# **CITS3001 Assignment Report – ThreeChess**

Lachlan D Whang (22466497), Tye Bridgeman (22326712)

There exist a wide variety of research and algorithms developed for the standard, 2-player game of Chess. 3-Player Chess is an interesting variant on this widely popular game which introduces a new degree of entropy and a vast number of new board layouts, and the difficulty involved with training intelligent agents to play a game of N-Player Chess grows exponentially as one observes the state-space explosion with increasing values of N. Thankfully, there exist a number of algorithms which can be used to generalize the agent learning process for any arbitrary number of players in a single game of chess. While a perfect game-playing agent may still be some distance away from our current efforts, this document will explore some of the suitable techniques which we can use to implement a rudimentary agent for 3-Player Chess.

**Greedy Algorithms**

All game-playing algorithms explored in this document work to maximize some utility based on the information available. Basic greedy algorithms constitute the most basic, game-playing agents whose moves are not random nor uninformed. At each turn, greedy agents make a move which maximizes the utility value in the short-term, in hopes that such moves will lead to maximized utility values in the long run. This utility value could be the number of pieces taken by the agent, or the total value of the points of pieces taken, or the total number of pieces remaining on the board that belong to the agent – the list goes on.

**Active Learning & Q-Learning**

The active reinforcement learning problem is characterized by a lack of a fixed policy, unknown reward and unknown transition models, with the latter 2 properties being present in Markov Decision Processes. An intelligent, game-playing agent in an active reinforcement learning environment needs to be able to select the “best” move based on the information it currently has access to at every turn, by calculating the utilities of states (and actions, depending on the algorithm), and selecting the action out of all possible actions in the current state that will produce the maximum known utility value.

Basic greedy algorithms can be applied to the active learning space. One simple greedy algorithm which undergoes the active learning process could involve the following 3 steps in the main game-playing loop.

1. Sense the current state and associated reward .
2. Call the **Adaptive Dynamic Programming** algorithm (a passive learning algorithm) to update the estimates for all previous utilities.
3. Choose the next, optimal action using the same equation for selecting an optimal action in Markov Decision Processes…

… where is the set of actions that can be performed in the current state .

The main problem with this basic, active learning, greedy algorithm is that it **does** **not** **converge on the optimal policy**. The algorithm actively exploits to the maximum possible extent permissible by the bounds of its knowledge, and exploration is minimal.

Q-Learning is the de-facto algorithm for solving the active reinforcement learning problem. The algorithm maintains a set of state-action pairs and their utilities, updating them after every move. Q-Learning is superior to basic greedy algorithms in that it balances **exploitation** with **exploration**, artificially inflating the utility values of state-action pairs that have been visited some number of times less than a predefined hyperparameter, such that when the algorithm attempts to select the action which maximizes this utility, it will explore the results of these actions. Exploration occurs most frequently towards the start of the learning process, and as the algorithm learns the outcome of more actions or moves, the -function responsible for the inflation of these utilities will start to present the actual, calculated utility value to the algorithm, instead of the inflated value, indicating a shift towards exploitation once the algorithm has “matured”.

The biggest challenge facing implementation of a Q-Learning agent is the method by which the algorithm’s knowledge will be **persisted** between training and playing sessions. Once the agent’s state is torn down at the end of a game, the information it has learnt needs to be stored. **I/O methods** are clearly the most straightforward and simplest way to persist this knowledge, but the functions are slow in comparison to the rest of the algorithm, and if the agent is unable to determine whether a position represents the end of a game, it will have to update stored utilities on disk after performing each move – a major detriment to its performance, although it is possible that this situation can be avoided depending on the information available.

**References**

Govindan, S. and Wilson, R., 2007. *Faculty & Research › Working Papers › A Decomposition Algorithm For N-Player Games A Decomposition Algorithm For N-Player Games*. PhD. Stanford Graduate School of Business.

Schrittwieser, J. et al., 2019. *Mastering Atari, Go, Chess And Shogi By Planning With A Learned Model*. PhD. Cornell University.