The way a human and the intelligent agent solve the maze are different. The main difference is the time it takes to solve the maze. A human can solve the maze in 1 or 2 attempts because there is a clear visual path to the end tile. A human can easily see the way in which a successful path is formed while the intelligent agent doesn’t. The intelligent agent repeatedly randomly goes from one tile to the next, giving a point deduction for every move until it fails or reaches the goal. It continues to do this over and over while saving the records. The intelligent agent continues hundreds if not thousands of times until it reaches a 100% win rate. For more complex mazes both human and intelligent agents have similar approaches by trying to find the goal in the highest effective way through multiple attempts. The difference is at what moves are made and why, a human will make a move knowing it approaches then closer to the goal, while the intelligent agent will do it randomly until it learns what moves are better at certain points in the maze.

The difference between exploration and exploitation when it comes to pathfinding is the reason behind the system making a move. During exploration the system will make moves in order to better understand its surroundings. While exploitation is making a move into promising areas based off of the exploration done. The ideal ratio between exploration and exploitation would be 1:9, 10% exploration and 90% exploitation. The reasoning is that exploration is mainly used to find areas that have potential and exploitation is the way the system knows if that potential was good or bad. If an area is seen as having potential, but doesn’t result well, it is better for the system to know this earlier on.

Reinforced learning can help determine a path to the goal because of the way it is built. The system is giving points based on each move it makes; reinforced learning gives the system the notion of maximising the cumulative reward, i.e. points. A pathfinding system using reinforced learning will prioritise gaining the maximum amount of points which in most cases results in the best route to the goal. A system like this could reduce the amount of time needed to get a 100% win rate.

A deep Q-learning system using neural networks would be implemented by having a means to save and replay previous attempts. The system would learn the current state environment, the action chosen, the reward, and the next state of the environment. After the system has experience it will be able to make better choices and gain a better result which will eventually lead to the system solving the maze and getting a 100% win rate.