# A Psychologically-Inspired Agent for Iterative Prisoner's Dilemma

**Rawad Al-Haddad and Gita Sukthankar**

School of Electrical Engineering and Computer Science
University of Central Florida
4000 Central Florida Blvd. Orlando, Florida, 32816
rawadh@cs.ucf.edu gitars@eecs.ucf.edu

**Abstract**

In this paper, a psychologically-inspired model for an Iterative Prisoner's dilemma (IPD) agent is proposed. This model is inspired by the "psychic apparatus" theory that was developed by Sigmund Freud in 1940. The model captures an agent with a true "character" by concurrently supporting the three constructs of personality: the Super-Ego, which represents the ideal part of the agent that always tries to elicit cooperation from opponents, the Id, which is characterized by its willingness to defect all the time to achieve instant gratification, and the Ego, which is the intelligent, realistic, part of the agent that relies on opponent-modeling techniques to decide on the best next move. These three constructs compete against each other in order to take control of the agent. This model was successfully prototyped and participated in a simulated IPD tournament along with other benchmark strategies. "FREUD", as the agent is called, achieved outstanding results in this mini-tournament by winning with a good margin. Our model represents a novel abstraction for IPD agent architecture that is potentially applicable to any decision-making task that requires evaluating the benefit of competitive vs. cooperative behavior.

## Introduction

The Iterated Prisoner's dilemma (IPD) has been widely used as a model of strategic interaction among self-interested, rational agents with the absence of centralized authority (Axelrod 1980). In addition, it has been considered as an abstract model of multi-agent decision-making tasks in which the individual agents seek to maximize their payoffs, which depend on the outcome of iterated interaction with other agents (Au and Nau 2006). These two important aspects, along with the fact that IPD provides a standardized environment for studying the evolution of cooperation among selfish agents (Axelrod and Hamilton 1981), have brought enormous attention to this game.

The prisoner's dilemma is a non-zero sum game; each round is played simultaneously between two agents where each agent has one of two choices: either to Cooperate (C) or Defect (D). The payoff matrix for the four possible scenarios in the traditional game is presented in Table 1.

Table 1: Prisoner's Dilemma Payoff Matrix

| | | Player 2 | |
|---|---|---|---|
| | | C | D |
| Player 1 | C | 3/3 | 0/5 |
| | D | 5/0 | 1/1 |

As seen in Table 1, the prisoner's dilemma presents an interesting paradox. Regardless of the opponent's move, each agent is compelled to defect as it will give the higher payoff; this will lead to mutual defection (DD) because both players are assumed to be rational and self-interested. However, the payoff if both agents cooperate (CC) is higher for each than the payoff they would have accrued had both agents defected. So, the paradox is that the Nash equilibrium (defection) is not Pareto optimal since there exists a Pareto improvement: a situation that can increase one agent's gain without decreasing the other's. The Stanford Encyclopedia of Philosophy defines the paradox as follows (Kuhn 2009):

> *"A group whose members pursue rational self-interest may all end up worse off than a group whose members act contrary to rational self-interest."*

The iterative aspect adds a strategic dimension to the game since the agents must consider the history of relationships while optimizing future gains. A rational agent should consider the history of the game to predict the best move against a specific opponent's strategy, and also should plan for future moves in order to maximize the expected payoff.

Examples of IPD-like situations in real life are numerous, including the interaction between the USA and Soviet Union during the cold war, advertising campaigns between competing companies, the brief unofficial cessations of hostilities between enemies in warfare, and many other examples listed in (Axelrod 1980).

## Related Work

Robert Axelrod organized the first IPD tournament back in 1980 (Axelrod 1980). Fifteen strategies participated in the tournament (14 submissions plus RANDOM), the tournament was held in a round-robin fashion with game lengths of 200 moves. The tournament results revealed many surprises. The first was the discovery that there is a subtle reason why a selfish agent should be nice to other agents. Surprisingly, niceness, defined by the requirement that an agent does not defect before its opponent does, was the trait that separated the good performers from the bad ones. The top eight agents by rank were all nice agents, whereas nasty agents occupied the bottom seven positions.

The second surprise was the winner: Tit-For-Tat (TFT), which was the simplest strategy in the tournament with only 4 lines of code. Most participants knew beforehand that TFT was the strategy to beat since it had won a preliminary unofficial tournament, and many of them tried to improve on the basic TFT strategy by making it less forgiving without success. In fact, all these modifications only made TFT less effective in IPD.

Axelrod repeated the tournament and added Tit-For-Two-Tats (TFTT), a more forgiving version of TFT. To his surprise, this strategy won the second tournament. Ironically, TFTT was actually sent to the participant as an example code fragment to show how to submit a strategy. The other would-have-won strategy was NICE DOWNING, a nice version of the original DOWNING (10th place finisher). All other aspects being the same, NICE DOWNING won the tournament. Axelrod thus concluded that niceness is an important quality in IPD tournaments.

After that, many efforts have sought to outperform TFT in IPD, the most influential one being Nowak's Pavlov strategy (win-stay lose-shift) (Nowak and Sigmund 1993). This strategy presented the self-monitoring procedure that inspired many other implementations. A self-monitoring agent maintains the same game play strategy as long as it is winning and only changes strategy if it starts losing.

Noisy IPD has gained more attention than the original noise-free IPD since it forces agents to deal with uncertainty, forgive unintended defects and re-engage in cooperative behaviors. Excellent analytical and experimental studies of these noisy environments can be found in (Bendor 1987, Molander 1985). Generosity, Contrition, and Win-Stay, Lose-Shift are identified as the three most common methods to deal with noise in IPD (Wu and Axelrod 1995). All these techniques inspired the design of FREUD as will be shown in the next section

## Agent Design

The "psychic apparatus" structural model (Freud 1940) remains one of the most notable models of human character in the field of psychology. The model divides the human character into three theoretical constructs. The first construct is the Super-Ego which represents all of our ideals and dreams. Second is the Id, which contains our instincts, basic drives, and the search for instant pleasure without reality consideration. The third construct is the Ego, which seeks long-term pleasure rather than short term one.

Similarly, FREUD is designed with these three constructs in mind. The Super-Ego is the cooperative part that always tries to stimulate cooperative behavior from the opponent. The Id is the always-defect "ALL-D" part of FREUD that will take any chance to run away with the temptation to defect payoff. Finally, the Ego is the realistic part that models the opponent in order to consciously devise the best next move. The three constructs operate in parallel; each presents its next move to FREUD decision-maker, which in turn decides which of the character parts to suppress and which to choose based on the history of their corresponding payoffs. Figure 1 depicts the detailed design of the proposed agent.

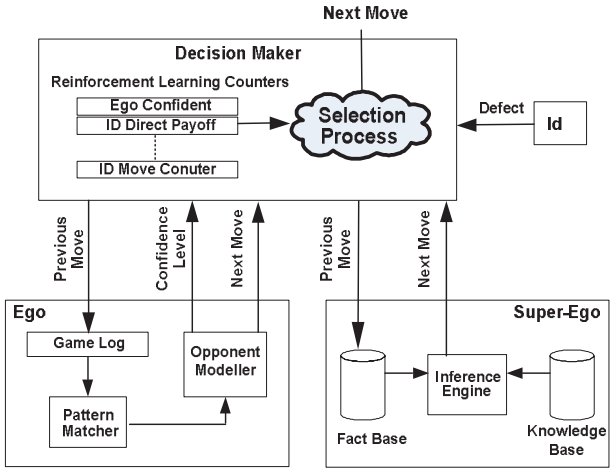


**Figure 1: FREUD System Architecture**

### The Super-Ego

The Super-Ego represents the idealistic construct of the agent. It follows a strict set of rules designed specifically to elicit cooperation from the opponent. Under any situation (noisy environment, exploitive opponent, etc…), this strategy will try to maximize the payoff of both agents by tending toward mutual cooperation. Despite the fact that this construct is the altruistic part of the agent, one should remember the main lesson of Axelrod's tournament: there are clear benefits to nice behavior in IPD; in the long run,

the nice agents were able to amass points when playing against other nice opponents and that helped them all to top the upper half of the ranking table.

The Super-Ego does not cooperate all the time, but it tends to do so in order to achieve long term cooperation of the opponent. Even when it defects, it has a very short memory of resentment and will revert to cooperative behavior as soon as the opponent shows any sign of cooperation. An example of the rule-set that the Super-Ego can implement:

**Be Nice**: The Super-Ego will never be the first to defect with certainty = 100%

*If Move_Counter=0 Then Next_Move = C (P=100%)*

*If Last_Move = CC Then Next_Move = C (P=100%)*

**Apologize**: The Super-Ego will always apologize after a defection that was uncalled for (This defection was certainly triggered by another construct of FREUD's character as we will see shortly):

*If Last_Moves = CC,DC Then NextMove = C (P=100%)*

**Accept Apology**: The Super-Ego will always accept the apology of the opponent for an already punished move; this will happen with a probability that decreases with every opponent defection (to avoid being exploited)

*If Last_Moves = CD,?C Then NextMove = C (P=ACCEPT_APOLOGY_PROB) (The ? is a wildcard that represents C or D)*

**Prevent Conflicts**: If the opponent defects without any reason, a bona fide cooperation is offered by the Super-Ego to prevent conflicts in the case of noisy environments; this will be driven by a probabilistic factor that will decrease each time the opponent defects without a reason.

*If Last_Moves = CC,CD Then NextMove = C (P=CONFLICT_PREVENTION _PROB)*

**Resolve Conflicts**: If the players are going through a mutual defection streak, the Super-Ego will offer a bona fide cooperation to resolve the conflict. Again, the conflict prevention is also driven by a probabilistic factor that decreases each time the opponent does not accept the conflict resolution offer.

*If Last_Moves = DD,DD Then NextMove = C (P=CONFLICT_RESOLUTION _PROB)*

**Build Trust**: For every Cooperative behavior shown by the opponent, the Super-Ego increases the cooperative probability. This will guarantee that this construct is capable of starting over and building trust after a couple of cooperative moves by the opponent.

To summarize, this construct is designed to be very optimistic in the way it realizes the world; it forgives easily and always tries to engage in cooperative behavior with little respect to the past. It is the part of the agent that always tells the opponent that "I'm ready to cooperate if you want to."

## The Id

The Id is the source of instant gratification for the agent; it introduces randomness and extra possibility of discovering naive opponents. Id will always opt for D as the next move. However, the Decision-Maker always assigns a low probability to the Id choice unless it proves its worth.

## The Ego

The Ego is the realistic part of the agent; it seeks a rational perception of the game history to maximize its long term payoff. This is done by modeling the opponent and looking for any systematic behavior that can predict the future. Therefore, the Ego starts the game as the weakest construct among the three, but with the game progress, it starts to gain more power to take over the agent decision-making process. This is a rational choice in this context as there is more reason to increase the dependability of the agent on the opponent modeling process when the opponent model gets stronger and more accurate.

The Ego implementation can be any type of opponent-modeling technique. In our implementation, a simple approach was adopted based on a pattern-matching scheme. The game log is stored as an even-length string of C/D pairs, with each pair representing the outcome of one move; the pairs are concatenated to the right of the string every time a move is recorded. The pattern matching technique can search through the string to find a set of "patterns" that resembles the very previous moves to the one that is to be executed next. The algorithm should consider all the hits and use them to establish a significance level of the prediction in term of three factors:

**The pattern match lengths**: Longer matching patterns have more significance than shorter ones. The pattern length gives an indication on how informative it is for the predicted move.

**The number of occurrences per pattern**: When the same pattern is matched more times, it becomes more evident that the predicted move is part of an agent strategy rather than a noisy or probabilistic action.

**The age of each match**: Recent matches are more reflective of the agent current strategy, especially with agents that keep updating their game plan according to some probabilistic factors.

The predicted move is analyzed by the Ego to determine the best response. This can be done by recalculating the payoff of the actual move that the agent did in all the confirmed matches. If the payoff exceeds a certain threshold (For instance, *3*number_of_moves* as this indicates a good payoff, where 3 is the reward of CC), the response is deemed to be effective and is used in the next move. If the aggregate payoff does not exceed this threshold, the "Ego" concludes that the response is not effective and thus considers the opposite one.

## The Decision-Maker

The Decision-Maker is responsible for accepting the input of the three constructs and deciding which one to use for

the next move. This is accomplished through 3 confidence factors that reflect the competition between the three constructs. At the beginning, the Super-Ego selection factor is close to 1 whereas the Ego selection factor is set to 0. The Id selection factor is set to small value to allow for casual defections, which can help in revealing certain characteristics in the opponents, like their forgiveness or provocability. Throughout the game, the Ego starts to gain more confidence in modeling and predicting the opponent, increasing its confidence level and giving more reason for the Decision-Maker to select its choice. The Super-Ego will be gradually suppressed if its performance in term of the gained payoff is below accepted.

The confidence levels are calculated by dividing the payoff achieved by each construct by the number of moves that the construct executed, or in other words, the confidence level can be defined as the average score that each construct has achieved up to the current move. The Super-Ego will have the lead in the first couple of moves which gives it a chance to kick off a healthy relation with the opponent (due to the cooperative nature of the Super-Ego). Once the random generator decides on selecting the Id move, the decision-maker will get the chance to probe the opponent's response. If the agent retaliates (for example, TFT would do), the confidence of the Id will decrease, reducing its chance to be selected later. On the other hand, if the opponent does not get provoked, the Id confidence will become higher and higher as it will be scoring an average of 5 points (DC outcome) which is certainly higher than what the Super-Ego could score with its C's. The Ego can be triggered manually after a predetermined number of moves if the Super-Ego's confidence is very close to the Id one.

## Experimental Setup

The evaluation plan in this paper is divided into two parts. The first aims to assess the performance of FREUD whereas the second tries to profile the dominance of FREUD's character constructs when they play against different type of opponents.

The performance analysis in this paper is based on the results of a small tournament that was created using the ipdlx simulator in (Humble 2004). As listed below, the tournament features different types of IPD agents that were selected carefully to reflect various aspects in an IPD opponent.

1- **RAND**: Plays D or C with 50% probability.
2- **ALL-D (Always defect):** Always D
3- **ALL-C (Always cooperate):** Always C
4- **TFT (Tit-for-Tat):** Starts with C, then repeats the last opponent's move.
5- **STFT (Suspicious TFT)**: TFT but with a D in the first move.
6- **TFTT (Tit-for-Two-Tats)**: TFT but plays D after two consecutive opponent defections.
7- **Pavlov:** divides game results into two groups: SUCCESS (3 or 5 points) and FAILURE (0 or 1 point). If its last result belongs to SUCCESS it plays the same move, otherwise it plays the other move.
8- **NEG (Negate):** plays the opposite of what the opponent did in the previous move.
9- **GRIM:** Cooperates until opponent defects. After that it keeps defecting until the end of the game.

Each participant has to play against all other participants and against its twin. The payoff matrix is the one listed in Table 1. Each game consists of 200 rounds, making 1000 the highest possible score in a game (200 * 5), 0 the lowest score in a game (200 * 0), and 600 a good score to achieve (200 * 3) because a game of mutual cooperation between the two players is considered above average in IPD.

## Results and Analysis

The tournament results have shown that a prototype of FREUD can outscore the other strategies. The tournament was run 20 times and FREUD won 17 of these runs. Table 2 lists the results of one of these tournaments, the table reports the strategies rank in the tournament, the match results for the strategies playing against FREUD, and the total strategies score.

Table 2: IPD Tournament Results

| Rank | Strategy | Result against FREUD | | Total Points |
|---|---|---|---|---|
| | | Opponent | FREUD | |
| 1 | FREUD | 562 | 562 | 5594 |
| 2 | GRIM | 273 | 198 | 5472 |
| 3 | TFTT | 331 | 771 | 5471 |
| 4 | TFT | 585 | 585 | 5397 |
| 5 | Pavlov | 220 | 195 | 4619 |
| 6 | ALL-D | 240 | 190 | 4272 |
| 7 | STFT | 580 | 575 | 4202 |
| 8 | RAND | 129 | 569 | 4163 |
| 9 | NEG | 36 | 961 | 3928 |
| 10 | ALL-C | 18 | 988 | 3918 |

The match logs were analyzed to see how FREUD handled different opponents and how the three constructs contributed to FREUD's strategy. The following analysis groups the matches that are similar in term of which construct has dominated over the other two.

### Id-Dominated Matches

Table 3 shows the constructs' profiles in the matches where Id dominated FREUD. The absence of the Ego is due to the fact that it was never utilized by FREUD in these matches, this is because the confidence level of the Id was too high to be ignored.

**Table 3: Construct profiles for Id-dominated matches (CL= Confidence Level)**

| Opp. | Super-Ego | | | Id | | |
|---|---|---|---|---|---|---|
| | Moves | Score | CL | Moves | Score | CL |
| ALL-C | 6 | 18 | 3 | 194 | 970 | 5 |
| ALL-D | 10 | 0 | 0 | 190 | 190 | 1 |
| RAND | 6 | 3 | 0.5 | 194 | 566 | 2.92 |
| NEG | 7 | 0 | 0 | 193 | 961 | 4.98 |
| Pavlov | 5 | 0 | 0 | 195 | 195 | 1 |
| GRIM | 21 | 15 | 0.715 | 179 | 183 | 1.02 |

Imagine the situation when the agent is playing against ALL-C. The Super-Ego will dominate until the Id throws a casual defection. Before that instance, the Super-Ego expected confidence is 3 (since both players cooperating yields 3 points for each). The casual defection will return a payoff of 5 (DC). Thus, FREUD will realize that listening to its Id will bring more utility than listening to its Super-Ego. Consequently, FREUD will suppress the Super-Ego and play a straight game of defections generated by its Id.

The previous example shows the importance of the Id. In Axelrod tournament, all "Nice" strategies scored 600 when played against each other. FREUD is not a NICE strategy as defined by Axelrod since it can defect first. This trait can be attributed to the Id part, and it allows the agent to discover naive opponents that do not retaliate to defections. It took FREUD 6 moves of Super-Ego's cooperation before it threw the first Id defect. After that the Id never surrendered its dominance over the remaining 194 moves as it was scoring a perfect 5 every move, amassing a total of 988 points, compared to the 600 points that a nice strategy would have scored against ALL-C.

The Super-Ego tried to elicit cooperation from ALL-D in the first moves by answering every defection with cooperation. ALL-D was not willing to cooperate and continued exploiting FREUD in the first 10 moves, until the Id was able to deliver its first defection, at which its confidence jumped to 1 (DD payoff) compared to the Super-Ego's confidence of zero. After that point, the Id dominated FREUD's strategy until the end of the game. This situation also occurred against PAVLOV and NEG, where FREUD recovered from constantly scoring the sucker payoff into either getting 1 point from every round (against ALL-D and Pavlov) or running away with the total 5 points of the temptation to defect payoff (against ALL-C and NEG)

The remaining two cases are also useful to analyze. First let us consider the FREUD vs. RAND match. It is known that the best strategy against RAND is to always defect as this will bring a utility of 3 on average (5 if RAND cooperated and 1 if RAND defected, averaging (5+1)/2 = 3 points per round). Playing ALL-C against RAND will average (0+3)/2 = 1.5 per round. In this particular case, RAND only cooperated once in the first 6 moves (when the Super-Ego was dominant), averaging a score of 0.5 per move. Once the Id defected, it kept its dominance over FREUD's character by converging as expected to a confidence level of nearly 3 (2.92 exactly)

Probing the opponent's provocability does not come without cost. GRIM, the never-forgiving agent, did not like the fact that FREUD defected first (the Id move), so it kept defecting until the end of the game, producing a note-worthy case in this study where FREUD kept trying to apologize with no avail. In this match, the Super-Ego played many moves trying to restore the cooperation of GRIM. The negative response from GRIM allowed the Id to dominate FREUD's strategy again and again to score an average of 1 point per round.

## Super-Ego-Dominated Matches

The matches in which the Super-Ego dominates FREUD provide a more interesting case for the advocates of the evolution of cooperation among selfish agents. In these matches, the Id was not able to prevail due to the strong response from the retaliating opponents. Table 4 shows the profile for these matches.

**Table 4: Construct profiles for Super-Ego-dominated matches (CL= Confidence Level)**

| Opp. | Super-Ego | | | Id | | |
|---|---|---|---|---|---|---|
| | Moves | Score | CL | Moves | Score | CL |
| TFT | 190 | 555 | 2.92 | 10 | 30 | 3 |
| STFT | 186 | 537 | 2.89 | 14 | 38 | 2.71 |
| FREUD | 181 | 543 | 3 | 19 | 19 | 1 |

An analysis of the game log against TFT reveals the learning capabilities of FREUD. When the Id first propagated its probing defect, TFT retaliated instantly, reducing the Id confidence level from 5 in the first move to 1 for the following move. This next move served as a lesson for FREUD to realize that TFT is serious against casual defections. As the difference between the Id and Super-Ego confidence levels remained close, the call went to the Super-Ego to decide on the next move. Luckily, TFT has a very short memory; its forgiveness cycle starts from only one cooperative move by the opponent, which caused the match to steer into a draw of 585-585, very close to the 600 points that a nice strategy would have achieved against TFT.

The same scenario was repeated against the suspicious TFT (a TFT variant that defects in the first move) indicating the robustness of FREUD's learning potential against eye-for-eye opponents. The bias to cooperate remained strong even with the sucker's payoff that the Super-Ego attained in its very first move. FREUD played ideally with its twin as the Id confidence never exceeded the threshold to ignite a defection streak.

### Ego-Dominant matches

The subtle nature of TFTT made it hard for FREUD to decide between its Id and Super-Ego. The Super-Ego started well until the Id generated its first defect, this action went unpunished by TFTT because it only retaliates after two consecutive defects. The next defect provoked TFTT, thus reducing the Id's confidence level, but not to the point that the Super-Ego was able to dominate. The cyclic behavior between the two constructs prompted a response from FREUD, who activated the Ego construct. The opponent modeling process started by pattern-matching the game log to look for opportunities to exploit the opponent. The Ego discovered that after each cooperation from FREUD, TFTT is willing to get the sucker's payoff twice before retaliating. This recurring trend led the Ego to execute a winning strategy by alternating between C and D, which allowed it to break the string of D's one move before TFTT would retaliate. This gave FREUD a payoff of consecutive 3s and 5s (one for mutual cooperation and the other for defection when TFTT cooperates).

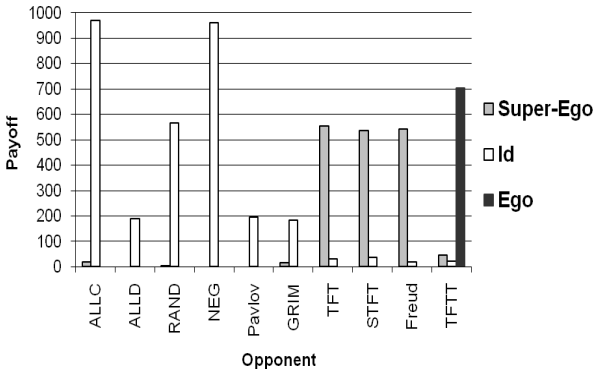


**Figure 2: Payoff of the Three Constructs**

Figure 2 shows the scores of the three constructs against all players. The Id dominated against completely naive or completely mean opponents like ALL-C and GRIM, respectively, whereas the Super-Ego excelled against the opponents who are heavy retaliators. Finally, the fact that the Ego dominated against TFTT demonstrates the capacity of this model to host different opponent-modeling techniques against non-trivial opponents.

## Conclusion

It is extremely important to mention that the contribution of this work does not come from the particular implementation of FREUD as presented in the experimental section; it is rather the idea of outlining a psychologically-inspired approach of modeling the IPD agents. The FREUD platform presents an abstract model that offers ample opportunities of fitting different implementation ideas in the three character constructs and the decision making process. The proposal of having an agent with three competing parts, each of which adopts a different agenda toward scoring high in IPD, has been successfully prototyped and tested against 9 simple benchmark strategies. The profiling of game logs gave a useful insight on how FREUD's character developed against various kinds of strategies. We believe that our future research should concentrate on more robust implementations of the three constructs, matches against stronger opponents, in addition to testing FREUD's resilience to noise.

**Acknowledgments** This work was partially supported by AFOSR award FA9550-09-1-0525

## References

R. Axerold, "Effective choice in the prisoner's dilemma," *Journal of Conflict Resolution*, 1980, Vol.24, No.1, pp. 3-25.

T.C. Au, and D. Nau, "Accident or intention: that is the question (in the Noisy Iterated Prisoner's Dilemma)," *Proc. International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, ACM, 2006, pp. 561-568.

R. Axelrod, and W. Hamilton, "The Evolution of Cooperation," *Science*, March 1981, Vol.211, No.4489, pp. 1390.

S. Kuhn "Prisoner's Dilemma", *The Stanford Encyclopedia of Philosophy*, (Spring 2009), Z. Edward (ed.),URL=<http://plato.stanford.edu/archives/spr2009/entries/prisoner-dilemma/>.

S. Freud, "An outline of psychoanalysis," *International Journal of Psychology*, 1940, pp. 144-207.

M. Nowak, and K. Sigmund, "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game," *Nature*, July 1993, pp. 56-58.

J. Bendor, "In Good Times and Bad: Reciprocity in an Uncertain World," *American Journal of Political Science*, August 1987, Vol. 31, No. 3, pp. 531-558.

P. Molander, "The Optimal Level of Generosity in a Selfish, Uncertain Environment," *The Journal of Conflict Resolution*, December 1985, Vol. 29, No. 4, pp. 611-618.

J. Wu, and R. Axelrod, "How to Cope with Noise in the Iterated Prisoner's Dilemma," *Journal of Conflict Resolution*, March 1995, Vol. 39, No. 1, pp. 183-189.

J. Humble, "IPDLX - IPDL library extension", (2004), URL= <http://www.prisoners-dilemma.com/>.