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]

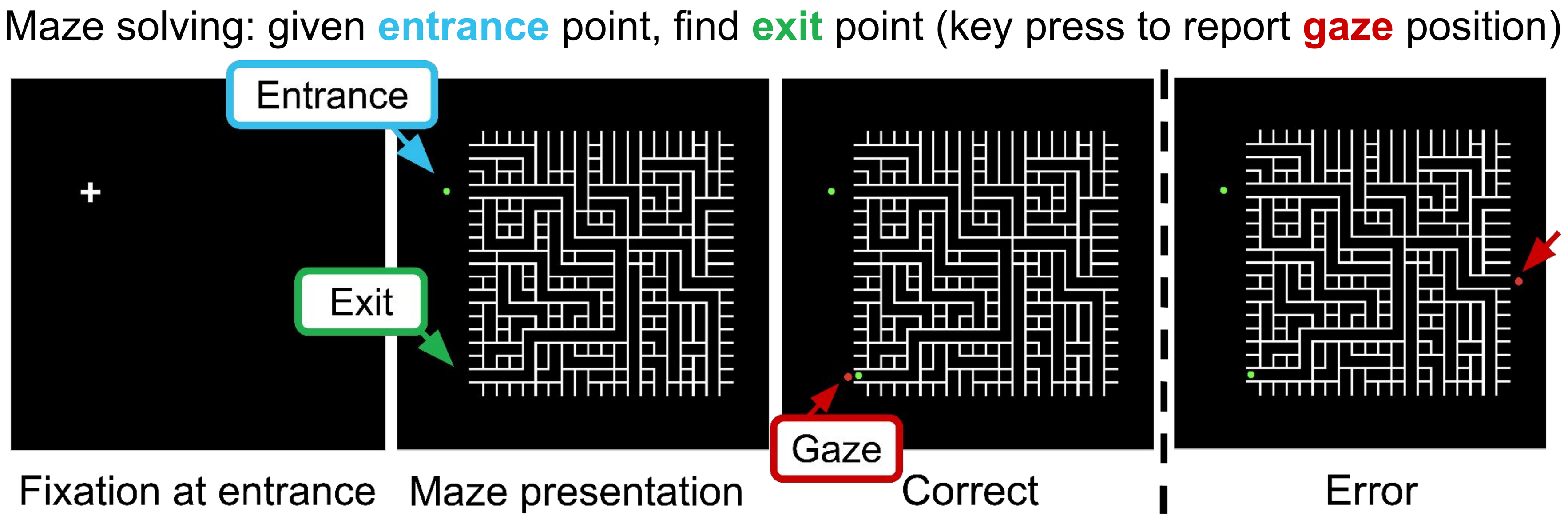
[1] McGovern Institute for Brain Research, Department of Brain & Cognitive Sciences, MIT

Abstract

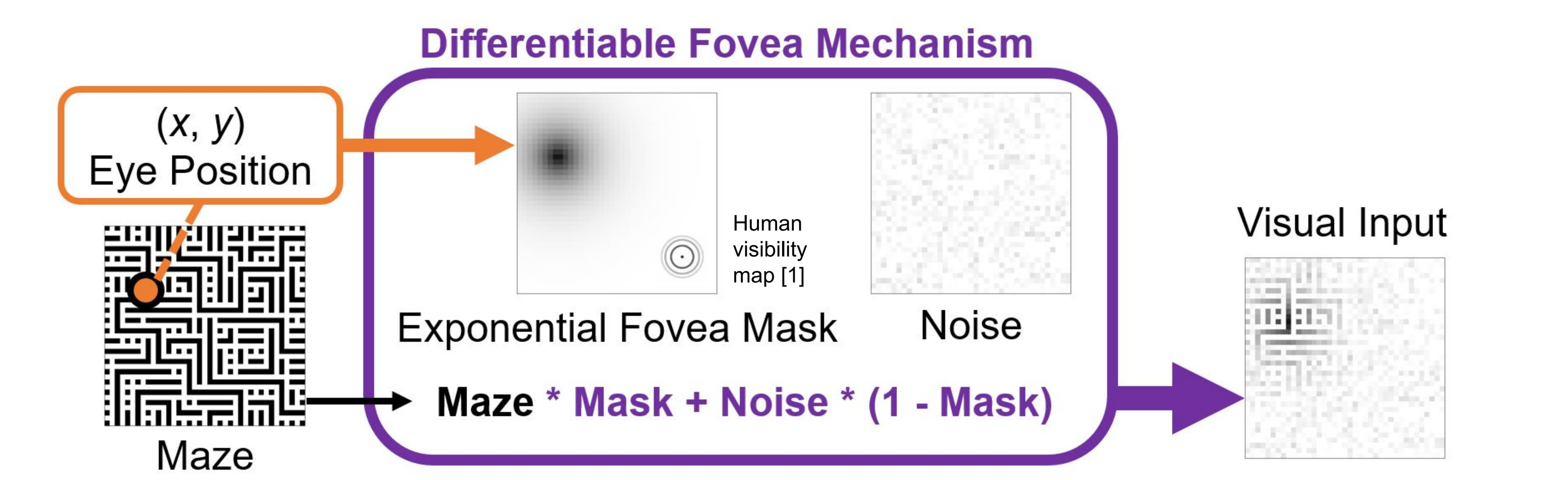
From smoothly pursuing moving objects to rapidly shifting gazes during visual search, humans employ a wide variety of eye movement strategies in different contexts. While eye movements provide a rich window into mental processes, building generative models of eye movements is notoriously difficult, and to date the computational objectives guiding eye movements remain largely a mystery. In this work, we tackled these problems in the context of a canonical spatial planning task, maze-solving. We collected eye movement data from human subjects and built deep generative models of eye movements using a novel differentiable architecture for gaze fixations and gaze shifts. We found that human eye movements are best predicted by a model that is optimized not to perform the task as efficiently as possible but instead to run an internal simulation of an object traversing the maze. This not only provides a generative model of eye movements in this task but also suggests a computational theory for how humans solve the task, namely that humans use mental simulation.

Task

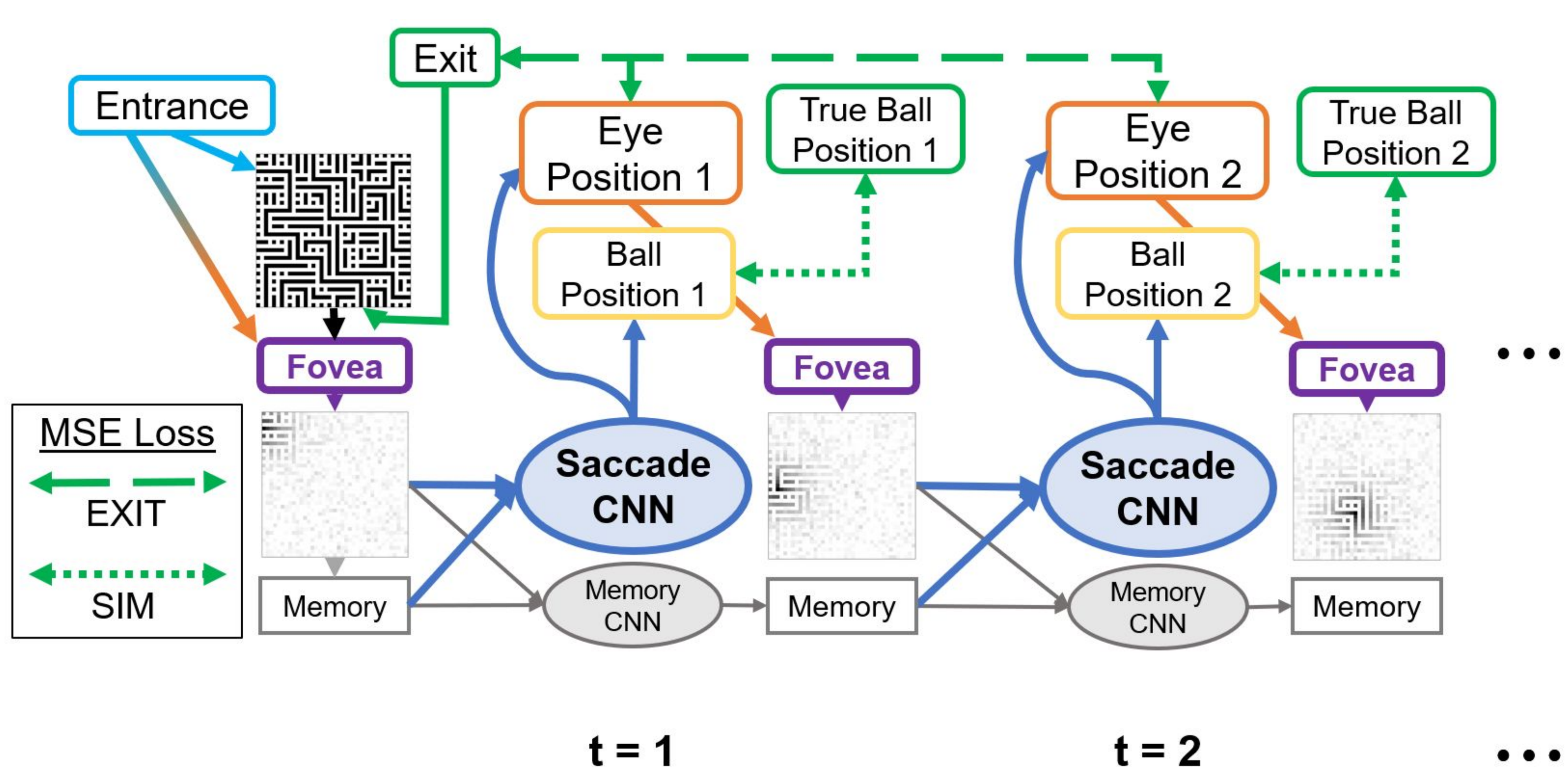
Maze solving: given entrance point, find exit point (key press to report gaze position)



Gaze Recurrent Neural Networks



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



Gaze RNN Models

3 gaze RNNs trained with different objective functions:

- **EXIT**: Reach exit in as few saccades as possible

Minimize $L_{\mathrm{EXIT}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i^{\mathrm{eye}} - p^{\mathrm{exit}})^2$

- **SIM**: Track an imaginary constant-velocity "ball" moving through the maze

Minimize $L_{\mathrm{SIM}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i^{\mathrm{ball}} - p_i^{\mathrm{ball}})^2$

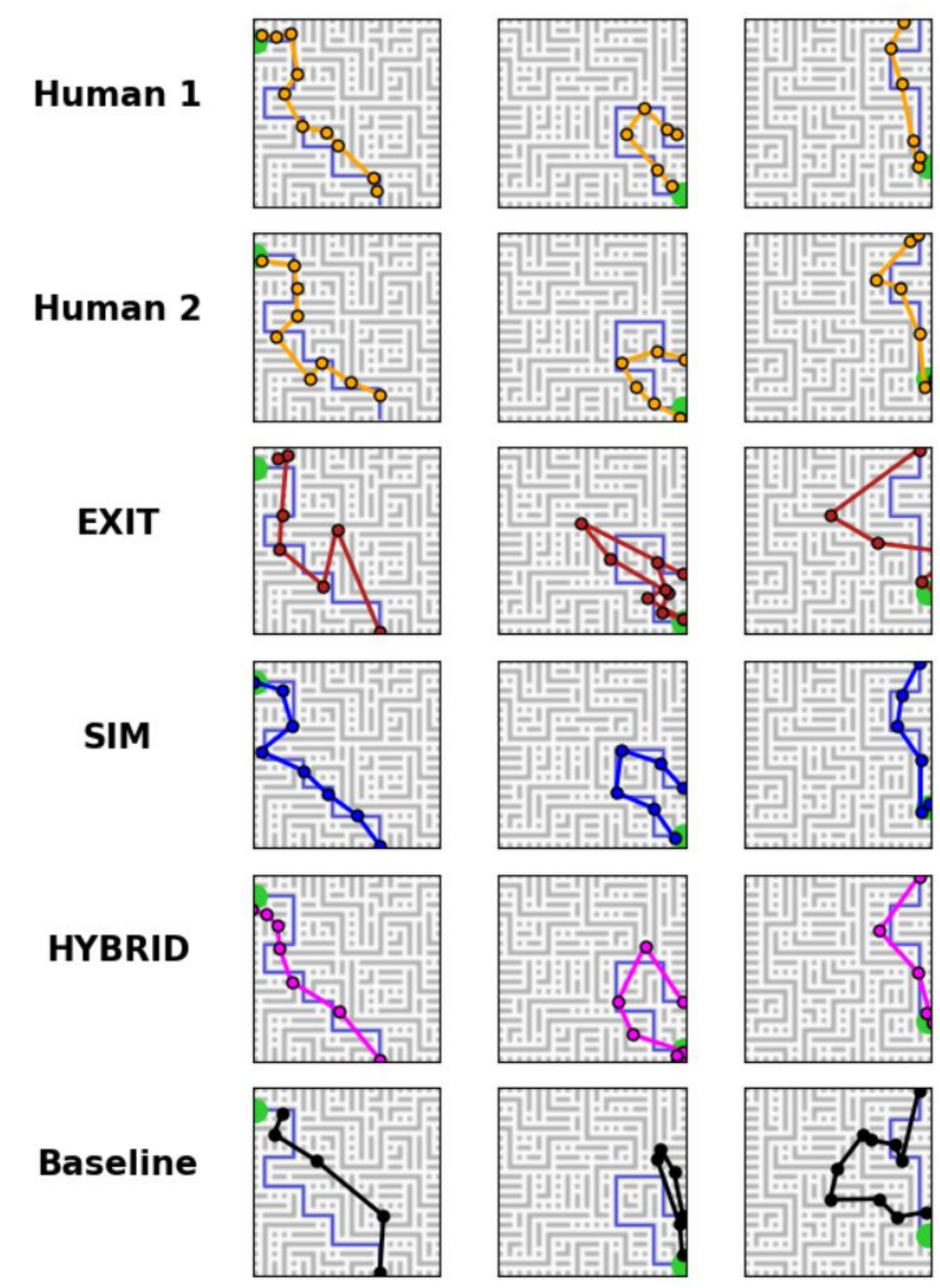
- **HYBRID**: Weighted sum of the two loss terms

Minimize $L_{\mathrm{HYBRID}} = \beta \cdot L_{\mathrm{EXIT}} + (1-\beta)\cdot L_{\mathrm{SIM}}$

where:

n	number of fixation positions
$\hat{p}_i^{\mathrm{eye}}$	model's i^{th} predicted eye position
$\hat{p}_i^{\mathrm{ball}}$	model's i^{th} predicted ball position
p^{exit}	maze's exit point
p_i^{ball}	i^{th} true ball position

Human and Model Behaviors



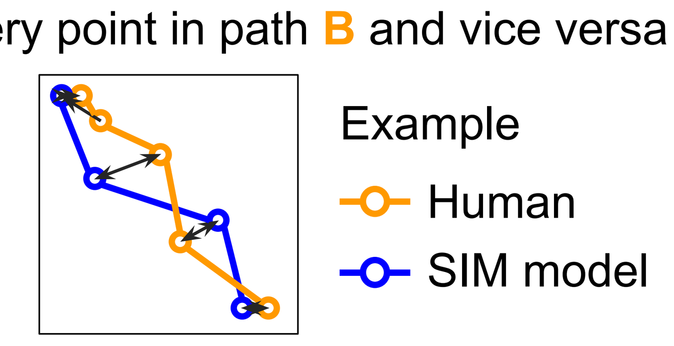
Baseline model

Construct eye paths by sampling saccade vectors from human distribution until the gaze position reaches the exit.

Metrics & Results

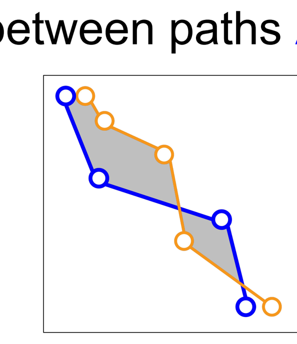
Nearest Neighbors Distance

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

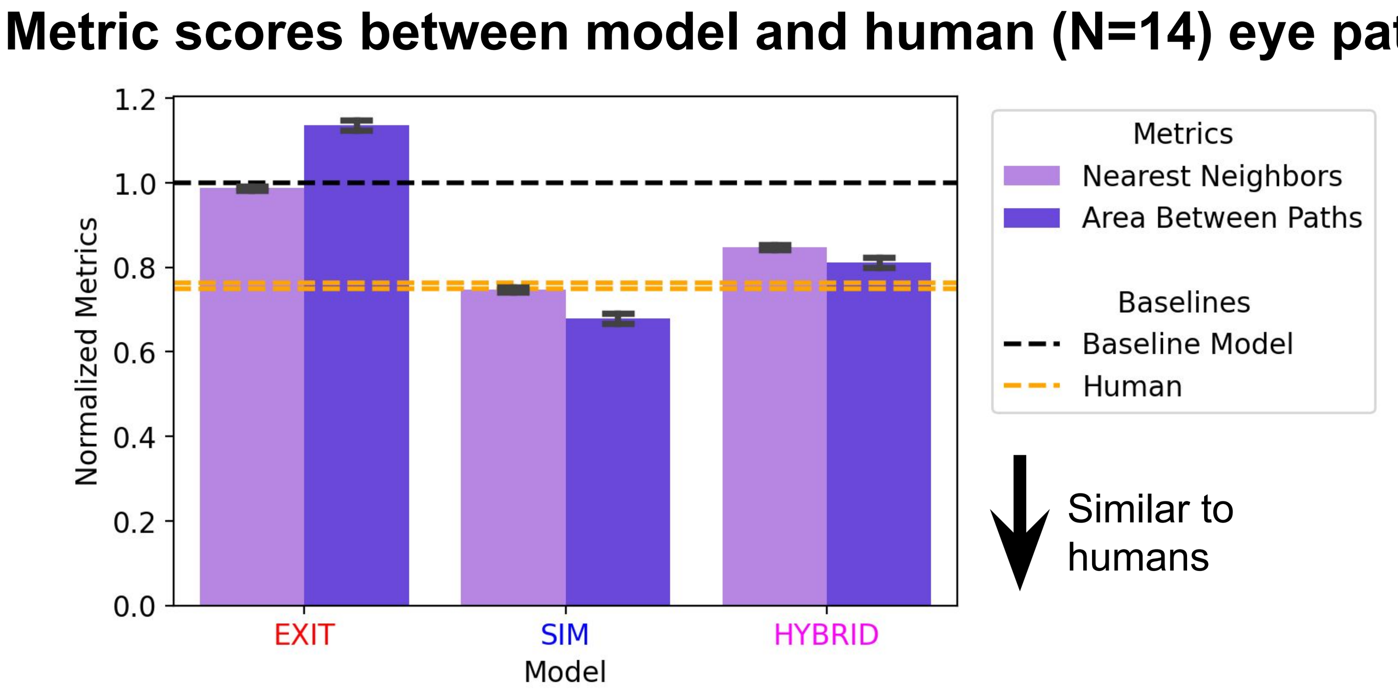


Area Between Paths

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

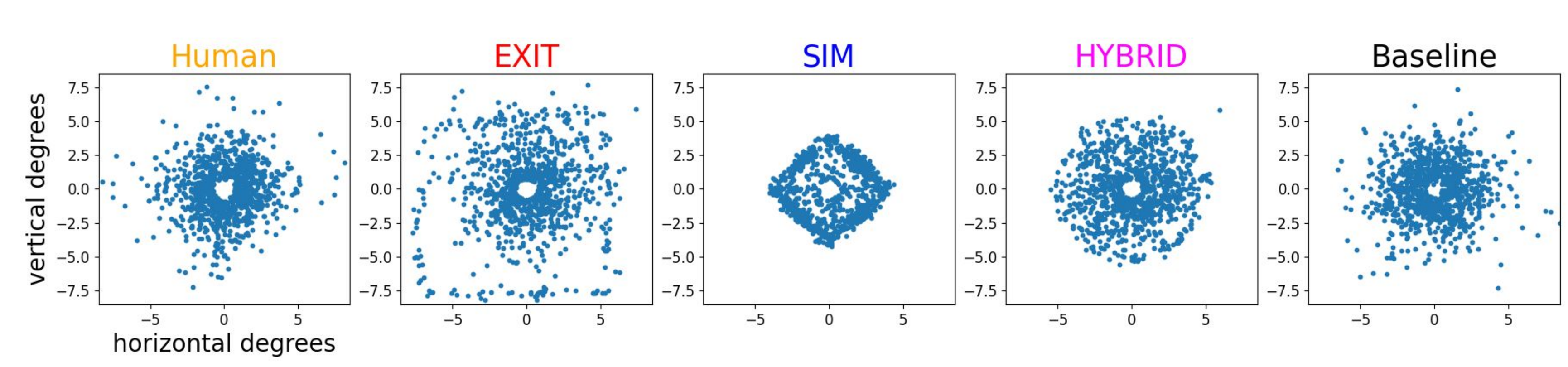


Metric scores between model and human (N=14) eye paths



- SIM model's eye paths are most similar to human eye paths
- EXIT model is comparable to the Baseline Model

Saccade Vector Distributions



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 a similar mental simulation when performing this task.

Future Directions

- Explore relationship between biological plausibility of fovea hyperparameters and model behavior.
- Introduce non-constant simulation speed.
- Allow gaze RNN to control fixation durations to model temporal dimension of eye movements.
- Apply our gaze RNNs to tasks beyond maze solving.

Reference [1] Najemnik & Geisler, Nature (2005)