## 1. Introduction

Hanabi is a cooperative card game with incomplete information. Each player is able to view every other player’s hand, but not their own. The goal of the game is to complete five separate stacks representing different coloured fireworks. A player has the ability to play/discard a card or give a hint to another player.

The inability to access all information within a game brings its unique challenges. This paper explores the possible strategies that help combat this limitation and explores the selected strategies for our own agents playing the game.

## 2. Literature Review Of Suitable Techniques

### 2.1 Incomplete Information And Its Effect On The Game

When participating in a game of Hanabi an agent is able to access every other player’s hands, but not its own. The agent has **incomplete information** when assessing what action will maximise total utility for the game.

Humans are able to alter their perception of their (inaccessible) hand and note which cards are seemingly worthless or of value based on the behaviour of others [Solving Hanabi]. An agent is limited in this respect.

An agent needs to acquire information from each action performed by every other agent in the game.

When a player gives a hint, it is important to recognise that this encounters a cost of one hint token. The information gained should be considered carefully. A hint gives two pieces of information. Those cards that *are*, and those cards that are *not*.

A discard can be interpreted that the player does not have enough information to play confidently or enough information has been given to discard a card.

Various articles offer different approaches to handle this restriction. [Solving Hanabi] offers an outline for a “self-recognition” strategy. This is built on top of a rule based agent as described below.

### 2.2 Rule Based Approach

When playing Hanabi, intuitively a number of rules become apparent when attempting to maximise the end score of the game.

When all fireworks are empty, any card with the value of “one” can be played

When all fireworks stacks are above a certain value, any cards below/equal to this value can be discarded

There is only a single five for each colour so they must not be discarded when in a player’s hand

2s, 3s and 4s all have duplicates – however if any one of these is discarded the remaining card must be treated appropriately.

These rules can be hard-code quite easily into a simple reflex agent. However, issues arise when introducing the concept of hints. Which hint will maximise the final payoff? How is an agent meant to interpret a hint? An agent is limited by the number of hint tokens available too.

The approach considered and implemented was that of a conservative/cautious agent which would follow a hierarchy of actions.

Play a card that is guaranteed to work

Give a hint to another players card which should maximise payout

Discard a card

Hints were prioritised by the amount of information that can be obtained from them. It was important that discarding was limited as it was a risk and could limit the maximum score obtained.

For every state, a utility was assigned to each card based on the information known. Therefore, the current agent was able to identify (based on the above rules) which of their cards and other players cards were important, and which were not.

Furthermore, hints to other players were checked against a memory bank so as to make sure agents weren’t wasting hint tokens on giving the same hints.

### 2.3 Self-Recognition Strategy

Since Hanabi is a cooperative game it is in each player’s own interest to play rationality. That is, to play so that each action attempts to maximise utility. To predict other agent’s actions, it is not unreasonable to simulate our own rational agent and record what actions it took. Any agents which are uncooperative will harm both everyone else’s utility and its own, an undesirable outcome.

The Self-Recognition Strategy assumes each agent has the same strategy. It also has a similar strategy to the rule-based agent with an additional step explored below.

Initially every possible hand our agent could have is generated. We then simulate the action the previous player would have played, given each of these hypothetical hands. If the hypothetical action is the same as the actual action the hand is kept in memory, otherwise it is discarded. This continues until all hypothetical hands have been exhausted. The two cards with the greatest chance of occurring within the agent’s hand is stored in the variables and . If then we estimate that this card is within the player’s hand. [Solving Hanabi] found that a threshold value () of 2.5 is optimal.

An agent with this strategy infers the state of their hand by using the assumed behaviour of others to interpret new information. This agent appears to act humanly.

### 2.4 Monte Carlo Tree Search

The Monte Carlo Tree Search

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8080417>

//~~~~~~~~~~~~~~~~~~` my work as below

## Rationale Of Selected Technique

Ultimately we settled on two agents. The primary agent implements a simple version of the outer-state strategy.

* Uses utility to prioritize playable cards and important other player cards to hint
  + Will prioritize playable cards in other players hands, if it finds none and has excess hint tokens it may give a hint that reveals the most information to a player
  + Else hint at random if too many hint tokens
  + Keeps memory of hints other players has received, able to give complimentary hints for other players needing information, will not give duplicate hints, only plays cards with certainty
* Self-recognition requires generating superset of hands and selecting subset of possible hands recursively, too much overhead
* Monte Carlo impractical for current implementation, cannot simulate other players hands as current players hand cannot be seen by itself due to restrictions=too much uncertainty as player is not allowed to see own hand and thus cannot simulate another player looking at the player’s hands, not much value in simulating round where one hand is absent, generalizing monte carlo breaks it
* When allowing risky moves with probability, you can only make as many mistakes as you have extra lives, which in this case is 2, so you’re not really gaining much from it, possible gain in score is limited by the shared limit of 2 mistakes per game across 3-5 players

## Implementation Description

## Validation

Solving Hanabi: Estimating Hands By Opponent’s Actions In Cooperative Game With Incomplete Information