Message – Cincinnati Friends Meeting
April 18, 2010

"The Prisoner's Dilemma"

by Donne Hayden

This week is the local NPR station's pledge drive week, WVXU 91.7 FM. I am a regular NPR listener—I was once devoted, now am a bit disgruntled at the extent of negative news coverage, although I am still inclined to be faithful. For years I have pledged support of my local NPR stations, wherever I happened to be. Last year, short of cash because of my many moves, I missed the fall pledge drive and haven't contributed since last spring. So my conscience is bothering me, and I'm going to have to call in. If you've ever listened to a pledge drive for public radio or television, you know that the only about 20% of those who listen to or watch broadcasts support the stations financially. In other words, 80% of those who use these services do not contribute to their financial support. To put it in the language of game theory, 20% are "cooperators," while 80% are "defectors." "Cooperators" respond to their consciences; defectors focus on what's good for them alone, procrastinating or figuring someone else will take care of NPR.

I'm thinking in terms of game theory, not because it comes naturally to me—it's quite alien to me—but because of a book I re-discovered in my most recent move. I culled a great many books from my shelves and donated them to the library, but one book that I decided to keep was *Metamagical Themas: Questing for the Essence of Mind and Pattern* by Douglas R. Hofstader. I acquired the book around 1985 when I was a member of one of those book clubs that automatically sent you a book each month unless you told them not to. Once when I forgot to tell them not to send that month's selection, I received *Metamagical Themas,* and, intrigued

by the title, kept the book. Though I moved it from place to place with me for the past 25 years, until this week, I hadn't actually read any of it. The 800-page book, a collection of the columns Hofstader wrote for the *Scientific American* from January 1981 to July 1983, covers a diversity of topics, ranging from, as he explains in his introduction: "sexism to music to art to nonsense, from game theory to artificial intelligence to molecular biology to the Cube, and more."[1]

One section of the book that gave me food for thought this week was called "The Prisoner's Dilemma and the Evolution of Cooperation." According to Hofstader, "The Prisoner's Dilemma" is a "very lifelike paradox" presented as a problem in logic, "discovered" in 1950 by Melvin Dresher and Merrill Flood of the RAND Corporation. No doubt you've all seen some version of it in police dramas on television. Here it is:

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 hat 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 go 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?"

---

[1] Douglas R. Hofstader, *Metamagical Schemas: Questing for the Essence of Mind and Pattern* (New York: Basic Books, Inc. Publishers, 1985), p. xxiv.

"Ah well, my friend—I'm afraid you'll both get four-year sentences then."[2]

So you see the dilemma. What would *you* do? Would you confess—defect, leave your accomplice twisting in the wind? Or maintain your claim to innocence hoping the other person will do the same so you can each spend only two years in jail? The latter would be an example of cooperation—i.e., working together for the good of both parties. *Pure logic*, however, tells you the best option is to admit your guilt, for then, if the other person maintains his innocence, you will go free; if he also confesses, however, you'll both go to jail for four years instead of five. Either way, if you confess, you avoid going to prison for five years. On a purely egocentric level, confessing, or "defecting" in game theory language, will get you either freedom or one-year less on your sentence. Maintaining your innocence, of course, would be best, IF you trusted the other person to do the same. But how many of us would trust that other person?

Game theorists spend years studying this paradox; they hold tournaments for the best strategies for dealing with this problem. They have elaborated on the problem, turning it into a situation in which the choice to cooperate or defect is repeated over and over again, which complicates the issue because the relationship is extended over time. Hofstader relates the dilemma to real-life situations such as this one:

Suppose I take my car in to get the oil changed. I know little about auto mechanics, so when I come in to pick it up, I really have no way to verify if they've done the job. For all I know, it's been sitting untouched in their parking lot all day, and as I drive off, they may be snickering

[2] Hofstader, p. 716.

behind my back. On the other hand, maybe I've got the last laugh, for how do they know if the check I gave them will bounce?[3]

Chances are, if you know you will see those people again, you will be inclined to be more cooperative. "On the other hand," Hofstader continues,

suppose I'm on my way across the country and have some radiator trouble near Gillette, Wyoming, and stop in town to get my radiator repaired there.. . . I'll probably never again need the services of this garage, and they'll never get another check from me. In the most crude sense, then, it's not in my interest to give them a good check, nor is it in theirs to fix my car. But do I really defect? Do I give out bad checks? No. Why not?

Consider this related situation. Late at night, I bang into someone's car in a deserted parking lot. I have the choice of leaving a note, telling the owner who's to blame, or scurrying off scot-free. Which do I do?[4]

This sort of dilemma is around us all the time. Do we cooperate or defect? Again, *pure logic* leads to the conclusion that your best option is to defect, to look out for Number 1. However, something unexpected happened in the first "tournament" based on this dilemma, in which fourteen strategists were invited to submit solutions to the problem. Hofstader says, "Even expert strategists from political science, sociology, economics, psychology, and mathematics made the systematic errors of being too competitive for their own good, not forgiving enough, and too pessimistic about the responsiveness of the other side."[5] Contrary to expectations, to "win" in a situation that involves a long-term relationship the most effective

[3] p. 731.
[4] *Ibid.*
[5] p. 722-723.

strategy is what game theorists call "TIT FOR TAT." In this strategy, your first move is to cooperate and thereafter, you do whatever your partner/opponent did the last time; in other words, you mirror what your opponent does. Even though TIT FOR TAT *does* defect when the opponent does, if the opponent then has a change of heart and cooperates, TIT FOR TAT responds in kind. Surprisingly, this strategy "won the tournament, not by beating the other player, but by eliciting behavior from the other player which allowed both to do well."[6] Hofstader summarizes the success of the very simple strategy and the "lesson of the first tournament" this way: "it is important to be nice ('don't be the first to defect') and forgiving ('don't hold a grudge after you've vented your anger').[7]

This was too simple. Mathematicians and game theorists couldn't believe such a straightforward and simple strategy was the best response. When TIT FOR TAT also won the second tournament, Hofstader comments:

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, provocable, and forgiving."[8]

As I was searching for information on game theory, I was stunned to find the website of an avowed *Christian* Game Theorist who explains how well the Prisoner's Dilemma fits with Christian morality. He says:

---

[6] p. 727.
[7] p. 723.
[8] p. 725.

According to the payoff matrix for this game, the best policy for minimizing losses is to testify against the other criminal. Since both criminals will realize this, the natural outcome is for both criminals to testify, thereby resulting in them both being punished for breaking the law. ***This outcome is in agreement with traditional Christian morality which teaches that we should not commit crimes and that we should tell the truth.*** (Emphasis mine)

And he left it there.

Did I *miss* something in Christianity?

What about Luke 6:27-31 where Jesus says,

"*Love your enemies . . . do good to those who hate you, [28]bless those who curse you, pray for those who abuse you. [29]If anyone strikes you on the cheek, offer the other also; and from anyone who takes away your coat do not withhold even your shirt. [30]Give to everyone who begs from you; and if anyone takes away your goods, do not ask for them again. [31]Do to others as you would have them do to you.*

What about Matthew 18:21-22:

[21] Then Peter came and said to him, 'Lord, if another member of the church sins against me, how often should I forgive? As many as seven times?' [22]Jesus said to him, 'Not seven times, but, I tell you, seventy-seven times.

The one thing we find so hard in the teachings of Jesus is his insistence on repeated forgiveness, and on eliminating the *provocable s*tage in the relationship. TIT FOR TAT says, "Do unto others *as* they do unto you." But Jesus was a game-changer. He said, "Do unto others as you *would have them* do unto you."

Now here is a paragraph from the book that stopped me. Remarking that application of the TIT FOR TAT strategy "can be on the scale of bacteria, small animals, large animals, or nations," Hofstader summarizes how cooperation can occur "in a world of raw egoism" and "unconditional defection."

*. . . 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 of migration, say) and propagate even in a hostile environment, provided they are defensive like TIT FOR TAT. Complete pacifists—Quaker-like programs—will not survive, however, in this harsh environmen*t.[9]

Perhaps this is true. But I have a sneaking suspicion that occasionally, the Great Game Master introduces a wild card into the game of life. One Jesus, one George Fox, one Gandhi, or Mother Teresa changes the game. Down through the millennia, individuals have come along who have required us humans re-consider how we play the game. Perhaps 20% of us at any given time are playing by different rules than those of logic and rationality or TIT FOR TAT. Rules that recharge an impulse to cooperate with others, to sacrifice selfish desires and to work for the greater good. May it be so. And may I find myself in that 20%.

---

[9] p. 729. Conscience leads us to be cooperators, but one game theory expert claims that "conscience is self-eliminating." Hofstader writes:

Be nice as long as the other is nice; retaliate when the other crosses you, but forgive when the other reaches out. Apparently leaving out the "provocable" step, however, makes cooperators at risk for extinction. In a later column, Hofstader returns to the issue of cooperating or defecting, quoting Garrett Hardin's article about over-population in which he says that "cooperators in the birth control game will breed themselves right out of the population."