Journal of Theoretical Biology 299 (2012) 21–24

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Launching “The Evolution of Cooperation”

Robert Axelrod *

Ford School of Public Policy, University of Michigan, Ann Arbor, MI 48104, USA

ARTICLE INFO

Available online 22 April 2011

Keywords:
Cooperation
Prisoner’s dilemma
Reciprocity
Interdisciplinary research
Generosity

ABSTRACT

This article describes three aspects of the author’s early work on the evolution of the cooperation. First, it explains how the idea for a computer tournament for the iterated Prisoner’s Dilemma was inspired by the artificial intelligence research on computer checkers and computer chess. Second, it shows how the vulnerability of simple reciprocity of misunderstanding or misimplementation can be eliminated with the addition of some degree of generosity or contrition. Third, it recounts the unusual collaboration between the author, a political scientist, and William D. Hamilton, an evolutionary biologist.

0. Introduction

This article is a response to three questions, starting with the one I am asked most often.

1. How did you get the idea for a computer tournament for the iterated Prisoner’s Dilemma?
2. After thirty years and nine thousand citations, do you see any problem with your early work on the evolution of cooperation?
3. What was it like for you as a political scientist to work with William Hamilton, one of the world’s leading evolutionary biologists?

1. Origins

Where did the idea for a computer tournament come from? In retrospect, I realize that it came from my interest in artificial intelligence, which started while I was in high school and an interest in game theory that started in college. In high school I came across an article about a checker-playing program that learned to improve its own play (Samuel, 1959). I was fascinated. Afterwards, I followed the development of computer chess through the 1960s, as well as the computer chess tournaments that began in 1970.

As an undergraduate math major in the early 1960s, I had a growing interest in international politics and especially the risk of nuclear war. While studying a standard text on game theory (Luce and Raiffa, 1957), I came across the iterated Prisoner’s Dilemma. To me, the Prisoner’s Dilemma captured the essence of the tension between doing what is good for the individual (a selfish defection) and what is good for everyone (a cooperative choice). In graduate school, while pursuing a PhD in Political Science, I read intriguing research on how human subjects played the game, and how game theorists were still arguing with each other about the best way to play the game (e.g. Rapoport and Chammah, 1965). My main interest in the game was its potential as a source of insights into international conflicts, including arms races and escalation of crises. My motivation for understanding how one should play the game was based on a desire to promote cooperation between players. While I could not have articulated it at the time, my implicit premise was that the understanding of the conditions under which even egoistic players would cooperate with each other could be used to promote cooperation by fostering just those conditions.

The literature on the iterated Prisoner’s Dilemma left me frustrated because there was no clear answer to the question of how to avoid conflict, or even how an individual (or country) should play the game. Apparently, my frustration in graduate school stayed with me while I started thinking about the problem again a dozen years later. Inspired by computer chess I wondered, what a good computer program for playing the iterated Prisoner’s Dilemma would look like? I thought, the way to test such a program would be to try it out on several more or less expert human players. One of the first I recruited was Professor James Coleman, a world-class sociologist and math modeler. Unbeknownst to him, the computer program I arranged for him to play with was TIT FOR TAT. After a few rounds I asked him how he thought he was doing. He said something like, “I don’t know, but I am doing better than the computer, so I guess I’m doing fine.” My immediate reaction was that if someone as smart as he was would use the other player’s score as a benchmark then the Prisoner’s Dilemma really was a pretty good tool for exploring the subtleties of strategic interaction.

After watching a few more people play with computer programs, I realized that there were several limitations in using people – even skilled people – to study how best to play the Prisoner’s Dilemma. One problem was that people get bored if

---

* Tel.: +734 763 0099; fax: +734 761 9181.
E-mail address: axe@umich.edu

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.04.015

they do manage to achieve mutual cooperation for an extended period of time, and then they try a defection or two to see what might happen. A more fundamental problem was that people don't stick to a consistent strategy so it is almost impossible to determine what works well and what doesn't. This led to the realization that the best way to explore effective play in the iterated Prisoner's Dilemma was to have a variety of computer programs play each other. That way I would know exactly what each player was doing and why.

But what programs should I try? I knew that what strategy works well depends, in an important way, on what other players are doing. This knowledge is what which led to the insight that to explore effective strategies I could invite experts to submit programs to a tournament, where each participant would know that the rules his or her program would be facing would be designed by others to do well. Thus each would provide the environment for the others.

Having based my expectations on computer chess, I was surprised that the winner was the simplest of all the strategies submitted, namely TIT FOR TAT. TIT FOR TAT begins with cooperation and then, as the name implies, simply does what the other player did on the previous move. I next recruited a much larger group of participants including computer hobbyists. Once again the winner was TIT FOR TAT. At this point I knew I was on to something. The most fascinating point was that TIT FOR TAT won the tournaments even though it could never do better than the player it was interacting with. Instead it won by its success at eliciting cooperation.

I analyzed how the highly successful strategies managed to elicit cooperation, and was then able to prove several theorems about the stability of TIT FOR TAT once established, and the relative ease with which a small cluster of TIT FOR TAT players can invade a hostile environment (Axelrod, 1981). I next worked on the biological applications with William Hamilton (see Section 3 below). I then decided to write a book aimed at a general audience, especially those interested in cooperation among people. My publisher told me that every formula would halve the readership, so I banished all but one equation to the appendix (Axelrod, 1984).

Since then, I have worked with a number of collaborators to develop the theory and applications from tumor cells (Axelrod et al., 2006) to international politics (Axelrod and Keohane, 1985). There have been over 9000 citations to my work on the evolution of cooperation,[1] many of them extending the basic paradigm in directions I had not even conceived of while doing this work. And the rate of citations for the early work has not yet peaked even after thirty years.

2. Generosity and contrition

I turn now to the second most common question I have been asked, "After thirty years and thousands of citations, do you see any flaws in your early work on the evolution of cooperation?" The question often comes in a more pointed form, "Do you still think simple TIT FOR TAT is the best[2] strategy to use?" The answer to both questions is that the simple reciprocity needs to be slightly modified to take into account the possibility that one's choice will be not be implemented correctly, or that one's action will not be perceived correctly by the other player.

In my original tournament formulation there was no possibility of misimplementation or misunderstanding—two sources of what is usually called "noise." Yet, some degree of noise is typical of most strategic interactions. In the context of the iterated Prisoner's Dilemma, noise can easily cause unintended defections. For example, if two players are both using the TIT FOR TAT strategy, even one misunderstanding can echo indefinitely. Molander (1985) showed that in the presence of even small amounts of noise, two such players would get the same average payoffs as two interacting RANDOM players.

I wish I had appreciated this basic flaw in simple reciprocity earlier. I did not entirely ignore the problem of noise, but I had introduced it in the tournament at the level of the strategy rather than at the level of the choice. In other words, I told the entrants that one of the strategies would be random, but I didn't provide for the possibility that a choice by one player would occasionally be misreported to the other.

I did have an intuitive understanding of the problem that a misunderstanding can cause. In fact, when I developed the policy implications of my work I explicitly stated that "in many circumstances the stability of cooperation would be enhanced if the response were slightly less than the provocation" (Axelrod, 1984, p. 187).

The simplest way to correct for the possibility of noise in the iterated Prisoner's Dilemma is to allow some percentage of defections by the other player to go unpunished. This is the strategy of Generous TIT FOR TAT. Generous TIT FOR TAT has indeed been found to be effective for dealing with misperception (e.g., Molander, 1985; Bendor et al., 1991; Godray, 1992; and Nowak and Sigmund, 1992).

There is another source of noise besides misperception, namely misimplementation. With misimplementation a player realizes that its intended choice was not the choice it executed. In that case, a player may choose not to respond to the defection by the other player which its own error evoked. This is a strategy called Contrite TIT FOR TAT (Sugden, 1986; Boyd, 1989; Boerlijst et al., 1997; Rand et al., 2009).[3]

Both the generous and contrite variants of TIT FOR TAT perform well when noise is added to the heterogeneous environment of the 63 rules of the second round of the Prisoner's Dilemma tournament (Wu and Axelrod, 1995). Contrite TIT FOR TAT does especially well when replicator dynamics are used to simulate future rounds of the tournament in which the rules that are less successful in dealing with noise are displaced by rules that are more successful.[4]

---

[1] The source is Science Citations Index. The breakdown is roughly 5,000 citations for the book *Evolution of Cooperation* (Axelrod, 1984), 2000 citations for article by Axelrod and Hamilton (1981), and 3000 citations for my other work of cooperation such as *The Complexity of Cooperation* (Axelrod, 1997) and work on cooperation among tumor cells (Axelrod et al., 2006). I owe a great debt to my collaborators: David E. Axelrod, Michael D. Cohen, Douglas Dion, Stephanie Forrest, Alan Grafen, John Holland, William D. Hamilton, Ross A. Hammond, Geoffrey Hodgson, Robert Keohane, Kenneth Pienta, Rick L. RIolo, and Jianzhong Wu.

[2] Of course, there is no one "best" strategy for the iterated Prisoner's Dilemma. For example, if a player cares little about future payoffs relative to the current payoff, it is best to always defect. Likewise, if the other player is unlikely to be responsive to your choices, the player should always defect. But if neither of these conditions holds, a player might do well by allowing for the possibility that mutual cooperation could be attained and will result in a high payoff to an egoistic player. See Axelrod (1984, p. 14f, 38–40, 176f, and 206–12).

[3] Contrite TIT FOR TAT is defined in terms of its three states: *contrite, content, and provoked*. It begins in *content* with cooperation and stays there unless there is a unilateral defection. If it was the victim while *content*, it becomes *provoked* until a cooperation by the other player causes it to become *content*. If it was the defector while *content*, it becomes *contrite* and cooperates. When *contrite*, it becomes *content* only after it has successfully cooperated.

[4] Another rule that has received attention is known under several names: Simpleton (Rapoport and Chammah, 1965), Pavlov, or Win Stay Lose Shift (Nowak and Sigmund, 1993). This rule starts with cooperation. From then on, it changes its choice if and only if it received one of the two lowest payoffs (i.e. the other player defected). Pavlov has the virtue of being evolutionarily stable if players do strictly better by cooperating than by alternating one exploiting the other (Fudenberg and

Thus, in the presence of noise, reciprocity still works well provided it is accompanied either by generosity (some chance of cooperating when one would otherwise defect) or contrition (cooperating after the other player defects in response to one's own defection). I wish I had understood that from the beginning.

3. Interdisciplinary collaboration[5]

My collaboration with Bill Hamilton is an unusual story. Although I am a political scientist by training, I have long been interested in evolutionary theory. But when I wanted to write about the evolutionary implications of my work on the Prisoner's Dilemma, I knew I was in over my head. I wrote to an entrant in one of my tournaments, Richard Dawkins. He pointed me to another evolutionary biologist, William Hamilton, who happened to be at my own university. I already knew of Hamilton's influential theory of inclusive fitness, so I gave him a call.

In his memoirs (Hamilton 2002), Bill describes his reactions to this phone call.

> One day in the Museum of Zoology at Ann Arbor there came a phone call from a stranger asking what I knew about evolutionarily stable strategies and for some guidance to relevant literature. (p. 118) …Now on the phone to me was someone out of political science who seemed to have just the sort of idea I needed. A live games theorist was here on my own campus! Nervously, and rather the way a naturalist might hope to see his first mountain lion in the woods, I had long yearned for and dreaded an encounter with a games theorist. How did they think? What were their dens full of? …Axelrod on the phone sounded nice and, very surprising to me, he was more than a bit biological in his manner of thinking. I sensed at once a possibility that the real games theorists might be going to turn out to be a kind of kindred to us [biologists]. (p. 120).

Had Bill known of my long-standing interest in evolutionary theory, he might not have been quite so surprised that my thinking was more than a bit biological. For example, in high school I wrote a computer simulation to study hypothetical life forms and environments. This early interest in evolution was nurtured during college by a summer at the University of Chicago's Committee on Mathematical Biology.

That first phone call led to a lunch where he suggested that we work together.

> Soon after the lunch again I proposed that the work seemed so interesting biologically we might try writing it up for a joint paper in *Science*; [Axelrod's] contribution would be the basic ideas plus the description of his tournaments, and mine to add a natural scientist's style and some biological illustrations. (p. 122).

I was delighted to accept Bill's invitation to collaborate. Despite coming from different disciplines, Bill and I shared not only mathematical training, a love of formal modeling. Bill had even published one paper using the Prisoner's Dilemma, although he was hoping to get away from that when I dragged him back.

Bill's proposed division of labor turned out to be a good description of how the collaboration developed. I gradually realized, however, just how much was included by Bill's modest formulation of adding "a natural scientist's style and some biological illustrations." Bill's naturalist's style included having at his fingertips an astonishing knowledge of species from bacteria to primates. His experience as a naturalist often gave him the capacity to check out the plausibility of an idea with pertinent examples right off the top of his head. It also helped him to generate surprising new ideas.

Here is how Bill saw us working together.

> That brilliant cartoonist of the journal *American Scientist*, Sidney Harris, has a picture where a mathematician covers the blackboard with an outpouring of his formal demonstration. … [I]t starts top left on the blackboard and ends bottom right with a triumphant 'QED'. Halfway down, though, one sees a gap in the stream where is written in plain English: 'Then a miracle occurs', after which the mathematical argument goes on. Chalk still in his hand, the author of this *quod est demonstrandum* now stands back and watches with a cold dislike an elderly mathematician who peers at the words in the gap and says: 'But I think you need to be a bit more explicit-here in step two.' I easily imagine myself to be that enthusiast with the chalk and I also think of many castings for the elderly critic. Yet how easy it is to imagine a third figure-Bob-in the background of the picture, saying cheerfully: 'But maybe he has something all the same, maybe that piece can be fixed up. What if….' (p. 123)

I shared Bill's surprise at how well we worked together. As he put it,

> I would have thought it a leg-pull at the time if someone had told me of a future when I would find it more rewarding to talk 'patterns' to political scientists rather than to fellow biologists. (p. 126)

Perhaps the most important thing we shared was our aesthetic sense.

> [A]n intuitive understanding between us was immediate. Both of us always liked to be always understanding new things and to be listening more than talking; both of us had little inclination for the social manoeuvring, all the 'who should-bow-lowest' stuff, which so often wastes time and adrenalin as new social intercourse starts. Bob is the more logical, but beyond this what we certainly share strongly is a sense for a hard-to-define aesthetic grace that may lurk in a proposition, that which makes one want to believe it before any proof and in the midst a confusion and even antagonism of details. Such grace in an idea seems often to mean that it is right. Rather as I have a quasi-professional artist as my maternal grandmother, Bob has one closer to him-his father. Such forebears perhaps give to both of us the streak that judges claims not in isolation but rather by the shapes that may come to be formed from their interlock, rather as brush strokes in a painting, shapeless or even misplaced considered individually, are overlooked as they join to create a whole…. (p. 122).

I see a further connection between art and modeling. My father painted to express how he saw the world that day, highlighting what was important to him by leaving out what was not. Likewise, I see my modeling as an expression of how I see some

---

(*footnote continued*)
Maskin, 1990). Moreover, the rule did well in a simulation that did not take account of discounting of payoffs over time (Nowak and Sigmund, 1993). However, in most biological applications, the discounting of payoffs is appropriate for two reasons: players tend to value payoffs less as the time of their attainment recedes into the future, and these is always some chance that the players will not meet again due to mobility or death (Axelrod, 1984, 12). Moreover, neither the basic nor the generous version of the strategy did well when noise was added to the heterogeneous environment of the second round of he Prisoner's Dilemma Tournament (Wu and Axelrod, 1995). In fact, with the usual payoff parameters it is easily exploited by a strategy that alternates between cooperation and defection (Rapoport and Chammah, 1965).

[5] This section is an adaptation of material from Axelrod (2006). Reprinted by permission.

social or biological dynamic, highlighting what I regard as important, and leaving out everything else.

Bill's disciplinary training as an evolutionary biologist and a naturalist proved essential to make our theoretical work compelling to biologists. He was adept at identifying pertinent biological examples so that biologists could see what we were talking about. While not all of his proposed applications have been borne out, he was able to demonstrate the potential relevance of computer tournaments for the major biological puzzle of why individuals cooperate with unrelated others. He was also able to explain what our contribution added to what was already understood about evolution. Specifically, he showed how our modeling work provides a solid foundation for many of the insights about altruism formulated years earlier by Robert Trivers (1971). Bill was also able to show how our model could be used by other evolutionary biologists to formulate and test new hypotheses about animal behavior, as well as explore dozens of variants of the simple iterated Prisoner's Dilemma.

When Hamilton sent Trivers a copy of our paper, Trivers wrote back that "my heart soared." He later wrote that, "For one wild moment, I kidded him, I actually believed there was progress in science!" (Trivers 2002, p. 53)[6,7]

Acknowledgments

For their editorial help, I thank David Axelrod, Vera Axelrod, Kenneth Pienta and Amy Saldinger.

References

Axelrod, R., 1981. Emergence of cooperation among egoists. American Political Science Review 75, 18–306.

Axelrod, R., 1984. Evolution of cooperation. Basic Books, New York.

Axelrod, R., 1997. The complexity of cooperation. Princeton University Press, Princeton, NJ.

Axelrod, R., 2006. Agent-based modeling as a bridge between disciplines. In: Leigh Tesfatsion, Judd, Kenneth (Eds.), Handbook of Computational Economics, vol. 2: Agent-Based Computational Economics. North-Holland, New York, pp. 84–1565.

Axelrod, R., 2008. Political science and beyond: presidential address to the American Political Science Association. Perspectives on Politics 6, 3–9.

Axelrod, R., Axelrod, D.E., Pienta, K.J., 2006. Evolution of cooperation among tumor cells. Proceedings of the National Academy of Sciences 103, 13474–13479.

Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. Science 221, 1390–1396.

Axelrod, R., Keohane, Robert O, 1985. Achieving cooperation under anarchy: strategies and institutions. World Politics 38, 54–336.

Bendor, J., Kramer, R.M., Stout, S., 1991. When in doubt… cooperation in a noisy prisoner's dilemma. Journal of Conflict Resolution 35, 691–719.

Boerlijst, M.C., Nowak, M.A., Sigmund, K., 1997. The logic of contrition. Journal of Theoretical Biology 185, 281–293.

Boyd, R., 1989. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. Journal of Theoretical Biology 136, 47–56.

Fudenberg, D., Maskin, E., 1990. Evolution and cooperation in noisy repeated games. American Economic Review Papers and Proceedings 80, 274–279.

Godray, H.C.J., 1992. The evolution of forgiveness. Nature 355, 206–207.

Hamilton, W.D., 2002. Narrow Roads of Gene Land: The Collected Papers of W.D. Hamilton, vol. 2, The Evolution of Sex. Oxford U Press, Oxford.

Hamilton, W.D., Axelrod, R., Tanese, R., 1990. Sexual reproduction as an adaptation to resist parasites (a review). Proceedings of the National Academy of Sciences 87, 3566–3573.

Hurst, L.D., Peck, J.R., 1996. Recent advances in understanding of the evolution and maintenance of sex. Trends in Ecology & Evolution 11, 46–52.

Lively, C.M., 2010. A review of Red Queen models for the persistence of obligate sexual reproduction. Journal of Heredity 101, S13–S20.

Luce., R.D., Raiffa, H., 1957. Games and Decisions. Wiley, New York.

Molander, P., 1985. The optimal level of generosity in a selfish, uncertain environment. Journal of Conflict Resolution 29, 611–618.

Nowak, M.A., Sigmund, K., 1992. Tit for tat in heterogeneous populations. Nature 355, 250–253.

Nowak, M., Sigmund, K., 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. Nature 364, 56–58.

Rand, D., Hisashi, G.O., Nowak, M.A., 2009. Direct reciprocity with costly punishment: generous tit-for-tat prevails. Journal of Theoretical Biology 256, 45–57.

Rapoport, A., Chammah, A.M., 1965. Prisoner's Dilemma: A study in Conflict and Cooperation. University of Michigan Press, Ann Arbor.

Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development 3, 211–299.

Sugden, R., 1986. The Economics of Rights, Co-operation and Welfare. Basil Blackwell, Oxford.

Trivers, R.L., 1971. The evolution of reciprocal altruism. Quarterly Review of Biology 46, 35–57.

Trivers, R.L., 2002. Natural Selection and Social Theory: Selected Papers of Robert Trivers. Oxford University Press, Oxford.

Wu, J., Axelrod, R., 1995. How to cope with noise in the iterated prisoner's dilemma. Journal of Conflict Resolution 39, 183–189.

---

[6] Bill Hamilton and I also collaborated on a proof of principle for his theory that sexual reproduction could be an adaptation to resist parasites (Hamilton et al., 1990). For a description of that collaboration see Axelrod (2006, p. 1578–82). For an assessment of the evidence on this and competing theories of sex, see Hurst and Peck (1996) and Lively (2010). In the end, Bill Hamilton gave his life to science. In 2000, despite the risks, he went to the jungles of central Africa to gather evidence needed to test a theory about the origin of AIDS. He contracted a virulent form of malaria that proved fatal.

[7] I later collaborated with an oncologist and a geneticist to explore the idea that cancer involves the cooperation premalignant tumor cells (Axelrod et al., 2006). For the story of this collaboration and some lessons on doing interdisciplinary research, see Axelrod (2008).