Ατομική Διπλωματική Εργασία

ΣΥΝΔΕΣΗ ΜΕΤΑΞΥ ΘΕΤΙΚΩΝ/ΑΡΝΗΤΙΚΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΚΑΙ ΠΡΟΔΕΣΜΕΥΣΗΣ ΣΕ ΕΝΑ ΥΠΟΛΟΓΙΣΤΙΚΟ ΜΟΝΤΕΛΟ ΑΥΤΟΕΛΕΓΧΟΥ

Χρίστος Κασουλίδης

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Connection between positive/negative emotions and precommitment in a computational model of self-control

Christos Kasoulides

Επιβλέπων Καθηγητής
Δρ. Χρίστος Χριστοδούλου

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου

Μάιος 2023

Acknowledgments

I would like to express my heartfelt gratitude to my supervisor, Dr. Chris Christodoulou, for his invaluable support and guidance throughout my research journey. He imparted valuable knowledge and I feel fortunate to have worked on such a fascinating topic under his mentorship. I also want to thank my family and close friends for their constant support, which kept me motivated and focused.

Abstract

The thesis investigates the influence of emotions and precommitment on self-control behavior using a computational model. Drawing from Banfield's (2006) work on precommitment and Nikodemou's (2020) research on emotional arousal, the study aims to understand how these factors interact and affect self-control. The computational model employs an iterated version of the Prisoner's Dilemma game called the Iterated Prisoner's Dilemma (IPD) game to simulate the conflict between the higher (prefrontal cortex) and lower (limbic system) brain regions involved in self-control. Two agents representing these regions are trained using the Q-Learning algorithm to achieve mutual cooperation (CC), a measure of successful self-regulation. The effects of emotions are simulated by the change of the values in IPD’s payoff matrix T (Temptation), R (Reward), P (Punishment), and S (Sucker) which are reinforcement signals the agents receive and are in turn affected by the specified intensity value and interval of change (number of rounds between each change of the payoff values). Additionally, precommitment is simulated by adding a differential bias, denoted by ψ, to the diagonal terms of the payoff matrix at the beginning of the game to simulate low (ψ=0.01) and high (ψ=0.9) levels of precommitment. The findings demonstrated that high precommitment coupled with positive emotions had a notable influence on facilitating self-control behavior, surpassing the impact of low precommitment. This effect was particularly pronounced when employing a larger interval of change and utilizing a smaller positive intensity value, resulting in a decreased magnitude at a less frequent rate. To address negative emotions and precommitment, the results indicated that precommitment can once again serve as an effective strategy for enhancing self-control by mitigating the impact of negative reinforcement on the agents. This effect was particularly enhanced in the cases of negative and positive punishment (decreasing R and increasing T) when high precommitment was paired with a sufficiently large interval of change and a small intensity value. In contrast, when explicit negative emotions were induced by providing a negative intensity value to S and P, the scenarios where high precommitment still managed to achieve self-control and provided the best results were those where a higher negative intensity value was given in a large interval of change.

Contents

Chapter 1

Introduction

1.1. Introduction

The purpose of this thesis is to investigate the effects of precommitment in combination with emotional arousal on self-control behavior. Self-control is a fundamental concept in cognitive psychology, referring to the ability to resist smaller sooner (SS) rewards in favor of larger, later (LL) rewards something crucial for individuals trying to achieve their long-term goals but challenging as it involves overcoming the temptation of an immediate reward. One example of this concept can be when a person opts for a healthy meal instead of an unhealthy one, resisting in this way the immediate pleasure of indulging in some junk food in favor of the long-term goal of remaining in good health. This conflict between the desire for immediate gratification and the need to pursue long-term goals has been heavily studied and associated with the interaction that takes place between the higher (prefrontal cortex) and lower (limbic system) parts of the brain, which are responsible for the cognitive processed responses (long-term) and intuitive processed (short-term) responses, respectively (Rachlin, 2000).

The second key aspect which is discussed in my thesis is the concept of precommitment and how it relates to self-control. Precommitment refers to the act of making a commitment in advance to follow a specific course of action to bias future choices towards the larger, later reward (desired goal). Banfield (2006) argues that exercising precommitment can help overcome self-control problems as it will help individuals resist the temptation of the smaller, immediate reward in favor of the long-term one. This is based on the idea that precommitment can make it easier to choose the larger, later reward by reducing the conflict between the higher and lower parts of the brain.

The third aspect is emotional arousal in self-control behavior. Nikodemou (2020) suggests that as self-control is a human process it is bound to be linked with emotions and that a person's emotional state can have a significant impact on their ability to exercise self-control. For example, positive emotions can promote self-control whereas negative emotions impair it. This is based on the idea that positive emotions can increase people's willingness to pursue their long-term goals, while negative emotions can make them more susceptible to giving in to the temptations of immediate rewards which is an implication that, as Nikodemou (2020) states, turns out to be more complicated for the reason that negative emotions can also have positive effects on self-control (e.g., feelings of fear and guilt) and positive emotions negative ones.

Despite the extensive research mentioned before on self-control in relation to precommitment and emotions separately, no studies have examined the interaction between precommitment and emotions in a computational model of self-control. This research gap is significant because it limits our understanding of how these factors interact with each other to influence self-control behavior.

To explore this interaction, I developed a computational model based on the findings of Nikodemou (2020) and Banfield (2006). The model uses a general-sum game of the Prisoner's Dilemma (Kavka, 1991), where each agent can insist on maximizing its own reward by defecting, but the best outcome for both of them is achieved only when the agents cooperate. The agents do not know what the other one will choose: to cooperate (C) or defect (D). All the combinations of states in which the agents can result are CC, CD, DC, and DD where the state of self-control is represented by the CC state. The agents are trained using the Q-Learning algorithm with the addition of rewards/punishers to simulate emotions and the use of a differential bias on the payoff matrix of the agents to simulate precommitment. Earlier studies by Banfield (2006) and Cleanthous (2010) also modeled self-control behavior using neural networks, but it was found that the use of biologically realistic spiking neural networks was not necessary for the successful modeling of self-control, as demonstrated by Christodoulou et al. (2010).

1.2. Thesis outline

Chapter 1 serves as an Introduction to the thesis. In Chapter 2, the Epistemological Background of the research is discussed, focusing on self-control behavior in neuroscience and psychology. The chapter presents the Prisoner's Dilemma (PD) and its variation, the Iterated Prisoner's Dilemma (IPD), along with the rationale for choosing it to simulate the self-control structure in our model. The Reinforcement Learning algorithm chosen for the model is also explained. Additionally, the chapter introduces the concepts of emotions and precommitment, providing definitions that aid in simulating their effects on our self-control model. Chapter 3 focuses on the design and implementation of the Q-Learning model, with detailed explanations of simulating positive and negative emotions and incorporating high and low levels of precommitment. In Chapter 4, the results of the simulations are presented, including comparisons between high and low levels of precommitment and highlighting key findings. Finally, Chapter 5 provides an overview of the study, explains the findings in relation to existing literature, and proposes future research directions facilitated by this study.

Chapter 2

Epistemological Background

2.1 Self-control

Self-control refers to the ability to regulate and alter one's own responses, including thoughts, emotions, and behaviors, in order to achieve internalized goals or standards. Self-control is a concept that has received great attention from psychologists, behavioral economists, and neuroscientists in the search for its understanding (Muraven et al., 1999) and although the exact mechanisms have not been fully defined, the importance of self-control has been established through numerous studies. For example, a study found that young people with high levels of self-control invest more time in academics, have higher school attendance and focus, and as a result, earn higher grades (Duckworth & Seligman, 2005). On the other hand, low self-control has been linked to a variety of problems, such as academic underachievement, unhealthy lifestyle, procrastination, and legal issues (Moffitt et al., 2011).

Despite the challenges of self-control, it is not a fixed trait and can be developed and improved with practice (Muraven et al., 1999). The study mentioned earlier found that students who performed simple self-control tasks regularly for two weeks showed significant improvements in their self-control compared to participants who did not practice self-control.

2.1.1 The top-down model of self-control

We can think of self-control as a trade-off between short-term and long-term goals and a classic example of this dilemma is the choice people face between a larger later (LL) reward and a smaller sooner (SS) one. When faced with this dilemma, people have to choose whether they want to disregard the larger future reward and settle with the more immediate smaller reward, saving them in this way the trouble of waiting but making them miss out on a larger reward.

This trade-off has also faced great attention from the field of cognitive neuroscience where self-control behavior is thought of as an internal process that occurs within a person’s mind - Figure 2.1 (Rachlin, 2000) and is modeled by two agents. The first one is the prefrontal cortex (higher brain) which is responsible for rational thinking, and it is involved in the process of self-control by assessing the potential consequences of different actions before deciding and the second one is the limbic system (lower brain) which is responsible for generating emotions and it is involved in self-control by selecting actions based on them.

During the model process shown in Figure 2.1, the prefrontal cortex (PFC) and the limbic system interact with each other to reach a decision for taking an action. After making this decision feedback is received from the external environment in the form of stimuli (reward or punishment), guiding the agents toward the desired outcome. The PFC may then use information from the stimuli about the potential consequences of different actions, while the limbic system may influence behavior by generating emotional responses to the stimuli.

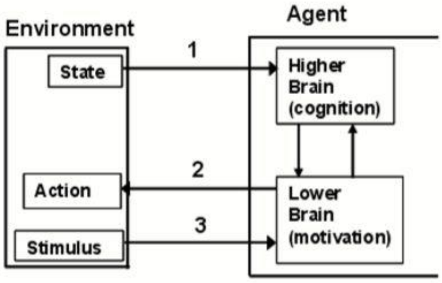


Figure 2.1: *A model of self-control behavior based upon Rachlin (2000). The PFC interprets the state of the environment (1) and with influence from the limbic system, an action is generated (2). The action is then rewarded or punished by stimuli from the external environment (3).*

2.1.2 Other Approaches to Self-control

While the top-down model of self-control is widely recognized and empirically verified, various approaches to self-control highlight other components of the process. The bottom-up model of self-control, for example, posits that self-control is predominantly driven by automatic processes of attention, perception, and response selection rather than deliberate decision-making processes (Hofmann et al., 2012). Self-control is viewed as a learned talent that is gained by repeated practice and feedback in this paradigm, rather than as a fixed attribute.

The strength model is another approach to self-control, which contends that self-control is a finite resource that can be exhausted through repeated use, lowering performance on subsequent tasks that call for self-control (Baumeister et al., 1998). This model places a strong emphasis on the need to conserve and replenish self-control resources through tactics like sleep, relaxation, and glucose ingestion.

Despite their differences, all these models acknowledge that self-control is crucial for achieving long-term objectives and controlling impulsive behaviors. Each model offers a distinctive viewpoint on the underlying self-control mechanisms and suggests various methods for enhancing self-control. The top-down model of self-control, however, is especially relevant to my approach because it places a strong emphasis on the use of deliberate decision-making processes and cognitive strategies to resist temptation and accomplish long-term objectives.

2.2 The Prisoner's Dilemma (PD)

One of the ways that self-control has been modeled is through the Prisoner's Dilemma (PD) game. In the PD game, two agents are faced with a choice between cooperating for a common good outcome or defecting for their personal benefit making the perfect way to model the internal conflict that happens between the two parts of the brain

The scenario on which the PD game is based is that two suspects are arrested for a crime and interrogated by the police. During the interrogation the suspects are given two options, to remain silent (cooperate) or to confess by testifying against the other suspect (defect). The outcome of the game then breaks down into four main scenarios that are dependent on the choices of both suspects. The first scenario is that if one of the suspects confesses and the other does not, the confessor will be freed, while the other will be sentenced to a longer prison term because of the testification against him (Cooperate – Defect). The second scenario can be seen as an exact mirror of the first one where one suspect defects and the other cooperates and as a result, the outcomes for each suspect are reversed (Defect – Cooperate). The third scenario occurs when both suspects confess and they will both receive a shorter prison sentence than if only one of them confesses, but still longer than if they both cooperated and remained silent (Defect – Defect). The last and most optimal scenario is when both suspects remain silent, in which case they will receive the shortest prison term (Cooperate – Cooperate). The dilemma arises from the fact that the best outcome, mutual cooperation, is not guaranteed as both suspects may choose to defect for personal gain.

Kavka (1991) applied the PD game to show how individuals experience value conflicts. The goal is always to maximize their reward. The outcome of the PD game is determined by the choices of both agents. If both agents choose to cooperate, the outcome will be a common good outcome with a larger reward (lower prison sentence) than if both choose to defect. However, if both choose to defect, the reward will be lower than if only one defects. The strategy that is chosen by the players and from which nobody wants to deviate is called a Nash equilibrium (Nash, 1950). In the PD game, the Nash equilibrium is mutual defection, which is considered the only Nash equilibrium.

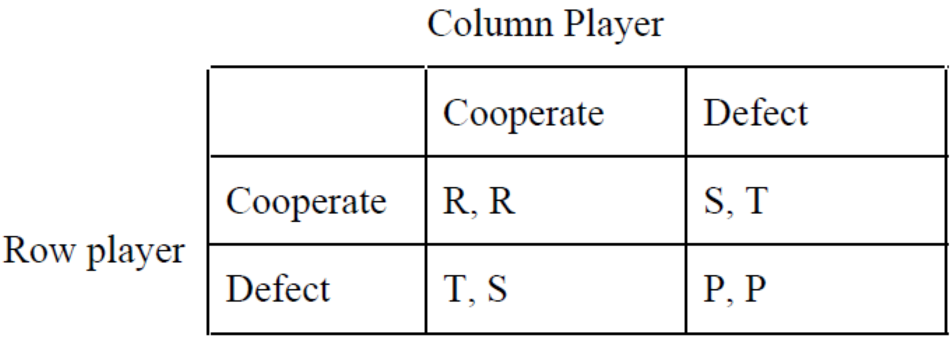
In Figure 2.2 we can see the payoff matrix of the Prisoner's Dilemma game, where a payoff is calculated based on the combination of actions performed by the agents. Based on the matrix there are four main outcomes of the PD game:

1. Both agents cooperate: In this case, both agents receive a Reward payoff (R).
2. Both agents defect: In this case, both agents receive a Punishment payoff (P).
3. One agent cooperates and the other defects: In this case, the agent who cooperated receives the Sucker's payoff (S) and the agent who defected receives the Temptation payoff (T).
4. One agent defects and the other cooperates: In this case, the agent who defects receives the Temptation payoff (T) and the agent who cooperated receives the Sucker's payoff (S).

The payoffs must satisfy the following ordering:

Temptation > Reward > Sucker's > Punishment **(1)**

This means that the temptation to defect (and receive the highest payoff) is greater than the reward for mutual cooperation, which is greater than the sucker's payoff for being the only one to cooperate, which is in turn greater than the punishment for mutual defection.



Column Player

		Cooperate	Defect
Row player	Cooperate	R, R	S, T
	Defect	T, S	P, P

Figure 2.2: *The matrix outlines the potential outcomes for each player based on their decision to either cooperate or defect. If both players choose to cooperate, they both receive a rewarding payoff (R). However, if they both defect, they both receive a punishing payoff (P). On the other hand, if one player chooses to cooperate while the other defects, they receive the Sucker's payoff and Temptation payoff, respectively.*

2.2.1 Iterated Prisoner's Dilemma (IPD)

A variation of the PD game that is used to model how individuals or agents handle decisions in environments where the consequences of each decision will not appear until later on in the game, is called the Iterated Prisoners Dilemma (IPD). The main difference between the two variations is that a game of IPD is played over multiple rounds allowing the players to learn from their previous mistakes before deciding to adapt and achieve the best outcome. This way the IPD can provide insights into how cooperation can occur in a complex environment.

During each round of the IPD, the agents must decide whether to cooperate or defect based on their own self-interest considering that the outcome of each round is affecting the reward of the next one. This way the IPD allows agents to follow different strategies for making decisions, such as always cooperating or always defecting.

One key characteristic of the IPD that distinguishes it from the standard PD game is the payoffs. The payoffs in the IPD satisfy the inequality,

$$2R > T + S \quad \textbf{(2)}$$

which means that mutual cooperation is the best outcome for both players overall. This is because mutual cooperation leads to a higher total reward (2R) than either player defecting (which results in a total reward of T+S).

This feature of the IPD is important because it allows for the possibility of cooperation to emerge between the two players over multiple rounds of the game. While the immediate temptation to defect may be strong, the possibility of gaining a higher total reward through mutual cooperation over time incentivizes players to cooperate with each other.

The IPD has been used to study how cooperation can emerge in complex environments, such as in the evolution of cooperation in biological systems. Axelrod and Hamilton (1981) conducted a famous tournament of IPD strategies, in which they invited researchers to submit their best strategies for playing the IPD. The winning strategy was a simple tit-for-tat strategy (cooperating during the first move and then mirroring the opponents last move), which showed that cooperation can emerge without central control or communication between agents.

In addition to the IPD, other general-sum games such as the Battle of the Sexes game (BoS) or Rubinstein's Bargaining Game (RBG) (Rubinstein, 1982) have been used to simulate interactions between the higher and lower brain. However, empirical evidence suggests that the PD game is currently the optimal option for modeling self-control internal conflict as shown in studies conducted by Banfield (2006), Cleanthous (2010), and Christodoulou et al. (2010).

2.3 Reinforcement Learning

Reinforcement Learning (RL) refers to a type of machine learning algorithm where we have an agent interacting with a dynamic environment through the use of reward signals. These signals simulate how humans learn new skills as they observe a new environment and think about how to act to achieve their desired goals which are to maximize these reward signals (Stone, 2010). This is the main point that distinguishes RL from other types of learning such as supervised and unsupervised where the models are trained through labeled data (supervised) and by learning to recognize patterns in a given data structure (unsupervised).

In RL, the agent does not know the optimal policy and must explore the environment to learn through trial and error. This exploration-exploitation trade-off is essential in RL, as the agent must balance between exploring new actions to learn and exploiting its current knowledge to achieve its goal (Woergoetter & Porr, 2008).

Finally, for our model, we use Temporal-difference (TD) learning which is a popular RL algorithm that combines the benefits of both dynamic programming (DP) and Monte Carlo methods. TD methods learn directly from raw experience and update estimates without waiting for the outcome like DP. The update rule of TD is based on the Bellman equation and can be expressed as follows (Sutton, 1988):

$$V(S_t) \leftarrow V(S_t) + a\,[\, R_{t+1} + \gamma\, V(S_{t+1}) - V(S_t)] \quad \textbf{(3)}$$

In equation 1, α refers to the constant step-size parameter, γ to the discount factor, R_{t+1} to the newly obtained reward, $V(S_t)$ to the current state’s estimated value, and $V(S_{t+1})$ to the next state's estimated value.

2.3.1 The Q-Learning Algorithm

One of the classic model-free algorithms for reinforcement learning from delayed reward is the Q-Learning algorithm (Watkins, 1989). It is one of the most basic methods to estimate Q-value functions and its update rule is based on the sum of two parts at the end of each step. The update rule is as follows:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \eta\, [\, R_{t+1} + \gamma\, \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\,] \quad \textbf{(4)}$$

where $Q(S_t, A_t)$ is the new Q-value of the state-action pair. At time t+1, it makes a useful update by using the observed reward R_{t+1} by taking action A_t, state S_t and the discounted estimate of optimal future reward. That is, the term $\max_a Q(S_{t+1}, a)$ returns the maximum Q-value in the next state S_{t+1} over all actions a. This future reward is discounted by the parameter γ ∈ [0, 1]. Values of the γ parameter closer to 0 indicate that the agent focuses on the immediate (or SS) rewards, whereas values closer to 1 indicate that the agent focuses on the future (or LL) rewards. The η term is the learning rate, which determines to what extent we weigh the two parts of the sum into the new Q-value.

Taking the maximum across all actions makes learning independent of the starting policy and allows keeping this policy throughout the whole learning process. This is the reason Q-Learning is an off-policy TD control algorithm, in contrast to the state, action, reward, state, action (SARSA) method (Rummery & Niranjan, 1994), which continuously updates its policy during learning (on-policy update). In other words, the Q-Learning update rule guarantees that the optimal policy and the optimal value function are found, a property that makes convergence control much easier.

The effectiveness of the Q-learning algorithm, when applied to non-spiking neural networks to simulate self-control, was first demonstrated by Vassiliades et al. (2011). Later, a simple Q-Learning model as well as the SARSA method were deployed by Georgiou (2015) to simulate self-control behavior and examine its relationship with consciousness. Georgiou's (2015) results showed that the agents learned to exercise self-control more easily with the Q-Learning algorithm in comparison with the SARSA algorithm. The SARSA method had good results only when it was combined with hill-climbing techniques. For the purposes of this thesis, we will only deploy the Q-learning algorithm since it is more effective, as previous work revealed, and it is more easily handled.

2.3.2 The ε-greedy policy

The exploration-exploitation dilemma is a crucial issue in reinforcement learning, as agents must balance exploring new options and exploiting their current knowledge to maximize rewards. One common solution to this problem is the ε-greedy policy (Sutton & Barto, 2018). This policy works by choosing an action based on the probability of exploring new options (ε) versus exploiting the current knowledge (1-ε).

When applying the ε-greedy policy, the agent selects an action randomly with probability ε, which may result in exploring new moves and payoffs. Alternatively, the agent selects the action with the highest estimated value (based on past experiences) with probability 1-ε, which means that it will exploit its current knowledge to maximize rewards. This way ε can determine the balance between exploring and exploiting past knowledge and is often set to a small enough value allowing that way for some exploration from the agent.

The ε-greedy policy is widely used in reinforcement learning and has been applied in many applications, including robotics (Kober et al., 2013) and game playing (Silver et al., 2016). Despite its popularity, the ε-greedy policy is not without limitations. For example, it may not be suitable for problems where the optimal action changes frequently, or in cases where exploration is essential to finding the optimal solution (Sutton & Barto, 2018).

2.4 Precommitment in Self-control

Precommitment is a strategy used to overcome short-term temptations in favor of long-term goals by deciding in advance to influence future behavior (Ainslie, 1975). The main idea is that binding oneself to a certain set of actions will make it much harder to deviate from that course later on. This is why precommitment is also thought of as a form of self-regulation, as it sets up barriers from temptations. It is also worth noting that precommitment can be viewed as costly, as it may restrict flexibility and impede spontaneous decision-making which can have implications for individual well-being and autonomy.

Concerning self-control, precommitment can be thought of as a way of overcoming the temptation to engage in short-term and impulsive behaviors and instead follow through on the long-term goals you have set. One example of this is a person who wants to lose weight can pre-commit to steering away from buying junk food or make a commitment to exercise regularly making it that way harder for them to deviate from their long-term goal.

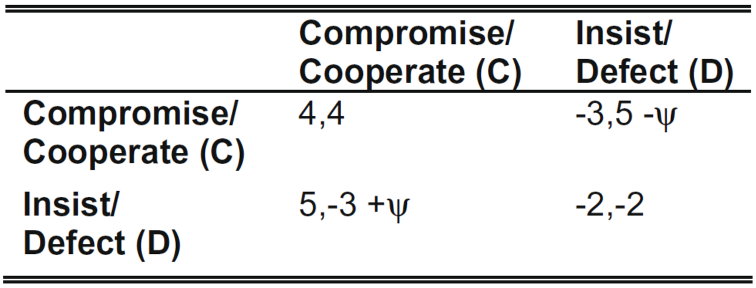
Finally, while precommitment can be an effective way to follow through on our long-term goals, it is important to acknowledge that it can also have limitations and costs associated with it. Thaler & Shefrin (1981) argue that precommitment can be costly because it involves modifying an individual's preferences in order to achieve a desired outcome thus limiting flexibility, which can have implications for individual well-being and autonomy. However, research by Banfield (2006) on the topic of precommitment and self-control using a computational model showed that increasing precommitment can lead to a higher probability of cooperating with oneself in the future and this way leading to self-control behavior. This suggests that the benefits of precommitment may outweigh the costs and that it can be an effective way to overcome short-term temptations and achieve long-term goals.

2.4.1 Precommitment implementation using a differential bias

This chapter will focus on Banfield's (2006) self-control model using a differential bias to simulate the effect of precommitment on the model. The experiment used a computational model that competed in games where the payoffs were not entirely unfavorable or entirely favorable to either part of the brain (PFC and the limbic system) and a payoff matrix was used to calculate the payoffs for two artificial neural networks (ANNs) during the game. The matrix included a differential bias, represented by the symbol ψ, which simulates the effect of precommitment on the model and is added only to the diagonal terms of the matrix (Figure 2.3).

The purpose of the experiment was to examine the effect of modeling a bias towards future long-term rewards as a variable bias applied to the payoff. This bias, with a value between 0 and 1, was assigned to the payoff matrix for both ANNs to calculate the differential payoff and is fixed for the duration of the trial.

The research results showed that by increasing precommitment, which biases choices towards a larger reward in the future, the probability of cooperating with oneself in the future increases suggesting that precommitment can enhance self-control as it promotes long-term goals. Furthermore, the study was able to show that the level of the differential bias had a high correlation with the level of cooperation between the ANNs. More specifically, as the differential bias increased, the level of cooperation also increased which suggests that the strength of one's commitment towards a long-term goal highly determines how effective the use of precommitment is as a strategy for achieving self-control.



	Compromise/ Cooperate (C)	**Insist/ Defect (D)**
Compromise/ Cooperate (C)	4,4	-3,5 $-\psi$
Insist/ Defect (D)	5,-3 $+\psi$	-2,-2

***Figure 2.3**: A bias towards future rewards implemented as a differential bias ψ applied to the payoff matrix.*

2.5 Self-control and Emotions

Nikodemou (2020) discusses the role of emotions in self-control, drawing on the cognitive arousal theory by Schachter and Singer (1962) and the work of computational neuroscientist Edmund T. Rolls (2018). Emotions are defined as complex mental constructions that motivate thoughts and behavior and are often indicators of how well a person adapts to challenges in their environment. According to Schachter and Singer (1962), emotions occur when a person feels aroused due to an event that is appraised as concern-relevant in a specific way.

Rolls' (2018) approach suggests that emotions are states elicited by rewards and punishers, or instrumental reinforcers, with rewards being anything for which one will work and punishers being anything that one will try to escape or avoid. For our computational model of self-control Nikodemou (2020) explains that the values of the payoff matrix, including Temptation, Reward, Sucker's, and Punishment values, serve as rewards and punishers that cause the agents' emotions. Positive emotions are elicited by positive rewards, such as a warm hug, while negative emotions are elicited by negative punishers, such as the death of a loved one.

Nikodemou (2020) notes that events unrelated to a task can also influence one's emotional state and impair self-control abilities, affecting overall performance. For example, a student who is experiencing psychological pain due to a loved one's declining health or being bullied may receive negative feedback or fail on other school-related assignments, affecting their emotions and behavior. These internally generated emotions can affect all decisions, actions, and behavior.

2.5.1 Self-control and Positive Emotions

As Nikodemou’s (2020) model suggests, positive emotions are a necessary component in computational models of self-control. Additionally, researchers propose that positive emotions also have a great impact on a person’s self-control as they can help restore their capabilities allowing them that way to exert self-control with greater ease (Baumeister et al., 2007; Ren et al., 2010; Tice et al., 2004).

More specifically, Nikodemou’s (2020) computational model of self-control is modeled in a way that incorporates positive emotions proving that they play a key role in determining the degree of self-control a person exhibits as they provide the ability to resist immediate temptations. This aligns with previous research that suggests that positive emotion can help individuals better control their urges and resist temptations to achieve their desired goals (Hofmann et al., 2012).

Furthermore, Nikodemou (2020) also notes that the context as well as the type of emotions also play a key role in experiencing self-control as people that experience positive emotions which are related to one’s goals can be beneficial for self-control but positive emotions which are unrelated to one’s goal can have the opposite effect. Two examples can be someone who is feeling excited about achieving a long-term goal (related emotion) and someone who is just feeling happy after eating a chocolate (unrelated emotion).

2.5.2 Self-control and Negative Emotions

While positive emotions can help people gain better control of their behavior by restoring their self-control resources, negative emotions can lead to their depletion (Baumeister et al., 2007; Nikodemou, 2020) making them a crucial component in a self-control model. However, their impact is complex as Nikodemou (2020) shows through her computational model of self-control.

In addition, Nikodemou's (2020) computational model of self-control incorporates negative emotions and suggests that the impact of negative emotions on self-control may depend on the context and the type of emotion experienced. For example, emotions like guilt and shame can be more effective at promoting self-control behavior than emotions such as anger and sadness which aligns with previous research that has found guilt and shame to be associated with increased self-control and better decision-making (Tangney et al., 2007). In addition, as mentioned before Nikodemou's (2020) model also suggests that negative emotion leads to a depletion of self-control resources making individuals more inclined to choose instant satisfaction over long-term rewards which is again something that is in line with previous findings (Baumeister et al., 2007).

Nonetheless, it is important to note that the impact of negative emotions on self-control is not always negative as there are some cases where negative emotions can enhance self-control, especially in cases where the emotions are related to an individual’s goals (Tamir, 2016). For example, when confronted with an obstacle that gets in the way of achieving your desired long-term goal, feeling angry or frustrated can motivate you to persist in overcoming it.

Chapter 3

Design and Implementation

3.1 Introduction

In this chapter, we will discuss a computational model that simulates the connection between emotions and precommitment in a self-control model. Similar to the work of Georgiou (2015), we are using the Q-Learning algorithm to develop an agent that competes in games with variable payoffs. However, we are also implementing the concept of precommitment as introduced by Banfield (2006) into the model. The precommitment is simulated by adding a differential bias, denoted by psi (ψ), to the diagonal terms of the payoff matrix used in the game. This allows us to investigate how emotions and precommitment interact to influence decision-making in the model. The programming language used for the implementation of the model is Java, and two main classes are being used: Player and MainProgram, which were developed by Georgiou (2015) and extended by Nikodemou (2020).

3.2 Simulating precommitment

Psychologists have long studied the concept of self-control and believe that precommitment is a useful technique to help individuals overcome self-control problems and stick to their goals or plans. Banfield's (2006) empirical research provides further support for this concept. This is why, in my implementation, I have incorporated the effect of precommitment using a parameter called "psi" on the model. The value of ψ is defined at the beginning of the testing and is fixed for the duration of the testing having each player learn to make decisions that are consistent with the level of precommitment (ψ = 0.01 for low level of precommitment and ψ = 0.9 for high level of precommitment). The decision to utilize such a small value for low precommitment (0.01) is based on the understanding that even a slight inclination towards immediate rewards can impact decision-making and self-control. This choice aligns with the empirical research conducted by Banfield (2006), where the same value was employed to study the impact of low precommitment.

The value of ψ represents a differential bias that is applied to the payoff matrix to calculate the differential payoff. As Banfield (2006) notes, low levels of precommitment can make it difficult for individuals to stick to their goals or plans, while high levels of precommitment can help individuals stay on track and achieve their desired outcomes. It is also worth noting that while precommitment affects the same two out of the four payoff values as emotions do it has a distinct meaning because it is added at the beginning of the game and remains fixed for the duration of the game in contrast with emotions which are added at different intervals throughout the game.

In the implementation, ψ is added or subtracted to the corresponding matrices of each agent using the equations lower.setPayoff_matrix(R, T-psi, S+psi, P) for the agent simulating the lower part of the brain and higher.setPayoff_matrix(R, S-psi, T+psi, P) for the agent simulating the higher part of the brain. By doing this, the player values more immediate rewards (higher S value) and fewer future rewards (lower T value) in the case of the lower player, and more future rewards (higher T value) and fewer immediate rewards (lower S value) in the case of the higher player.

This way, the behavior (C, D) is promoted, making it more similar to the reward for mutual cooperation for the agent simulating the higher part of the brain. Similarly, the behavior (D, C) is also promoted, making it more similar to the reward for mutual defection for the agent simulating the lower part of the brain. As a result, instead of four classes of rewards, the players are faced with just two, one with a tendency for cooperation and one with a tendency for defection, making it easier to choose the one for mutual cooperation.

3.3 Simulating emotions

For my research, I am following the implementation of Nikodemou's (2020) Q-learning model that simulates emotions by manipulating the values of the payoff matrix. The idea behind this approach is that positive emotions can promote self-control, while negative emotions can impair it.

To simulate the presence of positive emotions, Nikodemou (2020) incrementally increased the R payoff (positive reinforcement) and decreased the T payoff while incrementing the P and S payoffs (negative reinforcement). This means that cooperative behavior was more rewarded (higher R payoff), and defective behavior was more punished (lower T payoff). By doing this, the model was encouraged to learn cooperative behaviors and avoid defection.

Conversely, to simulate the presence of negative emotions, Nikodemou (2020) incremented the T payoff while decrementing the P and S payoffs (positive punishment) and decremented the R payoff (negative punishment). This means that cooperative behavior was punished more (lower R payoff), and defective behavior was rewarded more (higher T payoff). By doing this, the model was encouraged to learn defective behaviors and avoid cooperation.

Nikodemou (2020) experimented with different ratios in which the values of the payoff matrix were changed and the number of rounds that elapsed before changing the values again. By doing so, the model was able to simulate different emotional states such as happiness, anger, fear, and sadness.

Additionally, Nikodemou (2020) used different approaches to simulate emotions. For example, to simulate happiness, Nikodemou (2020) increased the Reward value for mutual cooperation. To simulate anger, Nikodemou (2020) increased the Temptation value for mutual defection. To simulate fear, Nikodemou (2020) decreased the Sucker value for being exploited by the opponent. To simulate sadness, Nikodemou (2020) decreased the Reward value for mutual cooperation.

This approach allows for the modeling of emotional states in decision-making scenarios, which can provide insight into the role of emotions in shaping behavior. It also provides a framework to investigate the effect of emotions on learning and decision-making processes.

3.4 The Q-Learning agents

The Q-Learning agents in the implementation are represented by instances of the Player class, and they are designed to simulate the interaction between the lower and higher parts of the brain through the IPD game. Each agent corresponds to one of these parts and has its own set of parameters, including the learning rate (η), epsilon value (ε) for the ε-greedy policy, discount factor (γ), Q-table for state-action pairs, payoff matrix, and current action taken by the agent. The discount factor is set differently for each agent to distinguish between them, as explained in Section 2.3.1.

The Q-table has dimensions of 4 rows and 2 columns, signifying the 4 possible states (CC, CD, DC, DD) and 2 actions (C, D) available to the agent. The Player class constructor initializes the Q-table and payoff matrix. Moreover, the class contains methods for setting and getting the values of each agent's parameters as shown in equation (3).

To address the exploration-exploitation dilemma, the e-greedy policy (Section 2.3.2) is used, which is implemented in the choose_action() method of the Player class. This method uses the epsilon value and the Q-values in the Q-table to determine whether the agent should choose a random action (explore) or exploit the Q-values to pick the most optimal action.

The update_Q() method is the central component of the Q-Learning algorithm and is responsible for updating the Q-values in the Q-table based on the rewards received from the IPD game. This method employs the update rule discussed in Section 2.3.1 to modify the values in the Q-table and promote learning for the agents.

3.5 The Q-Learning model

The Q-learning model is implemented in the MainProgram class, which initializes three arrays to hold the states, rewards, and overall payoff of the agents during learning. Each agent is created through the Player class constructor, where the payoff matrix values for both agents and the epsilon value for the epsilon-greedy exploration are set. For the first 500 rounds of the game, the value of epsilon for both agents is set to 0.1, which means that each agent will randomly select an action with a probability of 10%. However, for the next 500 rounds, the value of epsilon is set to 0, which means that both agents will always select the action that has the highest estimated value according to their Q-values. This change in the epsilon value allows the agents to explore the environment and learn from their experiences in the early stages of the game, while gradually shifting towards a more deterministic strategy as they gain more knowledge about the game. The discount factor (γ) is also set for each agent, with the agent representing the lower part of the brain (limbic system) having a discount factor of 0.1, which focuses on short-term rewards, and the agent representing the higher part of the brain (prefrontal cortex) having a discount factor of 0.9, which focuses on long-term rewards.

The parameters that define how the values of the payoff matrix change are then set. These parameters include the ratios for the Punishment value (ratio_P), the Reward value (ratio_R), the Sucker's value (ratio_S), the Temptation's value (ratio_T), the number of rounds (rounds) that need to elapse between each usage of the ratios and the ψ value (psi) used in the initialization of the matrix to simulate the existence of precommitment. It is important to note here that the update of each value of the payoff matrix takes place only if the two rules of the IPD game are satisfied.

The program then runs for 15 trials of 1000 episodes each, and the results are stored in text files for further analysis. At each round, both agents select an action based on their current state, which is used to update their Q-table. During learning, the state and reward obtained by each agent at each round are saved in the arrays. After 1000 episodes, the results are uploaded to two text files, payoffs.txt, and states.txt. The payoffs.txt file contains three columns that show the accumulated payoff of the lower brain agent, higher brain agent, and their sum, respectively. The states.txt file contains four columns that show the average percentage of the frequency of the CC, CD, DC, and DD states, respectively.

To analyze the data, two types of figures are produced. The first figure is a line plot with four traces, one for each outcome, which shows how the changing payoff values affect the outcomes throughout the rounds. For example, it shows whether one outcome is surpassed by the other and on which round, or how fast the system converges to a certain outcome. The second figure is a bar plot that shows the overall average outcomes, which is the percentage of the appearance of each state. This chart only shows the difference between the outcomes and which state dominated at the end.

Chapter 4

Results and Discussion

4.1 Introduction

This chapter tests the effectiveness of the methods described in Chapter 3 in simulating emotions and precommitment in a self-control model. Q-learning agents will compete in a game of IPD, exposed to emotional stimuli and precommitment in order to assess how emotions and precommitment impact the agents' decision-making and ability to exercise self-control.

4.1.1 Constant payoff matrix

This section establishes the baseline results for our future experiments, using a constant payoff matrix (Figures 4.1.1-4). The model of self-control we are using, as described in the thesis of Georgiou (2015), and enhanced later on by Nikodemou (2020), sets the learning rate and epsilon value of 0.1 for the first 500 rounds, and 0 for the remaining 500 rounds for both agents. In addition, the γ parameter, or the discount factor, is set to 0.1 for the lower agent and 0.9 for the higher agent as values of the γ parameter closer to 0 indicate that the agent focuses on the immediate (or SS) rewards, whereas values closer to 1 indicate that the agent focuses on the future (or LL) rewards. The initial payoff matrix values for all experiments are T=5, R=4, P=-2, S=-3, which correspond to "an internal conflict of moderate intensity" (Cleanthous, 2010). This indicates that the choices presented to the agents involve a moderate level of conflict, where neither the options are clear and straightforward (strong conflict) nor excessively complex or similar (weak conflict). The model runs 15 trials of the IPD game, with each game consisting of 1000 rounds (500 rounds before and 500 after). Two types of figures are produced for our experiment. Figure 4.1.1 presents the average of the outcomes and Figure 4.1.2 shows the average outcomes after 1000 rounds under low precommitment. Similarly, Figure 4.1.3 presents the average of the outcomes and Figure 4.1.4 shows the average outcomes after 1000 rounds under high precommitment. In analyzing the results, it is important to note that when the percentage of CC (Cooperate-Cooperate) states exceeds 50%, it indicates that self-control is achieved, as both agents consistently choose cooperation over defection.

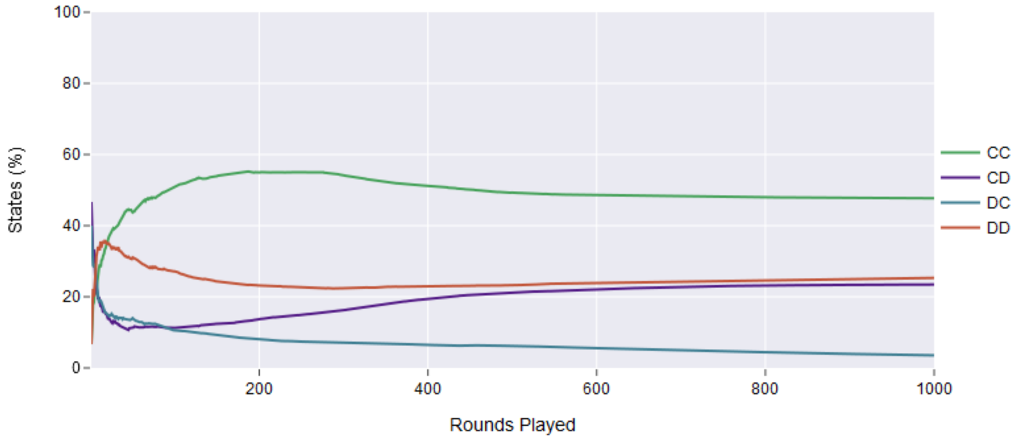


Figure 4.1.1: *Average outcomes CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game with the constant payoff matrix T=5, R=4, P=-2, S=-3, under low precommitment.*

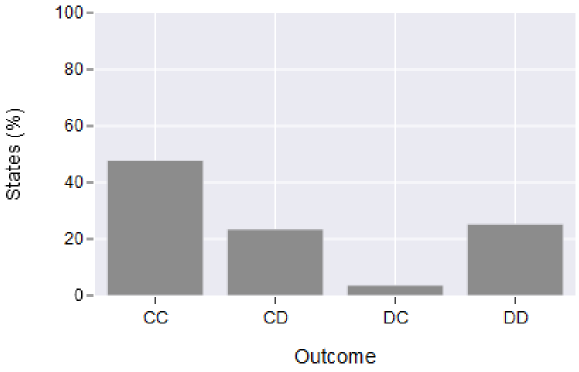


Figure 4.1.2: *Overall average outcomes of CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game with the constant payoff matrix T=5, R=4, P=-2, S=-3, under low precommitment.*

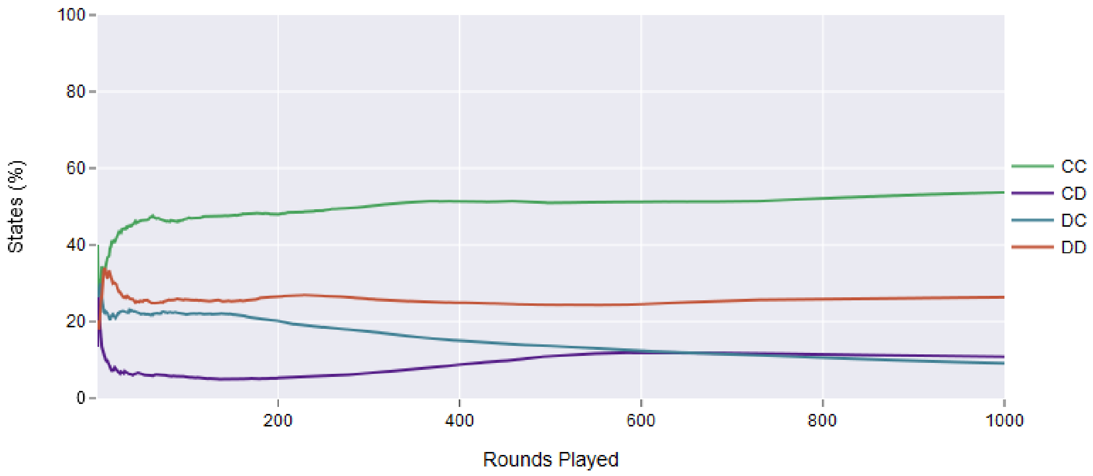


Figure 4.1.3: *Average outcomes CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game with the constant payoff matrix T=5, R=4, P=-2, S=-3, under high precommitment.*

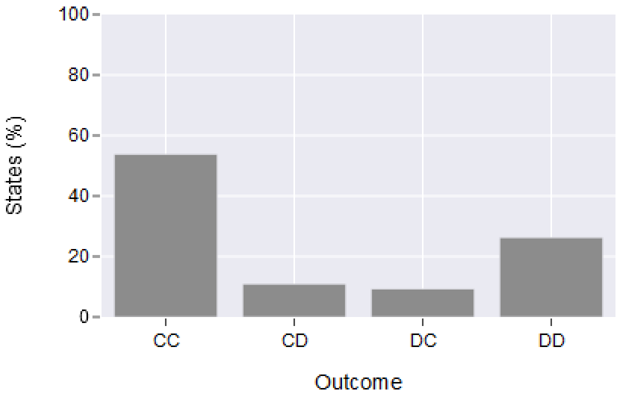


Figure 4.1.4: *Overall average outcomes of CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game with the constant payoff matrix T=5, R=4, P=-2, S=-3, under high precommitment.*

4.2 Simulating Precommitment in a Positive Emotional State

In 4.1.1, we used a constant payoff matrix to produce the baseline results, and the model is now ready to be tested with a non-constant matrix in order to test whether precommitment paired with positive emotions improve self-control or impair it. In order to simulate the level of precommitment the value of ψ will be alternating between 0.01 and 0.9 and for the presence of positive emotions, some payoff matrix values (T, R, P, S) will gradually change through the 1000 rounds of the IPD game. More specifically, positive emotional state is affected by the increment or decrement between each value (positive intensity value or negative intensity value) and the number of rounds between each change (interval of change). The same initial payoff matrix, learning rates and epsilon values as in section 4.1.1 are being used.

4.2.1 Increasing the Reward payoff

We implemented two different scenarios to simulate the increment of the Reward payoff in different levels of precommitment. By increasing the R value, we establish a stronger connection with positive emotions, as individuals are more motivated and rewarded for engaging in cooperative behaviors. The first scenario was a moderate positive intensity value (0.25) in moderate interval of change (25 rounds). The results were promising as they showed that the combination of a moderate positive intensity value, moderate increment, and high levels of precommitment (ψ=0.9) provides better results in contrast with low precommitment (ψ=0.01). By comparing the graphs in Figure 4.2.1 with the baseline results we can see that there is a slight increment in low precommitment without it managing to reach self-control (49.7%) while high precommitment achieves a percentage of 59.7% in CC states

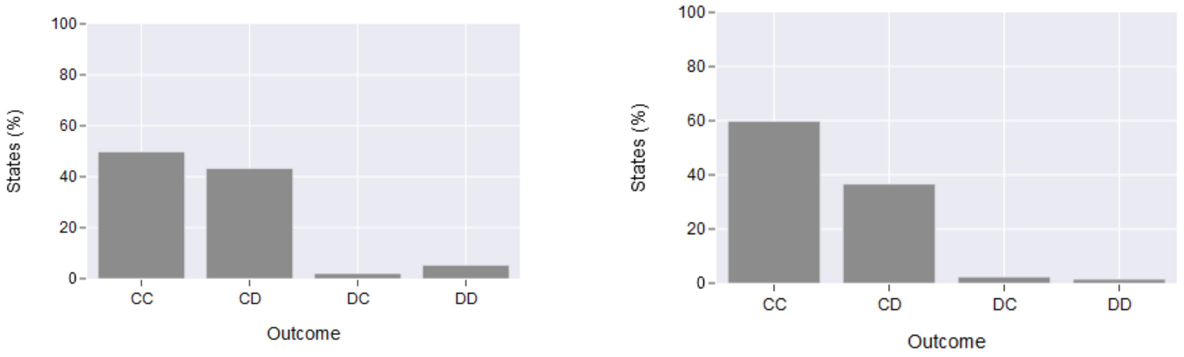


Figure 4.2.1: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.25, Interval=25 rounds.*

For the second scenario we tried a small positive intensity (0.1) in an infrequent interval (50 rounds). The results in Figure 4.2.2 showed an small change in low precommitment (ψ=0.01) with 50.5% in CC states, thus achieving self-control and a rather significant one in high precommitment (ψ=0.9) reaching 72.3% in CC states. Therefore, a smaller positive intensity value given infrequently paired with high precommitment status improves self-control behavior.

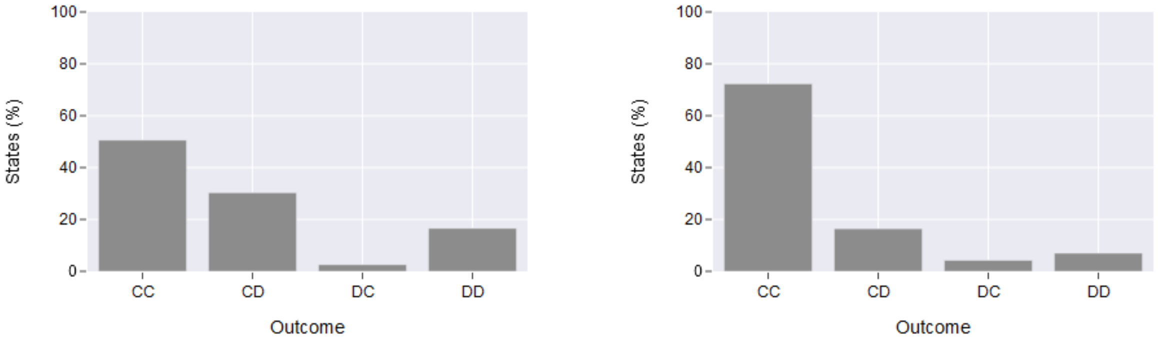


Figure 4.2.2: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=50 rounds.*

4.2.2 Increasing the Punishment payoff

We implemented three different scenarios to simulate the increment of the Punishment payoff in different levels of precommitment and thus simulating the decrement of negative emotional states. Increasing the Punishment value implies stronger penalties for defective behaviors thus discouraging individuals from engaging in actions that go against cooperative behavior. The first scenario was a small positive intensity value (0.1) in a small interval of change (10 rounds). The results showed that high interval with the combination of small positive intensity value, and high levels of precommitment (ψ=0.9) managed to achieve self-control rather than with low precommitment (ψ=0.01) where self-control behavior was not achieved (Figure 4.2.3). By comparing the graphs in Figure 4.2.3 with the baseline results we can see that there is a large decrement in low precommitment and a small decrement in high precommitment (52%).

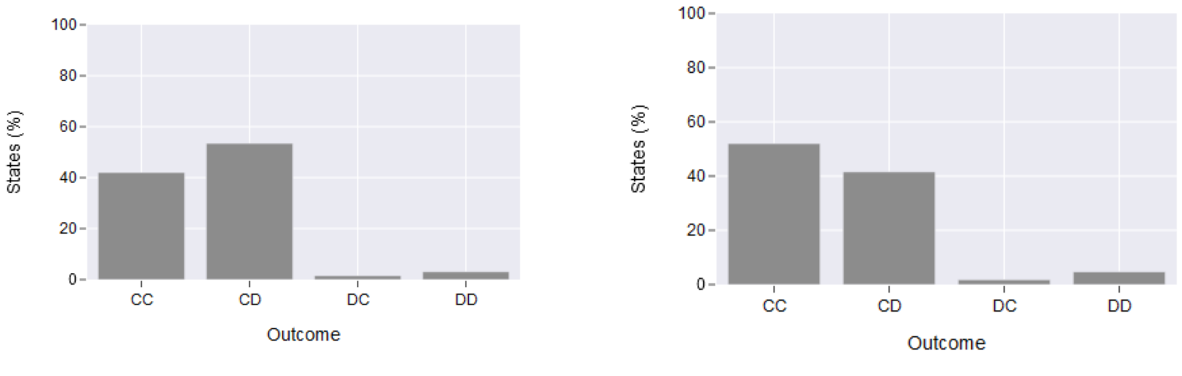


Figure 4.2.3: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000*

rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=10 rounds.

For the second scenario we tried a small positive intensity value (0.1) in a large interval of change (50 rounds). The results in Figure 4.2.4 showed a change in both low (53.4%) and high precommitment (61.0%) testings with both of them achieving self-control and exceeding the baseline results. Therefore, a smaller positive change given infrequently paired with high precommitment status achieves self-control behavior, but it is not as effective as increasing the R value paired with high precommitment.

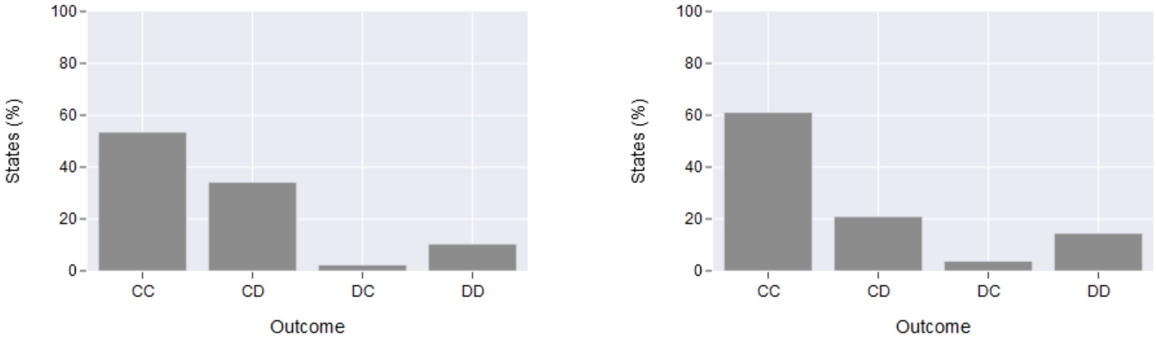


Figure 4.2.4: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=50 rounds.*

Lastly, for the third scenario we compared the average outcomes CC, CD, DC, and DD during the 1000 rounds between a small positive intensity value (0.1) in a high interval (50 rounds) and a large positive intensity value (0.5) in a high interval (50 rounds). Both outcomes were obtained in high precommitment status (ψ=0.9), and we can see that the model that received a smaller positive intensity value (Figure 4.2.5) in Punishment consistently provided better results than the one with a larger intensity value (Figure 4.2.6).

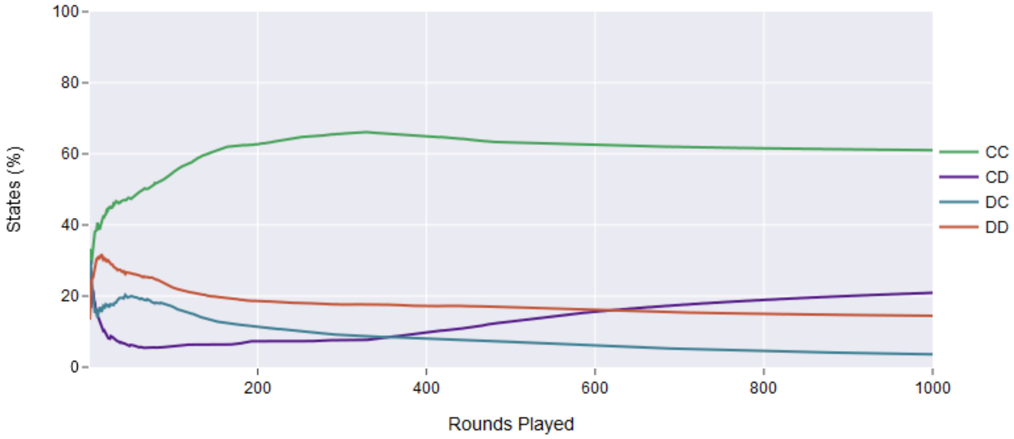


Figure 4.2.5: *Average of the outcomes CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game under high precommitment. Increasing the P payoff. Positive Intensity value=0.1, Interval=50 rounds.*

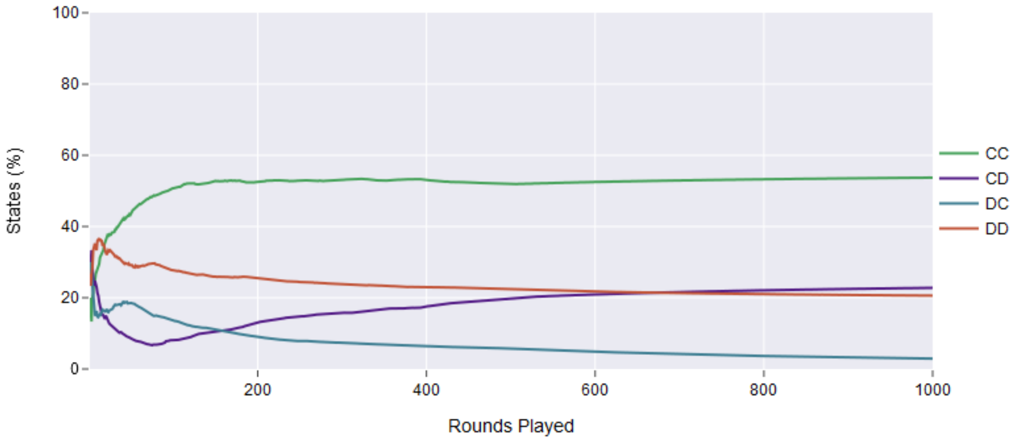


Figure 4.2.6: Average of the outcomes CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game under high precommitment. Increasing the P payoff. Positive Intensity value=0.5, Interval=50 rounds.

4.2.3 Increasing the Sucker's payoff

We implemented three different scenarios to simulate the increment of the Sucker's payoff in different levels of precommitment and thus simulating the decrement of negative emotional states. Increasing the Sucker's payoff (smaller negative value) promotes cooperation by reducing vulnerability and negative emotions associated with exploitation. The first scenario was a small positive intensity value (0.1) in a small interval of change (10 rounds). The results showed that high frequency with the combination of small intensity, and high levels of precommitment (ψ=0.9) managed to achieve self-control rather than with low precommitment (ψ=0.01) where we had a total self-control failure. By comparing the graphs in Figure 4.2.7 with the baseline results we can see that there is a large decrement in low precommitment and a small decrement in high precommitment (51.9%).

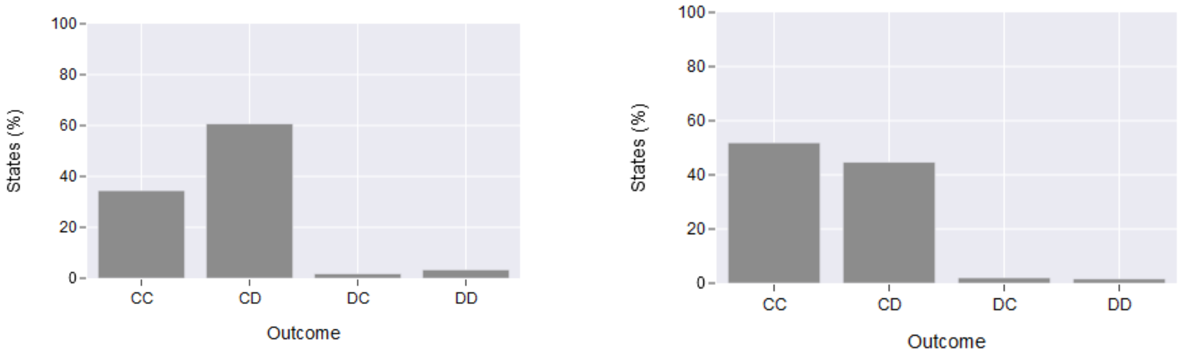


Figure 4.2.7: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* **(right)** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=10 rounds.*

For the second scenario we tried a small positive intensity value (0.1) in a large interval of change (50 rounds). The results in Figure 4.2.8 showed a significant change in both low (ψ=0.01) and high precommitment (ψ=0.9) testing with low making a big leap towards achieving self-control (45.3%) and in the case of high precommitment overpassing the baseline results with a percentage of 67% in CC states. Therefore, a smaller positive change given infrequently paired with high precommitment status achieves self-control behavior, but again it is not as effective as increasing the R value paired with high precommitment.

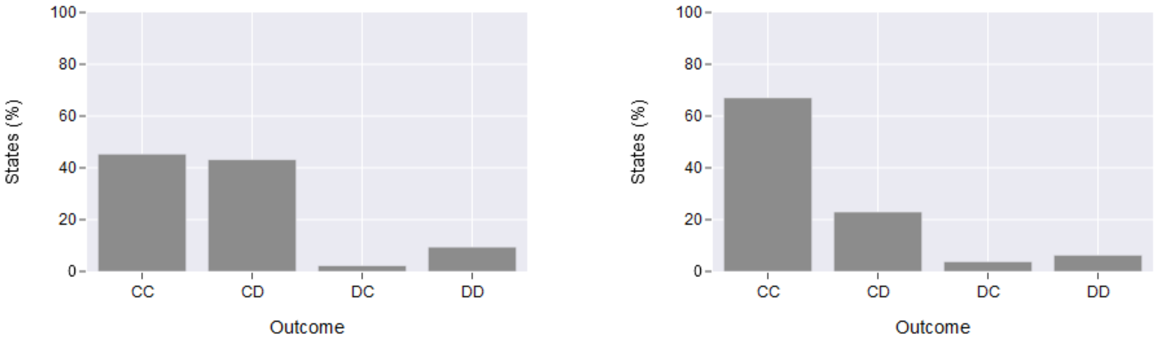


Figure 4.2.8: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=50 rounds.*

Lastly, for the third scenario we compared the average outcomes CC, CD, DC, and DD during the 1000 rounds between a small positive intensity value (0.1) in a high interval of change (50 rounds) and a large intensity value (0.5) in a high interval (50 rounds). Both outcomes were obtained in high precommitment status (ψ=0.9), and we can see that the model that received a smaller increment (Figure 4.2.9) in Sucker's consistently provided better results than the one with a larger increment (Figure 4.2.9).

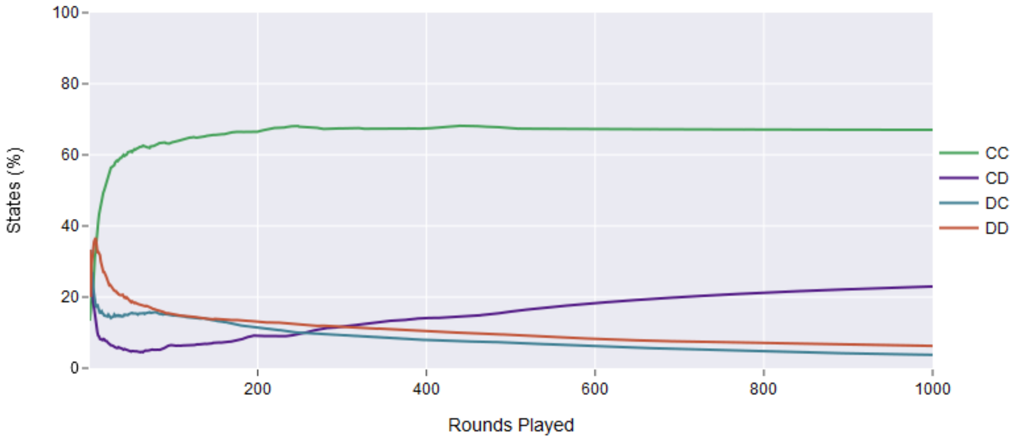


Figure 4.2.9: *Average of the outcomes CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game under high precommitment (ψ=0.9). Increasing the P payoff. Positive Intensity value=0.1, Interval=50 rounds.*

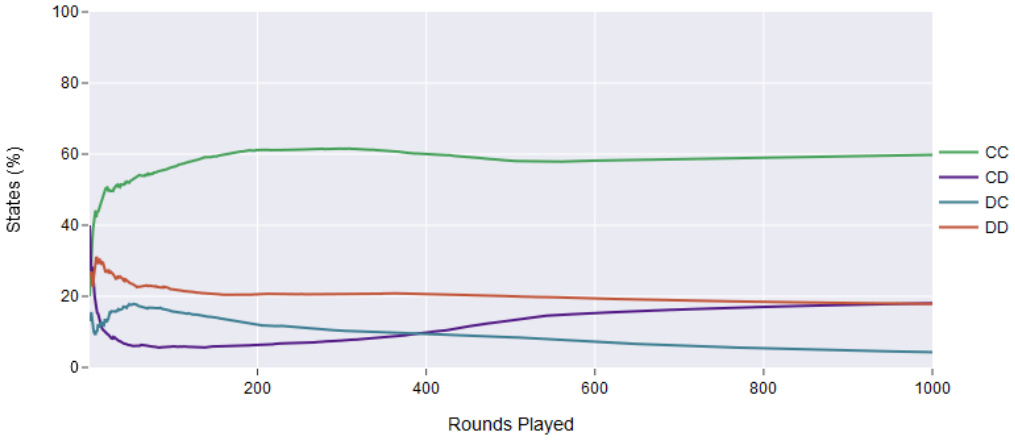


Figure 4.2.10: *Average of the outcomes CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game under high precommitment (ψ=0.9). Increasing the P payoff. Positive Intensity value=0.5, Interval=50 rounds.*

4.2.4 Decreasing the Temptation payoff

Another way of implementing positive emotional state is through the decrement of negative aspects as is Temptation (T payoff). Lowering the T payoff decreases the attractiveness of selfish behavior, leading to positive emotional states and encouraging cooperative actions. For the decrement of Temptation, we tried using the results obtained from Section 4.2.1 to see whether there is a correlation between increment and decrement of payoffs. That is why for our first scenario we used a moderate negative intensity value (0.25) in moderate interval of change (25 rounds). The results made it challenging to draw conclusive comparisons because even though self-control is achieved the graphs showed great similarity between the two levels of precommitment (Figure 4.2.11).

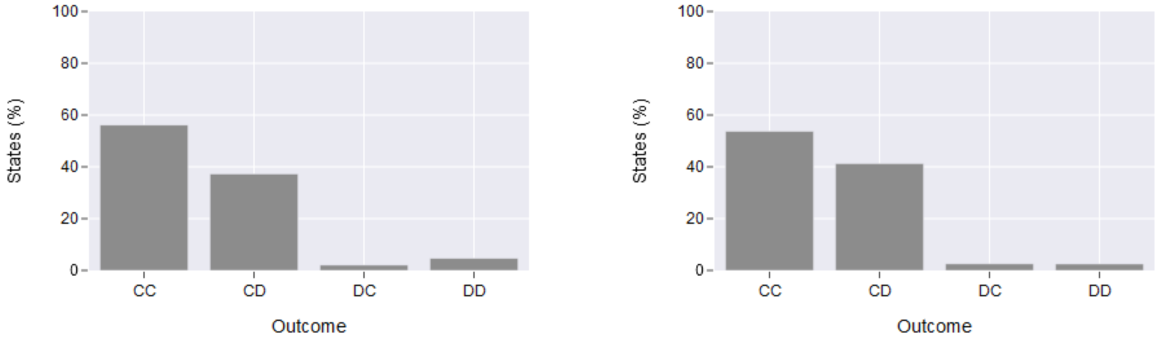


Figure 4.2.11: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.25, Interval=25 rounds.*

For the second scenario we tried a small negative intensity value (0.1) in a large interval of change (50 rounds). The results in Figure 4.2.12 showed an insignificant change in low precommitment (53.9%) and a rather significant one in high precommitment (ψ=0.9) reaching 65.7% in CC states. Therefore, a correlation between the small infrequent increment of Reward and the small infrequent decrement of Temptation paired with high precommitment appears to exist as it improves self-control behavior.

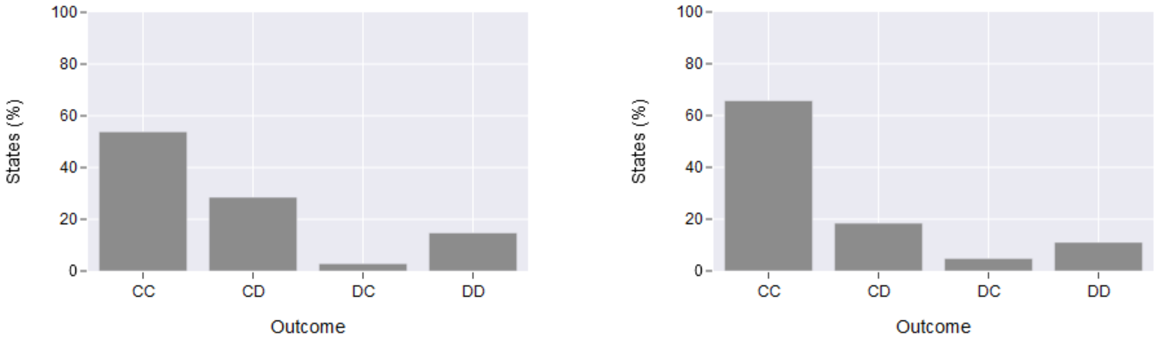


***Figure 4.2.12: (left)** Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01). **(right)** Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=50 rounds.*

4.2.5 Decreasing the Temptation and increasing the Punishment and Sucker's payoff

The last used method was a combination of the three previous methods for the decrement of negative emotion, that is decreasing the T and increasing the P and S, all at the same time while changing the precommitment status in each testing to see under which scenario we produce the best results. We set the intensity value to 0.1 for all the three values, since for all three it was the value that gave the best performance and a 50 rounds interval to provide an large interval of change. As expected the results showed the best performance out of all the methods for both low and high precommitment by reaching 60.4% and 75% respectively (Figure 4.2.13).

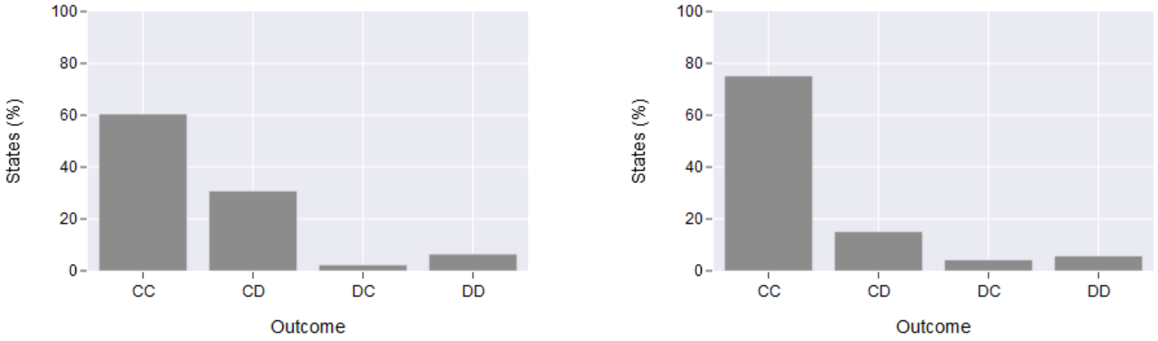


Figure 4.2.13: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.1 for T payoff, Positive Intensity value=0.1 for S, P payoffs Interval=50 rounds*

One could argue that the approach of increasing temptation and decreasing punishment and sucker's payoff follows a similar principle to incorporating precommitment into the lower agent, as described in Section 3.2. In order to gain a deeper understanding of the individual effects of precommitment and emotions, it would be valuable to establish a new baseline scenario where precommitment is entirely excluded. By comparing the results of this new baseline with the previous approach that involved precommitment, we can observe the significant impact of precommitment on achieving self-control. The baselines, which did not incorporate precommitment, managed to reach a success rate of 57.1% (Figure 4.2.14). This comparison underscores the importance of precommitment, as it clearly demonstrates that precommitment and emotions have distinct and complementary effects in promoting self-control.

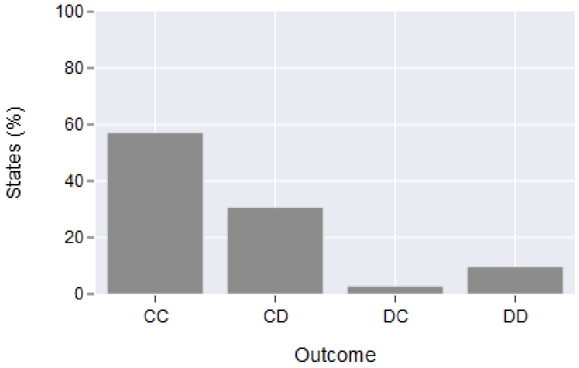


Figure 4.2.14: *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with no precommitment (ψ=0). Negative Intensity value=0.1 for T payoff, Positive Intensity value=0.1 for S, P payoffs Interval=50 rounds*

4.2.6 Summary and discussion on Precommitment in a Positive Emotional State

The first method used to test self-control was to increase the presence of positive emotions through positive reinforcement (increased R) at different levels of precommitment. The results indicated that positive reinforcement, when paired with high precommitment, had a significant impact on achieving self-control behavior compared to low precommitment (Figure 4.2.1). This effect was particularly pronounced when the interval of change (rounds) was sufficiently large, and the positive intensity value was small enough to provide a smaller magnitude at a more infrequent rate (Figure 4.2.2). The same pattern was observed when negative reinforcement methods were used to ease internal conflict (decreased T) and explicitly eliminate negative emotions (increased P, S) at different levels of precommitment. More specifically, a negative intensity value of 0.1 for T (Figure 4.2.12) and positive intensity value of 0.1 for P and S (Figures 4.2.4 & 4.2.8) was found to be the most effective when given infrequently in a high precommitment state, as values greater than this impaired self-control behavior and, in some cases, led to total self-control failure. Finally, as it was expected when combining the results of all three methods of negative reinforcement in both low and high precommitment self-control was achieved with high precommitment achieving the highest percentage of CC states of 75%

(Figure 4.2.13). Overall, the results showed that precommitment paired with positive emotions can be an effective strategy to improve self-control behavior.

Payoff Value	**Precommitment**	**Intensity**	**Interval**	**CC states**
Reward (R)	Low (ψ=0.01)	Positive (0.1)	50 rounds	50.5% ▲
Reward (R)	High (ψ=0.9)	Positive (0.1)	50 rounds	72.3% ▲
Punishment (P)	Low (ψ=0.01)	Positive (0.1)	50 rounds	53.4% ▲
Punishment (P)	High (ψ=0.9)	Positive (0.1)	50 rounds	61.0% ▲
Sucker (S)	Low (ψ=0.01)	Positive (0.1)	50 rounds	45.3% ▼
Sucker (S)	High (ψ=0.9)	Positive (0.1)	50 rounds	67.0% ▲
Temptation (T)	Low (ψ=0.01)	Negative (0.1)	50 rounds	53.9% ▲
Temptation (T)	High (ψ=0.9)	Negative (0.1)	50 rounds	65.7% ▲
P, S & T Combined	Low (ψ=0.01)	Positive (0.1) for P & S Negative (0.1) for T	50 rounds	60.4% ▲
P, S & T Combined	High (ψ=0.9)	Positive (0.1) for P & S Negative (0.1) for T	50 rounds	75.0% ▲

Figure 4.2.15: *Summary of the best results of each scenario tested in relation to the affected payoff value, the level of precommitment, intensity and interval of change. The outcome is given in the form of a percentage indicating the overall percentage of CC states achieved throughout the course of the game. The symbols "▲" and "▼" indicate whether self-control was achieved or not, respectively, with "▲" denoting successful self-control (percentage exceeding 50%) and "▼" indicating a lack of self-control.*

4.3 Simulating Precommitment in a Negative Emotional State

We have tested whether precommitment paired with positive emotions improves self-control or impairs it, now we can test the same regarding negative emotions. In order to simulate the level of precommitment the value of ψ will be again alternating between 0.01 and 0.9 and for the presence of negative emotions, some payoff matrix values (T, R, P, S) will gradually change through the 1000 rounds of the IPD game. More specifically, negative emotional state is affected by increment or decrement between each value (positive intensity value or negative intensity value) and the number of rounds between each change (interval of change). Again, the same initial payoff matrix, learning rates and epsilon values as in section 4.1.1 are being used.

4.3.1 Decreasing the Reward payoff

We implemented two different scenarios to simulate the decrement of the Reward payoff in different levels of precommitment. By decreasing the R value, the connection with positive emotions weakens, as individuals receive fewer rewards and motivation for engaging in cooperative behaviors diminishes. Following our previous findings on increasing the reward payoff we tested a small negative intensity value (0.1) in a large interval of change (50 rounds). The results were promising as they showed that high precommitment (ψ=0.9) is sufficient enough to counter the subtle decrement in positive emotional state when given in an infrequent rate (Figure 4.3.1) in contrast with low precommitment (ψ=0.01) where self-control was not sustained through the course of the game and was, in the end, not achieved (Figure 4.3.2). By comparing the graphs in Figure 4.3.1 with the baseline results we can see that there is a decrement in low precommitment with CC state reaching only 39.6% whereas high precommitment manages to achieve a percentage of 58.4% in CC states.

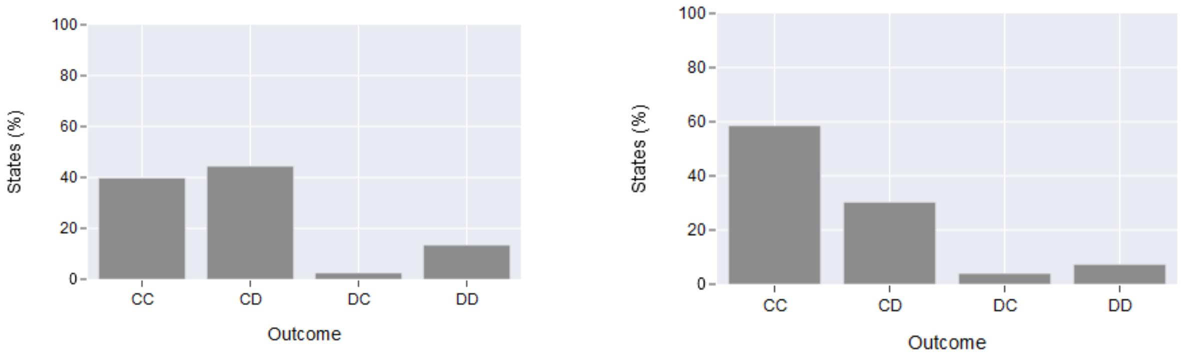


Figure 4.3.1: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.1, Interval=50 rounds.*

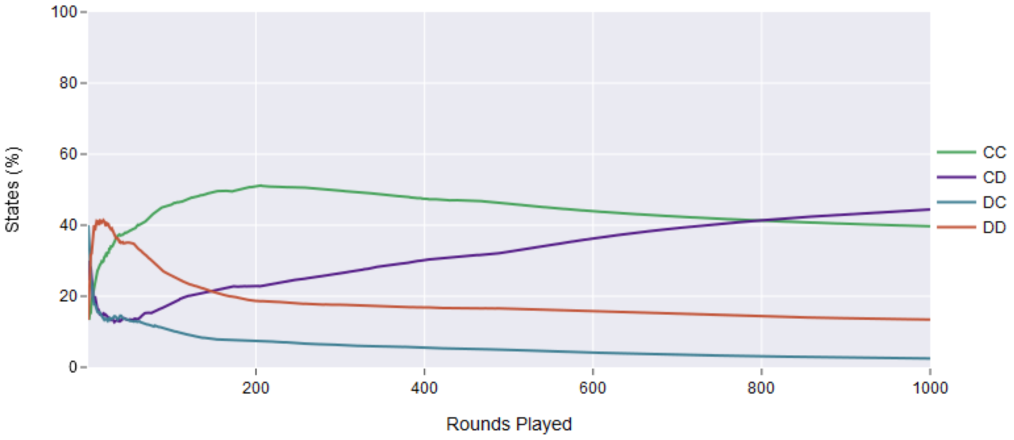


Figure 4.3.2: *Average of the outcomes CC, CD, DC, and DD during 1000 rounds of the Q-learning agents playing the IPD game. Decreasing the R payoff. Negative Intensity value=0.1, Interval=50 rounds in low precommitment (ψ=0.01).*

For the second scenario we tried a larger negative intensity value (0.5) in a large interval of change (50 rounds). The results in Figure 4.3.3 showed that larger increments, even when provided in an infrequent rate, have a great impact on self-control as both low and high precommitment scenarios were not able to sustain self-control and only achieved an overall percentage of 22.9% and 26.0% respectively. Therefore, a larger negative emotional change experienced in an infrequent rate cannot be overcome by high precommitment (ψ=0.9).

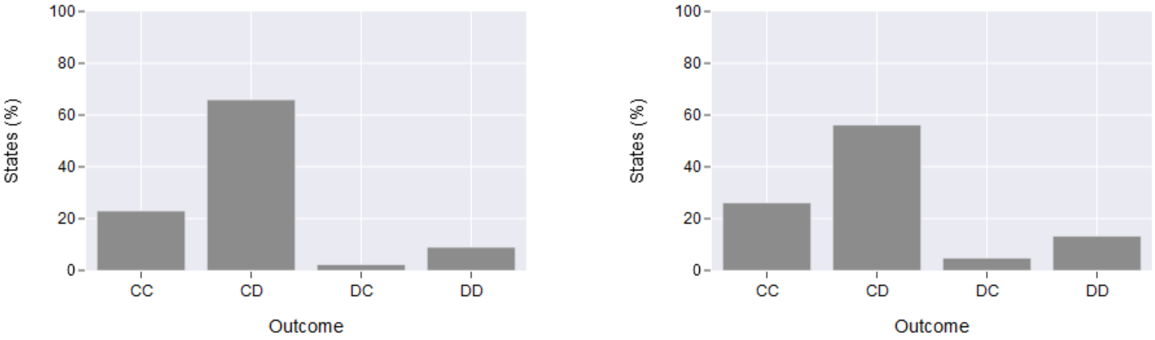


Figure 4.3.3: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.5, Interval=50 rounds.*

4.3.2 Decreasing the Punishment payoff

To provide the agent with a greater magnitude of negative signals we tried decreasing the Punishment payoff. Decreasing the Punishment payoff weakens the penalties imposed on defective behaviors, which can decrease the overall commitment to cooperative behavior. Since though there is small difference (1 point) between the Punishment's and Sucker's payoffs no values larger than 0.5 were tested as it will break the game's first rule after only a few changes. Firstly, we tested the negative intensity value of 0.1 in three different intervals (10, 25 and 50 rounds) under both low and high precommitment. The results showed that as the number of rounds increase, meaning the frequency rate gets lower, the results get better (Figure 4.3.4), something which is further enhanced by the addition of high precommitment (Figure 4.3.5).

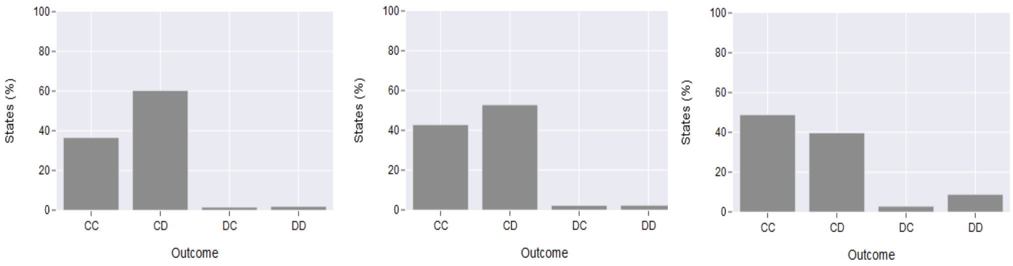


Figure 4.3.4: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.1, Interval=10 rounds.* ***(middle)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.1, Interval=25 rounds.* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.1, Interval=50 rounds. Low precommitment (ψ=0.01).*

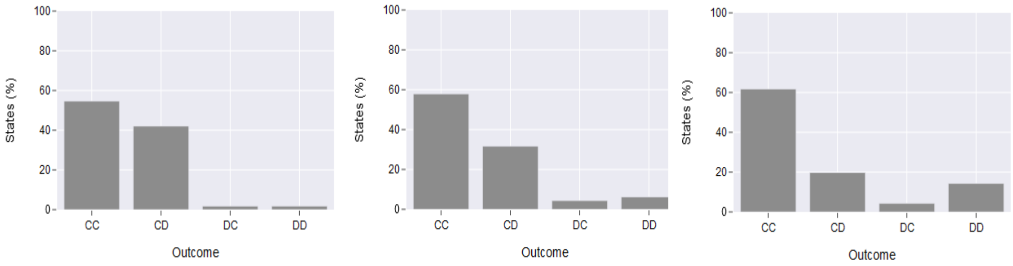


Figure 4.3.5: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.1, Interval=10 rounds.* ***(middle)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.1, Interval=25 rounds.* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.1, Interval=50 rounds. High precommitment (ψ=0.9).*

Secondly, we tested the negative intensity value of 0.25 in the three different intervals tested before (10, 25 and 50 rounds) under both low and high precommitment. The results showed that as the numbers of rounds increase, meaning the frequency rate gets lower, the results get better (Figure 4.3.6), something which is further enhanced by the addition of high precommitment (Figure 4.3.7).

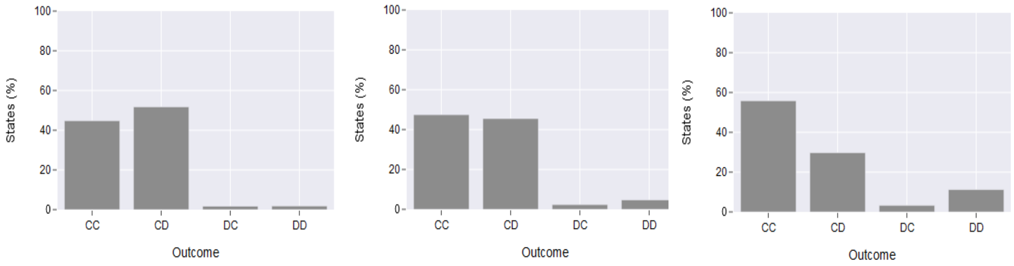


Figure 4.3.6: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.25, Interval=10 rounds.* ***(middle)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.25, Interval=25 rounds.* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.25, Interval=50 rounds. Low precommitment (ψ=0.01).*

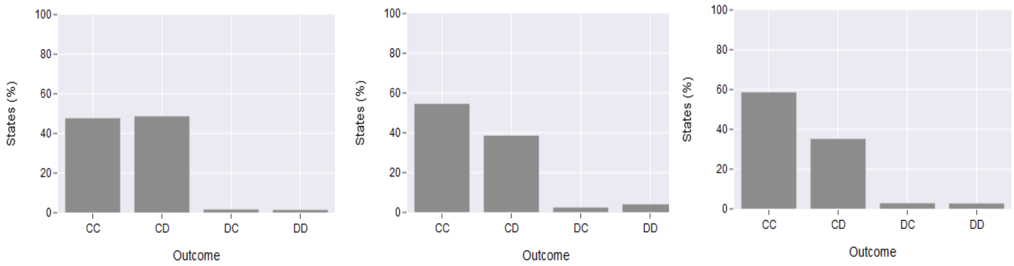


Figure 4.3.7: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.25, Interval=10 rounds.* ***(middle)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.25, Interval=25 rounds.* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.25, Interval=50 rounds. High precommitment (ψ=0.9).*

Finally, we tested the negative intensity value of 0.5 in the three different intervals tested before (10, 25 and 50 rounds) under both low and high precommitment. Even though the results were confusing when tested with low precommitment (Figure 4.3.8), after testing the same values in high precommitment (Figure 4.3.9), we could see that as the numbers of rounds increase, meaning the interval gets larger the results get better with low precommitment reaching a max percentage of 52.9% and high of 66.5 in CC states.

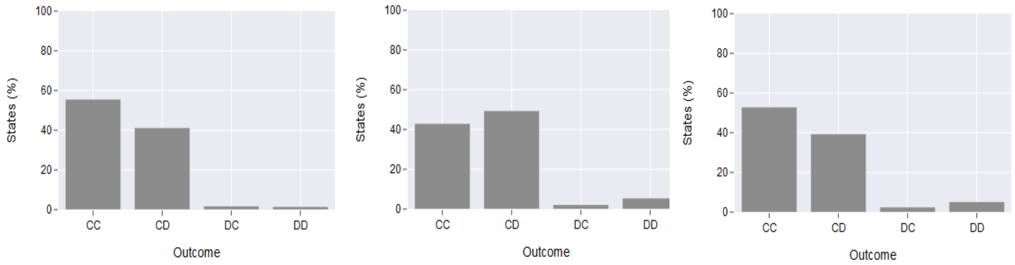


Figure 4.3.8: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.5, Interval=10 rounds.* ***(middle)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.5, Interval=25 rounds.* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.5, Interval=50 rounds. Low precommitment (ψ=0.01).*

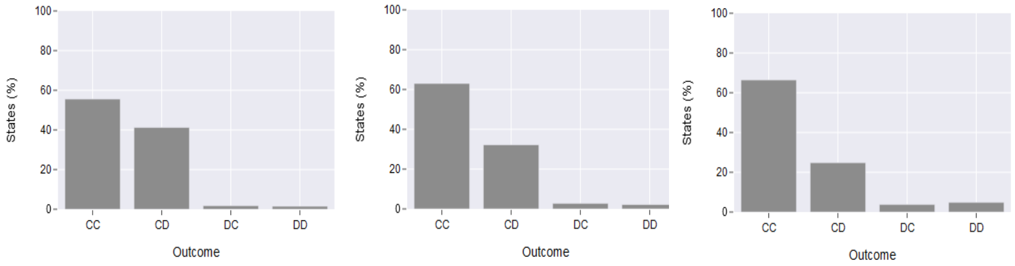


Figure 4.3.9: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.5, Interval=10 rounds.* ***(middle)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.5, Interval=25 rounds.* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with Negative Intensity value=0.5, Interval=50 rounds. High precommitment (ψ=0.9).*

4.3.3 Decreasing the Sucker’s payoff

To provide the agent with even greater magnitudes of negative signals we tried decreasing Sucker’s payoff. Decreasing the Sucker's payoff weakens the deterrent against exploitative behaviors, reducing the motivation for individuals to engage in cooperative actions. Since S is not restricted by any of the rules of the IPD game three different negative intensity values like 0.1, 0.5 and 1,0 were tested in a frequency rate of 10 and 50 rounds interval. Firstly, we tested the ratio of 0.1 in the two different intervals under both low and high precommitment. The results showed that high frequency with the combination of small negative intensity value, and high levels of precommitment (ψ=0.9) managed to achieve self-control rather than with low precommitment (ψ=0.01) where we had a self-control failure (Figure 4.3.10). By comparing the graphs in Figure 4.3.10 with the baseline results we can see that there is a large decrement in low precommitment (ψ=0.01) and a small decrement in high precommitment. However, when the interval of change is larger (Figure 4.3.11) low managed to make a leap toward self-control by reaching a percentage of 46.2% in CC states and high managed to achieve self-control with a percentage of 56%.

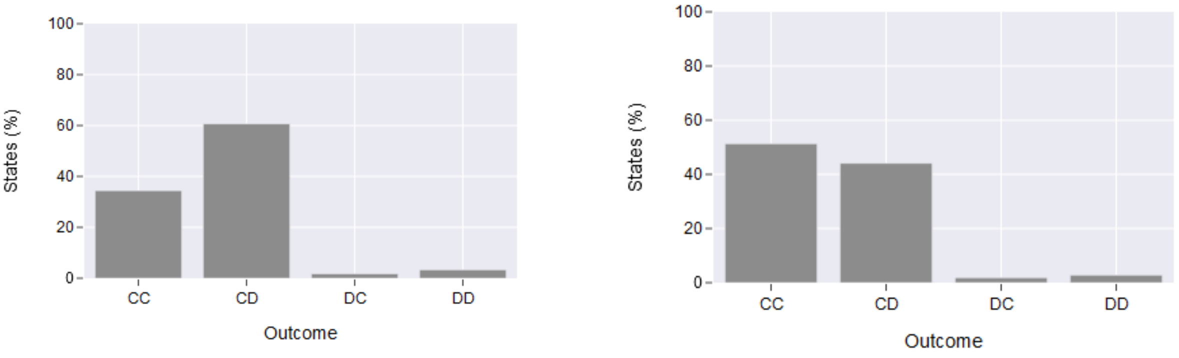


Figure 4.3.10: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.1, Interval=10 rounds.*

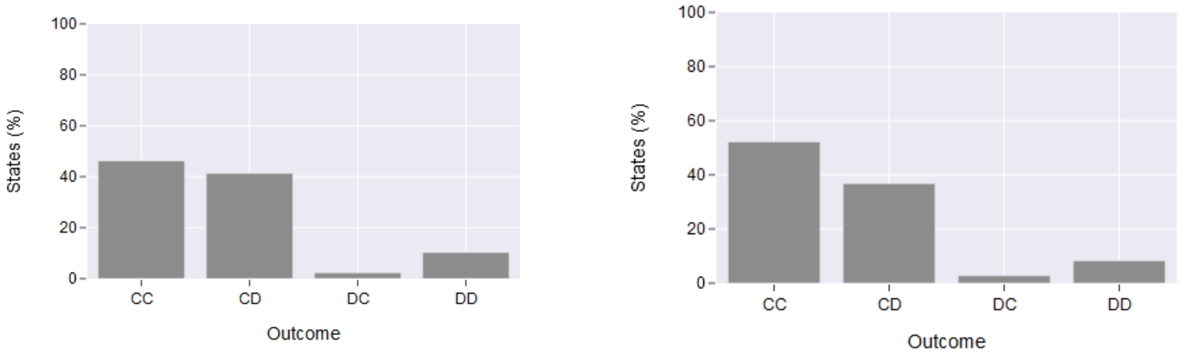


Figure 4.3.11: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.1, Interval=50 rounds.*

For the second scenario we tried a larger negative intensity value (0.5) in the two different intervals under both low and high precommitment. The results in Figure 4.3.12 showed that in small intervals of change neither of the two levels of precommitment could manage to achieve self-control behavior and instead led to a total self-control failure. In contrast in Figure 4.2.13 we can see for a large interval, low precommitment (ψ=0.01) did not manage to pass the threshold and achieve self-control (35.9%) whereas high precommitment reached percentage of 60% in CC thus achieved self-control. Therefore, a larger negative intensity value given infrequently (large interval) paired with high precommitment (ψ=0.9) status can achieve self-control behavior.

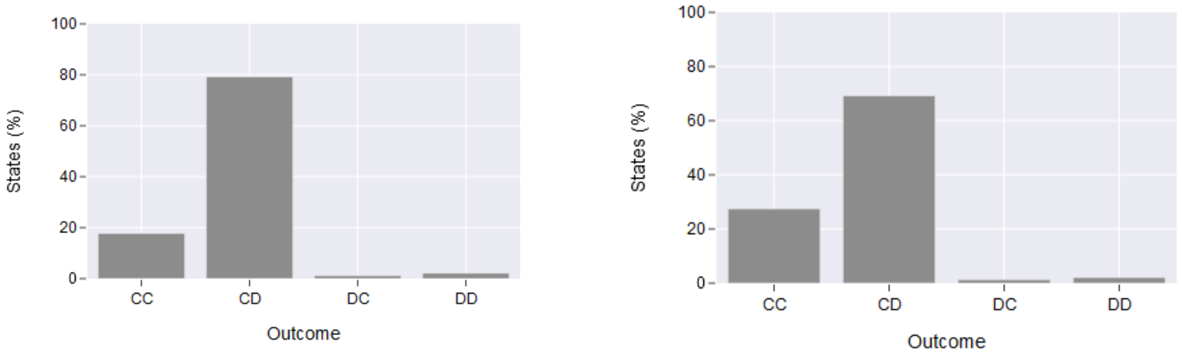


Figure 4.3.12: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.5, Interval=10 rounds.*

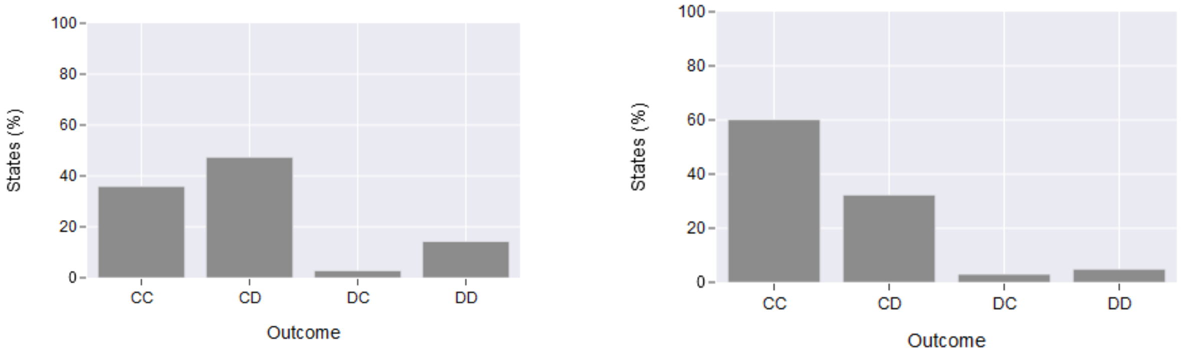


Figure 4.3.13: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=0.5, Interval=50 rounds.*

Lastly, we tried an even larger negative intensity value (1.0) in the two different intervals under both low and high precommitment to see whether this large increment in negative emotional state could still be handled by precommitment. The results in Figure 4.2.14 were not promising as they showed that for such a large increment neither of two precommitment levels could achieve self-control but instead led to a total self-control failure reaching as high as 84.2% in CD states. Similarly, in Figure 4.2.15 we can see that even though a larger interval of change provided better results it was still not enough to achieve self-control behavior with CC states reaching a max of 38.5%. Therefore, a very large negative intensity value given cannot be countered with precommitment.

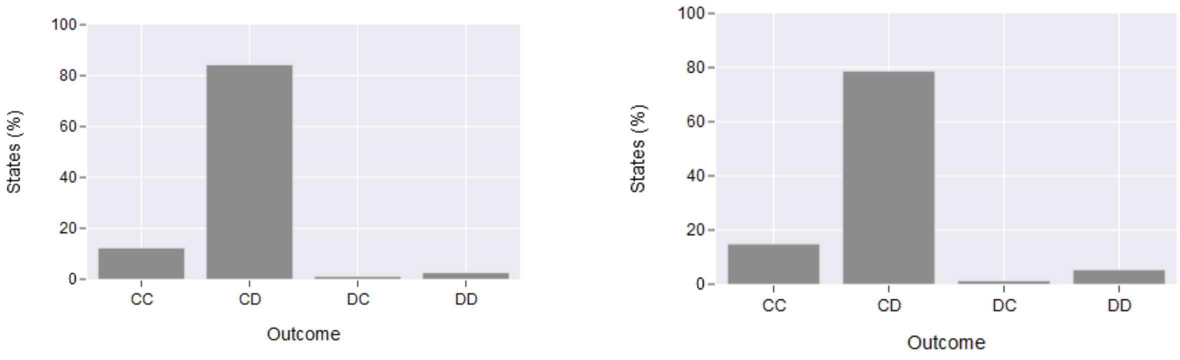


Figure 4.3.14: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=1.0, Interval=10 rounds.*

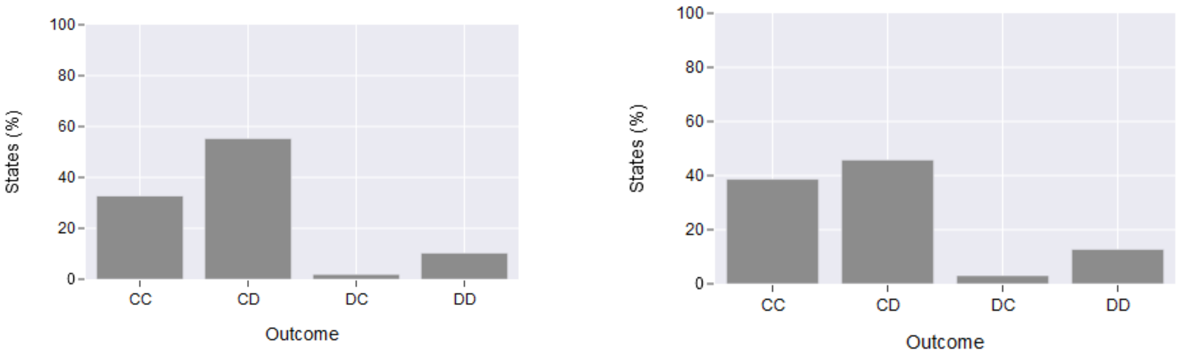


Figure 4.3.15: (left) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Negative Intensity value=1.0, Interval=50 rounds.*

4.3.4 Increasing the Temptation payoff

Another way of implementing negative emotional state is through the increment of negative aspects as is Temptation (T payoff). Increasing the Temptation payoff increases selfish behavior, as it makes it more tempting to prioritize immediate personal gains over cooperative actions. For our testing we used two different positive intensity values like 0.1 and 0.5 in intervals of change of 10 and 50 rounds. Firstly, we tested the intensity values of 0.1 in the two different intervals under both low and high precommitment. The results showed that large intervals with the combination of small positive intensity values, and high levels of precommitment (ψ=0.9) managed to achieve self-control (Figure 4.3.17) reaching a percentage 65.4% in high precommitment. In contrast, under small intervals of change (Figure 4.3.16) both low and high precommitment could not manage to overtake the negative impact of Temptation and thus did not achieve self-control.

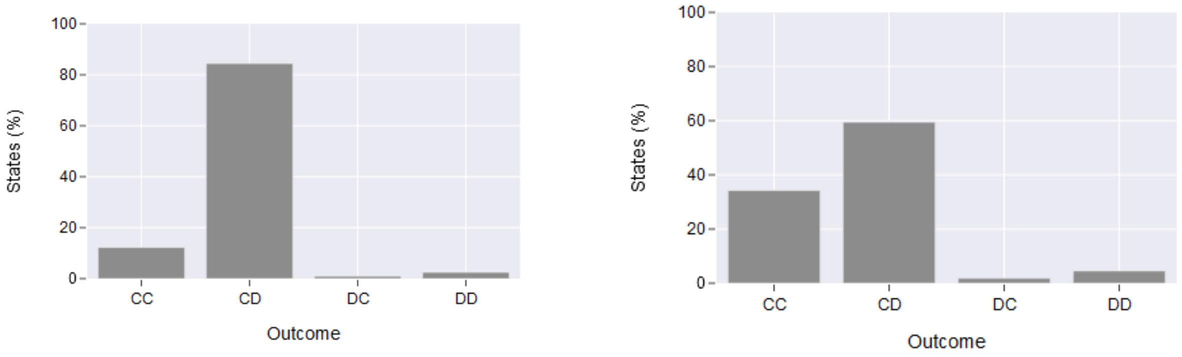


Figure 4.3.16: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=10 rounds.*

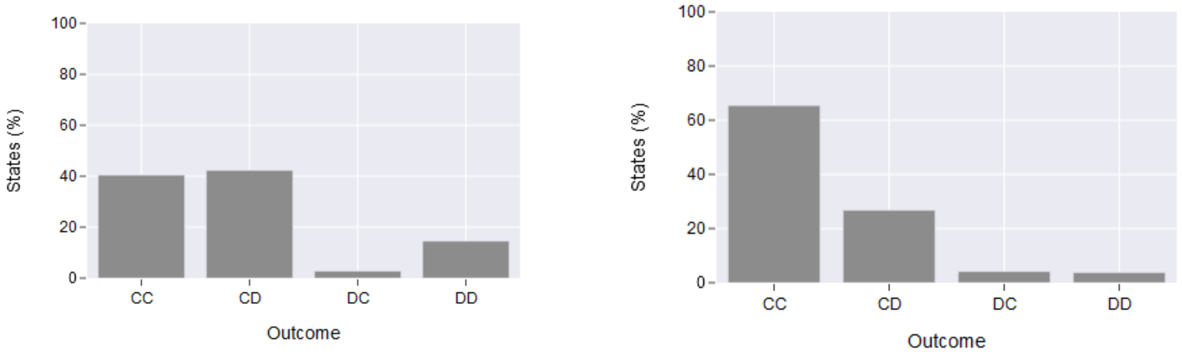


Figure 4.3.17: ***(left)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right)*** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1, Interval=50 rounds.*

For the second scenario we tried a larger positive intensity value (0.5) in the two different intervals under both low and high precommitment. The results in Figure 4.3.19 showed that even though there was an increment in CC states when the intervals of change were larger (50 rounds) and precommitment was high, none of the scenarios managed to achieve self-control behavior. Therefore, negative emotional state experienced through the increment of negative aspects can be overcome with precommitment only when the interval of change is low (10 rounds), and the positive intensity value is small enough (0.1).

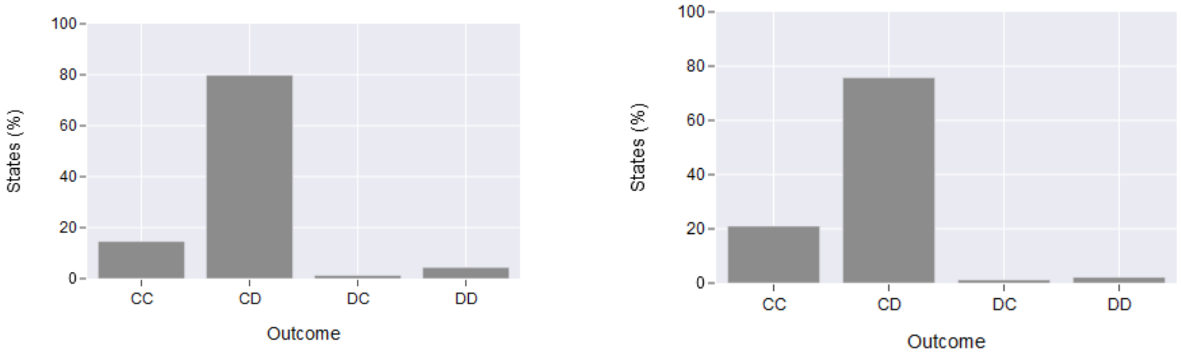


***Figure 4.3.18: (left)** Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01). **(right)** Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.5, Interval=10 rounds.*

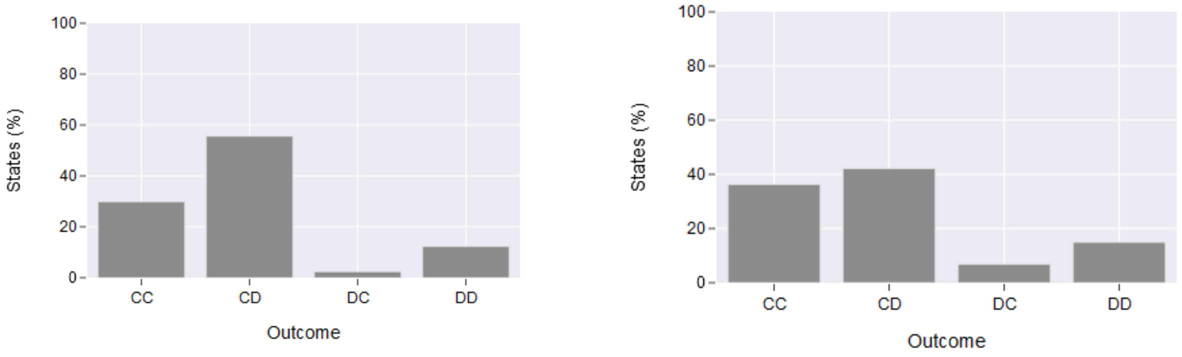


***Figure 4.3.19: (left)** Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01). **(right)** Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.5, Interval=50 rounds.*

4.3.5 Increasing the Temptation and decreasing the Punishment and Sucker's payoff

The last used method was a combination of the three previous methods for the increment of negative emotion, that is increasing the T and decreasing the P and S, all at the same while changing the precommitment status in each testing to see under which scenario we produce the best results. We set a positive intensity value to 0.1 for T and a negative intensity value of 0.5 for P,S, since they were the value that gave the best performance and a 50 rounds interval to provide an large interval of change. The results showed that high precommitment was able to prevail by achieving self-control with percentage of CC states reaching 57.8% (Figure 4.3.20) whereas low precommitment could't manage to achieve the same results with only achieving a percentage of 34.2%.

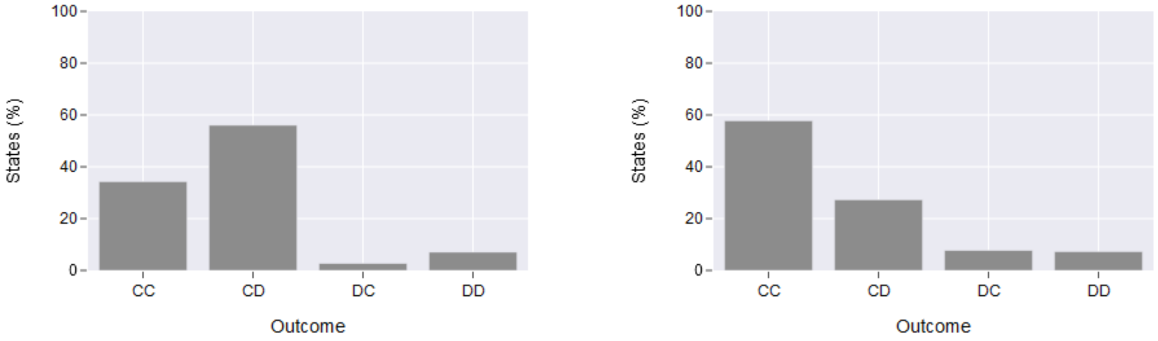


Figure 4.3.20:* (left)** *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with low precommitment (ψ=0.01).* ***(right) *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with high precommitment (ψ=0.9). Positive Intensity value=0.1 for T payoff, Negative Intensity value=0.5 for S, P payoffs, Interval=50 rounds*

One could argue that the approach of increasing temptation and decreasing punishment and sucker's payoff follows a similar principle to incorporating precommitment into the higher agent, as described in Section 3.2. In order to gain a deeper understanding of the individual effects of precommitment and emotions, it would be valuable to establish a new baseline scenario where precommitment is entirely excluded. By comparing the results of this new baseline with the previous approach that involved precommitment, we can observe the significant impact of precommitment on achieving self-control. The baselines, which did not incorporate precommitment, only managed to reach a success rate of 30.9% (Figure 4.3). This comparison underscores the importance of precommitment, as it clearly demonstrates that precommitment and emotions have distinct and complementary effects in promoting self-control.

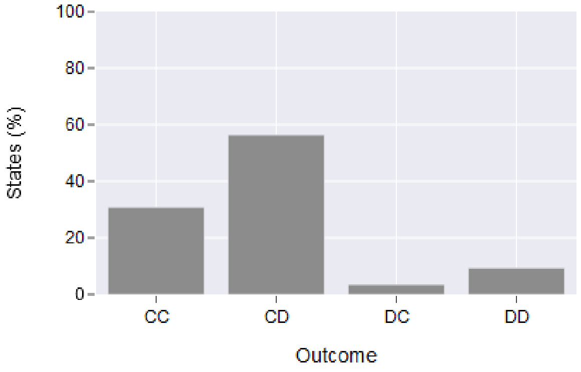


Figure 4.3.21: *Overall average outcomes after 1000 rounds of the Q-learning agents playing the IPD game with no precommitment (ψ=0). Positive Intensity value=0.1 for T payoff, Negative Intensity value=0.5 for S, P payoffs, Interval=50 rounds*

4.3.6 Summary and discussion on Precommitment in a Negative Emotional State

The first method used to test self-control involved increasing negative emotions through negative punishment (decreasing R) at different levels of precommitment. The results showed that negative punishment, when paired with high precommitment, can still lead to self-control behavior, but only if the interval of change (rounds) was sufficiently large and the negative intensity value was small enough to provide a smaller magnitude at a more infrequent rate (Figure 4.3.1). The same pattern was observed when a positive punishment method (increased T) was used to increase the presence of negative emotions. More specifically, the most effective approach was a positive intensity value of 0.1 for T and a negative intensity value of 0.1 for R, provided they were given in a large interval of change (50 rounds) under high precommitment (Figures 4.3.1 & 4.3.17). In contrast, larger intensity values in both scenarios impaired self-control and, in some cases, led to total self-control failure (Figures 4.3.2 & 4.3.19). When explicit negative emotions were induced by providing a negative intensity value to S and P, the scenarios where high precommitment still managed to achieve self-control and provide the best results were those where a higher negative intensity value (0.5) was given in a large interval of change (50 rounds) which interestingly contradicts the smaller positive intensity value that was used in T even though all three payoffs increment the presence of negative emotions (Figures 4.3.9 & 4.3.13). Finally, when combining the results of all three methods of negative reinforcement in both low and high precommitment the model successfully achieved self-control when precommitment was high, with a reaching percentage of 57.8% for CC states (Figure 4.3.20). In conclusion, these findings indicate that precommitment is crucial for achieving self-control in a negative emotional state, thus promoting precommitment may be an effective strategy for enhancing self-control when faced with negative reinforcements.

Payoff Value	**Precommitment**	**Intensity**	**Interval**	**CC states**
Reward (R)	Low (ψ=0.01)	Negative (0.1)	50 rounds	39.6% ▼
Reward (R)	High (ψ=0.9)	Negative (0.1)	50 rounds	58.4% ▲
Punishment (P)	Low (ψ=0.01)	Negative (0.5)	50 rounds	52.9% ▲
Punishment (P)	High (ψ=0.9)	Negative (0.5)	50 rounds	66.5% ▲
Sucker (S)	Low (ψ=0.01)	Negative (0.5)	50 rounds	35.9% ▼
Sucker (S)	High (ψ=0.9)	Negative (0.5)	50 rounds	60.0% ▲
Temptation (T)	Low (ψ=0.01)	Positive (0.1)	50 rounds	40.4% ▼
Temptation (T)	High (ψ=0.9)	Postive (0.1)	50 rounds	65.4% ▲
P, S & T Combined	Low (ψ=0.01)	Negative (0.5) for P & S Positive (0.1) for T	50 rounds	34.2% ▼
P, S & T Combined	High (ψ=0.9)	Negative (0.5) for P & S Positive (0.1) for T	50 rounds	57.8% ▲

Figure 4.3.22: *Summary of the best results of each scenario tested in relation to the affected payoff value, the level of precommitment, intensity and interval of change. The outcome is given in the form of a percentage indicating the overall percentage of CC states achieved throughout the course of the game. The symbols "▲" and "▼" indicate whether self-control was achieved or not, respectively, with "▲" denoting successful self-control (percentage exceeding 50%) and "▼" indicating a lack of self-control.*

Chapter 5

Conclusions and Future Work

5.1 Overview and conclusions

The thesis aimed to explore the potential effects of combining precommitment and emotional arousal on self-control in a computational model. Banfield's (2006) work on precommitment and Nikodemous's (2020) research on emotional arousal both provide valuable insights into self-control behavior. However, a gap was identified regarding how the combination of these two concepts could affect self-control behavior. To address this, a general-sum game called the Prisoner's Dilemma (PD) game was utilized to simulate the interaction between the upper and lower parts of the brain in an iterated version of the game known as the Iterated Prisoner's Dilemma (IPD). The objective of the game is to collaborate, which is indicative of successful self-regulation behavior. By incorporating both precommitment and emotional arousal, this computational model provides a platform to investigate how these two factors can impact self-control behavior.

Banfield (2006) argues that precommitment can be an effective strategy to overcome self-control problems by reducing the conflict between the higher and lower parts of the brain, which makes it easier to choose the larger, later reward. Emotional arousal is also a critical component of self-control behavior. Nikodemou (2020) suggests that emotions are inherently linked to self-control, with positive emotions promoting self-control and negative emotions impairing it. However, the relationship between negative emotions and self-control is more complex, as feelings of fear and guilt can have positive effects on self-control, while positive emotions can have negative effects. Considering the results of Nikodemou's (2020) and Banfield's (2006) research we can conclude that it is necessary to consider both precommitment and emotional arousal to gain a more comprehensive understanding of self-control behavior.

For my thesis, we employed the Q-Learning algorithm, similar to Nikodemou (2020), to train two agents representing the two parts of the brain, to cooperate and achieve the optimal solution (CC state). The simulations were initiated with the same starting values for the payoff matrix, learning factors, and epsilon values. However, the two agents representing the two parts of the brain had different discount factors, which distinguished them from each other. One agent focused on short-term rewards ($\gamma = 0.1$), while the other agent focused on long-term rewards ($\gamma = 0.9$). After the initialization phase, three critical factors changed. Firstly, we varied the precommitment status of the simulations, which determined whether the agents faced low precommitment (ψ=0.01) or high precommitment (ψ=0.9). This factor was added as the first step of the simulation. Secondly, we introduced intensity values, which were added to the values of the payoff matrix to simulate the presence of emotions. The intensity values could be positive or negative, depending on whether they were added or subtracted from the values. Finally, we varied the intervals of change, which were the rounds that needed to elapse between each addition or decrement of intensity values.

After running the simulation with different combinations of intervals of change (rounds) and intensity values under both low and high precommitment we came to some interesting conclusions regarding how the pairing of precommitment and emotions affects self-control behavior. In terms of the effects of positive emotions, the simulation revealed that the combination of precommitment and positive reinforcement can significantly improve self-control behavior. More specifically, positive reinforcement (increment of R) was found to be particularly effective when paired with high precommitment. This means that when individuals commit themselves to a goal or behavior, and positive emotions are present, they are more likely to achieve self-control behavior. Additionally, the intensity and frequency of positive reinforcement were also found to be essential, with a smaller magnitude of positive intensity given infrequently leading to the most significant impact on self-control behavior. Similarly, negative reinforcement methods were used to ease internal conflict and eliminate negative emotions, such as decreasing T and increasing P and S. The study found that negative intensity values of 0.1 for T and positive intensity values of 0.1 for P and S were the most effective when given infrequently (50 rounds) in a high precommitment state. When values greater than this were used, self-control behavior was impaired, and, in some cases, total self-control failure was observed. Finally, combining all three methods of negative reinforcement in both low and high precommitment states led to achieving self-control behavior with high precommitment achieving the highest percentage of CC states at 75%.

Now, turning to the testing of negative emotions and precommitment, the first method used involved increasing negative emotions through negative punishment (decreasing R) at different levels of precommitment. The findings revealed that negative punishment, when paired with high precommitment, can still lead to self-control behavior, but only if the interval of change (rounds) was sufficiently large and the negative intensity value was small enough to provide a smaller magnitude at a more infrequent rate. Similarly, positive punishment was used to increase the presence of negative emotions. The most effective approach for each case was a positive intensity value of 0.1 for T and a negative intensity value of 0.1 for R, provided they were given in a large interval of change (50 rounds) under high precommitment as larger intensity values in both scenarios impaired self-control. When explicit negative emotions were induced by providing a negative intensity value to S and P, the scenarios where high precommitment still managed to achieve self-control and provide the best results surprisingly were those where a higher negative intensity value (0.5) was given in a large interval of change (50 rounds) contrasting with the lower intensity value (0.1) needed to elicit negative emotions using the Temptation payoff. Finally, when combining the results of all three methods of negative reinforcement in both low and high precommitment, the model achieved self-control when precommitment was high, with a reaching percentage of 57.8% for CC states. These findings suggest that precommitment is crucial for achieving self-control in a negative emotional state, and promoting precommitment may be an effective strategy for enhancing self-control when faced with negative reinforcements.

In conclusion, this thesis aimed to investigate the impact of precommitment and emotional arousal on self-control behavior, using a computational model of the Iterated Prisoner's Dilemma game. The findings of this research suggest that both precommitment and emotional arousal are crucial components of self-control behavior, and their interaction can significantly influence the ability to regulate self-control behavior effectively. It is also noteworthy that despite the recognition that precommitment can sometimes be costly and may restrict flexibility, the results demonstrated that in the tested scenarios, precommitment not only managed to retain self-control but also enhance it. This outcome suggests that under certain conditions, precommitment can effectively help in overcoming self-control problems. Additionally, concerning emotions, the simulation results demonstrated that both positive and negative emotions, paired with high precommitment, can improve self-control behavior significantly. The study also found that the intensity and frequency of reinforcement methods, both positive and negative, play an essential role in shaping self-control behavior. Overall, the results of this research have significant implications for understanding self-control behavior and developing interventions to enhance self-regulation in order to help individuals and society as a whole to overcome self-control problems and achieve more positive outcomes.

5.2 Future Work

The present research managed to incorporate the findings of Nikodemou (2020) and Banfield (2006) regarding the effects of emotions and precommitment on self-control in order to show that the two concepts can interact to influence self-control behavior in a computational model of self-control as is the Iterated Prisoner's Dilemma. However, in order to assess cognitive adequacy of the model in relation to the results presented in this thesis, it would be beneficial to correlate the results with relevant psychological findings, that specifically explore the relationship of the combination of emotions and precommitment and its effect on self-control behavior. To the best of our knowledge, no such findings have been published, but further in-depth literature review is probably needed. Furthermore, as mentioned in previous chapters the initial payoff matrix values used for all experiments were T=5, R=4, P=-2, S=-3, which correspond to "an internal conflict of moderate intensity" (Cleanthous, 2010) leading to a gap in our understanding of how different types of conflict, specifically strong and weak, may impact this interaction.

Strong conflict refers to a scenario where there is a clear conflict between two options and the choice is relatively straightforward whereas weak conflict refers to a scenario where the choice is more complex as the choices are similar to each other. More specifically, regarding our implementation, prior research (Cleanthous, 2010) has established that the level of conflict among agents in the IPD game is directly proportional to the difference between the payoff for Temptation to Defect (T) and Sucker's Payoff (S). Additionally, the extent of conflict could also reflect the level of complexity involved in the task that agents need to accomplish for exercising self-control (CC state).

With this in mind, further research could explore the behavior of the model when using a payoff matrix with a strong conflict, such as T=15, R=4, P=-2, and S=-13, or one with a weak conflict, such as T=3, R=2, P=1, and S=0. Both of these scenarios maintain the rules of the IPD game (T>R>P>S, 2R > T+S), but the former involves a larger disparity between Temptation and Sucker's Payoff, while the latter features a smaller gap between these two rewards.

By exploring the effects of strong and weak conflict on precommitment and emotional arousal in self-control behavior, we may gain a better understanding of how individuals make decisions in different types of situations. For example, we might find that precommitment is more effective in situations with strong conflict, where the temptation to choose the smaller, immediate reward is much greater. Alternatively, we might find that emotional arousal has a greater impact on self-control behavior in situations with weak conflict, where the difference between the temptation and sucker payoffs is relatively small.

Furthermore, it would be beneficial to explore the simultaneous effects of positive and negative emotions on self-control behavior. The current research focused on either positive or negative emotions separately, but in real-world situations, individuals often experience a combination of both positive and negative emotions concurrently. Investigating how the interplay between positive and negative emotions influences self-control behavior can provide a more comprehensive understanding of the complex relationship between emotions and self-control. This could involve manipulating both positive and negative intensity values in the payoff matrix and examining their combined effects on precommitment and self-control behavior.

In addition to these directions, it would also be valuable to explore the effect of costly precommitment on self-control. Acknowledging that precommitment can have limitations and can be costly, we can incorporate the cost factor into the precommitment strategy used in the simulations. By varying the cost associated with precommitment and analyzing its impact on self-control outcomes, we can better understand the trade-offs and limitations involved. This investigation would provide insights into how a possible cost of precommitment influence its effectiveness in achieving long-term goals and overcoming short-term temptations, shedding light on the dynamics of self-control behavior.

Finally, another potential direction for future research is to explore the effectiveness of simulating precommitment using different values within the acceptable margins, ensuring this way that the conditions of the IPD game remain unbroken (equations 1 & 2). The original model allowed for precommitment values ranging from 0 to 0.99, but this range may be too restrictive to accurately reflect real-world scenarios. By expanding the range of precommitment values, we could gain a better understanding of how different levels of commitment affect self-control. For example, a precommitment value of 5.0 would represent a complete commitment to a particular action or behavior, a value of 2.5 would indicate a moderate level of commitment and 1.0 a low level.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. Psychological Bulletin, 82(4), 463-496.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. Science, 211(4489), 1390-1396.

Banfield, G. D. (2006). Simulation of Self-Control through Precommitment Behaviour in an Evolutionary System. Ph.D., Birkbeck, University of London. https://www.dcs.bbk.ac.uk/site/assets/files/1025/banfield.pdf

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource?. Journal of Personality and Social Psychology, 74(5), 1252-1265.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. Current Directions in Psychological Science, 16(6), 351-355.

Christodoulou, C., Banfield, G., & Cleanthous, A. (2010). Self-control with spiking and non-spiking neural networks playing games. Journal of Physiology, Paris, 104(3–4), 108–117.

Cleanthous, A. (2010). In search of self-control through computational modelling of internal conflict. Ph.D. University of Cyprus. https://gnosis.library.ucy.ac.cy/handle/7/39548

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in Predicting Academic Performance of Adolescents. Psychological science, 16(12), 939-944.

Georgiou, A. (2015). Μελέτη σχέσης αυτοελέγχου και συνειδητότητας (BSc Thesis). Department of Computer Science, University of Cyprus.

Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: an experience sampling study of desire, conflict, and self-control. Journal of Personality and Social Psychology, 102(6), 1318-1335.

Kavka, G. S. (1991). Is Individual Choice Less Problematic than Collective Choice? Economics & Philosophy, 7(2), 143–165.

Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. International Journal of Robotics Research, 32(11), 1238-1274.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences, 108(7), 2693-2698.

Muraven, M., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise. Journal of Social Psychology, 139(4), 446-457.

Nash, J. F. (1950). Equilibrium points in n-person games. Proceedings of the National Academy of Sciences of the United States of America, 36(1), 48–49.

Nikodemou, A. (2020). Positive and negative emotions in a computational model of self-control (BSc Thesis). Department of Computer Science, University of Cyprus.

Rachlin, H. (2000). The science of self-control. Cambridge, MA: Harvard University Press.

Ren, J., Hu, L., Zhang, H., & Huang, Z. (2010). Implicit Positive Emotion Counteracts Ego Depletion. Social Behavior and Personality: An International Journal, 38(7), 919–928.

Rolls, E. T. (2018). The Brain, Emotion, and Depression. Oxford, England: Oxford University Press

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. Econometrica, 50(1), 97-110.

Rummery, G. A., & Niranjan, M. (1994). On-line Q-learning using connectionist systems. (Technical Report No. 166). Cambridge, England: University of Cambridge, Department of Engineering.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological Review, 69(5), 379-399.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., . . . Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529 (7587), 484-489.

Stone, P. (2010). Reinforcement Learning. In C. Sammut & G. I. Webb (Eds.), Encyclopedia of machine learning (pp. 849-850). New York, NY: Springer.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. Machine learning, 3(1), 9-44.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). Cambridge, MA: MIT Press.

Tamir, M. (2016). Why do people regulate their emotions? A taxonomy of motives in emotion regulation. Personality and Social Psychology Review, 20(3), 199-222.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2007). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality, 72(2), 271-324.

Thaler, R. H., & Shefrin, H. M. (1981). An economic theory of self-control. Journal of Political Economy, 89(2), 392-406.

Tice, D. M., Baumeister, R. F., & Zhang, L. (2004). The Role of Emotion in Self-Regulation: Differing Roles of Positive and Negative Emotion. In P. Philippot & R. S.

Feldman (Eds.), The Regulation of Emotion (pp. 215–230). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Vassiliades, V., Cleanthous, A. and Christodoulou, C. (2011). Multiagent Reinforcement Learning: Spiking and Nonspiking Agents in the Iterated Prisoner's Dilemma. IEEE Transactions on Neural Networks, 22(4), 639-653.

Watkins, C. J. C. H. (1989). Learning from delayed rewards. Ph.D. University of Cambridge. http://www.cs.rhul.ac.uk/~chrisw/new_thesis.pdf

Woergoetter, F., & Porr, B. (2008). Reinforcement learning. Scholarpedia, 3(3), 1448.

Appendix A

The development of the basic Q-Learning model (Appendices A-B) is by Georgiou (2015).

Class MainProgram.java

```
import java.io.BufferedWriter;
import java.io.File;
import java.io.FileWriter;
import java.io.IOException;
import java.util.InputMismatchException;
import java.util.Scanner;

/**
 * Created by jeannettechahwan & annageorgiou on 20/01/15.
 */
public class MainProgram{

   public static Player lower;
   public static Player higher;

   public static double[][][] states_results;
   public static double[][][] payoff_results;
   public static double[] overall_payoff;

   public static double[] overall_payoff_neutral;
   public static double[][][] payoff_results_neutral;

   public static double T, R, P, S;

   public static int get_state(int action_lower, int action_higher) {
      int next_state;

      if(action_lower==0 && action_higher==0)
         next_state = 0;
      else if(action_lower==0 && action_higher==1)
         next_state = 2;
      else if(action_lower==1 && action_higher==0)
         next_state = 1;
      else
         next_state = 3;

      return next_state;
   }

   public static void update_states_results(int episode, int state, int trial){
      states_results[episode][trial][0] = states_results[episode-1][trial][0];
      states_results[episode][trial][1] = states_results[episode-1][trial][1];
      states_results[episode][trial][2] = states_results[episode-1][trial][2];
      states_results[episode][trial][3] = states_results[episode-1][trial][3];
      states_results[episode][trial][state]++;
   }
```

```
  public static void update_payoff_results(int episode, double lower_payoff, double
higher_payoff, int trial){
    payoff_results[episode][trial][0] = payoff_results[episode-1][trial][0];
    payoff_results[episode][trial][1] = payoff_results[episode-1][trial][1];

    payoff_results[episode][trial][0] += lower_payoff;
    payoff_results[episode][trial][1] += higher_payoff;

    // for neutral emotion
    if ((lower_payoff == S && higher_payoff == T) || (lower_payoff == T && higher_payoff == S))
{
      payoff_results_neutral[episode][trial][0] = payoff_results_neutral[episode - 1][trial][0];
      payoff_results_neutral[episode][trial][1] = payoff_results_neutral[episode - 1][trial][1];

    } else {
      payoff_results_neutral[episode][trial][0] = payoff_results_neutral[episode - 1][trial][0];
      payoff_results_neutral[episode][trial][0] += lower_payoff;

      payoff_results_neutral[episode][trial][1] = payoff_results_neutral[episode - 1][trial][1];
      payoff_results_neutral[episode][trial][1] += higher_payoff;
    }
  }

  public static void statistics(int numberOfTrials, int numberOfEpisodes){
    int e, t, s, p;
    double sumAll_states;
    double sumAll_payoffs;

    for(e=0; e<numberOfEpisodes; e++){
      for(t=0; t<numberOfTrials; t++){   // find sum of states visits from each trial of current
episode and save in last column
        for(s=0; s<4; s++){    // states
          states_results[e][numberOfTrials][s] += states_results[e][t][s];
        }

        for(p=0; p<2; p++){    // payoffs
          payoff_results[e][numberOfTrials][p] += payoff_results[e][t][p];
          payoff_results_neutral[e][numberOfTrials][p] += payoff_results_neutral[e][t][p];
        }
      }

      sumAll_states=0;
      sumAll_payoffs=0;

      for(s=0; s<4; s++){    // sum of state fields in last column (#trials+1)
        sumAll_states += states_results[e][numberOfTrials][s];
      }

      for(s=0; s<4; s++){    // normalise
        states_results[e][numberOfTrials][s] /= sumAll_states;
      }

      for(p=0; p<2; p++){    // sum of payoff fields in last column (#trials+1)
        sumAll_payoffs += payoff_results[e][numberOfTrials][p];
      }

      overall_payoff[e] = sumAll_payoffs/numberOfTrials;
    }
```

```
}

public static void print_statistics(int numberOfTrials, int numberOfEpisodes){
   int e, s, p;

   // states
   for(e=0; e<numberOfEpisodes; e++){
      for(s=0; s<4; s++){
         System.out.printf("%.4f ", states_results[e][numberOfTrials][s]);
      }
      System.out.println();
   }

   // payoffs
   for(e=0; e<numberOfEpisodes; e++){
      for(p=0; p<2; p++){
         System.out.printf("%.4f ", payoff_results[e][numberOfTrials][p]);
      }
      System.out.println();
   }
}

public static void printToFile_statistics(int numberOfTrials, int numberOfEpisodes){
   int e, s, p;

   // write results to file
   try {
      File file_s = new File("states_results.txt");
      File file_p = new File("payoff_results.txt");

      // if file doesnt exists, then create it
      if (!file_s.exists()) {
         file_s.createNewFile();
      }

      if (!file_p.exists()) {
         file_p.createNewFile();
      }

      FileWriter fw_s = new FileWriter(file_s.getAbsoluteFile());
      FileWriter fw_p = new FileWriter(file_p.getAbsoluteFile());
      BufferedWriter bw_s = new BufferedWriter(fw_s);
      BufferedWriter bw_p = new BufferedWriter(fw_p);

      for(e=0; e<numberOfEpisodes; e++){
         for(s=0; s<4; s++){
            bw_s.write(Double.toString(states_results[e][numberOfTrials][s]));
            bw_s.write(" ");
         }
         bw_s.write("\n");

         for(p=0; p<2; p++){
            bw_p.write(Double.toString(payoff_results[e][numberOfTrials][p]/numberOfTrials));
            bw_p.write(" ");
         }
         bw_p.write(Double.toString(overall_payoff[e]));
         bw_p.write(" ");
```

```
        bw_p.write(Double.toString(overall_payoff_neutral[e]));
        bw_p.write("\n");
      }

      bw_s.close();
      bw_p.close();
    } catch (IOException ex) {
      ex.printStackTrace();
    }
  }

  public static void main(String[] args) {
    int initial_state, current_state, next_state; // Takes values 0-3
    int i;
    int episodes_before = 500, episodes_after = 500;
    int trials = 15;
    states_results = new double[episodes_before + episodes_after][trials+1][4];
    payoff_results = new double[episodes_before + episodes_after][trials+1][2];
    overall_payoff = new double[episodes_before + episodes_after];

    payoff_results_neutral = new double[episodes_before + episodes_after][trials+1][2];
    overall_payoff_neutral = new double[episodes_before + episodes_after];

    double psi = Double.parseDouble(args[0]);
    double ratio_R = Double.parseDouble(args[1]);
    double ratio_T = Double.parseDouble(args[2]);
    double ratio_P = Double.parseDouble(args[3]);
    double ratio_S = Double.parseDouble(args[4]);
    int rounds = Integer.parseInt(args[5]); // interval rounds between the changes of the payoff
values


    T=5; R=4; P=-2; S=-3;


    for(int t=0; t<trials; t++) {
      T=5; R=4; P=-2; S=-3;
      // short-term
      lower = new Player();
      lower.setDiscount(0.1);
      lower.setLearning_rate(0.1);
      lower.setEpsilon(0.1);
      lower.setPayoff_matrix(R, T-psi, S+psi, P);

      // long-term
      higher = new Player();
      higher.setDiscount(0.9);
      higher.setLearning_rate(0.1);
      higher.setEpsilon(0.1);
      higher.setPayoff_matrix(R, S-psi, T+psi, P);

      initial_state = (int) (Math.random() * 4);
      current_state = initial_state;
      states_results[0][t][current_state]++;

      for (i = 1; i <= episodes_before; i++) {

        lower.setCurrent_action(lower.choose_action(current_state));
```

```
            higher.setCurrent_action(higher.choose_action(current_state));

            next_state = get_state(lower.getCurrent_action(), higher.getCurrent_action());

            lower.update_Q(current_state, next_state);
            higher.update_Q(current_state, next_state);

            //update lower player
            double R_lower = lower.getPayoff_matrix(0);
            double T_lower = lower.getPayoff_matrix(1);
            double S_lower = lower.getPayoff_matrix(2);
            double P_lower = lower.getPayoff_matrix(3);


            if ( i%rounds==0 &&
                  ((T_lower+ratio_T) > (R_lower+ratio_R)) &&
                  ((R_lower+ratio_R) > (P_lower+ratio_P)) &&
                  ((P_lower+ratio_P) > (S_lower+ratio_S)) &&
                  (2*(R_lower+ratio_R) > ((T_lower+ratio_T) + (S_lower+ratio_S)))
            ) {
              System.out.print("BEFORE Lower\n");
              R_lower = R_lower + ratio_R;
              T_lower = T_lower + ratio_T;
              S_lower = S_lower + ratio_S;
              P_lower = P_lower + ratio_P;
              lower.setPayoff_matrix(R_lower, T_lower, S_lower, P_lower);

            }
            //update higher player
            double R_higher = higher.getPayoff_matrix(0);
            double S_higher = higher.getPayoff_matrix(1);
            double T_higher = higher.getPayoff_matrix(2);
            double P_higher = higher.getPayoff_matrix(3);


            if ( i%rounds==0 &&
                  ((T_higher+ratio_T) > (R_higher+ratio_R)) &&
                  ((R_higher+ratio_R) > (P_higher+ratio_P)) &&
                  ((P_higher+ratio_P) > (S_higher+ratio_S)) &&
                  (2*(R_higher+ratio_R) > ((T_higher+ratio_T) + (S_higher+ratio_S)))
            ) {
              System.out.print("BEFORE Higher\n");
              R_higher = R_higher + ratio_R;
              T_higher = T_higher + ratio_T;
              S_higher = S_higher + ratio_S;
              P_higher = P_higher + ratio_P;
              higher.setPayoff_matrix(R_higher, T_higher, S_higher, P_higher);
            }

            current_state = next_state;

            update_states_results(i, current_state, t);
            update_payoff_results(i, lower.getPayoff_matrix(current_state),
higher.getPayoff_matrix(current_state), t);
        }

        // Set epsilon = 0 and initial payoff matrix
        lower.setEpsilon(0);
```

```
higher.setEpsilon(0);
T=5; R=4; P=-2; S=-3;
lower.setPayoff_matrix(R, T-psi, S+psi, P);
higher.setPayoff_matrix(R, S-psi, T+psi, P);

for (i = 1; i <= episodes_after; i++) {

    lower.setCurrent_action(lower.choose_action(current_state));
    higher.setCurrent_action(higher.choose_action(current_state));

    next_state = get_state(lower.getCurrent_action(), higher.getCurrent_action());

    lower.update_Q(current_state, next_state);
    higher.update_Q(current_state, next_state);

    //update lower player
    double R_lower = lower.getPayoff_matrix(0);
    double T_lower = lower.getPayoff_matrix(1);
    double S_lower = lower.getPayoff_matrix(2);
    double P_lower = lower.getPayoff_matrix(3);


    if ( i%rounds==0 &&
            ((T_lower+ratio_T) > (R_lower+ratio_R)) &&
            ((R_lower+ratio_R) > (P_lower+ratio_P)) &&
            ((P_lower+ratio_P) > (S_lower+ratio_S)) &&
            (2*(R_lower+ratio_R) > ((T_lower+ratio_T) + (S_lower+ratio_S)))
    ) {
        System.out.print("AFTER Lower\n");
        R_lower = R_lower + ratio_R;
        T_lower = T_lower + ratio_T;
        S_lower = S_lower + ratio_S;
        P_lower = P_lower + ratio_P;
        lower.setPayoff_matrix(R_lower, T_lower, S_lower, P_lower);

    }
    //update higher player
    double R_higher = higher.getPayoff_matrix(0);
    double S_higher = higher.getPayoff_matrix(1);
    double T_higher = higher.getPayoff_matrix(2);
    double P_higher = higher.getPayoff_matrix(3);


    if ( i%rounds==0 &&
            ((T_higher+ratio_T) > (R_higher+ratio_R)) &&
            ((R_higher+ratio_R) > (P_higher+ratio_P)) &&
            ((P_higher+ratio_P) > (S_higher+ratio_S)) &&
            (2*(R_higher+ratio_R) > ((T_higher+ratio_T) + (S_higher+ratio_S)))
    ) {
        System.out.print("AFTER Higher\n");
        R_higher = R_higher + ratio_R;
        T_higher = T_higher + ratio_T;
        S_higher = S_higher + ratio_S;
        P_higher = P_higher + ratio_P;
        higher.setPayoff_matrix(R_higher, T_higher, S_higher, P_higher);
    }

    current_state = next_state;
```

```
            update_states_results(episodes_before+i-1, current_state, t);
            update_payoff_results(episodes_before+i-1, lower.getPayoff_matrix(current_state),
higher.getPayoff_matrix(current_state), t);
         }
      }

      statistics(trials, (episodes_before+episodes_after));
      printToFile_statistics(trials, (episodes_before+episodes_after));

   }
```

Appendix B

Class Player.java

```
/**
 * Created by jeannettechahwan on 20/01/15.
 */

public class Player {

    private double learning_rate;
    private double epsilon;
    private double discount;
    private double[][] Q_table;
    private double[] payoff_matrix;
    private int current_action;

    public Player(){
        this.Q_table = new double[4][2];
        this.payoff_matrix = new double[4];
    }

    public void setDiscount(double value){
        this.discount = value;
    }

    public double getDiscount(){
        return this.discount;
    }

    public void setEpsilon(double value){
        this.epsilon = value;
    }

    public double getEpsilon(){
        return this.epsilon;
    }

    public void setLearning_rate(double value){
        this.learning_rate = value;
    }

    public double getLearning_rate(){
        return this.learning_rate;
    }

    public void setQ_table(int row, int column, double value){
        this.Q_table[row][column] = value;
    }

    public double getQ_table(int row, int column){
        return this.Q_table[row][column];
    }

    public void setPayoff_matrix(double value1, double value2, double value3, double value4){
        payoff_matrix[0] = value1;
        payoff_matrix[1] = value2;
        payoff_matrix[2] = value3;
        payoff_matrix[3] = value4;
    }

    public double getPayoff_matrix(int row){
        return this.payoff_matrix[row];
    }

    public void setCurrent_action(int action){
        this.current_action = action;
    }

    public int getCurrent_action(){
        return this.current_action ;
    }

    public int choose_action(int state){
        int action;
        double random = Math.random();

        if(random<epsilon){    //explore
            action = (int) (Math.random() * 2);    // values: 0(cooperate) or 1(defect)
        }

        else {    //exploit
            // find best known action
            if (this.getQ_table(state, 0) > this.getQ_table(state, 1))
                action = 0;
            else if(this.getQ_table(state, 0) < this.getQ_table(state, 1))
                action = 1;
            else
                action = (int) (Math.random() * 2);    // values: 0(cooperate) or 1(defect)
        }

        return  action;
    }

    public void update_Q(int current_state, int next_state){
        double max, Q;

        // find maximum value from Q table
        if(this.getQ_table(next_state,0) >= this.getQ_table(next_state,1))
            max = this.getQ_table(next_state, 0);
        else
            max = this.getQ_table(next_state, 1);

        // Calculate Q and set value in player's Q table
        Q = ((1 - this.learning_rate) * this.getQ_table(current_state, this.getCurrent_action()))
                + (this.learning_rate * (this.getPayoff_matrix(next_state) + (this.getDiscount() * max)));

        this.setQ_table(current_state, this.getCurrent_action(), Q);
    }

}
```