1 Overview

1.1 Environment

- The game is a single-agent environment.
- The game is played with a standard 52 card deck
- The goal of the game is to score as many points as possible

1.2 State space

- Each turn, you will have 8 cards in your hand.
- At the start of the game, you will have 3 Discards and 4 Plays (discussed below)
- Each turn, you will use some of the cards. New cards will be drawn from the remaining deck, and used cards are removed from the game.

1.3 Action space

- Each turn, you must use a Play or a Discard.
- If you Discard, you choose up to 5 cards. They are removed from the game, and you get that many replacement cards drawn randomly from the deck.
- If you Play, the same thing happens, but in addition to discarding the chosen cards, the value of the chosen cards are added to your total (hand values are explained below).

1.4 Rewards

- Each time you Play, the value of the hand is added to your total.
- Hand values are given by:

Hand	Value
Straight Flush	800
Four of a Kind	420
Full House	160
Flush	140
Straight	120
Three of a Kind	90
Two Pair	40
Pair	20
High Card	5

- To see a description of each hand, please see this link.
- The hand value is just the highest valued hand listed above. For example, playing 4 Jacks gives you just 420, it does not also count as a 3-of-a-kind in addition.
- Please note, you can play fewer than 5 cards (for example, if you play 4 cards, a Two Pair or four of a kind is possible).
- However, hands such as flushes and straights require 5 cards.

1.5 Terminal State

- An episode ends when you have no Plays left (Remaining discards are useless since they cannot give any reward).

2 Assignment

1. Write up the environment and test that it is working.
2. Use RLlib to get as high an expected score as possible with 1 Play and 1 Discard.
3. Use RLlib to get as high an expected score as possible with 4 Plays and 3 Discards.
4. Optional: Try to solve the Advanced version described below.

2.1 Advanced Rules

- The only difference in the advanced rules is in the rewards. First, the reward values are split between Chips and Multipliers. Here is the advance payout table:

Hand	Chips	Multiplier
Straight Flush	100	8
Four of a Kind	60	7
Full House	40	4
Flush	35	4
Straight	30	4
Three of a Kind	30	3
Two Pair	20	2
Pair	10	2
High Card	5	1

- You receive Chips x Multiplier, which by default will result in the same payoffs as the regular table.
- However, in the advanced version, scored cards add to the Chips.
- A scored card is any card needed for the hand. For example, if you play 4xJacks and an Ace, The 4 Jacks are scored for the Four of a Kind, while the Ace is not scored as it is not needed for the hand.
- Scored cards are worth their blackjack value (2-10 are worth their number, J,Q,and K are worth 10 each, and Aces are worth 11.
- The effect of this should be that it is optimal to bias yourself towards higher ranked cards.

2.2 Advanced Scoring Examples

- Hand 1

You play:	AK832 of hearts
Hand type:	Flush
Chips total:	35 + 11 + 10 + 8 + 3 + 2 = 69
Multiplier:	4
Total Hand Score:	69x4 = 276

- Hand 2

You play:	AAA82
Hand type:	Three of a Kind
Chips total:	30 + 11 + 11 + 11 = 63
Multiplier:	3
Total Hand Score:	63x3 = 189