Modeling the Basal Ganglia in Mice Through Reinforcement Learning in Maze Solving

Hooman Ramezani

Abstract

This report discusses the representation of the mouse brain's action selection mechanism with a deep reinforcement learning model. The model leverages the principle of Q-learning to represent the process of a mouse learning to solve a randomly generated maze and find a reward. Results indicate that while the model is able to effectively learn and solve randomly generated mazes, the model represents only a simplified abstraction of the mouse's physiological system. A more biologically representative model of the mouse's action selection and reinforcement learning mechanisms would exceed the computational resources by which this model is constrained.

Introduction

The purpose of this paper is to model the basal ganglia functionality from a mouse brain via deep reinforcement learning. The basal ganglia are known to provide a feedback loop with the cortex directly influencing motor functionality. Most of the neurons in the basal ganglia are inhibitory or dopaminergic. This is akin to the reward system of reinforcement learning. Since modeling the entire application of the basal ganglia is nearly impossible, this paper will focus on modeling the basal ganglia in a maze-solving application where a mouse is in a maze and it needs to find a "piece of cheese" (reward). This real-world test is commonly performed on mice and showcases well how the basal ganglia can be modeled with reinforcement learning, emulating learning through rewards [1]. In this case, the functionality and modeling of other related areas like the sensory cortex and the motor cortex will be abstracted away in terms of simulated actions. Overall the key functionality of the basal ganglia to be modeled through reinforcement learning will be its uses in action selection and learning.

Physiology Review

The basal ganglia play a significant role in action selection. The basal ganglia have involvement through two major pathways [4]. The first pathway is a direct pathway that facilitates action selection leading to outcomes associated with rewards [3]. The second pathway is the indirect pathway, which modifies action selection via inhibition of actions that are perceived to be negative or which are shown to lead to negative outcomes [3]. Note that the classification of outcomes that are considered rewarding is specified by input from the mouse cortex and the thalamus. These components provide information about the overview of the environment and the current goals and motivations of the individual [5]. These reward pathways are centered around dopamine release in the striatum of the basal ganglia [4]. Specifically, dopamine release in the striatum leads to modification of the activity in both the direct, and indirect pathways [4]. In the direct pathway, dopamine acts on D1 receptors causing an increase in the activity of the striatal neurons that project directly to the globus pallidus internus (GPi) and the pars reticulata (SNr) [4]. This activity disinhibits the thalamus since the striatal neurons will reduce the inhibitory influence of the globus pallidus and excite the motor cortex [4]. In the indirect pathway, dopamine acts on D2 receptors which leads to a decrease in the activity of striatal neurons that project onto the globus pallidus externus [4]. This activity disinhibits the subthalamic nucleus, which then sends excitatory signals to the GPi and the SNr. These signals inhibit the thalamus and by doing so, suppress movement [4]. These characteristics of excitation and inhibition of movement actions, depending on the mouse's perception of what actions will lead to rewards, are the specific system-level characteristics that are to be modeled by the network.

Moving onward, The hippocampus and visual cortex are two distinct regions in the brain with two different functions. However, integration is seen between these two brain regions in mice for important processes like spatial cognition, spatial navigation, and visual processing. The hippocampus plays an important role in learning and the construction of long-term memory [6]. Other studies have shown the hippocampus to be important for contextual information processing which serves as an important factor when it comes to spatial navigation [7]. The function of the visual cortex is to process visual stimuli to perform visual perception or object recognition tasks [8]. In terms of integration between the two regions, studies have shown hippocampus activation when mice view visual stimuli like objects or scenes indicating that there is integration between the two regions [9]. During spatial tasks, a synchronized firing pattern can be seen between the hippocampus and visual cortex of the mice seen in electrophysiological recordings [10]. When manipulating the neurons in the visual cortex, researchers observed that there can be modulation of the activity in the hippocampus based on the extent to which those neurons in the visual cortex were stimulated when it came to spatial memory tasks [11]. As previously mentioned the hippocampus plays an important role in spatial navigation, however, studies have shown visual cues are essential for hippocampal place cell firing in mice to provide an internal representation of the mice's surrounding environment [12]. These spatial representations that come from the visual cortex are essential for memory storage in the hippocampus [13]. Experiments have shown there be direct connections between (mainly) the primary visual cortex and other areas in the visual cortex like the secondary visual cortex and the posterior parietal cortex to the hippocampus via fluorescent tracers injected in the visual cortex to measure the uptake in the hippocampus [14]. This pathway is seen when the visual input from the visual cortex is sent along the parahippocampal cortex adjacent to the hippocampus and this information is relayed to the entorhinal cortex which is the main cortical input to the hippocampus [15].

The basal ganglia are also seen to be integrated with these two regions of the brain. The striatum in the basal ganglia receives direct inputs from the primary visual cortex which is integrated along with other sensory and motor inputs from other regions in the brain in order to output the proper motor commands [18]. This process indicates that the basal ganglia play an important role in visual perception and action planning [18]. As mentioned previously, the hippocampus is important for contextual information processing which is critical when signals from this region are sent to the striatum and the basal ganglia needs to select rewards-driven motor commands based on the current context [19]. This becomes especially prudent when it comes to spatial navigation and memory formation. Inputs are needed by the basal ganglia [19]. This all indicates that the basal ganglia, visual cortex, and hippocampus are all closely integrated with each region uniquely contributing to learning, memory, and decision-making.

Methods

Architecture & Hyperparameters:

The reinforcement model used in this paper is Q-learning. Q-learning is an approach that can learn the optimal solution but often forgoes safer paths like in the cliff walking example from class. The tradeoff is appropriate in this model as there is not much negative impact for the real or simulated mice when it walks into a wall.

The major hyperparameter that is analogous to a physiological function is mouse_exploration_factor, which is the frequency of exploration that the agent will do. Mice have been shown to have a balance of exploratory and task-centric behavior that is dependent on environmental conditions, previous experiences, age, and other characteristics of the specific mouse [1]. A study exploring the learning rate and the exploration of mice in a given maze found that mice initially exhibit exploratory behavior in a maze, but as they become more familiar with the layout they move towards task-centric behavior [1]. Specifically, mice are able to balance exploration and exploitation, with them exploring new paths when necessary but not wasting time going through paths that they know do not lead to the goal [1]. As such, mouse_exploration_factor is set to 0.1, where 10% of the agent's moves are exploratory. Setting it low allows for the agent to have some exploratory actions, but the vast majority of its moves are task-centric. This mirrors a mouse's ability to balance exploration with exploitation but the majority of its moves are geared toward achieving a reward.

Detailed specifically in the hyperparameters subsection of the results section, there is a set of hyperparameters designed to model mice's ability to assign value to specific actions over others. These hyperparameters add or subtract value from a reward function, based on whether actions taken by the agent move towards the goal (exploitation) or whether the actions involve an activity that does not directly move the agent towards the goal, such as exploration. These hyperparameters mirror mice's thalamic function in assigning value to potential actions based on the goals of the mouse in each environment [16]. Specifically, the thalamus is able to assign value based on prior experience and contextual information about what actions are thought to lead to rewards, and by doing so provide inhibitory and excitatory signals for specific mice actions [16].

<u>Training Data:</u>

The training data of the mouse brain model will be a solvable maze modeled by a 2D binary array. The array will contain 1's and 0's, this can then be used to represent a maze with walls, see the below figure for example.

```
[[1. 1. 0. 1. 1. 1. 0. 1.]
 [1. 0. 1. 1. 0. 1. 1. 1.]
 [1. 1. 0. 0. 1. 1. 0. 1.]
 [0. 1. 1. 0. 1. 0. 1. 1.]
 [1. 0. 1. 1. 1. 0. 1. 0.]
 [1. 0. 1. 0. 0. 1. 1. 1.]
 [1. 1. 0. 1. 0. 0. 1. 0.]
 [0. 1. 1. 1. 1. 1. 1. 1.]]
```

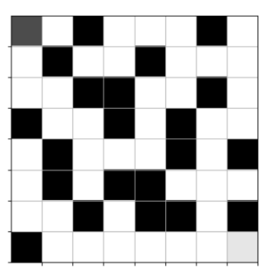
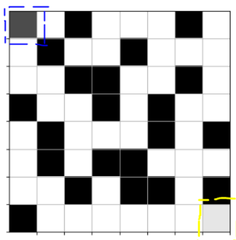


Figure 1. 2D Array and equivalent Maze Representation [DOJ, 2023]

For training our team will need a way to generate an unlimited amount of random solvable mazes. For this a depth-first search algorithm was applied. This method essentially goes through all cells in the maze flipping walls as it goes randomly until it reaches a desired cell, it backtracks when it gets stuck with nowhere else to go [2]. The maze generation was set up so that the starting point, cheese location, and total size of the maze were the same in all generations, this was chosen not because mice cannot solve complex mazes but in order to reduce the time to train and experiment. The generated code can be modified to support any sized maze with random start and cheese locations for future iterations.

<u>Simulating:</u>



In the figure to the left the square highlighted in blue represents the mouse and the square highlighted in yellow represents the cheese. To simulate going through the maze the location of the mouse is kept track of and the squares nearby are as well. This represents the mouse's visual field. The mouse can then act as an agent and have a list of four possible actions, go up, go down, go left, or go right. The reward (positive or negative) can then be given for moving further along, hitting a wall, and receiving the cheese. Rewards would diminish over time as the goal is to solve the maze as efficiently as possible, the mice would want to get the cheese as fast as possible. To promote the most direct solution a small negative reward of -0.04 is applied for each action.

<u>Training & Metrics:</u>

The goal is to train over a maze to allow the model to find an optimal policy that maximizes return for solving the maze. Some key metrics to be considered and compared for modeling the physiological system of a mouse would be win rate, loss, and episodes (how fast the mouse can get to an optimal policy).

Results

The training process was overall positive. Deep reinforcement learning with q-learning combined ultimately successfully trained a policy for our mice to traverse a randomly generated grid in order to reach the "cheese" (reward). Below in Figure 2, we have a matplotlib-generated gif that demonstrates a successful traversal of a grid in order to solve the maze. Black squares indicate regions in which the mouse cannot visit, the "dark grey" square represents the mouse, and the "light grey" squares indicate the path of the mouse. In each maze case, the location of the reward was always the bottom right-most corner of the maze.

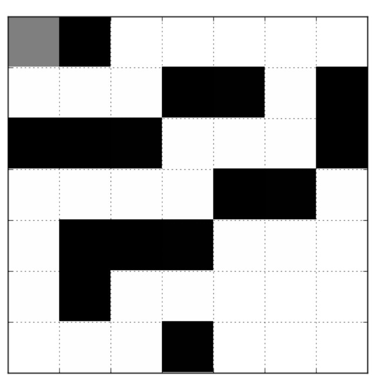


Figure 2: GIF Demonstrating Mouse Successfully Traversing a Maze. Start Top Left. Cheese (end): Bottom Right [HR, 2023]

Hyperparameters and Success Metrics

In order for training to be successful, a suite of hyperparameters had to be optimized in order to produce results in a reasonably efficient manner. Hyperparameters' final values for training were as follows:

Table 1: Hyperparameter Tuning Results [HR, 2023]

Hyperparemeter	What It Does	Final Value	Justification
epochs	How many times an entire dataset of samples is passed through the algorithm to update values and optimize.	1000	High values were decided to be okay, as the experiment automatically will stop when 100% win rate was chosen
max_memory	Maximum number of game experiences we keep in memory	8 * maze. size	Through trial and error, this value was found to train relatively quickly and be within GPU memory constraints
data_size	Number of samples we use in each training epoch. This is the number of episodes (or game experiences) that we randomly select from our experiences repository	32	Through trial and error, this value was found to train relatively quickly and be within GPU memory constraints
discount_factor	coefficient, usually denoted by γ, which is required for the Bellman equation for stochastic environments	0.95	Found as a common value to satisfy Bellman's equation for our type of problem through literature review
valid_move_penalty	Each new valid move will be penalized slightly. This will encourage the model to finish faster for a max return	0.05	Want small discouragement per move to encourage most efficient path
hit_wall_penalty	Hitting a wall applies a huge	0.75	Both penalties signify an

	penalty. This will encourage the mouse to not run into walls		invalid move that we want to strongly penalize
leave_maze_penalty	Leaving the maze applies a huge penalty. This will encourage the mouse to not leave the maze	0.75	
explore_visited_cell	Rat is penalized for going back to visited cells. Having to revisit a cell indicates a non-optimal path, this is penalized	0.25	Penalized, not as heavily since it is technically allowed but not encouraged
end_game_threshold	If the total return dips under this value, the game. The purpose is to avoid infinite loops from going in circles	-0.5 * maze size	Found experimentally as a good threshold to give up
reach_cheese_reward	Reward gained from getting to the cheese. A reward can only be gained from getting the cheese, this is the only real "reward" a mouse would get	1.0	Maximum reward possible
mouse_exploration_factor	exploration / exploitation split	0.1	Literature results on mice behavior while in a maze, show a balance between exploratory and task-centric behavior [1]. However, the majority of moves made were shown to be task-centric. Therefore, mouse_exploration_factor is set to 0.1 to reflect 10% exploration and 90% exploitation moves, which roughly mirrors mice.

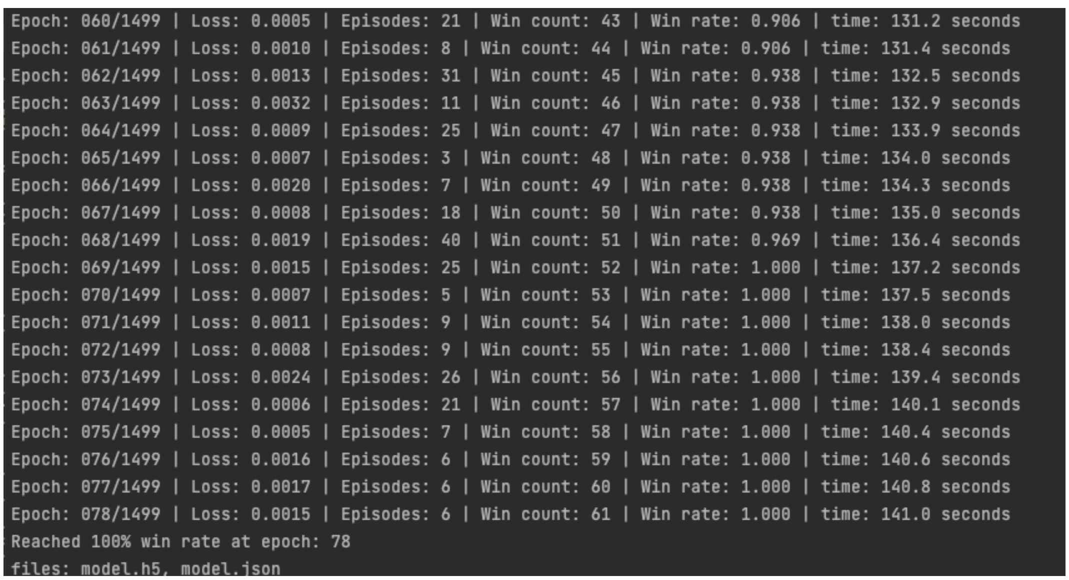
Table 2. Key Model Metrics of Success [HR, 2023]

Metric	**Explanation**
Loss	*MSE* Loss from *NN*
Episode's	Measure of Learning efficiency: A successful Q-learning model should be able to learn the optimal policy with as few iterations as possible. This can be measured by the total number of episodes required for the model to converge or

	by the time taken to achieve a satisfactory level of performance. Low Values are better.
Win Count	Total wins through the training process
Win Rate	Percent of episodes in an epoch that are won

Summary of Training

Final trained model was able to win 100% of the time, with minimum loss. Further testing of the model demonstrated a strong ability to generalize and find optimal paths, validated through a testing function we called completion_check where we calculated all paths to find the optimal one. The policy was able to find the optimal path within 100 epochs, and consistently win the game. Last 10 epochs contained 100% win rate, signaling the model is well-configured and signaling early stopping.



```
Epoch: 060/1499 | Loss: 0.0005 | Episodes: 21 | Win count: 43 | Win rate: 0.906 | time: 131.2 seconds
Epoch: 061/1499 | Loss: 0.0010 | Episodes: 8 | Win count: 44 | Win rate: 0.906 | time: 131.4 seconds
Epoch: 062/1499 | Loss: 0.0013 | Episodes: 31 | Win count: 45 | Win rate: 0.938 | time: 132.5 seconds
Epoch: 063/1499 | Loss: 0.0032 | Episodes: 11 | Win count: 46 | Win rate: 0.938 | time: 132.9 seconds
Epoch: 064/1499 | Loss: 0.0009 | Episodes: 25 | Win count: 47 | Win rate: 0.938 | time: 133.9 seconds
Epoch: 065/1499 | Loss: 0.0007 | Episodes: 3 | Win count: 48 | Win rate: 0.938 | time: 134.0 seconds
Epoch: 066/1499 | Loss: 0.0020 | Episodes: 7 | Win count: 49 | Win rate: 0.938 | time: 134.3 seconds
Epoch: 067/1499 | Loss: 0.0008 | Episodes: 18 | Win count: 50 | Win rate: 0.938 | time: 135.0 seconds
Epoch: 068/1499 | Loss: 0.0019 | Episodes: 40 | Win count: 51 | Win rate: 0.969 | time: 136.4 seconds
Epoch: 069/1499 | Loss: 0.0015 | Episodes: 25 | Win count: 52 | Win rate: 1.000 | time: 137.2 seconds
Epoch: 070/1499 | Loss: 0.0007 | Episodes: 5 | Win count: 53 | Win rate: 1.000 | time: 137.5 seconds
Epoch: 071/1499 | Loss: 0.0011 | Episodes: 9 | Win count: 54 | Win rate: 1.000 | time: 138.0 seconds
Epoch: 072/1499 | Loss: 0.0008 | Episodes: 9 | Win count: 55 | Win rate: 1.000 | time: 138.4 seconds
Epoch: 073/1499 | Loss: 0.0024 | Episodes: 26 | Win count: 56 | Win rate: 1.000 | time: 139.4 seconds
Epoch: 074/1499 | Loss: 0.0006 | Episodes: 21 | Win count: 57 | Win rate: 1.000 | time: 140.1 seconds
Epoch: 075/1499 | Loss: 0.0005 | Episodes: 7 | Win count: 58 | Win rate: 1.000 | time: 140.4 seconds
Epoch: 076/1499 | Loss: 0.0016 | Episodes: 6 | Win count: 59 | Win rate: 1.000 | time: 140.6 seconds
Epoch: 077/1499 | Loss: 0.0017 | Episodes: 6 | Win count: 60 | Win rate: 1.000 | time: 140.8 seconds
Epoch: 078/1499 | Loss: 0.0015 | Episodes: 6 | Win count: 61 | Win rate: 1.000 | time: 141.0 seconds
Reached 100% win rate at epoch: 78
files: model.h5, model.json
```

Figure 3: Demonstration of Successful Model Training [HR, 2023]

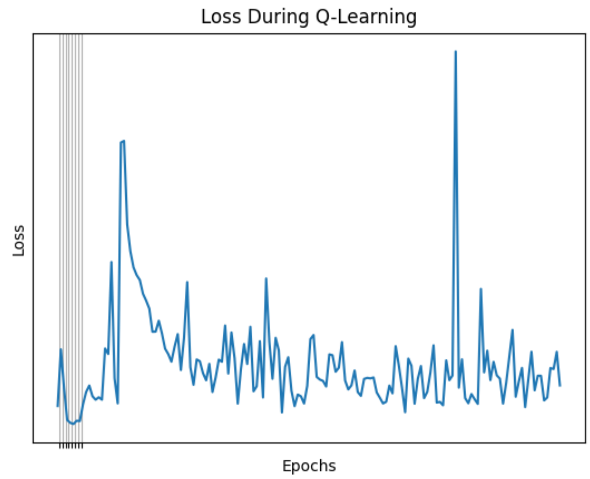


Figure 4: Loss Graph During Training Mouse on Mazes [HR, 2023]

Discussion

Before analyzing the results procured by the model it is important to take into consideration the key differences between the reinforcement Q-learning model used to solve the maze problem, and the basal-ganglia present in a mouse brain. Firstly, while the Q-learning model implemented using TensorFlow and Keras only has one direct pathway, the mouse brain has two pathways that affect action selection, (one direct and one indirect) [4]. Consequently, the machine learning model represents a simplification of the mouse brain's action selection mechanism. The resulting model is sequential and consists of an excitatory dense layer of neurons representing the mouse's visual cortex. The state of the current environment from the mouse's perspective is fed into the layer. The output of this layer is fed into a subsequent dense layer representing the basal ganglia itself. Outputs from the layer representing the basal ganglia are then fed into an output layer whose output is a simple numerical value corresponding to the appropriate action that the model predicts to be the best ranging from 0 to 3.

Contrasting the simple model to that of the mouse's action selection mechanism discussed in the aforementioned physiology review it is abundantly clear that while the model is abstractly biologically representative, it is not an exact representation. The reason for this simplification is attributed in large part to a lack of computing resources available. When constructing the model, the team experimented with more dense layers, and, hidden convolutional layers that were more biologically representative of the basal ganglia. These experiments resulted in training times far exceeding reasonable timeframes for feasible development given the computational resources available. The consequence of this resource limitation is that the model developed and discussed as part of this report should not be seen as a biologically representative model of the mouse's action selection mechanism, but rather an abstract biologically inspired representation of the mouse's action selection mechanism.

With that being said, the results of the model can still provide helpful insights into the model's physiological representation. As the model was trained, it became better at performing the maze, indicating that the model was learning with experience as intended. This is shown in Figure 4 presenting the loss incurred while training the model. It can be seen that the loss decreases steadily over time. While it has been suggested that consistent training of mice also yields similar results, recent studies have debated this model and have shown that mice appear to learn discontinuously and can have moments of “sudden insight” in which they solve a key element of a maze [17]. This suggests that the underlying search algorithm does not require a global memory of places visited and is largely explained by purely local turning rules, in contrast to the model [17].

The steady decrease loss over time in Figure 4 can be related to the hippocampus’ ability to aid with spatial navigation and memory which improves as the mouse becomes more aware of its surroundings when integrated with the visual cortex and basal ganglia, the mouse is able to make more efficient motor decisions from the basal ganglia over time. The visual cues represented in this model whether it was the walls or cheese represented rewards-based decision-making in the mouse to guide the decision-making process of the basal ganglia. Although this Q-learning model does not explicitly incorporate the hippocampus and rather has dense layers representing the visual cortex and basal ganglia, the improvement in win rate and decrease in loss indicates improvement in spatial navigation efficiency and accuracy over time which in mice is only possible with the aid of the hippocampus’ stored memory. Overall, the model provides a solid representation of basal ganglia’s reward-based decision-making process efficiency over time with the aid of the visual cortex and hippocampus.

Despite the debate about continuous or discontinuous learning, the model provides physiological insight into the learning rate of the basal ganglia when it comes to specific tasks. Therefore, the model can be modified to more accurately reflect mice learning rates found via empirical research. By doing this the model can represent a healthy baseline function for the learning rate when it comes to action selection. This can be useful information when it comes to understanding how modifications to the basal ganglia by diseases such as Parkinson’s or Huntington’s affect action selection. Specifically, making analogous changes to the model based on current research into the basal ganglia damage done by these diseases can allow the model to provide representative insight into how these diseases affect the basal ganglia’s ability to modulate action selectivity and may even provide mechanistic insight into these changes. Furthermore, the model is based on action selection in the specific task of pathfinding. Extrapolating the model to different tasks can provide information on the physiological mechanisms that underlie reward processing and decision-making in these tasks. This can provide holistic data on how the brain modifies functionality to different environments, as well as provide information on the physiological mechanisms underlying these modifications.

References

[1] M. Rosenberg, T. Zhang, P. Perona, and M. Meister, “Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration,” *eLife*, vol. 10, 2021.

[2] A. Kozlova, J. A. Brown, and E. Reading, “Examination of representational expression in maze generation algorithms,” *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, 2015.

[3]Freeze, B. S., Kravitz, A. V., Hammack, N., Berke, J. D., & Kreitzer, A. C. (2013). Control of basal ganglia output by direct and indirect pathway projection neurons. The Journal of Neuroscience, 33(47), 18531–18539. https://doi.org/10.1523/jneurosci.1278-13.2013

[4]Groenewegen, H. J. (2003). The basal ganglia and motor control. Neural Plasticity, 10(1-2), 107–120. https://doi.org/10.1155/np.2003.107

[5] Haber, S. N. (2016). Corticostriatal circuitry. Neuroscience in the 21st Century, 1721–1741. https://doi.org/10.1007/978-1-4939-3474-4_135

[6] Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychological Review, 99(2), 195-231.

[7] Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: is it spatial memory or a memory space? Neuron, 23(2), 209-226.

[8] Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology, 160(1), 106-154.

[9] Komorowski, R. W., Manns, J. R., & Eichenbaum, H. (2013). Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. Journal of Neuroscience, 33(24), 10075-10086.

[10] Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. Nature, 543(7646), 719–722. https://doi.org/10.1038/nature21692

[11] Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W. C., Jenvay, S., Miyamichi, K., Luo, L., & Dan, Y. (2020). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science, 370(6512), eaba4170. https://doi.org/10.1126/science.aba4170

[12] Aghajan, Z. M., Acharya, L., Moore, J. J., Cushman, J. D., Vuong, C., & Mehta, M. R. (2015). Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. Nature Neuroscience, 18(1), 121–128. https://doi.org/10.1038/nn.3884

[13] Cembrowski, M. S., Wang, L., Lemire, A. L., Copeland, M., & DiLisio, S. F. (2018). Spatial integration by the hippocampal-prefrontal circuit. Current Opinion in Neurobiology, 52, 127–135. https://doi.org/10.1016/j.conb.2018.05.006

[14] Zhou, M., Liang, F., Xiong, X. R., Li, L., Li, H., Xiao, Z., Tao, H. W., & Zhang, L. I. (2018). Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. Nature Neuroscience, 21(12), 1631–1640. https://doi.org/10.1038/s41593-018-0264-4

[15] Burwell, R. D. (2000). The parahippocampal region: corticocortical connectivity. Annals of the New York Academy of Sciences, 911(1), 25-42.

[16] Campus, P., Covelo, I. R., Kim, Y., Parsegian, A., Kuhn, B. N., Lopez, S. A., Neumaier, J. F., Ferguson, S. M., Solberg Woods, L. C., Sarter, M., & Flagel, S. B. (2019). The paraventricular thalamus is a critical mediator of top-down control of cue-motivated behavior in rats. ELife, 8. https://doi.org/10.7554/elife.49041

[17] Matthew Rosenberg, Tony Zhang, Pietro Perona, Markus Meister (2021) Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration eLife 10:e66175 https://doi.org/10.7554/eLife.66175

[18] Hong, S., & Hikosaka, O. (2008). The globus pallidus sends reward-related signals to the lateral habenula. Neuron, 60(4), 720-729.

[19] Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. Nature Reviews Neuroscience, 1(1), 41-50.

Individual Report(s)

<u>Hooman:</u>

Contribution	Time	Collaborators	Impact
Coding - Setup of Deep Learning System. ● Researched how mazes can be solved by deep learning systems ● Conducted literature review on Q-learning systems for maze solving ● Proposed deep q-learning	11 hours	Sammy	Major software development contribution to the ultimate end goal of a biologically representative maze traversal of mouse. Architecture trained

architecture for our use case and researched its implementation ● Implemented deep q learning in code, classes to explore maze, with help of sammy which included: ○ Q_maze class that contains the representation of maze ○ Play_game class to begin the game ○ Completion_check class to find optimal path for given maze ○ Build deep learning model to help with training and find optimal parameters ○ Qtrain function which featured the training loop ● Modified the proposed architecture to be more biologically representative ● Connected GPU cluster to mitigate issue with training, would take hours to train, vs minutes after integrating GPU into training			efficiently, optimally to find the best path. Crucial contribution in order for the deep learning component of the project to be finished.
Report - Results ● Consolidated results from training into tables and figures for the reader to understand and ideally be able to recreate ● Demonstrated the training process, model hyperparameters, training time, and how the architecture was configured to our needs	4 hours		Presents final hyperparameters, metrics, loss graph, training results, and visualizations from the training process. Demonstrates success in the project.
Report - Editing ● Edited all sections of report, focus on citations and accuracy of information	1 hour		Ensured accuracy of information and consistency, as well as citations throughout the report.

David Olive Jr:

Contribution	Time	Collaborators	Impact
Report - Introduction ● Provided an explanation of the goals with a clear and concise description of the system to be modeled and what method from SYDE 552 our group would be using to model it ● Required a high-level understanding of the project and undertakings	1 hours		Allows the reader of the report to easily understand right away what systems would be modeled and with what. This accounted for around 5% of the report.
Report - Methods ● Went through and explained the key methodological details for the system model with help from Hooman and Sammy who did much of the model architecture. Researched and justified appropriate metrics and hyperparameters based on the desired comparison. Also explained the data set used, how it was generated and how the model was simulated	4 hours	Hooman, Sammy	This section is useful to the reader as it provides clear steps and guidelines on how to reproduce the model to get similar results and carry on with the work. This accounted for around 20% of the report.
Coding - Dataset Generation & Simulation ● Conducted research and developed functions to generate random, solvable mazes. Also coded up the way for simulating actions within the maze and emulating a block of cheese and a simple mouse ● Required research into algorithms for generating solvable random mazes, worked with Hooman and Sammy to integrate into model training	8 hours	Hooman, Sammy	This piece of coding is crucial for the project as the model cannot be trained without this and creating a large enough dataset by hand would be impossible.

Group 7 Project Report

Sammy Robens-Paradise:

Contribution	Time	Collaborators	Impact
Report - Discussion ● Analyzed results derived from the training of the model and compared them to the biological system of the mouse's action selection mechanism ● Discussed limitations of the model relative to the mouse brain	4 hours	Mahima, Mohammad	Allows the reader to understand the conclusions that can be drawn from the model as well as understand the limitations of the methods
Report - Abstract ● Wrote the abstract outlining the general goal and conclusion of the report	25 min	-	Abstracts are a necessary component of any formal paper or report
Report - Editing ● Conducted edits of the report to ensure accuracy and consistency of claims	1.5 hours	-	Report editing is a key step in ensuring the accuracy of claims and ensuring readability of the report content
Coding - System architecture and reinforcement learning systems design. ● Constructed object-oriented class structure to generate, and maintain the state of a current maze ● Wrote helper function to determine the appropriate rewards that should be delivered to the mouse based on its various actions ● Wrote helper function to determine the validity of a mouses action ● Wote a helper function to get the	11 hours	Hooman	Core development of the infrastructure to create an executable python file that can train a reinforcement q-learning model on a maze randomly generated maze is key to the projects success.

current status of a maze game that the mouse is currently "playing" ● Constructed an "experience" class to maintain the current memory of the mouse over time so that it can effectively learn based on a given model ● Developed ergonomic interface to extract metrics from the code alongside Hooman so that results can be easily evaluated			

Mahima Ariyaratne:

Contribution	Time	Collaborators	Impact
Literature Review/ Report - Physiology Review: Basal Ganglia physiology and function ● Conducted literature review on the basal ganglia's role in action selection, and the underlying physiological mechanisms ● Searched for mechanistic information on the direct and indirect pathways and the specific structures involved ● Looked for dopamine involvement in the direct and indirect pathway and how it functions in reward processing ● Researched information on how the thalamus and cortex are involved in determining what actions are considered rewarding ● From there wrote a concise summary of the major points of this research in the physiology review section	7 hours	Mohammad	Provides a starting point for how the model should function, given that it is supposed to reflect the basal ganglia function of action selection. Also provides the reader a summary of the major points about the basal ganglia, which can be used for comparison when looking at the overview of the model.
Literature Review/ Report - Methods: Hyperparameters: Physiological justification ● Conducted literature review on	4 hours	David	Determines values for the hyperparameters that are

mice's ability to navigate mazes, specifically looking for information on exploration vs task-centric behavior ● Researched how the thalamus of mice assign value to specific functions ● From there wrote in the methods section on the physiological justification for specific hyperparameters which are analogous to basal ganglia function			physiologically representative of basal ganglia function. Also provides reader information on how the hyperparameters were determined and why they are set to those values.
Report - Discussion ● Discussed specific physiological insight from the results of the model	3 hours	Sammy Mohammad	Gives the reader information on why the model provides valuable data, and how the results can be used for future research.

<u>Mohammad Abrar:</u>

Contribution	Time	Collaborators	Impact
Literature Review/ Report - Physiology Review: Basal Ganglia physiology and function ● Conducted literature review on the physiological roles and mechanisms of the visual cortex and hippocampus ● Searched for how these two regions were related to one another relevant to the experiment in this paper ● Looked for the physiological connectivity of these two regions of the brain ● Researched information on these two regions of the brain are integrated with the basal ganaglia and how they aid in the decision making process	9 hours	Mahima	Gives the reader a solid understanding of the function and mechanisms of the visual cortex and hippocampus and also provides an understanding of how all 3 of these regions in the brain are integrated in order to perform different tasks.

• Found relevant research papers that provided experiments that would prove the links between these different regions of the brain • From multiple papers I provided a short, concise summary of papers that provided evidence to how these regions functions and/or how they interacted with other relevant regions of the grain			
Report - Discussion • Discussed how different aspects of the results and model related to the different relevant areas in the mouse brain • Analyzed how the results were an accurate representation of what would occur in a mouse brain if this experiment was actually done in a laboratory setting	3 hours	Sammy Mahima	Gives the reader context on how the results and model is related to the regions in the brain mentioned in the introduction and the degree of accuracy of that representation.

Images and Tables are attributed to authors unless otherwise specified.