The human in this scenario would look at the map of the maze. They would then decide what is the best path. Then they would move down then check the square when they didn’t find the treasure there, they would then move to a new square and repeat the next step until they found the treasure, or they reached the minimum score threshold. This process will help determine a path. The human will then see if there is a way to make the path better.

The intelligent agent then will use a rewards system to determine the best possible outcome based on the number-based system. This is then tested against all the other options and then finally the best possible choice is made for the computer and then it will start over and do the same thing until the system is finished or the minimum threshold has been reached, this will cause the system to shut down and the system will determine points and rewards. The system will then tell you whether you won or lost the game.

Both the human and the learning agent are both using similar styles. The system uses the point system to determine its way through the maze based on the number of points that it receives and if it is above the minimum score threshold and then it is repeated. Humans will analyze and check for squares and if they have the treasure. The human will skip the points where there is a block for the pirate. This will cause humans to be able to solve it faster. Humans will be less likely to try a random move and will instead stick to what has worked for it in the past.

The exploration-exploitation trade-off is a fundamental dilemma whenever you learn about the world by trying things out. The dilemma is between choosing what you know and getting something close to what you expect (‘exploitation’) and choosing something you aren’t sure about and possibly learning more (‘exploration’). (Yang, 2022) I feel that the best ratio would be the value of 0.5. I think that this is a good one because for the most part because we are trying to find where the treasure would be. The only issue becomes that we have no idea when we first start out where the treasure would even be. This causes us to need to do more exploration. When we have more values and a better understanding we could switch to a more exploitation-based system of finding the treasure but 0.5 would be the best for right now.

When it comes to reinforcement learning it is very similar to what would happen with a human player learning the game. The player would get rewarded every time that it makes a right move and would feel sad when it makes a wrong move. The reward is based on a points system in the AI and when there is a mistake points are taken away. Then over time the agent will learn how to earn the most points and therefore be the happiest. This will lead to more experimentation to see if there is a better path forward than the first one that was found. Then it will make a few calculations and adjust as needed. Then, when this is completed and based on the data it will start to win the game every time. Then this will cause it to feel accomplished and this will cause it to show that it can win the game.

Deep Q-learning was implemented by asking the agent to determine the best path forward. The agent would then select a direction. The system would then use the neural network to determine how many points it had earned, how that compared to other directions, how this change affected the overall point system and then determine if it wanted to do it again. When it made the decision, it then used deep Q-learning to decide the best path for it to precede and when it came across an answer that it didn’t like then it would find a better way to victory. This was repeated over several games and several attempts. Then it would take the failures and learn not to use those moves again and from there it would be able to solve the maze faster and a more efficient time frame

Beysolow II, T. (2019). *Applied Reinforcement Learning with Python* (1st ed., Vol. 1) [EBook]. Apress L.P. https://ebookcentral-proquest-com.ezproxy.snhu.edu/lib/snhu-ebooks/detail.action?docID=5880718#

Gulli, A., & Pal, S. (2017). *Deep Learning with Keras : Get to Grips with the Basics of Keras to Implement Fast and Efficient Deep-Learning Models* (1st ed., Vol. 1) [EBook]. Packt Publishing Limited. https://ebookcentral-proquest-com.ezproxy.snhu.edu/lib/snhu-ebooks/detail.action?docID=4850536

Yang, A. (2023, March 17). What is Exploration vs. Exploitation in Reinforcement Learning? *Medium*. https://angelina-yang.medium.com/what-is-exploration-vs-exploitation-in-reinforcement-learning-a3b96dcc9503#:~:text=The%20exploration%2Dexploitation%20trade%2Doff,more%20('exploration').