Kyle Cox

Professor McPartland

PHIL 126

20 May 2019

A Simulated Analysis of the Iterated Prisoner's Dilemma Inspired by Axelrod's Tournament

Background

Let us begin by discussing a different, but related, paradox. Newcomb's paradox presents us with a choice:

> There are two boxes before you, A and B. You may either open both boxes, or else just open B. You may keep what is inside any box you open, but you may not keep what is inside any box you do not open. The background is this.
>
> A very powerful being, who has been invariably accurate in his predictions about your behavior in the past, has already acted in the following way:
>
> He has put $1,000 in box A. If he has predicted that you will open just box B, he has in addition put $1,000,000 in box B. If he has predicted that you will open both boxes, he has put nothing in box B. (Sainsbury 69)

The matrix for the given outcomes is presented in Figure 1.

Figure 1. Newcomb Matrix.

		Genie Prediction	
		Box A and B	Box B
Your Choice	Box A and B	$1,000	$1,001,000
	Box B	$0	$1,000,000

Let us assume here that money is a measure of utility. Therefore, the question is which decision has the maximum expected utility (MEU). Taking money as a measure of utility allows us to compare Newcomb's Paradox to the Prisoner's Dilemma more accurately.

The crux of this paradox is a problem in causality. Consider that the powerful being—I prefer to call them a "genie"—is invariably accurate. It is reasonable to believe that, because the genie will not be incorrect, if you choose only Box B, you will win a million dollars, and if you choose both Boxes, you will win a thousand dollars. But consider what this implies about our logic: It supposes our choice impacts the decision of the genie (which already happened). That is, it implies a sort of backwards-causality that paradoxically seems both problematic and reasonable.

Consider another scenario: You have a friend, standing on the other side of the Boxes, who can see into them. You cannot. The friend will then advise you which decision to make: Box A, or both Boxes. This friend will invariably tell you to choose both Boxes. The approach I outlined in the previous paragraph I will call Strategy 1, and the approach from this paragraph I will call Strategy 2.

Strategy 1 supposes that a match in decisions is most likely: If you choose to take both Boxes, the genie will have predicted such, and if you choose to take only Box B, the genie will also have predicted such. If a match guaranteed, then our MEU matrix is reduced to two options: Take Box B and win one million dollars, or take both Boxes and win one thousand dollars. The Newcomb Matrix is reduced to the following:

Figure 2. Newcomb Strategy 1.

		Genie Prediction	
		Box A and B	Box B
Your Choice	Box A and B	$1,000	
	Box B		$1,000,000

Strategy 2 supposes that whatever decision the genie has made is in the past, and we should entirely ignore our beliefs about what the genie may have predicted we would do because our decision now cannot impact his past prediction. Then, disregarding which choice the genie made, our decision to choose both boxes yields one thousand more dollars each time. This is described by the following matrix for net gain:

Figure 3. Newcomb Strategy 2.

Your Choice	Box A and B	+$1,000
	Box B	+$0

The distinction between these two strategies and their respective matrices is central to the Prisoner's Dilemma and our evolutionary investigation of it. Strategy 1 is to assume you will match with your partner, and Strategy 2 is to disregard what your partner may have done and choose what will have the greater net utility.

The Prisoner's Dilemma goes as follows: You and friend have been arrested for some illicit deed. You are held in separate rooms, and you are given the following information, as Adapted from Sainsbury:

1. If we both remain silent, … we would then each get a year in jail.
2. If we both confess, we shall both get [three] years in jail.
3. If one remains silent and the other confesses, the one who confesses will get off scot-free (for turning state's evidence), and the other will go to jail for [five] years.
4. The other prisoner is also being told all of (1)−(4).
5. Each prisoner is concerned only with getting the smallest sentence for himself.
6. Neither has any information about the likely behavior of the other, except that (5) holds of him and that he is a rational agent. (Sainsbury 83)

The matrix we used for our Prisoner's Dilemma simulation is the following:

Figure 4. Prisoner's Dilemma Matrix.

		Partner Choice	
		Cooperate	Defect
Your Choice	Cooperate	(3, 3)	(0, 5)
	Defect	(5, 0)	(1, 1)

The first element of the tuple contained in each quadrant gives your payoff, and the second gives your partner payoff. Numbers are positive and represent a utility value, though ordinarily they would be described by number of years sentenced to prison, in which case a lower value would be more desirable.

The numbers used are mostly arbitrary. We can also present the matrix by the following,

Figure 5. Abstract Prisoner's Dilemma Matrix.

		Partner Choice	
		Cooperate	Defect
Your Choice	Cooperate	a	b
	Defect	c	d

(Kuhn)

where the variables represent your payoff (the first element in the tuple in Figure 4). All that matters here is that $c > a > d > b$.

The two strategies from Newcomb's Paradox provide us an interesting framework to analyze the Prisoner's Dilemma. The particularly paradoxical part is that if both parties do act rationally, then the result is that they have a worse utility yield than if they had simply cooperated. This is the consequence of Strategy 2, which is the most reasonable strategy for the Prisoner's Dilemma. The essence of the Dilemma is that you *cannot* know what the other partner is going to choose. This is a tenet that is left ambiguous in Newcomb's Paradox, in which there seems to be some way to infer what the genie is going to predict We do not have this luxury in the Prisoner's Dilemma. Therefore, we are left to adopt Strategy 2, and assume that we cannot know what the other is going to choose.

From here we conclude that it is reasonable to defect always. Defecting yields a better payout than either cooperating in both the event that the opponent cooperates (5 vs. 3) and the event that the opponent defects (1 vs. 0). This corresponds to Figure 3, which describes Strategy 2 for Newcomb's Paradox.

This is where the paradox that we previously mentioned occurs. Strategy 2 is the rational strategy for both players. Consequently, both players will defect. So both players will yield 1, whereas if they had simply cooperated they would have each yielded 3. This is essentially the end of the Prisoner's Dilemma: the harsh reality that rationality does not yield the optimal results. However, had the players adopted Strategy 1 (assuming a match), they would have yielded a higher utility. We wonder, then, if there are circumstances under which rational players would be inclined to mutually cooperate, rather than defect.

Here, Axelrod proposes an *iterated* Prisoner's Dilemma. It is important that the game be played for an *unknown*, finite amount of times, for if the number of games to be played is known, players will ultimately defect every iteration. Axelrod explains,

> If the game is played a known finite number of times, the players still have no incentive to cooperate. This is certainly true on the last move since there is no future to influence. On the next-to-last move neither player will have an incentive to cooperate since they can both anticipate a defection by the other player on the very last move. Such a line of reasoning implies that the game will unravel all the way back to mutual defection. (Axelrod 10).

Our Program

Using a computer simulation, we are able to define a finite number of games without the player "knowing." Therefore, there is no reason to defect every time. We explored six different heuristics. Three of these I call *dumb* heuristics for their simplicity:

1. Defector: Defects every game.
2. Cooperator: Cooperates every time.
3. Random: Chooses randomly between cooperating and defecting. Random has a parameter value, *c-strength*, which takes a value 0-100 that is a percentage likelihood of

choosing to cooperate. For example, a Random player with *c-strength* = 75 will cooperate 75% of the time. The default *c-strength* value is 50.

Our final three heuristics explored a more human, psychological strategy. We considered the possibility that players would develop a “reputation,” and certain players would use this reputation to inform their decisions. This reputation was represented in the code as a list containing all of a player’s past moves. The three *smart* heuristics are as follows:

4. Tit-For-Tat: This was the heuristic that performed best in Axelrod’s studies. It is also simple in nature. Tit-For-Tat cooperates on the first game, then mimics its opponents last move for the following rounds.
5. Random Reputation: Randomly selects an element from the opponent’s list of moves (uniform distribution) and uses it as its move. This has the effect of mimicking an opponent’s strategy over the course of the entire match. For example, if an opponent has defected 75 percent of the time, Random Reputation will do the same. Random Reputation also cooperates first round.
6. Majority Reputation: Selects as its move the move its opponent has made the majority of the time. If the opponent has chosen to defect and cooperate an equal number of times (including the first round, with zero played moves), Majority Reputation cooperates by default.

The reason I discussed Newcomb’s Paradox in depth in the background is because it poses a dilemma: Should we employ Strategy 1 (assume a match is most likely) or Strategy 2 (assume we cannot know the opponent’s move)? In the Prisoner’s Dilemma, we observe that, though it is rational to employ Strategy 2, it ultimately leads to an inferior yield (for both parties) compared to mutual collaboration. The next step is then to consider how Strategy 1 may be implemented in the Prisoner’s Dilemma, and this iterated version of the game, where certain heuristics have different methods of predicting their opponent’s move in an effort to match their move, is our attempt to do that. Further, we are able to explore which methods of opponent prediction (Tit-For-Tat, Majority Reputation, or Random Reputation) are most effective.

We created ten players of each heuristic (for a total of 60) and had them each play each other in iterated Prisoner’s Dilemma games of ten rounds each. Scores were assigned according to the matrix in Figure 4. At the end of this tournament, sums of scores for each heuristic were calculated. Then, maintaining a total population of 60, the number of players granted to each heuristic was redistributed according to their success. 10 random players were added to the population by default. This we called one evolution. The data below presents the redistributed population after the evolution. Random populations are excluded because they were always equal to ten. I performed twenty different trials.

Figure 6. One evolution.

	Cooperator	Defector	T4T	MajRep	RandRep
Trial 1	10	11	10	11	11
2	12	7	11	11	10
3	7	11	12	10	10
4	10	8	12	9	12
5	9	7	10	10	11
6	9	9	9	9	11
7	10	9	9	12	11
8	9	12	9	10	10
9	9	9	11	11	11
10	9	8	10	11	10
11	11	10	10	10	11
12	11	7	12	9	10
13	12	6	11	10	11
14	11	7	9	11	12
15	12	7	9	11	11
16	7	11	11	10	11
17	12	7	10	11	11

18	11	8	11	11	9
19	10	8	11	10	11
20	8	10	12	10	10
Avg.	9.95	8.6	10.45	10.35	10.7

We observe a number of things. First, the Cooperator performs more successfully than the Defector, but does not exceed the average population for each heuristic. This is likely because, while the Cooperator does maximize its score with the *smart* heuristics—MajRep, RandRep, and T4T—it will lose to Defectors and Random players, and thus it is more likely to decline in population over time.

The most important observation is that all of the smart heuristics are successful in two ways: (1) their population increases, and (2) they defeat (lessen) the defecting population. If we were to iterate this evolution (which unfortunately we were not able to do in code), we might extrapolate that the defecting population will approach zero as the evolutions approach infinity. Further, if the smart heuristics are left to compete among themselves (after the elimination of the Randoms and Defectors), they will all converge to cooperation and effectively maximize their collective utility. Observe the following data. Figure 7 presents a table for each of the smart heuristics. The tables contain the sum of their scores against each opponent after a 20-round iterated Prisoner’s Dilemma.

Figure 7. Smart Heuristic General Game Performance.

	Defect	Coop	Rand(c=50)	T4T	RandRep	MajRep
T4T	19	60	49	60	60	60
Opp	24	60	49	60	60	60
Sum	43	120	98	120	120	120

	Defect	Coop	Rand(c=50)	T4T	RandRep	MajRep
RandRep	19	60	49	60	60	60
Opp	24	60	64	60	60	60
Sum	43	120	113	120	120	120

	Defect	Coop	Rand(c=50)	T4T	RandRep	MajRep
MajRep	19	60	42	60	60	60
Opp	24	60	72	60	60	60
Sum	43	120	114	120	120	120

What we see here is that the smart players all compare equally in games against the dumb players and each other (differences can be attributed to randomness). While their point total may be marginally less than the Defectors (it is impossible to beat a defector), they succeed in lowering the Defectors total yield compared to other heuristics. A Cooperator, by contrast, would grant the Defector the highest possible yield in the game. In an evolutionary scenario, then, while smart players may lose individual games, they are successful in driving out Defectors and converging the population toward mutual cooperation. This was in fact the question of interest to Axelrod: Under what situation does a population converge to cooperation, when, on an individual level, it is not the most rational? Further data explore this question further. We can compare how the smart players perform against Random players with different *c-strengths* (likelihoods of cooperating), to see whether these heuristics do truly reward players who are more likely to cooperate, and thus converge the population toward cooperation under evolution.

The following data describe how the smart heuristics compete against random players with 20%, 40%, 60%, and 80% probabilities of cooperating. They played iterated games of 2000 rounds, and the final sum was divided by 100 to compare to previous data. More rounds were played with the random groups to compensate for variability in the random players strategies.

Figure 8. Smart Heuristic Random Game Performance.

	Rand(20)	Rand(40)	Rand(60)	Rand(80)
T4T	31.06	40.2	48.96	54.96
Opp	31.11	40.25	48.96	55.01
Sum	62.17	80.45	97.92	109.97

	Rand(20)	Rand(40)	Rand(60)	Rand(80)
RandRep	30.67	41.36	48.8	56.03
Opp	31.12	38.91	48.6	53.69
Sum	61.79	80.27	97.4	109.72

	Rand(20)	Rand(40)	Rand(60)	Rand(80)
MajRep	35.19	53.11	39.61	47.69
Opp	16.24	12.06	75.31	68.19
Sum	51.43	65.17	114.92	115.88

Tit-For-Tat and Random Reputation perform roughly the same. This is because, in competition against random players, they are essentially the same heuristic: Pulling the final element from a randomly generated list is the same as pulling a random element from a randomly generated list. Comparing these two heuristics to Majority Reputation, however, we see some interesting patterns. While Majority Reputation punishes lower percentage cooperators (LPCs) and rewards higher percentage cooperators (HPCs) more than the other two heuristics, it also rewards HPCs that are closer to 50% *c-strength*. As long as one can ensure that Majority Reputation will cooperate every time (in other words, as long as you cooperate greater than or equal to 50% of the time), one will be able to exploit it and defect up to 50% of the time while knowing Majority Reputation will cooperate.

The best strategy, then, to converge to cooperation is either Tit-For-Tat, as Axelrod also demonstrated, or our Random Reputation heuristic. While Tit-For-Tat performed equally as well as Random Reputation against a variety of random players, in the evolutionary trials (Figure 6), Random Reputation actually performed better than Tit-For-Tat. However, I am inclined to choose Tit-For-Tat over Random Reputation as the best strategy to ensure convergence to cooperation for the following reason. Consider we have a heuristic that is learning and changing its strategy as the game progresses. For example, the first ten turns the heuristic is random, and then it picks up on Random Reputation's strategy, and realizes that if it cooperates the rest of the game Random Reputation will also start to cooperate more often. So, for all the following turns, the heuristic only cooperates. Random Reputation would lack the ability to ever converge to 100% cooperation with this heuristic, because it always weights past turns as much as later turns. It does not reward a hypothetical "intelligent" heuristic for becoming more cooperative over the course of the game to the extent that Tit-For-Tat does.

Though Tit-For-Tat first appears too simple to be the most effective strategy, its power stems from its simplicity. The elegance of Tit-For-Tat is that its strategy is easy to learn as an opponent, and the learning of this strategy is beneficial for both parties. Even if a heuristic were to try and exploit Tit-For-Tat—say, cooperate and then defect the next round, knowing Tit-For-Tat would cooperate—the payoff will be less than had the heuristic cooperated, because either the heuristic cooperates the next round while Tit-For-Tat defects, (5 + 0 = 5 < 6 over two rounds), or the heuristic defects again, ensuring a maximum of 7 points (less than 9) over three rounds. Tit-For-Tat is largely the best strategy *because it cannot adapt*. This illustrates an interesting point in game theory. Tit-For-Tat, much like national foreign policy, for example, ensures mutual destruction. Though, if we were bombed, it would not be strictly rational to respond by bombing the one who bombed us (we do not benefit from this act), our holding of this strategy ensures that we will not be bombed. Tit-For-Tat is successful for the same reason: The most intelligent opponent will be wise enough to never defect, because Tit-For-Tat is steadfast (to the point of irrationality) in its strategy, and defecting will always ensure mutual destruction.

Works Cited

Axelrod, Robert. *The Evolution of Cooperation*. Basic Books, 2006.

Kuhn, Steven, "Prisoner’s Dilemma", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/sum2019/entries/prisoner-dilemma/>.

Sainsbury, R.M. *Paradoxes*. 3rd ed., Cambridge University, 2009.