Robert Axelrod, ***The Evolution of Cooperation*** **(1984), chs. 1-2**

Chapter 1: 'The Problem of Cooperation'

Axelrod's aim in this book is "to develop a theory of cooperation that can be used to discover what is necessary for cooperation to emerge. By understanding the conditions that allow it to emerge, appropriate actions can be taken to foster the development of cooperation in a specific setting". This theory is based upon an "investigation of individuals who pursue their own interests without central authority to force cooperation" (6).

The Prisoners' Dilemma forms the book's basis for analysis. Two actors interacting can either cooperate or defect. In a 2x2 matrix, if both cooperate, the result is "reward for mutual cooperation" (R). If one cooperates and the other defects, the result is a "sucker's payoff" (S) for the former and "temptation to defect" (T) for the latter. And vice-versa. If both defect, the result is "punishment for mutual defection" (P) for both.
The Prisoners' Dilemma is defined by:
1) The preference rankings are T, R, P, S
2) R is greater than the average of T and S.
Therefore, it is better to defect of you think the other player will cooperate, and to defect if you think the other player will defect. It pays to defect. but the same holds for the other player. Therefore, they will both get less than if they had both cooperated. Individual rationality leads to a worse outcome for both players. Hence the dilemma (7-10).

The issue therefore becomes the discovery of the precise conditions that are necessary for cooperation to emerge. The ability to recognize and remember allows the history of a particular interaction to be taken into account by a player's strategy. What makes cooperation possible is the fact that they may meet again in the future. But the pay off of the next move always counts less than the pay off of the current move.

This leads to *Proposition 1*:

"If the discount parameter, w [the weight of the next move relative to the current move], is sufficiently high, there is no best strategy independent of the strategy used by the other player".
i.e. It would be better to co-operate with someone who will reciprocate that cooperation in the future, but not with someone whose future behavior will not be very much affected by this interaction.

The simplicity of this framework makes it possible to avoid many restrictive assumptions that would otherwise limit analysis: pay offs are neither necessarily comparable nor symmetric; pay offs can be measured in relative terms; cooperation is not necessarily desirable to the rest of the world; decisions are not necessarily rational or conscious (17).

Chapter 2: Computer Tournament

The Prisoners' Dilemma gets players in a tangle because they are not only dependent on what the other is doing but on what the other expects them to do. Axelrod decides on a computer tournament to deal with this non-zero sum game because it allows players to take into account 1) other strategies and 2) the history of other strategies.

Game theorists were invited to submit their favorite strategy, and each of these decision rules was paired off with each of the others to see which would do best overall. the results of the first tournament were published with an analysis of why strategies did well, and then more entrants were invited to participate. In both tournaments, the simplest of all strategies submitted, TIT FOR TAT (which cooperates on the first round and then does whatever the other did on the previous round), won.

In the first tournament, being "nice" - cooperating on the first round - was a good tactic. Nice strategies did well with nice strategies, so they were distinguished from one another by their performance with those that were not nice. To be "forgiving" (not holding a defect against your opponent in the long term) was a strength: it minimized echo effects in an environment of mutual power. TIT FOR TAT is a forgiving strategy (if you defect and then cooperate, it will then cooperate with you). In the first tournament, in fact TIT FOR TWO TATS (which retaliates after two defections) would have won because it was more forgiving. But it does pay to retaliate.

In the second tournament, TIT FOR TATS strengths of being nice, forgiving and retaliatory were advertised. Strategies tended to diverge from TIT FOR TAT by either becoming more nice and forgiving or more exploitative. The former strategies, like TIT FOR TWO TATS did not do as well because they were exploited in the more competitive environment by the latter strategies. But exploitative strategies failed because they were not nice or forgiving enough. Again TIT FOR TAT won.

When tested over many generations, the survival of the fittest, TIT FOR TAT proved to be a robust strategy: it was most successful with other successful performers. It was noted that not being "nice" destroys the very environment it need for its own success. In the light of *Proposition 1*, TIT FOR TAT benefits from its own non-exploitability and clarity of behavior, while also gaining cooperation by not exploiting other rules.

A balance of *niceness*, *retaliation*, *forgiveness* and *clarity* are the key properties that make a decision rule successful. For TIT FOR TAT:
"Its niceness prevents it from getting into unnecessary trouble. Its retaliation discourages others from persisting whenever defection is tried. Its forgiveness helps restore mutual cooperation. Its clarity makes it intelligible to the other player, thereby eliciting long-term cooperation.