

Stochastic Collusion Summary

John Maloney
malo0052@umn.edu

The authors of [1] propose an alternative analysis of the evolution of cooperation in the two-person Iterated Prisoner's Dilemma game.

In the one-shot Prisoner's Dilemma game, rational choice theory¹ stipulates that two rational (or utility maximizing) players will each choose to defect leading to the next-to-worst possible payout for each player. In order for the socially optimal outcome of mutual cooperation to be obtained, there must be a sufficiently high probability that the two players will meet again. In this case, the theory predicts that two rational players will choose to cooperate. The standard explanation given for this outcome is that the two players apply forward-looking logical deduction and come to the conclusion that mutual cooperation is the best course of action.

This explanation for cooperation in the Iterated Prisoner's Dilemma game assumes that the participants have adequate cognitive abilities and the desire to do the following:

1. Estimate the probability of meeting their opponent in the future
2. Estimate the probability of being recognized by their opponent in a future interaction
3. Logically deduce that cooperation is the best course of action
4. Have confidence that their opponent also has adequate cognitive abilities to come to the same conclusion

In addition, the rational choice explanation assumes that cooperation emerges as a by-product of players seeking private gain. However, in human society, statements of group solidarity often justify acts of cooperation.

As an alternative to the rational choice model, the authors in [1] suggest a stochastic learning based model. In this model, the players are adaptive, backward looking and reactive rather than purposive, forward-looking and preemptive. In this model, the game payoffs act as positive or negative rewards that reinforce or attenuate a player's tendency to cooperate. Modeling players as having a tendency to cooperate means that players occasionally choose the unexpected action. The authors find that these random unexpected actions are critical to the ability of players to escape social traps.

The authors find that highly cooperative contestants can be drawn into a stable non-cooperative social trap. However, the stochastic nature of the model allows for the chance that a fortuitous sequence of unexpected actions can lead the two actors out of the trap and into a regime of mutual cooperation. The authors find that, within the stochastic learning model, the evolution of cooperation is dependent on the length of the sequence of steps that must be coordinated in order to escape social traps.

¹ https://en.wikipedia.org/wiki/Rational_choice_theory

The authors propose a *stochastic learning model of social exchange*. The model is based on the Bush-Mosteller stochastic learning model for binary choice [2]. In that model, the probability $p^{(t)}$ that an action will be selected at time t is given by the following:

$$p_{t+1} = p_t + O(1 - p_t)$$

where O is a positive constant less than one.

This model can be adapted to the Prisoner's Dilemma game as follows. Assume the following standard payout matrix for the game:

	Cooperate (C)	Defect (D)
Cooperate (C)	R, R	S, T
Defect (D)	T, S	P, P

where the payouts conform to the following constraints: $T > R > 0 > P > S$ and $2R > T + S$. This payout matrix can be reduced to two parameters: magnitude (σ) and severity (γ) defined as follows:

$$\sigma = \frac{(R - P)}{2}$$

$$\gamma = \frac{(T - S)}{R - P}$$

For their analysis, the authors further simplify the payout matrix by instituting the following additional constraints on the payouts:

$$R = -P$$

$$T = -S$$

$$1 > T > R > 0$$

Given this simplification, $\sigma = R$ and $\gamma = T/R$ and the payout matrix can be reformulated as follows:

	Cooperate (C)	Defect (D)
Cooperate (C)	σ, σ	$-\sigma\gamma, \sigma\gamma$
Defect (D)	$\sigma\gamma, -\sigma\gamma$	$-\sigma, -\sigma$

Let $O_{t,i}$ be the payout received by player i at time t . Given the constraints imposed on the payouts, $1 > O_{t,i} > 0$ when there is mutual cooperation (R) or the player unilaterally defects (T) and $0 > O_{t,i} > -1$ when there is mutual defection (P) or the player unilaterally cooperates (S).

Let $C_{i,t}$ be the action taken by player i at time t defined as follows:

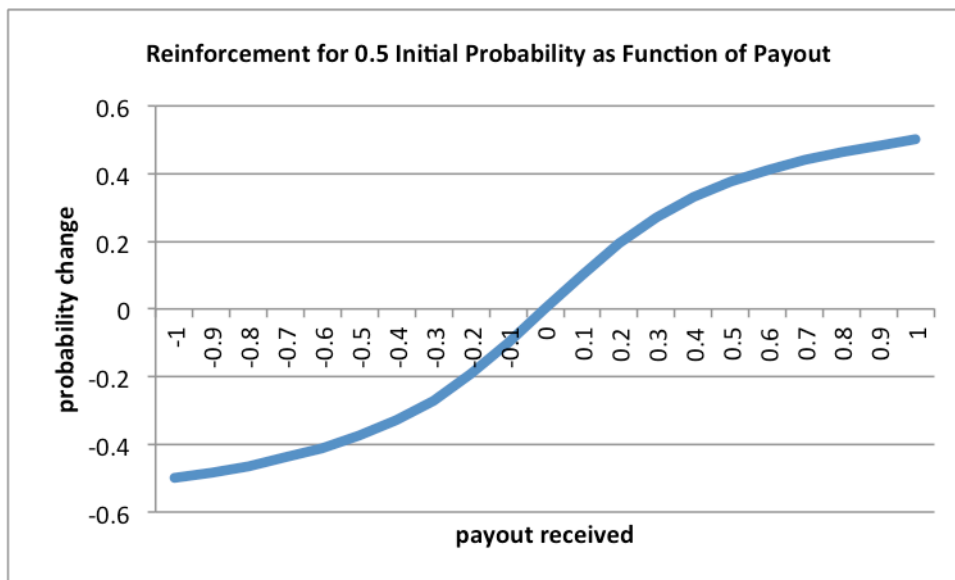
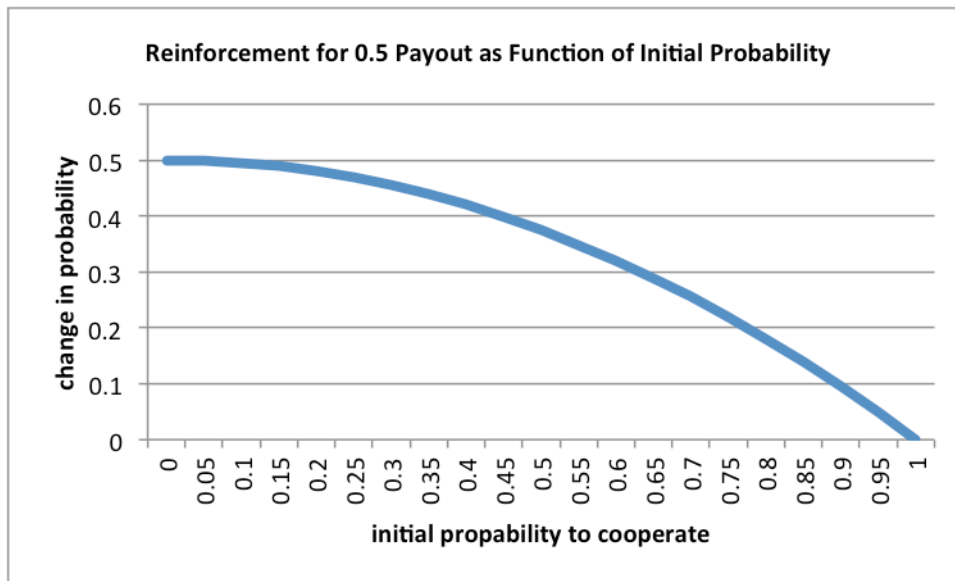
$$C_{t,i} = \begin{cases} 1, & \text{player } i \text{ cooperated at time } t \\ 0, & \text{player } i \text{ defected at time } t \end{cases}$$

Then the probability $p_{t,i}$ that player i will cooperate at time t is defined as follows:

$$p_{t+1,i} = p_{t,i} + \left[O_{t,i} \cdot \left(1 - p_{t,i}^{|O_{t,i}|^{-1}} \right) \cdot C_{t,i} \right] - \left[O_{t,i} \cdot \left(1 - p_{t,i}^{|O_{t,i}|^{-1}} \right) \cdot (1 - C_{t,i}) \right]$$

This adapted form of the model includes two reinforcement terms. The first of these applies when the player chooses to cooperate at time t and increases (decreases) the probability that the player will cooperate at time $t+1$ if a positive (negative) payout was received at time t . The second term applies when the player chooses to defect at time t and decreases (increases) the probability that the player will cooperate at time $t+1$ if a positive (negative) payout was received at time t .

As seen in the following two charts, the reinforcement generated by the model decays with both the initial probability to cooperate and the size of the payout received. The first chart shows how the reinforcement provided, given a constant payout of 0.5, decays as the initial probability to cooperate $p_{t,i}$ increases. The second chart shows how the reinforcement provided, given a constant initial probability to cooperate of 0.5, decays as the absolute value of the payout received approaches one.



Given this model, the symmetric moves CC and DD provide positive cooperation reinforcements while the asymmetric moves CD and DC provide negative cooperation reinforcements. Let p_e be the probability that a player cooperates at equilibrium and let $q_e = (1 - p_e)$ be the probability that a player defects at equilibrium. The reinforcements received by a player for each possible move and the probabilities of each move are given in the following table:

Move	P(Move)	Reinforcement
CC	$P(CC) = p_e^2$	$\sigma (1 - p_e^{1/\sigma})$
DD	$P(DD) = q_e^2$	$\sigma (1 - q_e^{1/\sigma})$
CD	$P(CD) = p_e q_e$	$-\sigma \gamma (1 - p_e^{1/\sigma \gamma})$
DC	$P(DC) = q_e p_e$	$-\sigma \gamma (1 - q_e^{1/\sigma \gamma})$

At equilibrium, the positive cooperation reinforcements received by a player are balanced by the negative cooperation reinforcements received leading to no net change in the player's probability of cooperation. This is captured in the following equation:

$$p_e^2 \sigma (1 - p_e^{1/\sigma}) + q_e^2 \sigma (1 - q_e^{1/\sigma}) - p_e q_e \sigma \gamma (1 - p_e^{1/\sigma \gamma}) - q_e p_e \sigma \gamma (1 - q_e^{1/\sigma \gamma}) = 0$$

In computer simulations involving two agents using the model described above and playing a repeated Prisoner's Dilemma game, the authors find that two equilibriums exist along with a threshold that determines which equilibrium the agents will obtain. A non-cooperative equilibrium exists where the agents are caught in a "self defeating rut" in which asymmetric moves are too common for the agents to develop the trust required for mutual cooperation to prevail. A cooperative equilibrium exists where the agents have built up enough trust to withstand the occasional asymmetric move. A threshold exists such that if both agents' probability to cooperate is above the threshold then the agents will be pushed into the cooperate equilibrium. Below this threshold, the agents are pushed into the non-cooperative equilibrium.

Agents can escape from the non-cooperative equilibrium through a process the authors call *stochastic collusion*. Stochastic collusion occurs when the agents are able to string together a sequence of synchronized symmetric moves that allow them increase their probability of cooperation to the point where they cross the threshold and get pulled into the cooperative equilibrium. Although, the agents select their actions at random, it appears as if the agents "are clever strategists who have finally engineered a tacit collusion". Reducing the number of coordinated moves required to reach the threshold increases the chances of stochastic collusion occurring. Increasing the size of probability change that occurs with each move reduces the required number of moves. Increasing the magnitude (σ) of the payouts increases the size of the probability change and therefore increases the chances that the agents will escape the non-cooperative social trap.

1 References

- [1] Macy, M. W., "Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange," *American Journal of Sociology*, vol. 97, no. 3, pp. 808-843, 1991.
- [2] Bush, R. R., and F. Mosteller, *Stochastic Models for Learning*, New York: Wiley, 1955.