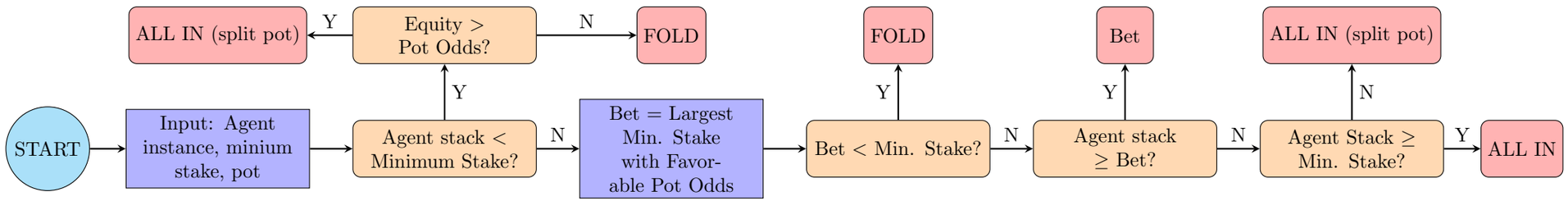
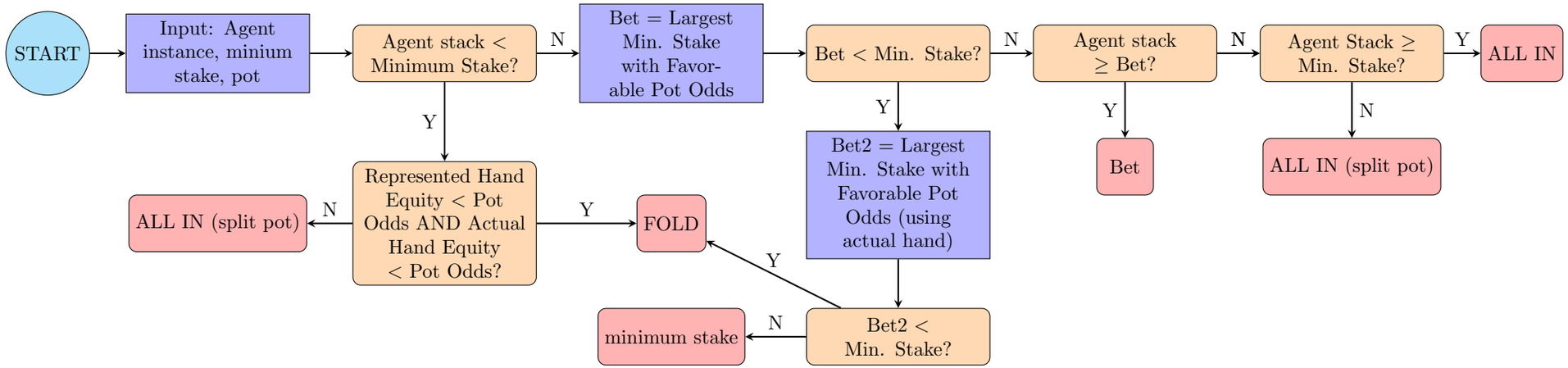
1 Determine Action Flowcharts

1.1 RATIONAL AND ONE CHOICE RANDOM



For one choice random, equity is chosen randomly when the cards are dealt (as opposed to a table lookup). Otherwise the two strategies determine their action the same way.

1.2 BLUFFER

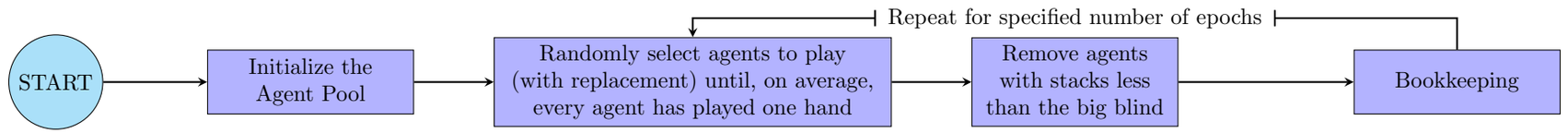


The bluffer uses their represented hand to determine their action. In a case where their represented hand would direct them to fold, they only do so if both their represented *and* their actual hand are getting unfavorable pot odds.

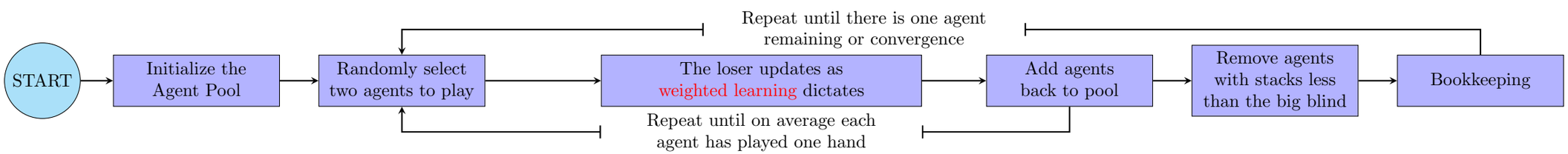
2 Dynamics Flowcharts

Each of these flow charts shows one iteration of each dynamic. Plots of the results of simulations are averages of 100 or 500 iterations. The learning dynamics also include the death dynamic.

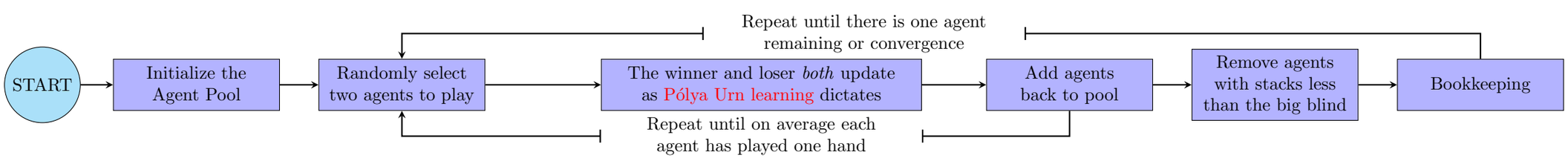
2.1 Death Dynamic



2.2 Weighted Learning



2.3 Pólya Urn Learning



3 Agent Class Description

Below is a table with a breakdown of how each strategy utilizes the attributes of the agent differently. The **bold** attributes must be provided upon instantiation of the agent.

Attribute	Description
ID	String identifier for the agent
Stack	Integer chip stack of the agent
Strategy	A dictionary for which the keys are the possible strategies and the associated value is the agents confidence in that strategy. For simulations in which agents to not learn, the values of the dictionary will remain the same throughout the simulation. When determining their strategy on a hand, it will be chosen to be the strategy for which the agent has the highest confidence. For simulations in which learning occurs, the confidence associated with each strategy is updated for the agent as they learn.
Hand	The agent's *actual* hand, used to determine whether they win the hand. It is updated at the start of every hand as they are dealt a new hand
Act	The most recent play made by the agent.
Previous Bet	Total money invested into the current pot. This attribute is reset every hand they play.
Strategy Used	The strategy which the agent will use on the current hand to determine its action using the Strategy attribute
Rank	The rank of the agent's *actual* hand
Represented Hand	The hand with which the agent will make rational plays. If bluffing, the Represented Hand attribute will differ from the Hand attribute. Otherwise, they will be the same.
Bluffing	One if bluffing, zero otherwise.
Odds	The Agent's perceived equity. If bluffing, it is the equity of their represented hand. If one choice random or forgetful random, this will be a random number. If rational, this is the proper preflop equity of the hand.
Previous Stack	At the start of each hand, this is set to the agents stack to allow for the amount of money won/lost on a hand to be tracked.
Hands Survived	The number of hands that the agent has played and not gone broke on.

4 Strategies

Strategy	Identifier	Description
Rational	rational	The base rational agent will take its odds to be the true probability that its cards will win the pot. Note that since the poker being played only involves preflop betting, their equity is a simple table lookup. They will have a bluff probability of zero. Other implementations of "rational" behavior were tried using expected value or a basic multiplication of equity and stack, but ultimately it was decided that determining action of the agent using pot odds and maximizing bets given those pot odds was the clearest while still being effective.
Bluffer	bluffer	On each hand, when the agent is dealt their cards they will decide whether or not to bluff, doing so with probability equal to their bluff probability attribute (have an equal probability of bluffing across all possible starting hands). If they do not bluff, they play exactly like a rational agent. If bluffing, the hand that they represent is the one with ranking of their bluff aggression attribute (lower ranking is better hand). From that point they determine their action in a similar but not identical manner to that of a rational agent dealt the represented hand. They will have a non-zero bluff probability, and a value must be provided for bluff aggression when initializing the agent. For the complete/incomplete information learning, the agents that pick the bluffing strategies will bluff 100% of the time.
One Choice Random	one choice random	This is the implementation of random that was chosen. At the start of each hand, the agent will randomly pick their equity to be a value between zero and one. From that point, their odds will be that value and not reflect the strength of their hand. These agents go on to play rationally but just use this randomly chosen value as their equity. Bluffing and random play are not combined on agents, so these agents have a bluff probability of zero.
Forgetful Random	forgetful random	Each time the agent must determine their action, the agent picks their equity to be a value between zero and one. Thus, their perceived equity can change throughout the hand. After picking their equity, they proceed to play rationally but with the randomly chosen equity instead of the equity that properly reflects the strength of their hand. Again, bluffing and random play are not combined on agents, so these agents have a bluff probability of zero.
Wheel Spin Random	wheel spin random	When it is an agent's turn to act, the possible actions are fold, call, or raise (they don't check preflop). These agents will randomly choose their behavior by randomly choosing from the feasible actions when it is their turn to act with equal probability. If they choose to raise, the amount they raise is chosen to be a random amount from the minimum stake plus one to the agent's stack. If the agent does not have enough money to pay the minimum stake, they will have a 50-50 chance to go all in or to fold. They will not have the possibility of bluffing.

5 Dynamics

Dynamic	Description
Death	This death dynamic follows basic intuition of what it means to no longer be fit to compete in poker. For the simulations, the money that an agent is taken to be their fitness, as the goal of poker is to maximize money earned and the agents' goal should be to maximize their fitness. Then, it makes sense that an agent "dying" would correspond to them having no money, which would mean they have no fitness. Thus, the death dynamic involves filtering all agents unable to pay the minimum fee to play (the blinds) out of the pool of agents.
Complete Information Weighted Learning	Under this dynamic, the agents can learn to play differently by playing more hands. Each agent maintains a confidence level in the various possible strategies. When an agent loses, they will lose confidence in the strategy they used in the hand they lost proportional to the percent of their stack that they lost. If an agent loses enough money that they can no longer play, they don't update, as they will no longer play any more hands. When updating, they will lose confidence in the strategy they just lost with and that confidence will get transferred to the strategy that the opponent used. This scenario has complete information in the sense that the loser gets to know what strategy the winner used.
Incomplete Information Weighted Learning	This dynamic also enables agents to employ different strategies in different hands. Each agent will maintain a confidence in the possible strategies in the simulation. When an agent loses, they will update their strategy by losing confidence in their used strategy an amount proportional to the percent of their stack that they lost. An agent that loses enough of their stack to prevent them from playing any more hands does not update. When updating, they will lose confidence in the strategy they just lost with and that confidence will get distributed uniformly amongst the strategies that they did not use. The information in this scenario is incomplete because the losing agent does not get to learn with knowledge of the strategy used by the winning agent.
Pólya Urn Complete Information Learning	Unlike weighted learning, both the winning *and* losing agents update after the hand. Each agents "carries around an urn" that contains one marble for their assigned strategy at the start of the simulation. When an agents plays a hand, they randomly remove a marble from their urn and play the strategy on that marble. Once the hand is finished, the winning agent puts back the marble they removed prior to the hand, plus an additional marble of that strategy. The losing agent puts back the marble they removed prior to the hand, and with complete information they add a marble for the strategy that the winner used.
Pólya Urn Incomplete Information Learning	Again, both the winning *and* losing agents update after the hand. Each agents "carries around an urn" that contains one marble for their assigned strategy at the start of the simulation. When an agents plays a hand, they randomly remove a marble from their urn and play the strategy on that marble. Once the hand is finished, the winning agent puts back the marble they removed prior to the hand, plus an additional marble of that strategy. The losing agent puts back the marble they removed prior to the hand, and with incomplete information they randomly pick from the strategies that they did not use and add a marble for that chosen strategy.