

## Morality and the Selfish Gene



Prepared by  
Jim Fulton

For the  
November 15, 2009 Service at  
**All Souls Unitarian Universalist Community**

Contents

Morality and the Selfish Gene 1	
1. The Prisoner's Dilemma - Modified	2
2. The Prisoner's Dilemma - Iterated	3
3. The Prisoner's Dilemma - Strategies	3
4. The Simulation Tournament	4
5. The Second Simulation Tournament	7
6. The Ecological Tournament	8
7. The Researcher's Lessons	11
8. More Lessons and Questions	12
9. Appendix 1. The Original Prisoner's Dilemma	13
10. Appendix 2. Related Sermons by Jim Fulton	14

## Morality and the Selfish Gene

Evolution is usually described as the survival of the fittest, the province of the “selfish gene”. We’re all in competition for resources and mates; those who win in that competition survive to have offspring and descendants. How is that in such a dog-eat-dog world our species, and others, could have developed a concept of ethical behavior, a set of rules that restricts the way we engage in that competition? How can such a set of rules make us more fit?

Surely, you say, it has something to do with the fact that we’re a social species; we have to live together. And you’re right, as far as it goes, but there seems to be a chicken-and-egg problem here: are we ethical because we’re a social species, or are we a social species because we developed a sense of ethical behavior.

A couple weeks ago, Rev. Jim Anderson offered a very optimistic outlook.<sup>1</sup> He suggested that social interdependency was planted by evolution in our very souls: “The purpose of life for humans is to simply live in relationship with each other.” You might not have caught part of what he said in explanation: “Using computer game theory, scientists have found that in the long run it is always best to first be kind and cooperate. If the other person betrays your kindness, you respond with a negative consequence, and then return to a stance of kindness and cooperation.” I did notice this remark, because I was already planning today’s discussion.

Now evolution is a notoriously difficult subject for experimental science. The time span required for traits to evolve, especially human traits, is typically far longer than allowed by the academic publish-or-perish cycle. So playing empirical what-if games in the laboratory doesn’t work. Evolutionists normally follow the geologists around, looking for confirmation of their hypotheses in the fossil record.

To some extent, though, evolutionists can explore the viability of some of their hypotheses using computer simulations. They develop models of the evolutionary process, or some significant part of it; then they set up certain scenarios to see what happens. Such simulations are not conclusive, of course, but they can help in determining whether a given hypothesis is worthy of further exploration.

Such a simulation was reported in *Scientific American* in 1983 by Douglas Hofstadter, the author of *I Am a Strange Loop*, the book we are reading in our Sunday morning religious education class.<sup>2</sup> In this article, Hofstadter distills and analyzes the findings of Robert Axelrod published in 1984, regarding an intriguing set of computer simulations comparing strategies for resolving the so-called *Prisoner’s Dilemma*.<sup>3</sup> (Hofstadter, by

---

<sup>1</sup> Rev. Jim Anderson, The Religion of our Genes, All Souls UUC, November 1, 2009, [http://www.asuuc.org/Home\\_files/Sermons/2009/1101-ReligionOfOurGenes/The\\_Religion\\_of\\_our\\_Genes\\_JimAnderson.pdf](http://www.asuuc.org/Home_files/Sermons/2009/1101-ReligionOfOurGenes/The_Religion_of_our_Genes_JimAnderson.pdf)

<sup>2</sup> Douglas R. Hofstadter, “The Prisoner’s Dilemma – Computer Tournaments and the Evolution of Cooperation”, “Metamagical Themas”, *Scientific American*, May 1983. Reprinted in Hofstadter, *Metamagical Themas*, Basic Books, New York, 1985, pp. 715-734.

<sup>3</sup> Robert Axelrod, *The Evolution of Cooperation*, Basic Books, New York, 1984.

the way, didn't have a time-machine to jump ahead to 1984 to examine Axelrod's publication; rather he had access to a pre-publication draft.)

Today I will summarize Hofstadter's article, from which I will quote extensively, so extensively that once we get started, you can assume the words are Hofstadter's unless I tell you otherwise. Yes, I know, I giving you a report of a report of a report, so there's a significant likelihood of the kind of distortion you find in the child's game of passing a secret from one child to the next in a long chain. But I think you'll find the results interesting and perhaps a little troubling.

## 1. The Prisoner's Dilemma - Modified

Hofstadter begins:

*Life is filled with paradoxes and dilemmas. Sometimes it even feels as if the essence of living is the sensing – indeed, the savoring – of paradox. Although all paradoxes seem somehow related, some paradoxes seem abstract and philosophical, while others touch on life very directly. A very lifelike paradox is the so-called “Prisoner's Dilemma”, discovered in 1950 by Melvin Dresher and Merrill Flood of the RAND Corporation. Albert W. Tucker wrote the first article on it, and in that article he gave it the now-famous name. I shall here present the Prisoner's Dilemma – first as a metaphor, then as a formal problem.<sup>4</sup>*

*.... Assume you possess copious quantities of some item (money, for example), and wish to obtain some amount of another item (perhaps stamps, groceries, diamonds). You arrange a mutually agreeable trade with the only dealer of that item known to you. You are both satisfied with the amounts you will be giving and getting. For some reason, though, your trade must take place in secret. Each of you agrees to leave a bag at a designated place in the forest, and to pick up the other's bag at the other's designated place. Suppose it is clear to both of you that the two of you will never meet or have further dealings with each other again. [Now if you're wondering why this is called “the **prisoner's** dilemma”, it's because Axelrod and Hofstadter used a somewhat different story than the original, which you can find in an Appendix at the end of this sermon. Now back to Hofstadter:]*

*Clearly, there is something for each of you to fear: namely, that the other one will leave an empty bag. Obviously, if you both leave full bags, you will both be satisfied; but equally obviously, getting something for nothing is even more satisfying. So you are tempted to leave an empty bag. In fact, you can even reason it through quite rigorously this way: “If the dealer brings a full bag, I'll be better off having left an empty bag, because I'll have gotten all that I wanted and given away nothing. If the dealer brings an empty bag, I'll be better off having left an empty bag, because I'll not have been cheated. I'll have gained nothing but lost nothing either. Thus it seems that no matter what the dealer **chooses** to do, I'm better off leaving an empty bag. So I'll leave an empty bag.”*

---

<sup>4</sup> Throughout this report, I will use italics to indicate text quoted from Hofstadter, and will remove italics where he uses them.

*The dealer, meanwhile, being in more or less the same boat (though at the other end of it), thinks analogous thoughts and comes to the parallel conclusion that it is best to leave an empty bag. And so both of you, with your impeccable (or impeccable-seeming) logic, leave empty bags, and go away empty-handed. How sad, for if you had both just cooperated, you could have each gained something you wanted to have. Does logic prevent cooperation? This is the issue of the Prisoner's Dilemma....*

## 2. The Prisoner's Dilemma - Iterated

*Let us now go back to the original metaphor and slightly alter its conditions. Suppose that both you and your partner very much want to have a regular supply of what the other has to offer, and so, before conducting your first exchange, you agree to carry on a lifelong exchange, once a month. You still expect never to meet face to face. In fact, neither of you has any idea how old the other one is, so you can't very sure of how long the lifelong agreement may go on, but it seems safe to assume it'll go on for a few months anyway, and very likely for years.*

*Now, what do you do on your first exchange? Taking an empty bag seems fairly nasty as the opening of a relationship – hardly an effective way to build up trust. So suppose you take a full bag, and the dealer brings one as well. Bliss – for a month. Then you both must go back. Empty, or full? Each month, you have to decide whether to defect (take an empty bag) or to cooperate (take a full one). Suppose that one month, unexpectedly, your dealer defects. Now what do you do? Will you suddenly decide that the dealer can never be trusted again, and from now on always bring empty bags, in effect totally giving up on the whole project forever? Or will you pretend you didn't notice, and continue being friendly? Or – will you try to punish the dealer by some number of defections of your own? One? Two? A random number? An increasing number, depending on how many defections you have experienced? Just how mad will you get?*

*This is the so-called **iterated** Prisoner's Dilemma. It is a very difficult problem....*

*It is obvious that in a **collective** sense, it would be best for both of you to always cooperate. But suppose you have no regard whatsoever for the other. There is no "collective good" you are both working for. You are both supreme egoists. Then what? ... This is what is meant by "egoism". It means you have no feeling of friendliness or goodwill or compassion for the other player; you have no conscience; all you care about is [profit]<sup>5</sup>....*

## 3. The Prisoner's Dilemma - Strategies

*Well, what would be your best strategy? It can be shown quite easily that there is no universal answer to this question. That is, there is no strategy that is better than all other strategies under all circumstances. For consider the case where the other is playing ... the strategy of defecting each round. In that case, the best you can possibly do is*

---

<sup>5</sup> I have omitted sections where Hofstadter formalized the problem of the Prisoner's Dilemma. On those occasions where he uses notation from those sections, I have replaced that notation by language that expresses his intent, and enclosed those replacements in square brackets.

to defect each time as well, including the first. On the other hand, suppose the other player is using the Massive Retaliatory Strike strategy, which means “I’ll cooperate until you defect and thereafter I’ll defect forever.” Now if you defect on the very first move, then you’ll [unfairly profit from that one and never benefit again] until one of you dies. But if you had waited to defect, you could have benefited from a relationship of mutual cooperation, amassing many [fair trades] beforehand. Clearly that bunch of [fair trades] will add up to more than the single [unfair profit] if the game goes on for more than a few moves. This means that against the [Always Defect] strategy, [Always Defect] is the best counterstrategy, whereas “Always cooperate unless you learn that you or the other player is just about to die, in which case defect” is the best counterstrategy against Massive Retaliatory Strike. This simple argument shows that how you should play depends on who you’re playing.

The whole concept of the “quality” of a strategy takes on a decidedly more operational and empirical meaning if one imagines an ocean populated by dozens of little beings swimming around and playing Prisoner’s Dilemma over and over with each other. Suppose that each time two such beings encounter each other, they recognize each other and remember how previous encounters have gone. This enables each one to decide what it wishes to do this time. Now if each organism is continually swimming around and bumping into the others, eventually each one will have met every other one numerous times, and thus all strategies will have been given the opportunity to interact with each other. By “interact”, what is meant here is certainly not that anyone knocks anyone else out of the ocean, as in an elimination tournament. The idea is simply that each organism gains zero or more points in each meeting, and if sufficient time is allowed to elapse, everybody will have met with everybody else about the same number of times, and now the only question is: Who has amassed the most points? Amassing points [i.e., profiting the most from the trades] is truly the name of the game.

It doesn’t matter if you have “beaten” anyone, in the sense of have gained more from interacting with them than they gained from interacting with you. That kind of “victory” is totally irrelevant here. What matters is not the number of “victories” rung up by an individual, but the individual’s **total point count** – a number that measures the individual’s overall viability in this particular “sea” of many strategies. It sounds nearly paradoxical, but you could lose many – indeed, all – of your individual skirmishes, and yet come out the overall winner.

As the image suggests very strongly, this whole situation is highly relevant to questions in evolutionary biology. Can totally selfish and unconscious organisms living in a common environment come to evolve reliable cooperative strategies? Can cooperation emerge in a world of pure egoists? In a nutshell, **can cooperation evolve out of non-cooperation?** If so, this has revolutionary import for the theory of evolution, for many of its critics have claimed that this was one place that it was hopelessly snagged.

## 4. The Simulation Tournament

Well, as it happens, it has now been demonstrated rigorously and definitively [remember, I am quoting Hofstadter here] that such cooperation can emerge, and it was done through a computer tournament conducted by political scientist Robert Axelrod of

the Political Science Department and the Institute for Public Policy Studies of the University of Michigan in Ann Arbor. More accurately, Axelrod first studied the ways that cooperation evolved by means of a computer tournament, and when general trends emerged, he was able to spot the underlying principles and prove theorems that established the facts and conditions of cooperation's rise from nowhere.... Furthermore, he and evolutionary biologist William D. Hamilton have worked out and published many of the implications of these discoveries for evolutionary theory. Their work has won much acclaim – including the 1981 Newcomb Cleveland Prize, a prize offered annually by the American Associations for the Advancement of Science for “an outstanding paper published in Science”.

There are really three aspects of the question “Can cooperation emerge in a world of egoists?” The first is: How can it get started at all? The second is: Can cooperative strategies survive better than their noncooperative rivals? The third one is: Which cooperative strategies will do the best and how will they come to predominate?

To make these issues vivid, let me describe Axelrod's tournament and its somewhat astonishing results. In 1979, Axelrod sent out invitations to a number of professional game theorists, including people who had published articles on the Prisoner's Dilemma, telling them that he wished to pit many strategies against one another in a round-robin Prisoner's Dilemma tournament, with the overall goal being to amass as many points as possible. He asked for strategies to be encoded as computer programs that could respond to the 'C' [cooperate] or 'D' [defect] of another player, taking into account the remembered history of previous interactions with that player. A program should always reply with a 'C' or 'D', of course, but its choice need not be deterministic. That is, consultation of a random-number generator was allowed at any point in a strategy.

Fourteen entries were submitted to Axelrod, and he introduced into the field one more program called RANDOM, which in effect flipped a coin (computationally simulated, to be sure) each move, cooperating if heads came up, defecting otherwise. The field was a rather variegated one, consisting of programs ranging from as few as four lines to as many as 77 lines (of Basic [a programming language]). Every program was made to engage each other program (and a clone of itself) 200 times. No program was penalized for running slowly. The tournament was actually run five times in a row, so that pseudo-effects caused by statistical fluctuations in the random-number generator would be smoothed out by averaging.

The program that won was submitted by the old Prisoner's Dilemma hand, Anatol Rapoport, a psychologist and philosopher from the University of Toronto. His was the shortest of all submitted programs, and is called TIT FOR TAT. TIT FOR TAT uses a very simple tactic:

Cooperate on move 1;  
thereafter, do whatever the other player did the previous move.

That is all. It sounds outrageously simple. How in the world could such a program defeat the complex stratagems devised by the other experts?

Well, Axelrod claims that the game theorists in general did not go far enough in their analysis. They looked “only two levels deep”, when in fact they should have looked

three levels deep to do better. *What precisely does this mean? He takes a specific case to illustrate his point. Consider the entry called JOSS.... JOSS's strategy is very similar to TIT FOR TAT's, in that it begins by cooperating, always responds to defection by defecting and nearly always responds to cooperation by cooperating. The hitch is that JOSS uses a random-number generator to help it decide when to pull a "surprise defection" on the other player. JOSS is set up so that it has a 10 percent probability of defecting right after the other player has cooperated.*

*In playing TIT FOR TAT, JOSS will do fine until it tries to catch TIT FOR TAT off guard. When it defects, TIT FOR TAT retaliates with a single defection, while JOSS "innocently" goes back to cooperating. Thus we have a "DC" pair. On the next move, the 'C' and 'D' will switch places since each program in essence echoes the other's latest move, and so it will go: CD, then DC, CD, DC, and so on. There may ensue a long reverberation set off by JOSS's D, but sooner or later, JOSS will randomly throw in another unexpected D after a C from TIT FOR TAT. At this point, there will be a "DD" pair, and that determines the entire rest of the match. Both will defect forever, now. The "echo" effect resulting from JOSS's first attempt at exploitation and TIT FOR TAT's simple punitive act lead ultimately to complete distrust and lack of cooperation.*

*This may seem to imply that both strategies are at fault and will suffer for it at the hands of others, but in fact the one that suffers from it most is JOSS, since JOSS tries out the same trick on partner after partner, and in many cases this leads to the same type of breakdown of trust, whereas TIT FOR TAT, never defecting first, will never be the initial cause of a breakdown of trust. Axelrod's technical term for a strategy that never defects before its opponent does is **nice**. TIT FOR TAT is a nice strategy, JOSS is not. Note that "nice" does not mean that a strategy never defects! TIT FOR TAT defects when provoked, but that is still considered being "nice".*

*Axelrod summarizes the first tournament this way:*

A major lesson of this tournament is the importance of minimizing echo effects in an environment of mutual power. A sophisticated analysis must go at least three levels deep. First is the direct effect of a choice. This is easy, since a defection always earns more [on **that** occasion] than a cooperation. Second are the indirect effects, taking into account that the other side may or may not punish a defection. This much was certainly appreciated by many of the entrants. But third is the fact that in responding to the defections of the other side, one may be repeating or even amplifying one's own previous exploitative choice. Thus a single defection may be successful when analyzed for its direct effects, and perhaps even when its secondary effects are taken into account. But the real costs may be in the tertiary effects when one's own isolated defections turn into unending mutual recriminations....<sup>6</sup>

*Axelrod not only analyzed the first tournament, he even performed a number of "sub-junctive replays" of it, that is, replays with different sets of entries. He found, for instance, that the strategy called TIT FOR TWO TATS, which tolerates two defections be-*

---

<sup>6</sup> Axelrod. Quoted sections in Hofstadter's article were taken from an early draft of that book.

*fore getting mad (but still only strikes back once), would have won, had it been in the line-up....*

*In summary, the lesson of the first tournament seems to have been that it is important to be **nice** (“don’t be the first one to defect”) and **forgiving** (“don’t hold a grudge once you’ve vented your anger”). TIT FOR TAT possesses both these qualities, quite obviously.*

## **5. The Second Simulation Tournament**

*After this careful analysis, Axelrod felt that significant lessons had been unearthed, and he felt convinced that more sophisticated strategies could be concocted, based on the new information. Therefore he decided to hold a second, larger tournament. For this tournament, he not only invited all the participants in the first round, but also advertised in computer hobbyist magazines, hoping to attract people who were addicted to programming and who would be willing to devote a good deal of time to working out and perfecting their strategies. To each person who entered, Axelrod sent a full and detailed analysis of the first tournament, along with a discussion of the “subjunctive replays” and the strategies that would have won. He described the strategic concepts of “niceness” and “forgiveness” that seemed to capture the lessons of the tournament so well, as well as strategic pitfall to avoid. Naturally, each entrant realized that all the other entrants had received the same mailing, so that everyone knew that everyone knew that everyone knew that ....*

*Altogether, 62 entries were received, and generally speaking, they were of a considerably higher degree of sophistication than those in the first tournament. The shortest was again TIT FOR TAT, and the longest was a program from New Zealand, consisting of 152 lines of Fortran. Once again, RANDOM was added to the field, and with a flourish and a final carriage return, the horses were off! Several hours of computer time later, the results came in.*

*The outcome was nothing short of stunning: TIT FOR TAT, the simplest program submitted, won again. What’s more, the two programs submitted that had won the subjunctive replays of the first tournament now turned up way down in the list....*

*This may seem horribly nonintuitive, but remember that a program’s success depends entirely on the environment in which it is swimming. There is no single “best strategy” for all environments, so that winning in one tournament is no guarantee of success in another. TIT FOR TAT has the advantage of being able to “get along well” with a great variety of strategies, while other programs are more limited in their ability to evoke cooperation....*

*[T]he majority of participants in the second tournament really had not grasped the central lesson of the first tournament: the importance of being willing to initiate and reciprocate cooperation. Axelrod feels so strongly about this that he is reluctant to call two strategies playing against each other “opponents”; in his book he always uses neutral terms such as “strategies” or “players”. He even does not like saying they are playing **against** each other, preferring “with”. In this article, I have tried to follow his usage, with occasional departures. One very striking fact about the second tournament is the*

success of “nice” rules: of the top fifteen finishers, only one (placing eighth) was not nice. Amusingly, a sort of mirror image held: of the bottom fifteen finishers, only one was nice!

Several non-nice strategies featured rather tricky probes of the opponent (sorry!), sounding it out to see how much it “minded” being defected against. Although this kind of probing by a program might fool occasional opponents, more often than not it backfired, causing severe breakdown of trust. Altogether it turned out to be very costly to try to use defections to “flush out” the other player’s weak spots. It turned out to be more profitable to have a policy of cooperation as often as possible, together with a willingness to retaliate swiftly against any attempted undercutting. Note, however, that strategies featuring **massive** retaliation were less successful than TIT FOR TAT with its more gentle policy of **restrained** retaliation. Forgiveness is the key here, for it helps to restore the proverbial “atmosphere of mutual cooperation” (to use the phrase of international diplomacy) after a small skirmish.

“Be nice and forgiving” was in essence the overall lesson of the first tournament. Apparently, though, many people just couldn’t get themselves to believe it, and were convinced that with cleverer trickery and scheming, they could win the day. It took the second tournament to prove them dead wrong. And out of the second tournament, a third key strategic concept emerged: that of **provocability** - the notion that one should “get mad” quickly at defectors, and retaliate. Thus a more general lesson is: “Be nice, provokable, and forgiving.”

Strategies that do well in a wide variety of environments are called by Axelrod **robust**, and it seems that ones with “good personality traits” - that is, nice provokable, and forgiving strategies - are sure to be robust. TIT FOR TAT is by no means the only possible strategy with these traits, but it is the canonical example of such a strategy, and it is astonishingly robust.

## 6. The Ecological Tournament

Perhaps the most vivid demonstrations of TIT FOR TAT’s robustness were provided by various subjunctive replays of the second tournament...

Undoubtedly the most significant and ingenious type of [these] that Axelrod tried was the **ecological tournament**. Such a tournament consists not merely of a single subjunctive replay, but of a whole cascade of hypothetical replays, each one’s environment determined by the results of the previous one. In particular, if you take a program’s score in a tournament as a measure of its “fitness”, and if you interpret “fitness” as meaning “number of progeny in the next generation”, and finally, if you let “next generation” mean “next tournament”, then what you get is that each tournament’s results determine the environment of the next one - and in particular, successful programs become more copious in the next tournament. This type of iterated tournament is called “ecological” because it simulates ecological adaptation (the shifting of a **fixed** set of species’ populations according to their mutually defined and dynamically developing environment), as contrasted with evolution via mutation (where new species can come into existence).

*As one carries out an ecological tournament through generation after generation, the environment gradually changes. In a paraphrase of how Axelrod puts it, here is what happens. At the very beginning, poor programs and good programs alike are equally represented. As time passes, the poorer ones begin to drop out while the good ones flourish. But the rank order of the good ones may now change, because their "goodness" is no longer being measured against the same field of competitors as initially. Thus success breeds ever more success - but only provided that the success derives from interaction with other similarly successful programs. If, by contrast, some program's success is due mostly to its ability to milk "dumber" programs for all their worth, then as those programs are gradually squeezed out of the picture, the exploiter's base of support will be eroded and it will suffer a similar fate.*

*A concrete example of ecological extinction is provided by HARRINGTON, the only non-nice program among the top fifteen finishers in the second tournament. In the first 200 generations of the ecological tournament, while TIT FOR TAT and other successful nice programs were gradually increasing their percentage of the population, HARRINGTON too was increasing its percentage. This was a direct result of HARRINGTON's exploitative strategy. However, by the 200th generation, things began to take a noticeable turn. Weaker programs were beginning to go extinct, which meant fewer and fewer dupes for HARRINGTON to profit from. Soon the trend became apparent: HARRINGTON could not keep up with its nice rivals. By the 1,000th generation, HARRINGTON was as extinct as the dodos it had exploited...*

*Needless to say, TIT FOR TAT fared spectacularly well in the ecological tournament, increasing its lead ever more. After 1,000 generations, not only was TIT FOR TAT ahead, but its rate of growth was greater than that of any other program. This is an almost unbelievable success story, all the more so because of the absurd simplicity of the "hero". One amusing aspect of it is that TIT FOR TAT did not defeat a single one of its rivals in their encounters. This is not a quirk; it is in the nature of TIT FOR TAT. TIT FOR TAT simply **cannot** defeat anyone; the best it can achieve is a tie, and often it loses (though not by much).*

*Axelrod makes this point very clear:*

*TIT FOR TAT won the tournament, not by beating the other player, but by eliciting behavior from the other player which allowed both to do well. TIT FOR TAT was so consistent at eliciting mutually rewarding outcomes that it attained a higher overall score than any other strategy in the tournament....*

*[Axelrod] gives examples from everyday life in which this principle holds. Here is one:*

*A firm that buys from a supplier can expect that a successful relationship will earn profit for the supplier as well as the buyer. There is no point in being envious of the supplier's profit. Any attempt to reduce it through an uncooperative practice, such as by not paying your bills on time, will only encourage the supplier to take retaliatory action. Retaliatory action could take many forms, often without begin explicitly labeled as punishment. It could be less prompt deliveries, lower quality control, less forth-*

*coming attitudes on volume discounts, or less timely news of anticipated market conditions. The retaliation could make the envy quite expensive. Instead of worrying about the relative profits of the seller, the buyer should worry about whether another buying strategy would be better.*

[Hofstadter continues:] *Like a business partner who never cheats anyone, TIT FOR TAT never beats anyone - yet both do very well for themselves....*

*Fortunately, in an environment where there are programs that cooperate (and whose cooperation is based on reciprocity), being unresponsive is a very poor strategy, which in turn means that [ALWAYS DEFECT] is a very poor strategy. The single unresponsive competitor in the second tournament was RANDOM, and it finished next to last. The last-place finisher was responsive, but its behavior was so inscrutable that if **looked** unresponsive....*

*One way to explain TIT FOR TAT's success is simply to say that it **elicits cooperation**, via friendly persuasion. Axelrod spells this out as follows:*

*Part of its success might be that other rules anticipate its presence and are designed to do well with it. Doing well with TIT FOR TAT requires cooperating with it, and this in turn helps TIT FOR TAT. Even rules that were designed to see what they could get away with quickly apologize to TIT FOR TAT. Any rule that tries to take advantage of TIT FOR TAT will simply hurt itself. TIT FOR TAT benefits from its own nonexploitability because three conditions are satisfied:*

- 1. The possibility of encountering TIT FOR TAT is salient;*
- 2. Once encountered, TIT FOR TAT is easy to recognize; and*
- 3. Once recognized, TIT FOR TAT's nonexploitability is easy to appreciate.*

[Hofstadter continues:] *This brings out a fourth "personality trait" (in addition to niceness, provocability, and forgiveness) that may play an important role in success: recognizability, or straightforwardness. Axelrod chooses to call this trait **clarity**, and argues for it with clarity:*

*Too much complexity can appear to be total chaos. If you are using a strategy that appears random, then you also appear to be unresponsive to the other player. If you are unresponsive, then the other player has no incentive to cooperate with you. So being so complex as to be incomprehensible is very dangerous.*

[Hofstadter continues:] *How rife this is with morals for social and political behavior! It is rich food for thought....*

## 7. The Researcher's Lessons

*In his book, Axelrod takes pains to spell out the answers to three fundamental questions concerning the temporal evolution of cooperation in a world of raw egoism. The first concerns **initial viability**: How can cooperation get started in a world of unconditional defection - a "primordial sea" swarming with unresponsive [ALWAYS DEFECT] creatures? The answer (whose proof [Hofstadter] omit[s] here) is that invasion by small clusters of conditionally cooperating organisms, even if they form a tiny minority, is enough to give cooperation a toehold. One cooperator alone will die, but small clusters of cooperators can arrive (via mutation or migration, say) and propagate even in a hostile environment, provided that they are defensive like TIT FOR TAT. Complete pacifists - Quaker-like programs - will **not** survive, however, in this harsh environment.*

*The second fundamental question concerns **robustness**: What type of strategy does well in unpredictable and shifting environments? We have already seen that the answer to this question is: Any strategy possessing the four fundamental "personality traits" of niceness, provocability, forgiveness, and clarity. This means that such strategies, once established, will tend to flourish, especially in an ecologically evolving world.*

*The final question concerns **stability**: Can cooperation protect itself from invasion? Axelrod proved that it can indeed. In fact, there is a gratifying asymmetry to his findings: Although a world of "meanies" (beings using the inflexible [ALWAYS DEFECT] strategy) is penetrable by cooperators in clusters, a world of cooperators is **not** penetrable by meanies, even if they arrive in clusters of any size. Once cooperation has established itself, it is permanent. As Axelrod puts it, "The gear wheels of social evolution have a ratchet."*

*The term "social" here does not mean that these results necessarily apply only to higher animals that can think. Clearly, four-line computer programs do not think - and yet, it is in a world of just such "organisms" that cooperation has been showed to evolve. The only "cognitive abilities" needed by TIT FOR TAT are: (1) recognition of previous partners, and (2) memory of what happened last time with this partner. Even bacteria can do this, by interacting with only one other organism (so that recognition is automatic) and by responding only to the most recent action of their "partner" (so that memory requirements are minimal). The point is that the entities involved can be on the scale of bacteria, small animals, large animals, or nations. There is no need for "reflective rationality"; indeed TIT FOR TAT could be called "reflexive" (in the sense of being as simple as a knee-jerk reflex) rather than "reflective".*

*For people who think that moral behavior toward others can emerge only when there is imposed some totally external and horrendous threat (say, of the fire-and-brimstone sort) or soothing promise of heavenly reward (such as eternal salvation), the results of this research must give pause for thought. In one sentence, Axelrod captures the whole idea: Mutual cooperation can emerge in a world of egoists without central control, by starting with a cluster of individuals who rely on reciprocity.*

## 8. More Lessons and Questions

That ends Hofstadter's article, except for a postscript that, along with some tangents and technical paragraphs, I have not included. I think you can see why I found it provocative. I'd like to close with some reflections of my own.

Axelrod's research can be interpreted as offering an empirical basis for refinements to some of our most cherished moral principles, namely those that cluster under the general category of *The Golden Rule*, which in one form or another is common to religions world wide<sup>7</sup>. The research confirms the economic benefits of *niceness*, that is, the unwillingness to defect, to cheat, to harm another. The Golden Rule, reciprocity, love turn out to be in our own self-interest. But not blind love or reciprocity. The lesson from the research is to love your enemies, but punish their transgressions first, in moderation, and then forgive them.

Now I want to acknowledge that there are those, perhaps some of you here, who think that we cannot learn anything of value from computer simulations such as these. Human beings, these critics might say, are totally unlike these mere computer games; our behavior is much more complex. I myself am more optimistic about the lessons from this research. I'm inclined to think that the research reveals some fundamental properties of all systems, even human systems, in an economically competitive environment.

And for those of you who still doubt, I want to ask: Are the punitive and exploitative strategies you are employing instead really working? Is the world you live in because of those strategies full of love and trust? Or is it one of dysfunction. It has been said that the essence of madness is to continue doing the same thing, thinking that next time the results will be different. Maybe it's time to give love, provokable love, a try.

***This I believe. This I choose.***

***Namaste!***

---

<sup>7</sup> Click on the images in the [Phoenix Circle](http://www.asuuc.org/Home_files/Special/PhoenixCircle/index.html) to see pages showing versions of the Golden Rule for various faiths. ([http://www.asuuc.org/Home\\_files/Special/PhoenixCircle/index.html](http://www.asuuc.org/Home_files/Special/PhoenixCircle/index.html))

## Appendix 1. The Original *Prisoner's Dilemma*

*'In case you're wondering why it is called "Prisoner's Dilemma", here's the reason. Imagine that you and an accomplice (someone you have no feelings for one way or the other) committed a crime, and now you've both been apprehended and thrown in jail; and are fearfully awaiting trials. You are being held in separate cells with no way to communicate. The prosecutor offers each of you the following deal (and informs you both that the identical deal is being offered to each of you – and that you both know that as well!): "We have a lot of circumstantial evidence on you both. So if you both claim innocence, we will convict you anyway and you'll both get two years in jail. But if you will help us out by admitting your guilt and making it easier for us to convict your accomplice – oh, pardon me, your alleged accomplice – why, then, we'll let you out free. And don't worry about revenge – your accomplice will be in for five years! How about it?" Warily, you ask, "But what if we both say we're guilty?" "Ah, well, my friend – I'm afraid you'll both get four-year sentences, then.'*

*'Now you're in a pickle! Clearly, you don't want to claim innocence if your partner has sung, for then you're in for five long years. Better you should both have sung, then you'll only get four. On the other hand, if your partner claims innocence, then the best possible thing for you to do is sing, since you're out scot-free! So at first sight, it seems obvious what you should do: Sing! But what is obvious to you is equally obvious to your opposite number, so now it looks like you both ought to sing, which means Sing Sing for four years! At least that's what logic tells you to do. Funny, since if both of your had just been illogical and maintained innocence, you'd both be in for only half as long! Ah, logic does it again.'*<sup>8</sup>

---

<sup>8</sup> Hofstadter, p. 716.

## Appendix 2. Related Sermons by Jim Fulton

1. *Universalism*,  
[http://www.asuuc.org/Home\\_files/Sermons/2009/0315-Universalism/Universalism.pdf](http://www.asuuc.org/Home_files/Sermons/2009/0315-Universalism/Universalism.pdf)
2. *I Am ...*,  
[http://www.asuuc.org/Home\\_files/Sermons/2009/0920-IAm/I%20Am%20....pdf](http://www.asuuc.org/Home_files/Sermons/2009/0920-IAm/I%20Am%20....pdf)
3. *I Believe - A Universalist Catechism*,  
[http://www.asuuc.org/Home\\_files/Sermons/2009/0927-IBelieve/UniversalistCatechism.pdf](http://www.asuuc.org/Home_files/Sermons/2009/0927-IBelieve/UniversalistCatechism.pdf)
4. *Phoenix Communion*,  
[http://www.asuuc.org/Home\\_files/Sermons/2009/1004-PhoenixCommunion/PhoenixCommunion-0910.pdf](http://www.asuuc.org/Home_files/Sermons/2009/1004-PhoenixCommunion/PhoenixCommunion-0910.pdf)
5. *Sin and the Universalist*,  
[http://www.asuuc.org/Home\\_files/Sermons/2009/0524-Sin&Universalist/Sin&Universalist.pdf](http://www.asuuc.org/Home_files/Sermons/2009/0524-Sin&Universalist/Sin&Universalist.pdf)
6. *On Being Part of Something*,  
[http://www.asuuc.org/Home\\_files/Sermons/2009/1025-PartOfSomething/PartOfSomething.pdf](http://www.asuuc.org/Home_files/Sermons/2009/1025-PartOfSomething/PartOfSomething.pdf)