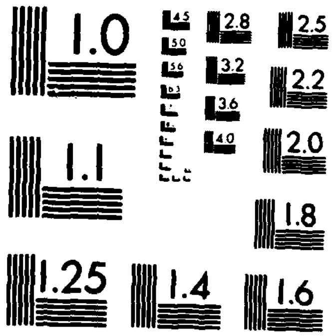
AD-A151 801 MORE ON THE EVOLUTION OF COOPERATION(U) GEORGIA INST OF TECH ATLANTA SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING J J BARTHOLDI ET AL. DEC 84 1/1

UNCLASSIFIED N00014-80-K-0709 F/G 5/4 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

MORE ON THE EVOLUTION OF COOPERATION

GEORGIA INSTITUTE
OF TECHNOLOGY
ATLANTA, GEORGIA 30332

December 1984

MORE ON THE EVOLUTION OF COOPERATION

John J. Bartholdi, III
C. Allen Butler
Michael A. Trick
Georgia Institute of Technology

Abstract

This paper discovers new structure in the suggestive "world" created by Axelrod, which is based on iterated play of the Prisoner's Dilemma game, and was studied to reveal how cooperative behavior can arise in a world of egoists. One of Axelrod's conclusions is that the viability of a strategy depends on how heavily the future is discounted. Our results explain in additional detail the nature of this dependence, and suggest how a specific cooperative strategy, TIT FOR TAT, might evolve from a world of defectors.

Key words; Prisoner's Dilemma; cooperative behavior; Tit-for-Tat.

This research was supported by the National Science Foundation under grant #ECS-8351313 and the Office of Naval Research under grant #N00014-80-k-0709.

Submitted to *J. Conflict Resolution*

0. Introduction.

This paper discovers new structure in the suggestive "world" created by Axelrod [1-5]. This world is based on iterated play of the Prisoner's Dilemma game, and was used to study how cooperative behavior can arise in a world of egoists.

The world is populated by creatures that interact pairwise through plays of the Prisoner's Dilemma game, which is given by the following payoff matrix.

	Cooperate	Defect
Cooperate	R,R	S,T
Defect	T,S	P,P

where $T > R > P > S$ and $R > (T+S)/2$.

All creatures are egoists, who attempt to maximize their expected score. At each encounter each creature has an incentive to defect, since that is the dominant strategy regardless of what the other player does; yet if both players defect, then both are worse off than if they had cooperated with each other. This apparent paradox expresses the tension between individual and social rationality.

In this world creatures can communicate only through plays of this game. Each creature is imbued with a "strategy", which is a mapping from past experience (*i.e.* a history of cooperations and defections) to a probability of cooperation on the next play of the game.

The *discount factor* is the fundamental parameter of the world; it is a number $0 \leq w < 1$, and may be thought of as the probability of two creatures meeting again for another interaction (another play of Prisoner's Dilemma). If w is close to 0, creatures are unlikely to encounter each other again and so there is greater incentive to defect. On the other hand, if w is close to 1, creatures are likely to have a durable relationship, and the incentive to cooperate is increased. One of Axelrod's conclusions is that the viability of a strategy depends on how heavily the future is discounted. Our results explain in additional detail the nature of this dependence, and furthermore suggest how a specific cooperative strategy, TIT FOR TAT (TFT), might evolve from a world of defectors.

TIT-FOR-TAT is the strategy which, for each opponent, cooperates at the first encounter, and in subsequent encounters remembers and plays that opponent's most recent play. TFT has proven remarkably successful in computer tournaments [1,2]. Moreover it has a certain human appeal: it is simple and forthright; it is never the first to defect; it resists exploitation by defecting immediately after the opponent defects; and it is forgiving in that it cooperates immediately after the opponent cooperates [1-5]. Our analysis shows how the very structure of the world can ensure the evolution of such a human strategy, even from a world of egoists who practice defection. In particular we shall imagine a chronology for this world in which the value of w begins near 0 and increases toward 1, and we shall suggest how cooperative strategies in general, and TFT in particular, might spontaneously arise and establish themselves as the durability of relationships (as measured by w) increases. This can be taken to suggest that behavior to which some sort of "morality" might be imputed can spontaneously emerge from a world of selfishness.

1. When is there a best strategy?

The total discounted score expected to accrue to strategy B when playing against strategy A is denoted by $V(B|A)$. A strategy S^* is defined to be a *best* strategy if it does at least as well against any conceivable opponent as would any other strategy: that is $V(S^*|A) \geq V(B|A)$ for any strategies A and B. (Note that "best" is a very strong property: it is defined independently of what strategies are actually practiced by other creatures in the world.) Axelrod observed that if $w=0$ then ALL D, the strategy which defects on every play, is a best strategy, and that if $w > (T-R)/(T-P)$, then there is no best strategy independent of the strategies used by the other players [3,4,5]. Theorems 1 and 2 supplement this:

LEMMA 1: *ALL D is the only strategy which can be a best strategy.*

PROOF: (By contradiction.) Assume that S^* is a best strategy that is not ALL D. Then there must exist some strategy A that can induce S^* to cooperate, say on the *nth* play. Let X be the strategy that performs exactly as A for the first n-1 plays, and then defects ever after. Then

$$V(\text{ALL D}|X) - V(S^*|X) \geq w^n(P-S) > 0,$$

so that S^* is not a best strategy. ■

THEOREM 1: *ALL D is the best strategy for all w in the region*

$$0 \leq w \leq \min\left\{\frac{T-R}{(T-P)+(T-R)}, \frac{P-S}{T-S}\right\} = w_D. \qquad (1)$$

PROOF: Let B be any strategy different from ALL D. We shall compare V(ALL D|A) and V(B|A) for all strategies A. First note that against any strategy A with which B never cooperates, V(ALL D|A) = V(B|A). Next, since B is not ALL D, there must exist some strategy with which B eventually cooperates. Let A be any such strategy and assume that B first cooperates with A on move *n*.

Since the first n-1 moves of B against A are identical to the first n-1 moves of ALL D against A, the *nth* move of A against B must be identical to the *nth* move of A against ALL D. Now if A defects on move n, then ALL D gets P and B gets S; if A cooperates on move n, then ALL D gets T and B gets R. From move n+1 on, the worst that ALL D can do is to get P at each move, and the best that B can do is to get T at each move. Consequently,

$$V(\text{ALL D}|A) - V(B|A) \geq w^n \min\{T-R, P-S\} + \frac{w^{n+1}}{1-w}(P-T),$$

and the latter term is ≥0 only for w as in (1). Thus ALL D is a best strategy in this region, and by Lemma 1 ALL D must be the unique best in this region. ■

THEOREM 2: *There is no best strategy for any w in the region*

$$w_D = min\left\{\frac{T-R}{(T-P)+(T-R)}, \frac{P-S}{T-S}\right\} < w < 1.$$

PROOF: (By contradiction.) By Lemma 1, if a best strategy exists it must be ALL D. Now consider the following strategies:

MG: ("Mindlessly Grateful") plays D, but responds to any C with uninterrupted C's ever after;

MR: ("Massive Retaliation") plays C, but responds to any D with uninterrupted D's ever after;

CI: ("Completely Impressionable") plays first move C and then repeats opponent's first move forever;

PI: ("Perversely Impressionable") plays first move C and then plays the opposite of the opponent's first move forever.

Now for ALL D to be best means that V(ALL D|MG) = $P/(1-w)$ must be at least as large as V(MR|MG) = $S + wT/(1-w)$; but this is so exactly when $w \leq (P-S)/(T-S)$.

Similarly, ALL D best means that V(ALL D|CI) = $T + wP/(1-w)$ must be at least as large as V(PI|CI) = $R + wT/(1-w)$; and this is so exactly when $w \leq (T-R)/[(T-P)+(T-R)]$. Therefore ALL D can be a best strategy only when w is no larger than the minimum of these two bounds, which, when simplified, = w_D. ■

Theorems 1 and 2 say that the range of discount factors w is partitioned into two regions:

$0 \leq w \leq w_D$	:the "Region of Despair", within which ALL D is the uniquely best strategy;
$w_D < w < 1$	:the "Region of Hope", within which there is no strategy that is best independently of the other strategies present.

Of course the exact value of w_D depends on T, R, P, and S. For appropriate values of these payoffs, w_D can be arbitrarily close to 0, so that for some versions of the Prisoner's Dilemma game the Region of Despair can be arbitrarily small. However, w_D is always less than 1/2, so that for *any* version of the game, the Region of Despair is smaller than the Region of Hope.

For any value of w, the expected number of meetings between two creatures is $1/(1-w)$. If the expected number of meetings is at least 2, this corresponds to $w \geq 1/2 > w_D$, so w is within the Region of Hope. Thus there is a possibility of cooperation whenever one expects to meet another creature again - and this holds independently of the values T, R, P, and S.

2. The Evolution of Tit-for-Tat.

In the Region of Despair, ALL D is invincible: no other strategy can do better than it. Moreover, any strategy that may be the first to cooperate will do strictly worse than ALL D, no matter which other strategies populate the world. This raises the question of how cooperative behavior can establish itself as a basis for viable strategies. Axelrod and Hamilton [4] give an imaginative chronology suggesting how TFT can establish itself as a viable strategy in a world of defectors once individuals playing TFT are present in sufficient numbers. However, the chronology does not explain how TFT could ever evolve from and attain sufficient numbers beginning from a primordial world where w is small. In this section we supplement their chronology and suggest a means whereby cooperative behavior in general, and TFT in particular, might evolve from a primordial world of defectors.

A strategy B is *collectively stable* if $V(B|B) \geq V(A|B)$ for any strategy A. This means that, on the average, individuals using the community strategy B, who interact mostly with other B's, can do at least as well as an invading individual using some other other strategy A. Stability is important because it enables a community of individuals using the same strategy to remain viable even in the presence of a rogue using a strategy that may be stronger ($V(A|B)$ may be larger than $V(B|A)$).

For a given strategy B, let us define its *region of stability* to be the set of w for which B is collectively stable. For example, in [3,4,5] it is shown that the region of stability of ALL D is $0 \leq w < 1$, and the region of stability of TFT is

$$\max \left\{\frac{T-R}{T-P}, \frac{T-R}{R-S}\right\} \leq w < 1.$$

The viability of a strategy is indicated by the size and structure of its region of stability. Thus the effectiveness of ALL D is reflected by its region of stability spanning all $0 \leq w < 1$. (We note in passing that it might be interesting to classify strategies by the structures of their regions of stability.)

The following is implicit in [3,4,5].

OBSERVATION 1: *No strategy which may be the first to cooperate can be stable for $0 \leq w < (T-R)/(T-P)$. In particular, no such strategy can be stable in the Region of Despair.*

This follows since otherwise such strategies would be vulnerable to invasion by ALL D.

The region of stability of TFT is entirely contained within the Region of Hope. Since TFT can never be stable for w in the Region of Despair, it is puzzling how TFT could evolve from a primordial world where w is small. It seems that TFT must evolve from some less cooperative ancestor that is better adapted to a world of defection.

We suggest that the "missing link" may be *SUSPICIOUS* TFT (STFT), that defects on the first play, and afterwards plays TFT. Three observations make this plausible. First, TFT is very similar to STFT, and may be derived from it by a simple mutation that is independent of the logic of the strategy (reciprocity). The second observation is that STFT is stable throughout the Region of Despair, and so can be viable in the primordial world:

THEOREM 3: *Suspicious TFT is stable for all discount factors*

$$0 \leq w \leq \min \left\{\frac{P-S}{R-S}, \frac{P-S}{T-P}\right\}.$$

PROOF: The argument is similar to that which establishes the region of stability for TFT [3,4,5]. Since STFT remembers only its last encounter, if any strategy can invade STFT, then either ALL C or else ALTERNATING C AND D can invade. But neither of these can invade exactly when the conditions of the theorem hold. ■

Finally, the regions of stability of TFT and STFT can overlap. This provides a stable "bridge" along which evolution can proceed from the Region of Despair to the Region of Hope.

THEOREM 4: *The regions of stability of TFT and STFT overlap if-and-only-if $R+P \geq T+S$.*

PROOF: Assume that $R+P \geq T+S$. Then,

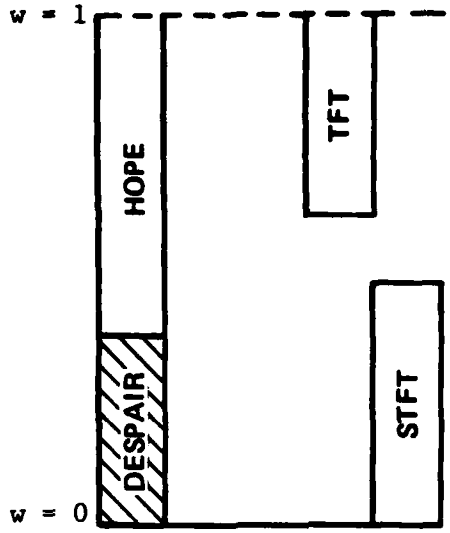
$$\begin{aligned}
R^2 - P^2 &\geq (T+S)(R-P) && \Rightarrow \\
R^2 - P^2 &\geq TR + RS - TP - PS && \Rightarrow \\
TP - ST - P^2 + PS &\geq TR - ST - R^2 + RS && \Rightarrow \\
(P-S)(T-P) &\geq (R-S)(T-R) && \Rightarrow \\
\frac{P-S}{R-S} &\geq \frac{T-R}{T-P} && \Rightarrow \\
\min\left\{\frac{P-S}{R-S}, \frac{P-S}{T-P}\right\} &\geq \max\left\{\frac{T-R}{R-S}, \frac{T-R}{T-P}\right\},
\end{aligned}$$

so that the regions of stability overlap.

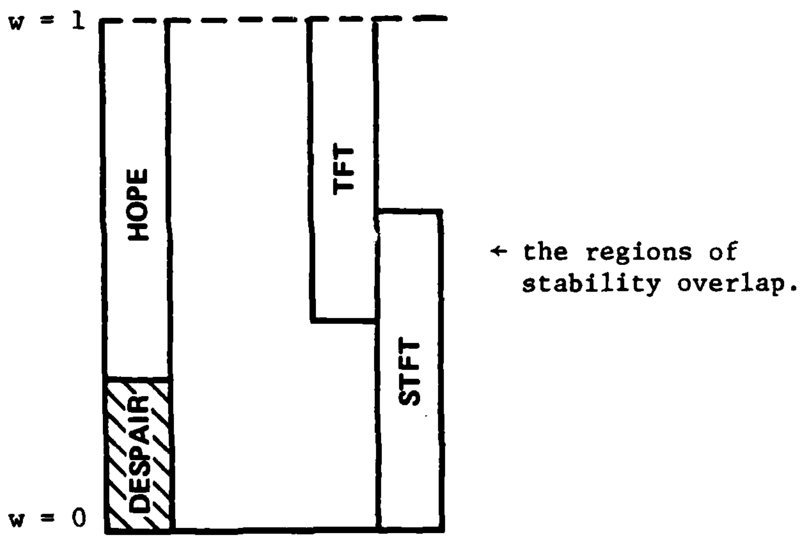
The converse follows by similar calculations. ∎

We can imagine $R+P$ to indicate the aggregate value of synchronous play in the Prisoner's Dilemma, and $T+S$ the value of asynchronous play. The relative values of synchronous and asynchronous play determine some of the important structure of the world. In particular, whether $R+P$ is greater than or less than $T+S$ determines endpoints of the Region of Despair and of the regions of stability of STFT and TFT. (See Figure 1.)

The condition $R+P > T+S$ has important implications for TFT. First, such a situation seems to favor strategies which practice reciprocity. This may be seen by considering a creature's immediate incentive to defect, that is, its immediate net gain should it cease cooperating with the other player and instead defect on the next play. If the other player has been cooperating, that incentive is $T-R$, while if the other player has been defecting, the incentive is $P-S$. The condition $R+P > T+S$ implies that the immediate incentive to defect from a defecting other player is greater than the incentive to defect from a cooperating other player. Thus the structure of the world may apply evolutionary pressure that favors the practice of reciprocity.



When R + P < T + S...



When R + P > T + S...

Figure 1: The structure of the world is determined by the values of T, R, P, S, and w.

An additional implication of $R+P > T+S$ is that TFT is stable throughout the region $(T-R)/(T-P) \leq w < 1$, and by Observation 1 TFT is *maximally stable* among strategies which may be the first to cooperate. Furthermore, within the region $(T-R)/(T-P) \leq w \leq (P-S)/(R-S)$, both STFT and TFT are collectively stable so that a community practicing one of these strategies can resist invasion by an individual practicing any other strategy.

The stability of STFT in this region enables a community practicing STFT to successfully resist invasion by a single TFT. However, communities may succumb to stronger forms of invasion. In particular, a community practicing strategy B may be invaded by a cluster of individuals practicing strategy A if the individuals of that cluster interact mostly with themselves and not randomly with the larger community. Thus a *ρ-cluster of A invades B* if $\rho V(A|A) + (1-\rho)V(A|B) > V(B|B)$, where ρ is the proportion of interactions of an individual using strategy A with another such player [3,4,5]. Consequently, a ρ-cluster of TFT can invade STFT when

$$\rho\left(\frac{R}{1-w}\right) + (1-\rho)\left(\frac{S+wT}{1-w^2}\right) > \frac{P}{1-w},$$

that is, when

$$\rho > \frac{P-S-w(T-P)}{R-S-w(T-R)}.$$

Now the last expression is monotonically decreasing in w, so that as w increases within the region $(T-R)/(T-P) \leq w \leq (P-S)/(R-S)$ STFT becomes more vulnerable to invasion by ρ-clusters of TFT (*i.e.* subject to invasion for smaller values of ρ). Moreover, for appropriate values of T, R, P, and S, ρ can be an arbitrarily small positive value. (For example, let T and R be large relative to P and S.)

A chronology for the evolution of TFT might be as follows. In the primordial world w is small and ALL D is the uniquely best strategy. Mutations occur which produce other strategies, but because of the small value of w, the only strategies which are viable are those which never cooperate first. Consequently most individuals appear to practice ALL D.

Strategies with memory evolve, perhaps because they are better able to take advantage of occasional cooperators.

If $R+P > T+S$, then strategies practicing reciprocity are favored to evolve, as discussed earlier. This enables the evolution of STFT, a defector that practices reciprocity. Since it is never the first to cooperate, STFT can survive in a highly discounted world of defectors, among whom it is indistinguishable.

A simple mutation of STFT, from defecting on the first play to cooperating on the first play, produces TFT, a cooperator that practices reciprocity. Since the mutation is so simple, mutants practicing TFT are produced relatively often. However, for low values of w, TFT is not stable and so is not a viable strategy.

In the primordial world there occur backwaters and eddies in which, at least locally, $w > w_D$. STFT can still be stable there, even though w is within the Region of Hope. Furthermore, where $R+P > T+S$, both STFT and TFT can be collectively stable for appropriate values of w.

For still larger values of w, STFT becomes increasingly vulnerable to invasion by clusters of its mutant, TFT. For the right values of w, T, R, P, and S (which may occur locally), STFT must succumb to invasion by even very small clusters of TFT.

For $(P-S)/(R-S) < w < (T-R)/(R-S)$, STFT is no longer stable; from the details of Theorem 3, communities of STFT can be invaded by individuals that appear to practice ALL C. However, while individuals that appear to practice ALL C can invade STFT, they cannot establish themselves since they in turn can be invaded by individuals practicing STFT! There is complex interaction in this region, with neither STFT nor apparent practicioners of ALL C able to establish themselves. Meanwhile, ρ-clusters of TFT can continue to successfully invade communities of STFT.

For $w > (T-R)/(R-S)$ even individuals practicing TFT can successfully invade the dying parent community of STFT. Moreover, within its region of stability TFT can resist invasion by any other strategy, even if the invading strategy comes in clusters [3,4,5].

Once communities of TFT are established, they provide hospitable environments for other strategies that are also "nice" (never the first to defect), but possibly weaker. Thus other cooperators are able to live symbiotically within the community.

Finally, once communities of cooperators are established, they might gradually learn to extract greater reward R from mutual cooperation. As R increases, T-R decreases, and w_D decreases. Thus the structure of the world may be changed (at least locally) to be still more favorable to cooperative behavior.

ACKNOWLEDGEMENTS

We thank Terry Connolly and Steven Hackman for helpful comments.

REFERENCES

1. R. Axelrod (1980a). "Effective Choice in the Prisoner's Dilemma", *Journal of Conflict Resolution* **24**(1): 3-25.

2. ---------- (1980b). "More Effective Choice in the Prisoner's Dilemma", *Journal of Conflict Resolution* **24** (3): 379-403.

3. ---------- (1981). "The Emergence of Cooperation Among Egoists", *The American Political Science Review* **75**: 306-318.

4. ---------- (1984). *The Evolution of Cooperation*, Basic Books.

5. ---------- and W.D. Hamilton (1981). "The Evolution of Cooperation", *Science* **211**: 1390-1396.

END

FILMED

5-85

DTIC