**Assignment 1 - The Matching Pennies game**

* **Describe 2 possible strategies to play the matching pennies game**

**The Rescorla-Wagner model**

The Rescorla-Wagner model can be seen as a general way of playing the matching pennies game. The model can be mathematically formulated as an expectation that gets updated on each trial based on the previous expectation and then a weighted prediction error:

(1)                                               x(t + 1) = x(t) + 𝛼 (r - x(t))

Where x(t+1) is the expectation (probability of choosing 1) on the next trial, x(t) is the expectation on the current trial, and r  is the actual outcome which is either 0 or 1.

The equation above explains that the expectation in the next trial depends on the expectation in the current trial and 𝛼 (the agent’s learning rate), multiplied by the prediction error, which is the difference between the predicted outcome and the actual outcome. Thus the learning rate, which differentiates between individuals, determines how much the prediction error affects the expectation for the next trial.

This model can be used to describe a win-stay-lose-shift agent as that would amount to an agent that has a learning rate of 1. Furthermore a random bias agent would be an agent that uses the Rescorla wagner model with a learning rate of 0, and would therefore not update his expectation and stay with the expectation he “gets” on the first trial. One could also have an agent that produces responses that are opposite to what the agent should expect given the previous trials. The extreme example of such an agent would be a win shift lose stay agent, which could be a strategy used if one believes that the opponent believes that you follow a Rescorla wagner model.

We have decided to use the following two strategies:

Strategy 1:

Differing learning rates:

Our first agent consists of a Rescorla-Wagner model with different learning rates for winning and for losing. The learning rate for losing is higher than for winning, allowing the agent to update expectations more when it loses, and perform smaller updates to expectations when the agent wins. Given what we know of loss aversion this might be a strategy that some humans would display. Concretely a learning rate of 0.2 and 0.5 will be used for winning and losing respectively.

Strategy 2 :

Shifting strategies:  
The first half of the trials include a specific strategy, such as random bias i.e. a Rescorla wagner model with learning rate of 0 and the second half a win stay lose shift strategy  i.e. a Rescorla wagner model with learning rate of 1, and for the second half of the trials, the agent includes another strategy such as a random bias. The random bias does not include any memory, as the choice is random and not dependent on the previous trials. This can be done by setting the alpha to 0, and thus not letting the agent learn from prior trials. This kind of agent could resemble a participant that tries to either confuse his opponent or are just lazy or unwilling to play for the first half of the trials and then tries to win for the last half of the trials. To make the distinction clear we made the bias for the first 60 trials 0 such that he always just responds 0 in the first 60 trials.

Implementation of the strategies:

The implementation for the first strategy can be seen in the github code but goes as follows:  
The agent is initialized with a random bias which is the probability of answering one on the first trial i.e. a value of either 0 or 1 is  sampled from a Bernoulli distribution with probability equal to the bias. Then this bias (x(t))  is updated based on the equation (1), which takes into account whether the agent won or lost while respecting a different learning rate for winning and losing.

The implementation for the second strategy is as for the first however here we first loop through half of the trials with one learning rate which is 0 because the agent first displays random bias behavior and then a learning rate of 1 which is a win stay lose shift behavior.

Results:

First we test our strategy one agent against a random bias agent of probability 0.7, we do this over 100 consecutive agents and plot the percent of wins on the y-axis with a shaded area of 2 times the standard error. For the first 4 plots our agents (i.e. strategy 1 and 2 are going to be the matchers that is the red color)

Chart, line chart

Description automatically generated

Figure 1: displays the result of running 100 agents of strategy one versus a random bias agent with a bias of 0.7.

Next we test strategy one versus a win-stay-lose-shift agent.

Chart

Description automatically generated

Figure 2: displays the result of running 100 agents of strategy one versus a win stay lose shift agent.

Next we ran our strategy two versus the random bias of 0.7 and the win stay lose shift agent a 100 times:

Chart, line chart

Description automatically generated

Figure 3; displays strategy 2 vs a random bias agent with a bias of 0.7

Line chart

Description automatically generated

Figure 4; displays strategy 2 versus a win stay lose shift agent:

Lastly we plot strategy 1 versus strategy 2, here we set strategy 2 as being the matcher and strategy 1 as the non-matcher:

Chart, line chart

Description automatically generated

Figure 5; displays the two implemented strategies against each other through 100 agents playing versus each other.

Discussion:

We can see that strategy one fares quite well against the random bias agent, meaning that it seems to consistently win after around 10 trials. However against the win-stay lose shift agent it does not fare as well and it's pretty much random which agent wins throughout all trials. This makes sense given that strategy one tries to update his expectation based on what the other agent does, however because of the very sporadic behavior of the other agent it cannot find a pattern and it therefore becomes more or less random because the expectation will center around 0.5 for strategy 1.

From the plots for strategy 2, it is very evident that this agent uses two strategies. The agent performs really well against the random bias for the first 60 trials as here the strategy 2 agent chooses 0 every time and the random bias agent chooses 1 in 70% of the cases meaning that strategy 2 agent wins around 70% of the times. However after the 60 trials of choosing 0 agent 2 shifts strategy and begins to lose more which is evident from the plot given that the two curves begin to converge after 60 trials. This pattern is even more evident when strategy 2 agent plays against a win stay lose shift agent. Here strategy 2 agent loses the first 60 trials as he always picks 0 and the win-stay-lose-shift agent always will choose 1 and stay on 1 because he never loses. After the 60 trials we see that the lines converge again meaning that strategy 2 agent begins to win more and more trials.

Lastly when playing the two agents strategy one versus strategy two against each other we see a similar pattern of that when strategy 2 played against a win-stay-lose-shift, however with a less steep raise in wins for the opponent of strategy 2. This is because strategy 1 quickly learns  the fact that strategy 2 only picks 0 for the first 60 trials; however it is not instantaneous as with the win-stay-lose-shift. After the 60 trials we again see the same pattern as versus the win stay lose shift however with a more smooth curve compared to the jagged one for the win-stay-lose-shift.