from doing the same.

- Heuristic Model: Previously, the lab developed a heuristic model that predicts participants' moves in 4-in-a-row and estimates each participants' depth of planning (DOP).
- “Think-aloud” method: ask participants to say everything that go through their mind when doing a task. It is used by De Groot (1946) to study how chess experts plan when they solve chess puzzles.

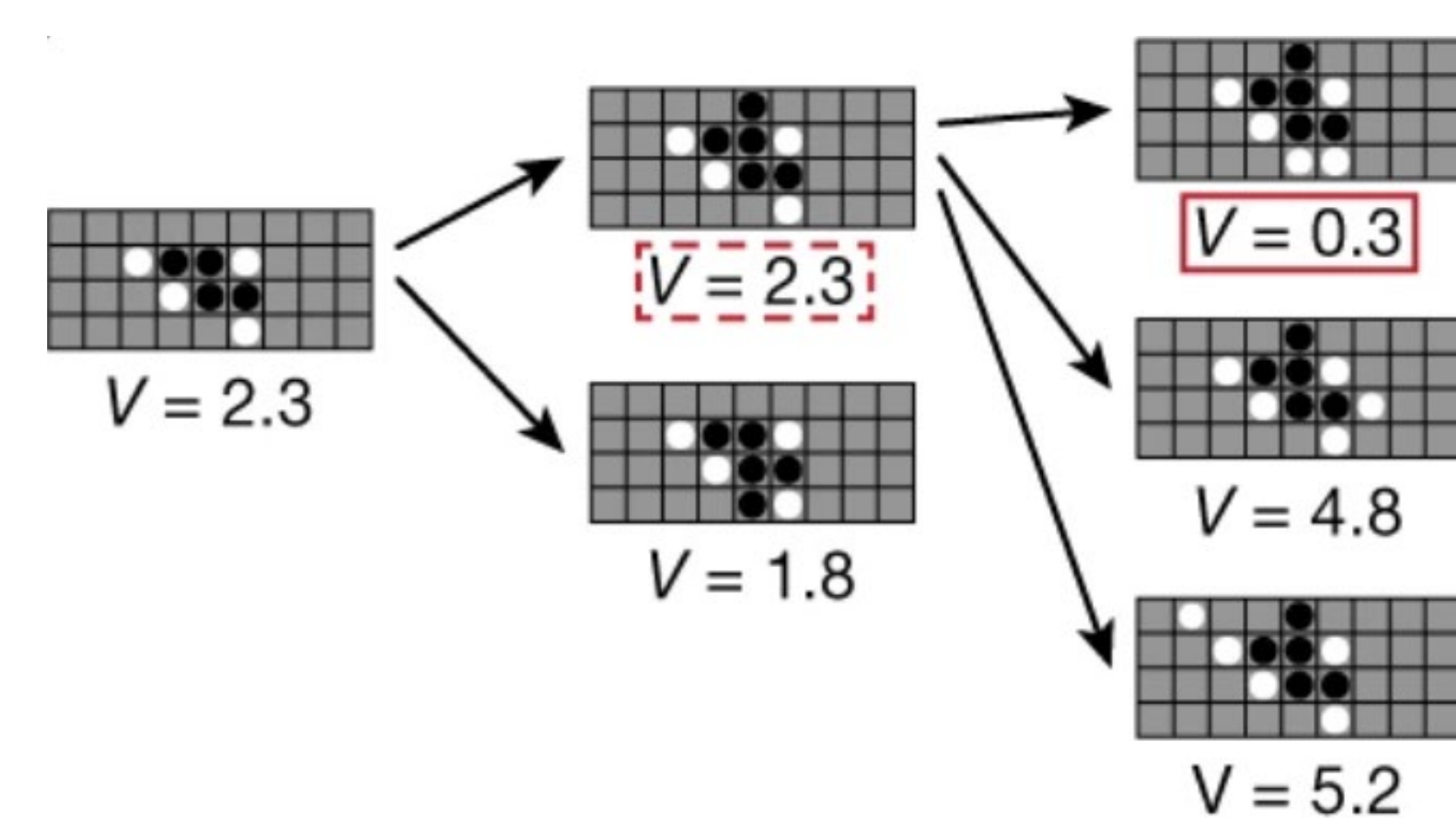
## Research Question

- Research Question: How deeply do people plan?
- Aim: Comparing the heuristic model with the traditional “think-aloud” method in DOP
- Q1: Does the DOP obtained from the heuristic model correlate with the planning depth obtained from narrations?
- Q2: Does any of the metrics from the participant's narration correlate with the Elo (the playing strength of the participant)?  
=> to identify key planning factors that predict a participant's performance in free play.

## Method

Heuristic Model:

- Heuristic function is a weighted combination of board features that maps a board state to a value estimate



- Pipeline:
- Fit the individual free-play data  
=> the weights of the heuristic function
- Stimulate best-first heuristic search on each board state  
=> DOP for each board state and each individual

“Think-aloud” Method:

- Ask participants to say everything that goes through their mind when they are solving 4-in-a-row puzzles.
- Code their narrations into a set of metrics, including planning depth, features on the board, planning sequence, high-level strategy, etc.

Participants:

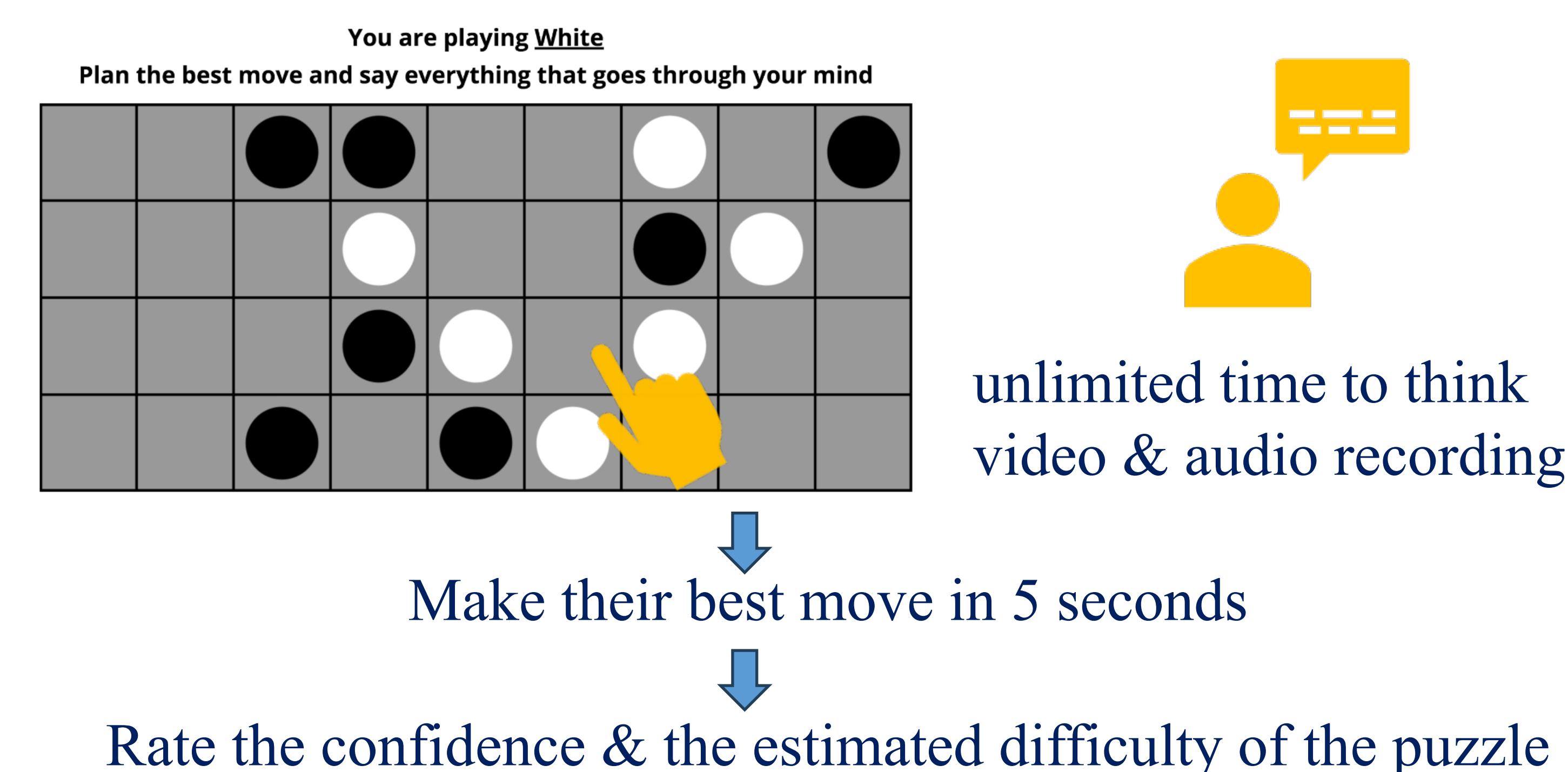
- Recruitment: Posters, NYU paid SONA system, and NYU psychology major student mailing list.
- Sample size: 34 participants (18 females, 2 non-binary, 14 males,  $M=23.32$  years, range= 18-50 years)

## Experiment Design

Stage 1: Free-play (~ 25min, 40 rounds)

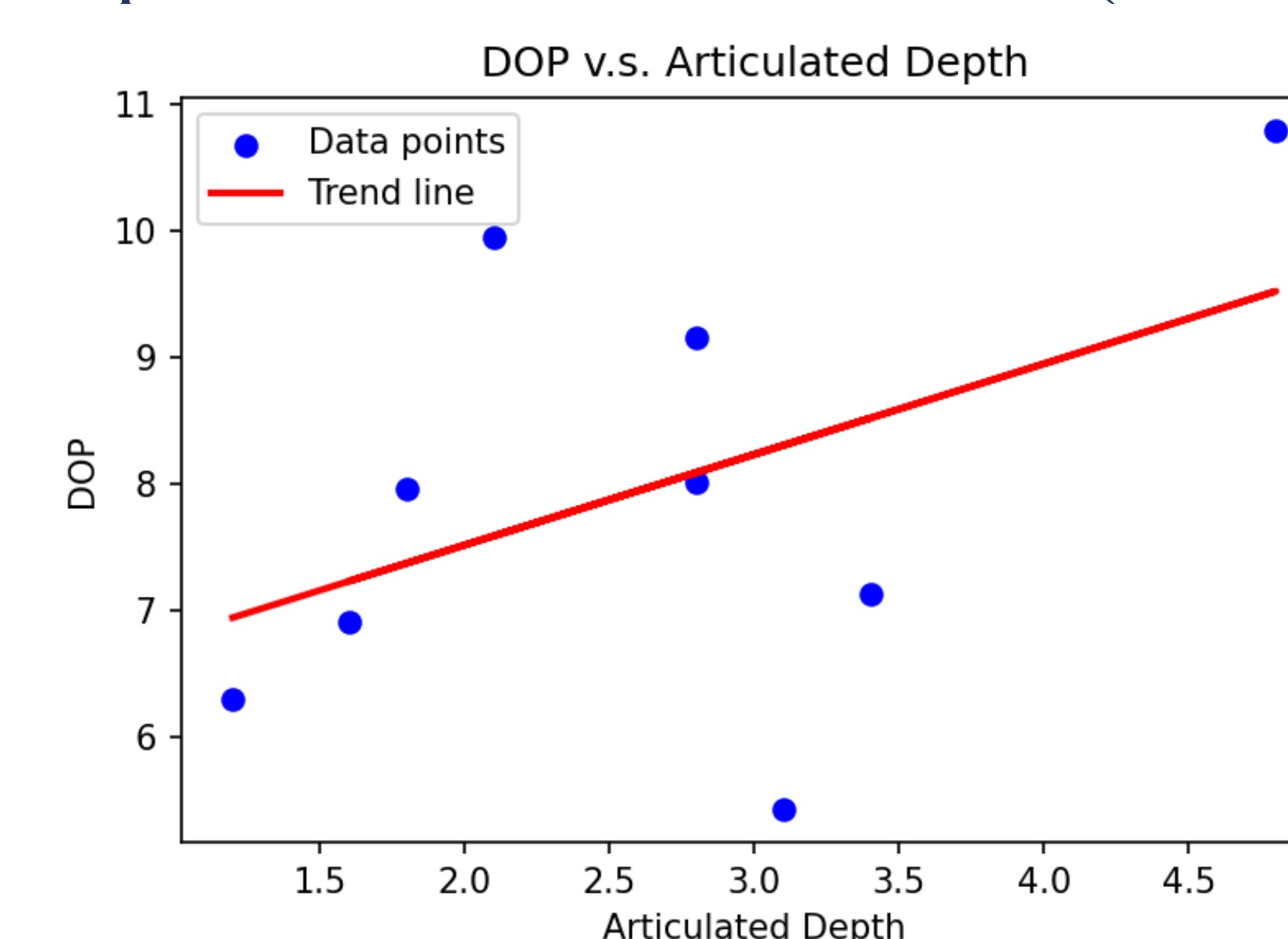


Stage 2: Think-aloud in puzzle solving (~ 40min, 10 puzzles)



## Result

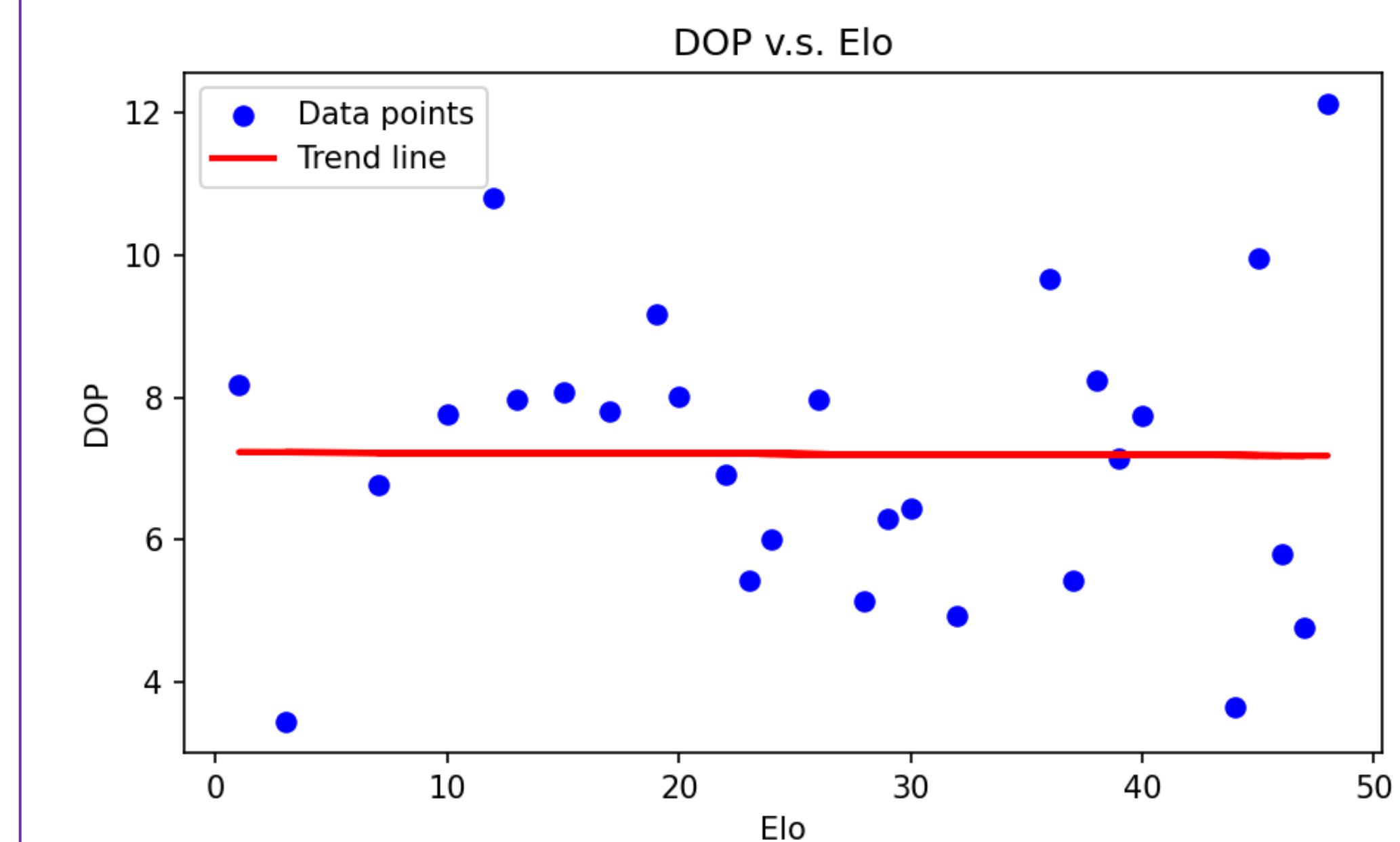
We have just done the data collection and not yet done analyzing the data. We finished the analysis of the narration of 9 participants. We averaged the maximum depth of narration across 10 puzzles to get the articulated depth ( $M=2.622$ ,  $SD= 1.036$ ) for each participant. We compared them with their DOP ( $M=7.967$ ,  $SD= 1.644$ ).



We did a Pearson's R correlation test between DOP and Articulated depth

$R=0.451$   
 $p\text{-value}=0.223$   
The sample size is too small.

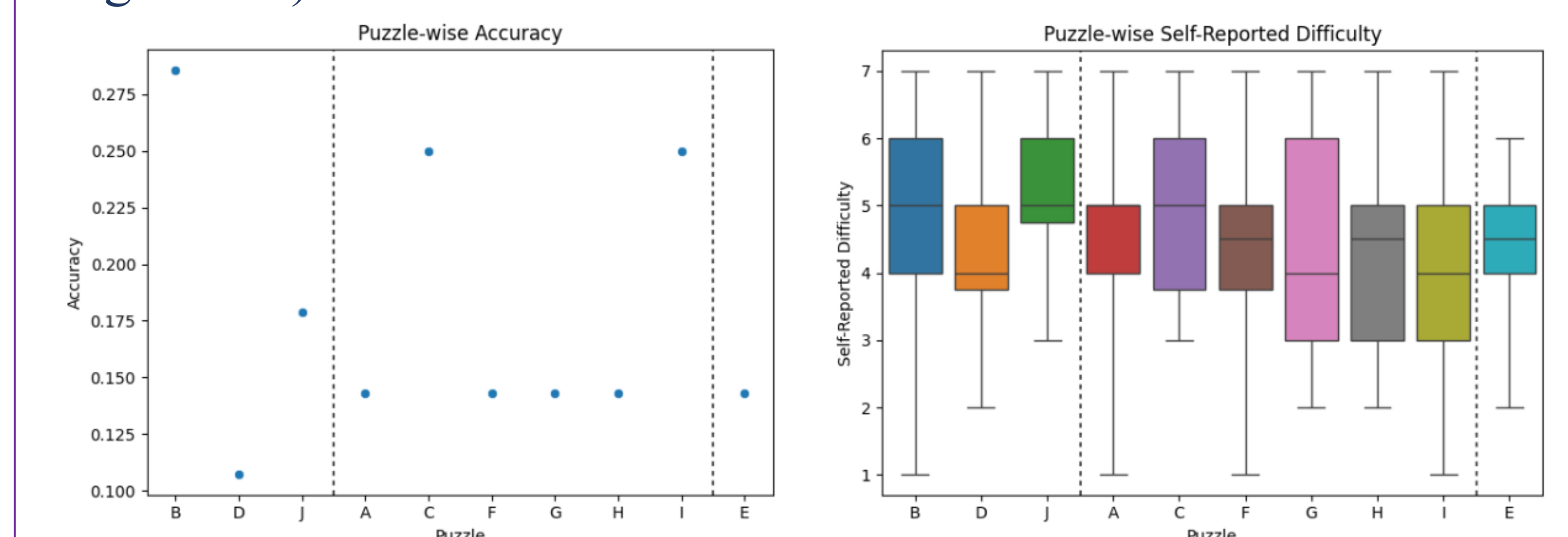
We fitted the free-play data of 28 participants to get the average DOP for each participant ( $M=7.201$ ,  $SD= 2.013$ ). We also calculated the Elo ranking of these 28 participants with their free-play data.



We did a Pearson's R correlation test between DOP and Elo

$R=-0.0069$   
 $p\text{-value}=0.97$   
No correlation.

Puzzle-wise analysis: The puzzles are designed to have only one best move that leads to a forced win. Puzzles B, D, and J are win-in-3, Puzzles A, C, F, G, H and I are win-in-4, and Puzzle E is win-in-5. The total 34 participants on average solved 1.79 puzzles ( $SD = 1.398$ , range = 0-5).



No correlation is found between win-in-n and accuracy or between win-in-n and self-report difficulty

## Conclusion

- We designed a new think-aloud task and puzzles with adequate complexity and difficulty. Puzzle-wise analysis indicates that the depth of puzzles (win-in-n) not related to the participants' performance and self-report difficulty.
- Based on the results from 28 participants, there is no significant correlation between DOP and Elo.
- Based on the results from the 9 participants, there is no significant correlation between DOP and planning depth obtained from narrations.

## Acknowledgment

I would like to express my deepest appreciation to Wei Ji Ma Lab. I would like to say thank you to Jeroen Olieslagers, Jordan Lei, and Nastaran Arfaei for their advice. I'm extremely grateful to Jonathan Winters for his IT support. Last but not least, I want to send special thanks to Jieyu Li, who designed a powerful puzzle solver that helped me to validate and improve the puzzles.