Modeling Human Eye Movements with Neural Networks in a Maze-Solving Task

Jason Li[1], Nicholas Watters[1], Yingting (Sandy) Wang[1], Hansem Sohn[1], Mehrdad Jazayeri[1]

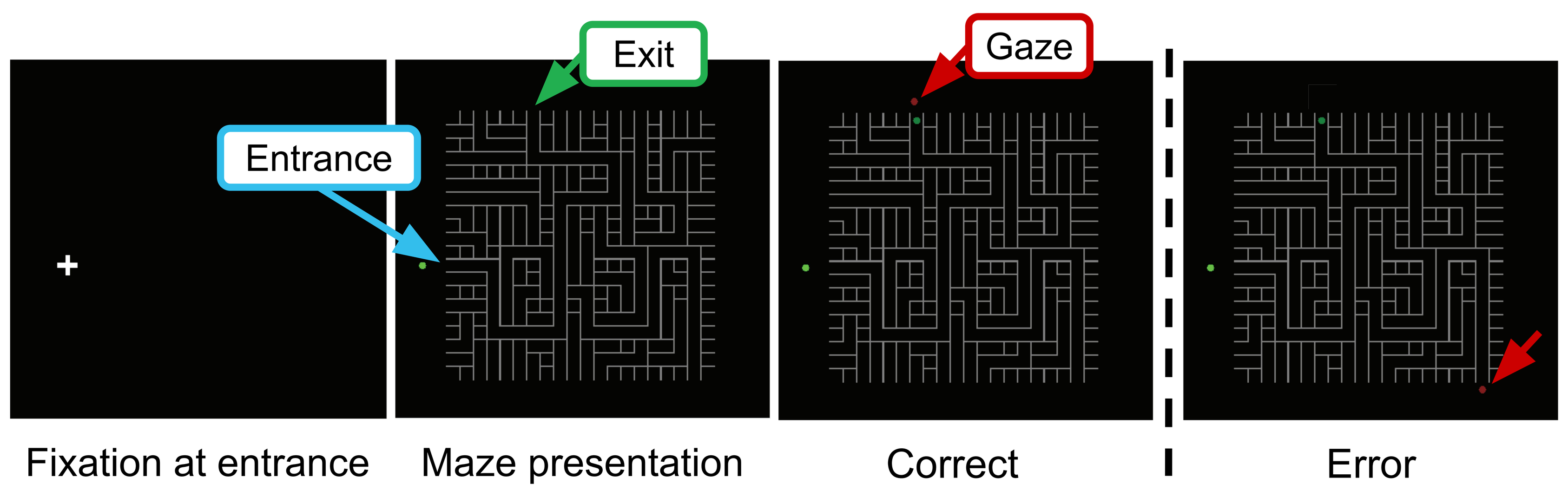
[1] McGovern Institute for Brain Research, Department of Brain & Cognitive Sciences, MIT

Abstract

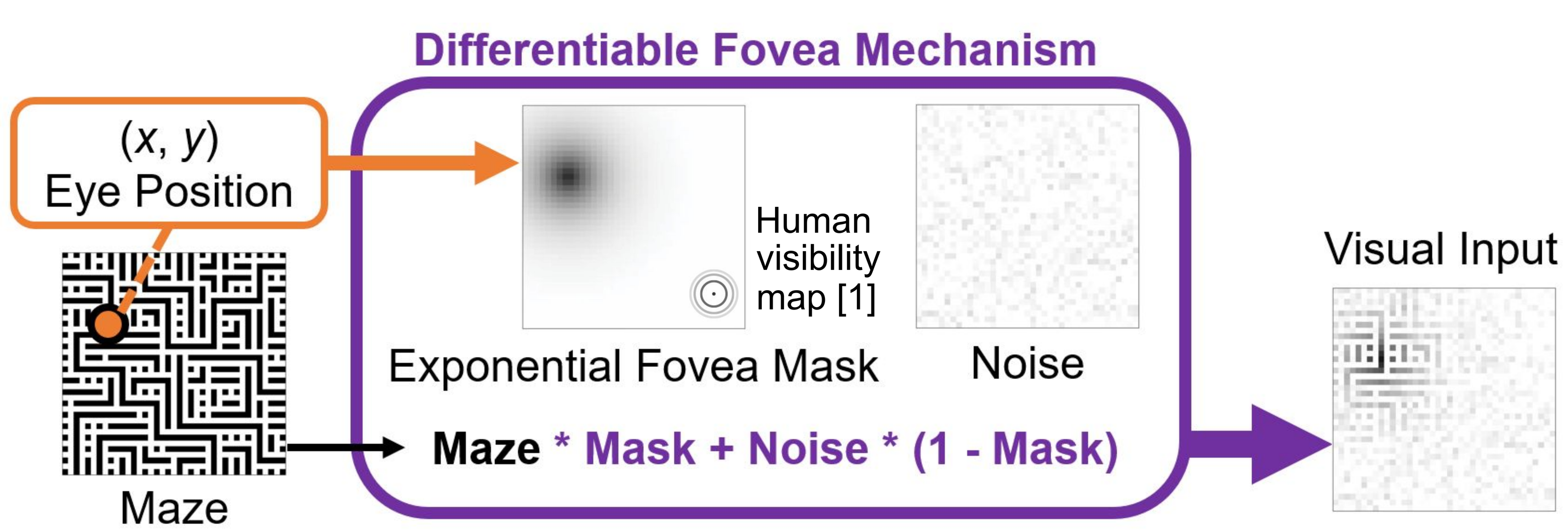
While solving visual tasks, humans employ a wide variety of context-dependent eye movement strategies. There is growing interest in understanding the computational logic of gaze control as it may serve as a window into the underlying mental processes. However, building models of human eye movements has proven notoriously difficult. Here, we tackle this problem in the context of a virtual maze task. We compared eye movement data collected from humans (*N=12*) to deep generative models with a novel differentiable architecture built to solve the same task with different objective functions. We found that human eye movements did not follow the patterns generated by a model optimized to perform the task with the smallest number of eye movements. Instead, eye movements were better captured by a model that was optimized to run an internal simulation of an object traversing the maze. These results not only provide a generative model of eye movements in a rich visual task but also provide tantalizing evidence that humans rely on mental simulations to solve maze tasks.

Task

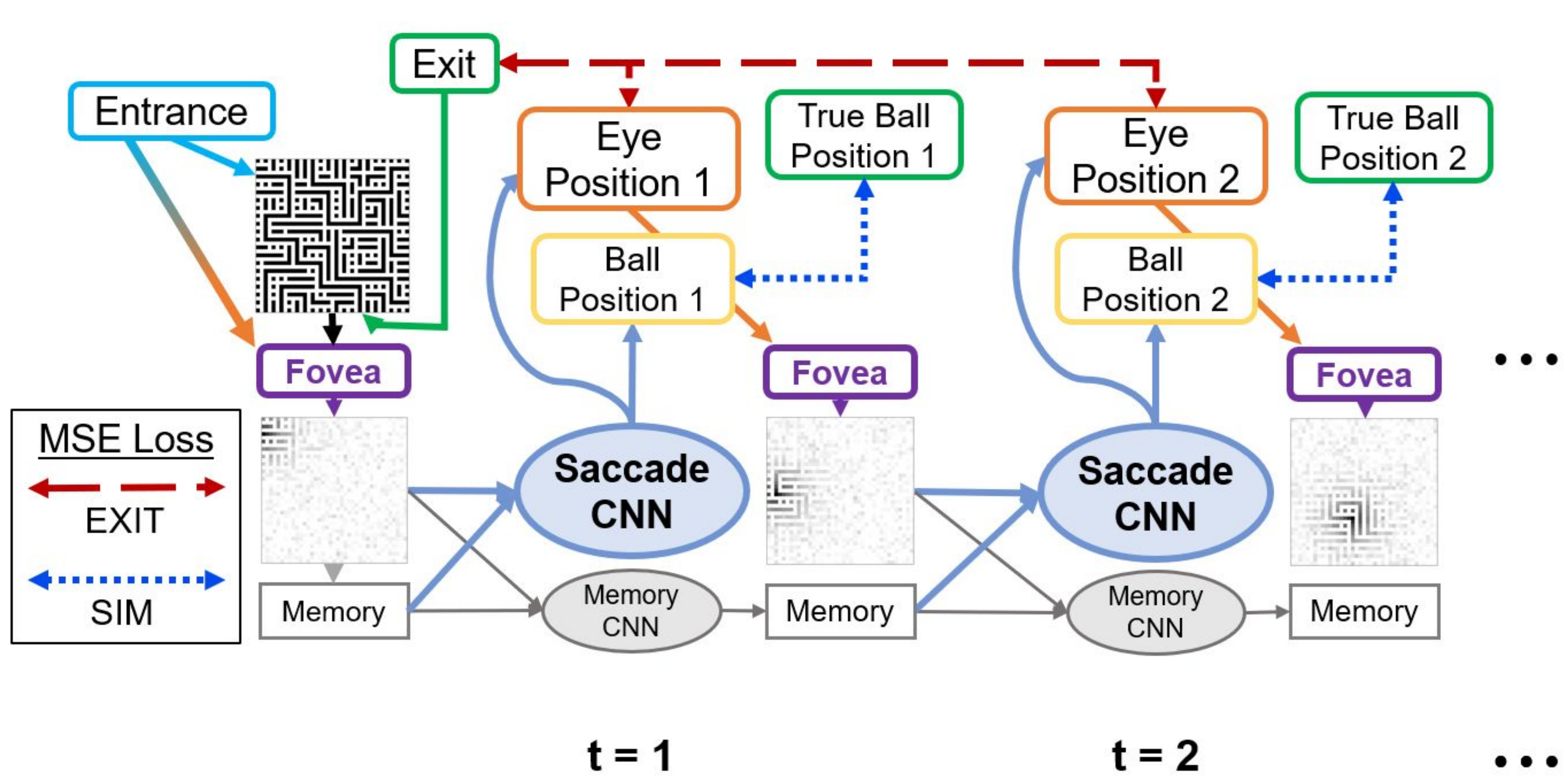
Maze solving: given **entrance** point, find **exit** point (a key press to report **gaze** position)



Gaze Recurrent Neural Networks



Convolutional neural net generates new eye position at each step;
eye position → **differentiable fovea mechanism** → next visual input



Models & Example Behavior

3 gaze RNNs trained with different objective functions:

- **EXIT**: Reach exit in as few saccades as possible
- **SIM**: Track an imaginary "ball" moving through the maze
- **HYBRID**: Weighted sum of the two loss terms

EXIT	Minimize $L_{\text{EXIT}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i^{\text{eye}} - p^{\text{exit}})^2$
SIM	Minimize $L_{\text{SIM}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i^{\text{ball}} - p_i^{\text{ball}})^2$
HYBRID	Minimize $L_{\text{HYBRID}} = \beta \cdot L_{\text{EXIT}} + (1-\beta)\cdot L_{\text{SIM}}$

where n is the number of fixation points,

$\hat{p}_i^{\text{eye}}$ is the model's i^{th} predicted eye position,

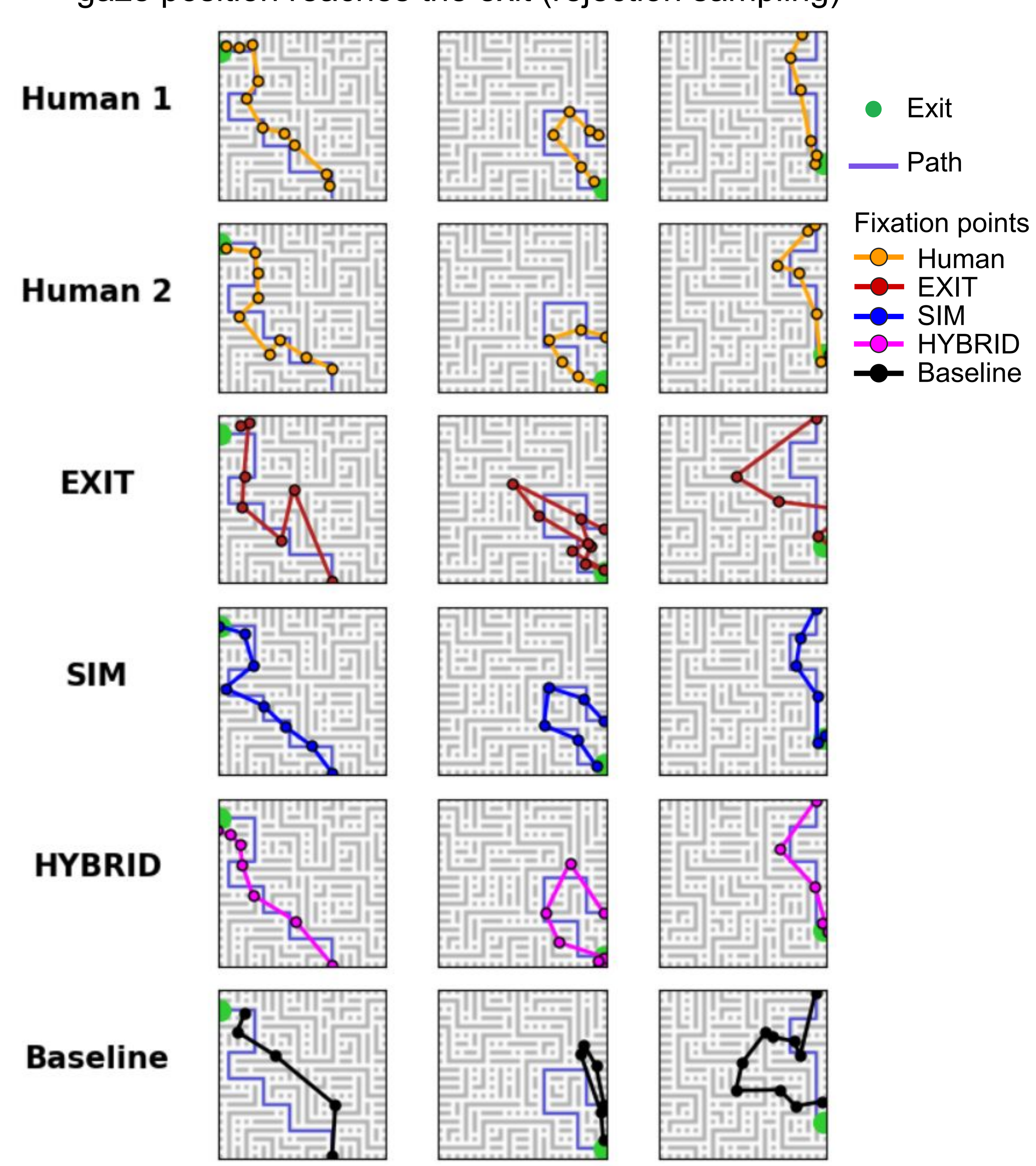
$\hat{p}_i^{\text{ball}}$ is the model's i^{th} predicted ball position,

p^{exit} is the maze's exit point, and

p_i^{ball} is the i^{th} true ball position.

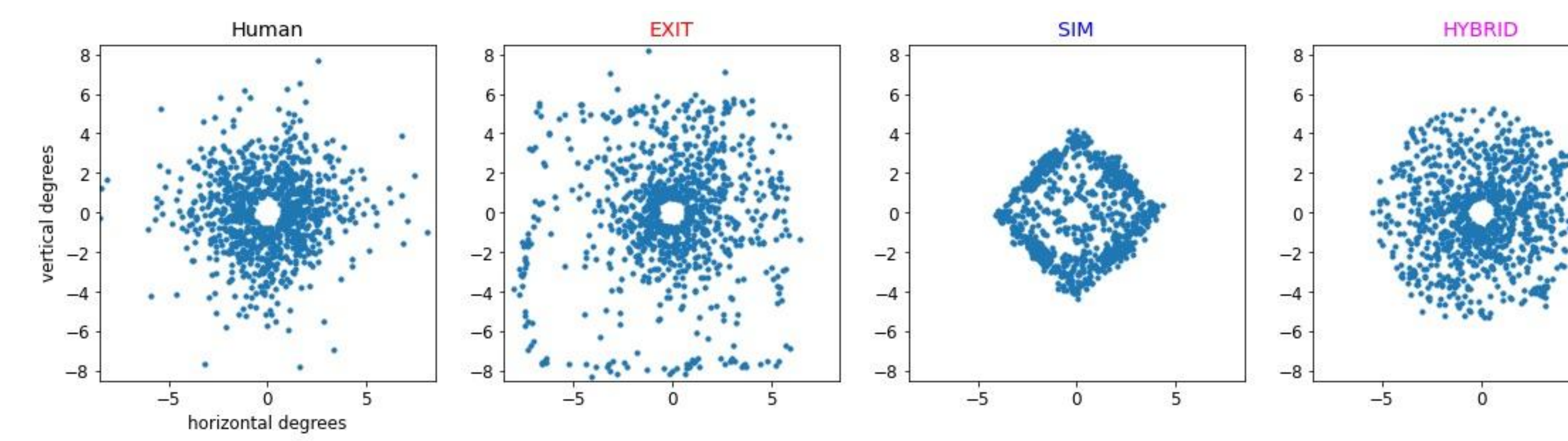
Baseline model

- Randomly sample saccade vectors from human data until the gaze position reaches the exit (rejection sampling)



Metrics & Results

Saccade Vector Distributions

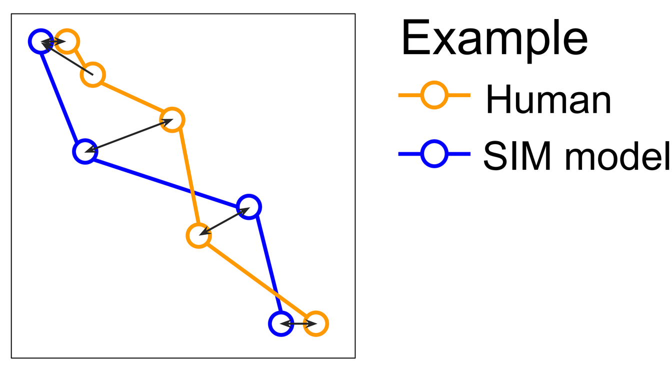


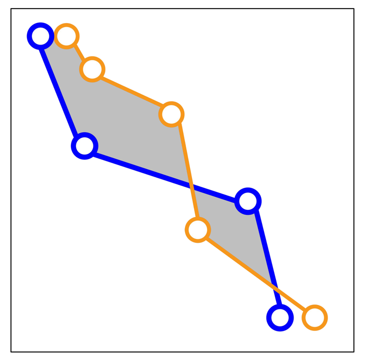
Nearest Neighbors Distance

Mean of the nearest point in path **A** to every point in path **B** and vice versa

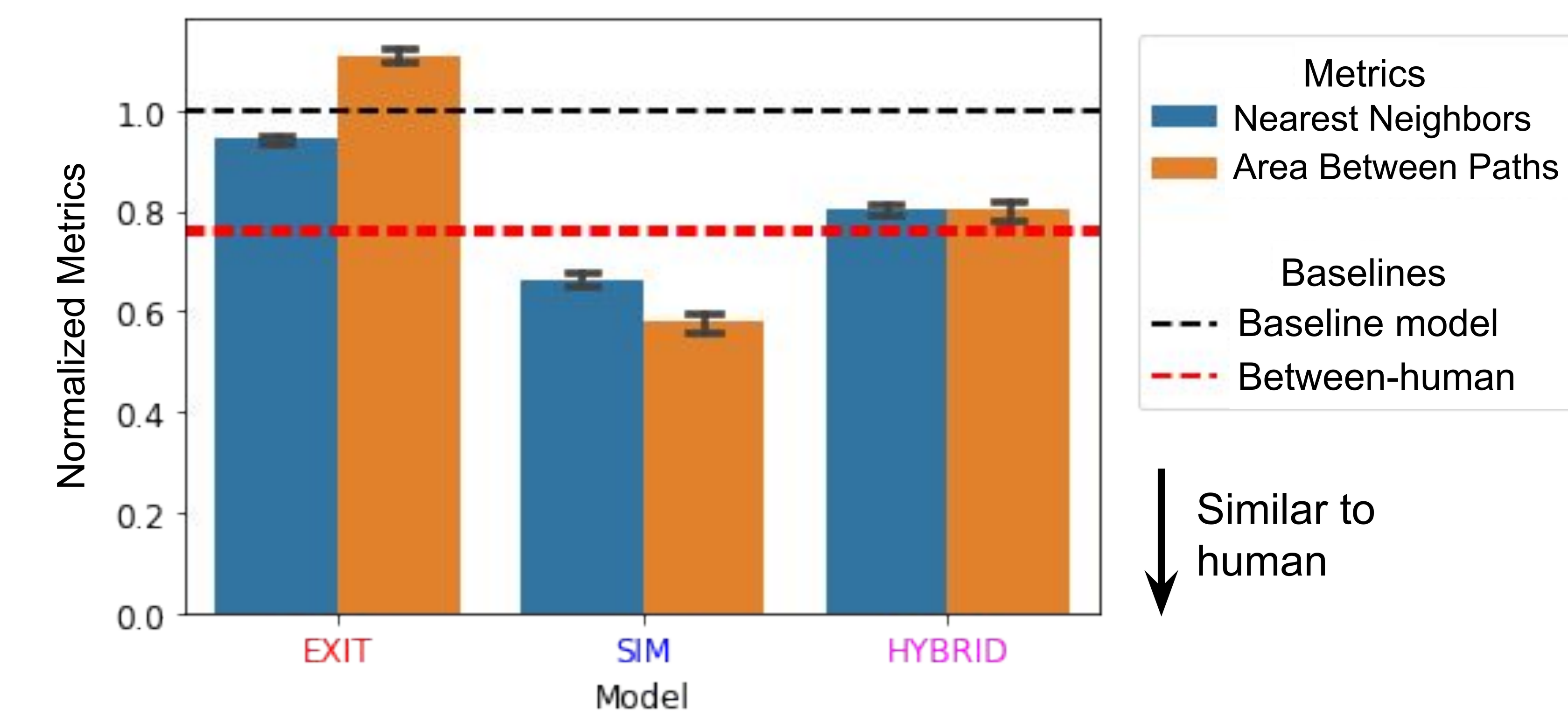
Area Between Paths

Total plane area of the polygon(s) formed between paths **A** and **B**





Metric scores between model and human eye paths



- Simulation model is most similar to human eye paths

Conclusions

- In a maze-solving task, a gaze RNN trained to run an internal simulation better matches human behavior than a model trained to solve the task as efficiently as possible.
- Humans may employ mental simulation when performing this task.

Future Directions

- Explore relationship between biological plausibility of fovea hyperparameters and model behavior.
- Apply our gaze RNNs to tasks beyond maze solving.

Reference

[1] Najemnik & Geisler, Nature (2005)