2024

Analyzing Tit-For-Tat's Performance in Axelrod's Tournament

CS825: GAME THEORY MINI PROJECT

CALVIN HEW: 202369487

Introduction

The motivation for this study stems from an article posted by Ferguson (2024) who contests the widely celebrated dominance of the Tit-For-Tat (TFT) strategy in Axelrod's tournament. His findings identified a strategy known as permanent retaliation to outperform TFT and claims Axelrod's approach is flawed – describing it as "neither complete nor objective".

Studies such as Bendor et. al. (1991), Nowak and Sigmund (1993), Diekmann and Mitter (2012) are all literature studies with a similar sentiment of disproving TFT's competency and providing an alternative answer to the most effective strategy to play for Iterated Prisoner's Dilemma (IPD).

Although this is a fair criticism of Axelrod's approach, there is also an equally justifiable argument against these studies which misinterpret Tit-For-Tat's effectiveness. An abundant number of studies involving simulations of Axelrod's tournament focus on either: (i) finding strategies which "beat" TFT or (ii) proposing the new best performing strategy.

This study aims to point out the flaws in the perspective of a "best" strategy. As a strategy's performance is intertwined with the environment it is evaluated in, there does not exist an optimal approach that performs well under every circumstance.

A number of simulations of the Axelrod tournament will be conducted to analyse the performance of Tit-For-Tat (TFT) and attempt to understand how it "won". An emphasis will be placed on providing potential explanations for situations where it performs effectively vs. poorly, given the results of the simulations. Granted, this study is not aiming to reproduce Axelrod's results but rather, propose an idea of how performance changes depending on the environment and why the concept of "beating" TFT is rather irrational.

The discussion will be structured under three main headings:

1) Discussing and simulating TFT's performance in Axelrod's First Tournament
2) Discussing and simulating TFT's performance in Axelrod's Second Tournament
3) Discussing and exploring TFT under different Tournament Formats and Objective Criterion

Axelrod’s Tournament and Tit-For-Tat

Background

Firstly, it may be helpful to discuss the foundation of Axelrod’s tournament: the iterated Prisoner’s Dilemma (IPD). As the name suggests, the IPD is taking the stage game of Prisoner’s Dilemma (PD) and playing it for a repeated number of times. When playing a single round of PD, the Nash Equilibrium is to act in one’s own self-interest and defect as it leads to best outcomes from either player’s perspective (Schimit et. al., 2015). However, Axelrod’s tournament produced results which states the opposite is true for IPD – that co-operation is the better strategy in the long run (Axelrod and Hamilton, 1981).

Axelrod’s tournament essentially involved a simulation of 15 different strategies which played IPD against each other with the goal of maximizing number of points won across all matches against all players (Rapoport et. al., 2015). Much to the surprise of many, TFT was the best performing strategy of Axelrod's tournament – a strategy which always begins with co-operating and continues to do so until the other player defects. In other words, it is a reactive strategy that copies the other player’s last move (Milinski, 1987).

Interestingly, Axelrod shared a theory that a variation of TFT known as Tit-For-2-Tats would have outperformed TFT had it entered in this pool (Shah, 2021). However, the question is: would the original TFT be second place then? Or rather, how effective are the original strategies after an extra strategy has been added? After all, it is widely accepted that the effectiveness of a single strategy is dependent on the pool of all other strategies (Rapoport et. al., 2015). Certainly, this dependency is acknowledged in literature (Hoffman, 2020; Rapoport et. al., 2015) with Axelrod (1980) himself even claiming that: “had only the entries which actually ranked in the top half been present, then TFT would have come in 4th”.

Despite acknowledgement that the performance of any given strategy is contingent on the pool of competing strategies, there is a tendency to overlook that the environment is also shaped by this pool. For example, Ferguson (2024)’s approach deviates significantly from the original Axelrod tournament because the pool of strategies does not reflect the original in any way. This is not to question the validity of his findings, but using a completely different pool of strategies has implications on the credibility for disproving Tit-For-Tat’s effectiveness.

A more credible attempt to challenge TFT would be theoretically done in a similar environment as the original tournament. This includes rules such as the number of rounds, number of repetitions etc. but should also factor in the pool of strategies.

Following these guidelines, this study will conduct a simulation of Axelrod’s first tournament to:

(i) Identify whether Axelrod’s hypothesis of Tit-For-2-Tats proves to be true: that it is more effective than baseline TFT.

(ii) Identify the consequence of adding an additional strategy to the pool – how this will affect the performance of other strategies.

Simulating Axelrod's First Tournament

Method

The implementation involved using the `axelrod` Python package which contains several useful prebuilt classes and functions to simulate an Axelrod tournament.

For this simulation, Tit-For-2-Tats was added to the original pool of 15 strategies from the first tournament. An important detail to note is that the number of rounds was fixed at 200 with 5 repetitions of matches played to replicate the original first tournament (Kaznatcheev, 2015).

Moreover, a seed parameter is used to ensure the simulation's behavior remains consistent across different runs, but also provides the opportunity to test different instances/runs. This is particularly useful as there is a random strategy amongst the original pool of 15 which adds a component of uncertainty. The relevant code is presented in the provided python notebook file under the section: **Axelrod's First Tournament**.

Results

The first instance of the tournament (left of the screenshot) proved Axelrod's claim to be true; Tit-For-2-Tats does indeed outperform the original TFT. Notice that the original TFT dropped to 4th place. One would initially expect a drop to 2nd place, but this result reinforces the argument that performance is contingent on the competing strategies in the same pool.

More importantly, the ranking (and therefore performance) of many of the other strategies has also changed. This is particularly true in the case of the top half: Tidemand and Chieruzzi dropped from 2nd place all the way to 7th place whereas Stein and Rapoport shot up to 2nd place from 6th place. As shown by these examples, the addition of a new strategy can result in a significant performance drop or increase of an existing one. The set of results on the right represents the second instance:

```python
print("Ranking:")
for pos, name in enumerate(tour1.ranked_names):
    print(f'{pos+1}: {name}')
```

✓ 0.0s

```
Ranking:
1: Tit For 2 Tats
2: First by Stein and Rapoport: 0.05: (D, D)
3: First by Grofman
4: Tit For Tat
5: First by Shubik
6: First by Nydegger
7: First by Tideman and Chieruzzi: (D, D)
8: First by Davis: 10
9: Grudger
10: First by Downing
11: First by Graaskamp: 0.05
12: First by Feld: 1.0, 0.5, 200
13: First by Joss: 0.9
14: First by Tullock
15: First by Anonymous
16: Random: 0.5
```

```python
print("Ranking:")
for pos, name in enumerate(tour2.ranked_names):
    print(f'{pos+1}: {name}')
```

[148] ✓ 0.0s

```
Ranking:
1: First by Stein and Rapoport: 0.05: (D, D)
2: Tit For 2 Tats
3: First by Grofman
4: First by Shubik
5: Tit For Tat
6: First by Tideman and Chieruzzi: (D, D)
7: First by Nydegger
8: Grudger
9: First by Davis: 10
10: First by Graaskamp: 0.05
11: First by Downing
12: First by Feld: 1.0, 0.5, 200
13: First by Joss: 0.9
14: First by Tullock
15: Random: 0.5
16: First by Anonymous
```

Figure 1.0: Ranking of Axelrod's first tournament with Tit-For-2-Tats added to the pool

Discussion

As mentioned, the first instance of the tournament produced results that confirmed both Axelrod's hypothesis for Tit-For-2-Tats and its effect on the original TFT, but how reproducible are these results?

The second instance shows very inconsistent results compared to the first, but why is this? Besides the aforementioned degree of uncertainty concerning each individual tournament, other major factors are actually suggested by the original author of the `axelrod` package (Knight, 2015):

(i) The 5 repetitions conducted in original tournament is not enough.
(ii) The python implementation of these strategies was solely based on the description of provided by Axelrod – leading to some assumptions/ambiguity in the design.

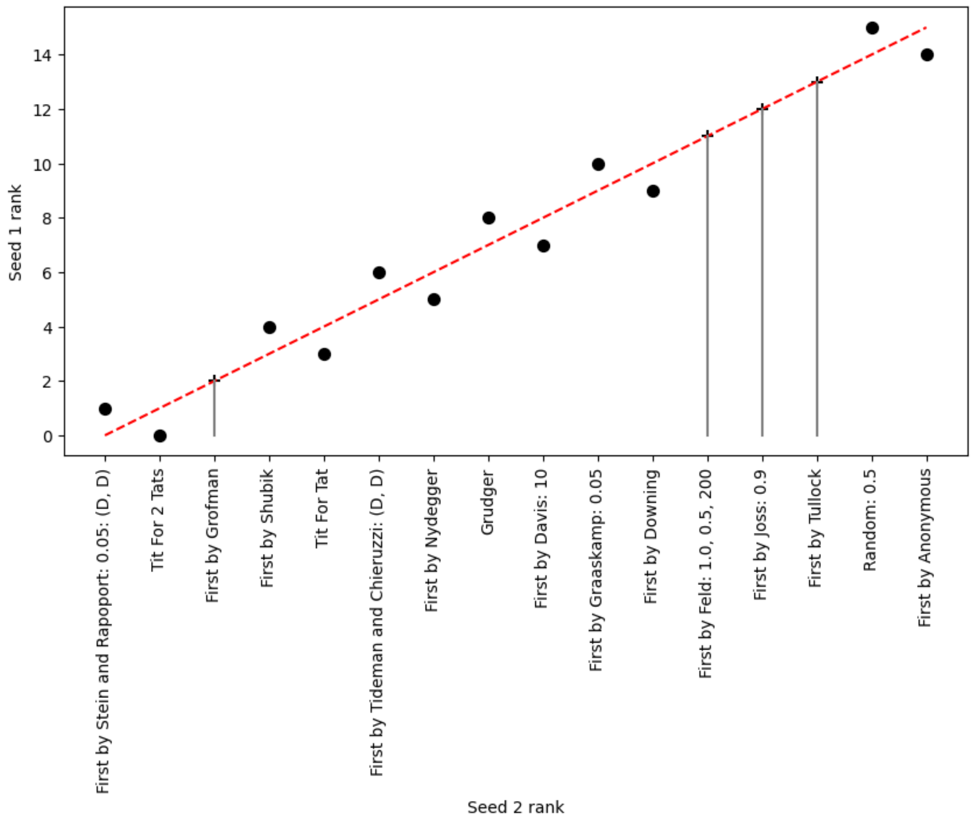

Figure 2.0: *Plot of the differences in the ranking from the first vs. second instance. Any strategy intersecting with the fit line had the same rank in both tournaments.*

Although there is no official source, a discussion on StackExchange (2015) claims the original FORTRAN implementation has been provided by Axelrod at some point in time. Despite this, the intended behavior of the original code could have also been lost in translation when porting from FORTRAN to Python.

This has important implications for simulating the Axelrod tournament and using these results to compare against the original rankings. This is not a fault in Axelrod's approach, but it does raise questions about the reproducibility of such simulated tournaments – including and beyond findings of this study.

Regardless, the simulation was tested with a number of different seeds and although Tit-For-2-Tats did not always come out on top, it did consistently outperform the original TFT. However, does this prove the original TFT is beaten?

Another important point to consider is that studies similar to this one is conducted with the benefit of hindsight. The original tournament was conducted without knowledge of what other strategies would be in the pool and this may be an important factor in considering TFT's effectiveness.

Moreover, there is also criticism that the pool of strategies introduce bias in Axelrod's study as they were manually picked as opposed to more rational approaches like using an evolutionary algorithm as suggested by Dyer (2004). However, it can be argued that a strategy's performance is still ultimately dictated by the pool of strategies and so reducing bias would not eliminate this dependency.

As a result, the following conclusions that can be drawn are:

- Tit-For-2-Tats is consistently more effective that the baseline TFT under the environmental conditions of the first Axelrod tournament (same strategy pool, same number of rounds etc.).
- However, it can be argued that the strategy pool is also part of the environment and so contesting TFT's effectiveness in an objective manner is difficult.
- The addition of an extra strategy can greatly influence the ranking in the first Axelrod tournament - suggesting effectiveness of a given strategy is dependent on the competing strategies within the existing pool.

Axelrod's Second Tournament

Background

Axelrod held a second tournament following the results of the first tournament, challenging participants once again to submit a strategy to determine which one plays IPD most effectively. There were 3 key differences to the sequel tournament when compared to the first.

Firstly, Axelrod disclosed details such as the results and strategies from the first tournament as supplementary information to the participants of the second. Rapoport et. al. (2015) believed this factor helped participants make more informed submissions. This information also likely played a sizeable role over the number of "nice" strategies submitted this time around as co-operative strategies saw major success in the last tournament (Shah, 2021).

Secondly, having the value fixed at 200 led to some undesirable strategic behavior of an interaction known as end-game effects. For example, some strategies may end up betraying the other player towards the end of the game because it knows there is less likelihood of retaliation. One proposed solution to end-game effects is to implement a stochastic method to decide on the final round of a game instead of using a fixed value. Indeed, the second tournament uses a probabilistic value of 0.00346 to determine the end of a round (Kaznatcheev, 2015). Apart from strategies which exploited end-game effects, how might this stochastic approach affect the performance of a given strategy?

Lastly, the pool of strategies expanded from just 15 in the first tournament to over 60 submissions this time (Kopelman, 2020)– which obviously has an impact on the effectiveness of a given strategy. Or does it?

To everyone's surprise, the winner of the tournament was TFT once again. (Axelrod, 1980). This suggests that either its performance is not easily influenced by the competing strategies, or the competing strategies favored TFT once again. Recall that a high number of "nice" strategies were submitted in this second tournament which may hint towards the latter reasoning.

Now what about the performance of Tit-For-2-Tats? It was one of the strategies submitted by one of the tournament's participants. Despite a promising result in the simulation of the first tournament, it only placed 24th in this tournament (Player, 2024). This leaves many questions to be explored. Was Tit-For-2-Tats more sensitive to influx of additional strategies compared to the original TFT? Perhaps it was negatively affected by the stochastic configuration of rounds to be played. Alternatively, there is a possibility that it is a combination of both.

As such, a simulation of Axelrod's second tournament will be conducted to understand the new environment in which the original TFT supposedly flourishes whilst Tit-For-2-Tats struggles in. The main objectives of this section will be to:

(i) Identify the performance of TFT under Axelrod's Second tournament setup.
(ii) Identify the performance of Tit-For-2-Tats under Axelrod's Second tournament setup.

Simulating Axelrod's Second Tournament

Method

Once again, implementation of this simulation will be built on the foundation of the `axelrod` Python package. Admittedly, the tools available for simulating the second tournament are a lot more limiting compared to before. Furthermore, there are significantly more strategies to consider and many of them are comparatively more complex and ambiguous.

In the end, only less than half of the full set of submitted strategies (30/63) have been implemented. It is beyond the scope of this project to implement 30+ strategies and the objective is not to necessarily replicate results. Running a simulation of tournaments under such conditions will likely generate different results from the original but can still provide valuable implications. The relevant code is presented in the provided python notebook file under the section: **Axelrod's Second Tournament**.

As before, the parameters are similar with the use of a seed and 5 repetitions for ensuring consistency. However, this implementation differs in that there is no longer a fixed length for the number of rounds and so requires a new parameter: **prob_end** which is set to 0.00346 as required. The 30 available strategies were then played against each other in a simulated IPD.

Results

Once again, the simulation was run for several instances. Figure 3.0 shows the results for the TFT strategies when played in tournaments seeded 1 and 2. A more detailed ranking is provided in the figure below labelled Figure 4.0. The results fail to reflect results in Axelrod's actual second tournament (Axelrod, 1980) as both TFT and Tit-For-2-Tats performed worse than expected.

```
print(f"Tit For Tat Ranking:", second.ranked_names.index("Tit For Tat")+1)
print(f"Tit For 2 Tats Ranking:", second.ranked_names.index("Tit For 2 Tats")+1)
✓ 0.0s

Tit For Tat Ranking: 7
Tit For 2 Tats Ranking: 28
```

```
print(f"Tit For Tat Ranking:", second2.ranked_names.index("Tit For Tat")+1)
print(f"Tit For 2 Tats Ranking:", second2.ranked_names.index("Tit For 2 Tats")+1)
✓ 0.0s

Tit For Tat Ranking: 6
Tit For 2 Tats Ranking: 27
```

Figure 3.0: Ranking of TFT and Tit-For-2-Tats playing Axelrod's second tournament

The base TFT was ranked either 6/30 or 7/30 (depending on the iteration) and can be considered as a decent strategy, but does not compare with the original results crowning it as the top performer. However, the most surprising result was arguably the drop in Tit-For-2-Tats' performance. This strategy was considered middle of the road at rank 29/63 in the original implementation. However, in this case of this simulated run, it drops to near the bottom of the ranking at 27/30 or 28/30 (depending on the iteration).

```
Ranking:
1: Second by Tideman and Chieruzzi
2: Second by Weiner
3: Second by Borufsen
4: Second by Kluepfel
5: Second by Cave
6: Tit For Tat
7: Second by RichardHufford
8: Second by Harrington
9: Second by Getzler
10: Second by WmAdams
11: Second by Tranquilizer
12: Second by Grofman
13: Second by Eatherley
14: Second by Leyvraz
15: Second by Rowsam
16: Second by Tester
17: Second by Black
18: Second by Yamachi
19: Second by White
20: Second by Mikkelson
21: Soft Go By Majority
22: Second by Champion
23: Second by Gladstein
24: Second by Colbert
...
27: Tit For 2 Tats
28: Win-Stay Lose-Shift
29: Grudger
30: Random: 0.5
```

```
Ranking:
1: Second by Borufsen
2: Second by Tideman and Chieruzzi
3: Second by Tranquilizer
4: Second by Weiner
5: Second by Cave
6: Second by Kluepfel
7: Tit For Tat
8: Second by WmAdams
9: Second by Getzler
10: Second by RichardHufford
11: Second by Harrington
12: Second by White
13: Soft Go By Majority
14: Second by Mikkelson
15: Second by Leyvraz
16: Second by Rowsam
17: Second by Black
18: Second by Eatherley
19: Second by Grofman
20: Second by Yamachi
21: Second by Champion
22: Second by Gladstein
23: Second by Tester
24: Second by Appold
...
27: Win-Stay Lose-Shift
28: Tit For 2 Tats
29: Grudger
30: Random: 0.5
```

Figure 4.0: Ranking of all strategies playing Axelrod's second tournament

Comparing Results to the first simulation:

- TFT performs worse in this simulated Axelrod's tournament compared to the first simulated tournament: **Rank 6/30** vs **Rank 4/16.**
- Tit-For-2-Tats performs significantly worse in this simulated Axelrod's tournament compared the first simulated tournament: **Rank 27/30** vs. **Rank 1/16.**

Comparing Results to the actual results of the second tournament:

- TFT performs worse in this simulated Axelrod's tournament compared to the actual second tournament: **Rank 6/30** vs **Rank 1/63.**
- Tit-For-2-Tats performs significantly worse in this simulated Axelrod's tournament compared to the actual second tournament: **Rank 27/30** vs **Rank 29/63.**

Discussion

Ultimately the results do not reflect that of the Second Axelrod's tournament for the reasons likely pertaining to the limitations of the `axelrod` python library as discussed in methods. It is worth noting that TFT still performed relatively well in the incomplete second tournament.

Missing half the pool changes not only the number of individual interactions a strategy partakes in, but more importantly, it changes the entire environment it plays in. It can be inferred that a number of these missing strategies may have been integral to fostering a co-operative environment due to the drop in both TFT and Tit-For-2-Tats' performance.

Looking at the full ranking presented in Figure 4.0, the top 5 strategies consistently performed effectively across different instances of the tournament. In a sense, this set of strategies can be interpreted as those which "beat" TFT just as Ferguson (2024) claimed his strategy did. In both simulations, the "rules" did technically reflect that of Axelrod's, but a drastic change in the entire pool of strategies has led to the deterioration of both TFT strategies.

Instead of looking at how a strategy fits within a specific environment, it may be more beneficial to look at how an environment benefits/fosters certain types of strategy. For example, Axelrod (1980) popularized the following four traits that were common amongst all strategies which succeeded in his tournaments (all of which imply a co-operative environment):

- Niceness: not being the first to defect.
- Provocable: willing to punish other players if they defect (avoids being exploited).
- Forgiveness: willing to restore co-operation if the other co-operates.
- Clarity: predictability makes it easier to cooperate with.

Perhaps it is best explained by Glynatsi and Knight (2020) but the goal of studying IPD tournaments should be to identify properties pertaining to successful strategies as opposed to proposing a single "best" strategy.

The subtly of this difference is important as there is a tendency for literature to view strategy through a perspective of how well it performs without acknowledging the environmental factors which allowed it to succeed in the first place. This often leads to the misunderstanding of an optimal or "best" approach. As a result, there is reason to approach studies which claim to outperform TFT with skepticism. It is not to discredit or discourage such exploratory works, but rather, to question the underlying purpose of their study.

For example, there are a number of studies utilizing sophisticated methods like the Zero-Determinant (ZD), Reinforcement learning (RL), hidden Markov models and even finite-state machines (FSM) have all been used in an attempt to "win" or be the best strategy (Glynatsi, and Knight, 2020). In particular, Vincent and Fryer (2021) trained an FSM with an evolutionary algorithm to impressively obtain a high score in the Axelrod tournament. However, does this strategy perform as effectively consistently across different pools of competing strategies? Or more clearly, under what circumstances does this strategy stay as the best vs. when performance starts dropping.

A consensus appears to exist amongst researchers who acknowledge that no strategy can be considered the "best" in every situation. Thus, it is fascinating that studies still obsess over this pursuit of not only "beating" TFT but achieving the best score. Perhaps it is simply within human nature to be competitive.

Tournament Format and Objective Criterion

Background

Whilst this study has analyzed the incoherent ideology of finding the “best” strategy considering rankings are dependent on the pool of competing strategies, other decisions regarding the Axelrod tournament’s setup have been overlooked. For one, how is **tournament** defined in this context?

Rapoport et. al. (2015) proposes the design of tournaments involves three components being: the format, the objective criterion and pool of participants.

The format selected in the IPD tournaments proposed by Axelrod is a single stage round-robin (Schimit et. al., 2015) – a configuration where points are accumulated from playing against all strategies in the pool. The advantage of this format is that it allows for a fair assessment of a strategy’s overall performance as it must interact with each one of the other strategies.

On the other hand, the objective criterion is more subject to scrutiny. Axelrod defines the criteria of success for a given strategy as the maximisation of points across all pairwise interactions. Whilst this scoring is not inherently wrong, an extremely valuable point made by Rapoport et. al. (2015) suggests that strategies submitted for the first tournament appeared to be built based on winning individual interactions as opposed to maximizing total points considering all encounters.

Since the tournament setup appeared to favor co-operation, this reasoning could also explain why TFT performed as well as it did and why defective strategies were almost always outperformed. So this leads to an interesting point of: how does TFT fare in terms of individual interactions?

```
print("Ranking:")
for pos, name in enumerate(tour1.ranked_names):
    print(f'{pos+1}: {name}')
✓ 0.0s
Ranking:
1: Tit For 2 Tats
2: First by Stein and Rapoport: 0.05: (D, D)
3: First by Grofman
4: Tit For Tat
5: First by Shubik
6: First by Nydegger
7: First by Tideman and Chieruzzi: (D, D)
8: First by Davis: 10
9: Grudger
10: First by Downing
11: First by Graaskamp: 0.05
12: First by Feld: 1.0, 0.5, 200
13: First by Joss: 0.9
14: First by Tullock
15: First by Anonymous
16: Random: 0.5
```

```
ranked_avg_wins
[146] ✓ 0.0s
{'First by Tideman and Chieruzzi: (D, D)': 12.0,
 'First by Joss: 0.9': 11.4,
 'First by Stein and Rapoport: 0.05: (D, D)': 11.0,
 'First by Feld: 1.0, 0.5, 200': 10.4,
 'First by Tullock': 9.0,
 'First by Downing': 8.0,
 'First by Graaskamp: 0.05': 5.0,
 'Grudger': 5.0,
 'Random: 0.5': 4.6,
 'First by Anonymous': 4.0,
 'First by Davis: 10': 4.0,
 'First by Shubik': 3.0,
 'First by Grofman': 1.0,
 'Tit For 2 Tats': 0.0,
 'Tit For Tat': 0.0,
 'First by Nydegger': 0.0}
```

Figure 5.0: Ranking of strategies based on points maximization (left) vs. based on no. of wins (right)

Recall the simulation of the first Axelrod tournament. A parameter that can be accessed from a tournament's result set is **wins -** *the number of encounters that each strategy managed to win* against other players. An average number of wins was calculated over the 5 repetitions and presented in Figure. 5.0 on the right-hand side. As shown, both TFT and Tit-For-2-Tats did not win a single interaction despite their strong performance in Axelrod's original tournament.

This may leave individuals dissatisfied with acknowledging TFT's effectiveness. A strategy that accumulates the highest final score but never wins can be considered consistent and not necessarily the "best". However, this is the realization that this study hopes to convey.

Axelrod's tournament is a not a competition in the traditional sense of finding a definitive winner of IPD, but rather to inform readers of what strategies may flourish in a tournament setting. The weakness in his approach is perhaps a lack of clarity in presenting the reasonings behind the configurations of his tournament.

Rapoport et. al. (2015) shows that changing these configurations will undeniably change the environment in which a given strategy interacts in, and consequently it's performance. His approach utilizes a two-stage round robin which resulted in TFT being outperformed by two other strategies. The important distinction in his study to others challenging TFT's dominance is that he acknowledges that the result is contingent of this tournament setup.

One unexplored avenue of tournament setup for IPD is the format known as the knockout tournament. The knockout tournament is typically the format best associated in sports or competitions where winning competitors progress on to the next round whereas losing competitors are eliminated until a single winner has been decided.

Researchers seem to unanimously agree on using round robin formats compared to knockout formats as it avoids dependency on an individual interaction (Dyer, 2004). However, with the competitive undertone in finding the "best" strategy that exists in much of the IPD research, it may be worthwhile proposing an approach that fosters this innate desire to compete.

Simulating a Knockout Tournament

Method

As the Axelrod tournament only considered round robin formats, the `axelrod` package does not provide any support for knockout style tournaments. Thus, an implementation of a knockout tournament must be coded manually – this is labelled in the Python notebook under the heading **Knockout Tournament.**

A decision was made to eliminate both strategies if the result of the interaction is a draw. This is to ensure other strategies can continue playing until a winner is determined. However, if all interactions of competing strategies in each round results in a draw state, then no winner can be decided.

The first approach involves simply running a knockout tournament on the initial set of strategies used in Axelrod's tournament. The simulation was tested on both the objective criterion of total points accumulated and number of wins.

A second approach proposes a two-stage format which mixes both round robin and knockout. The top half of the best performing strategies from the initial simulation of Axelrod's first tournament were utilized as the round robin results. From these, the top 8 strategies participated in the knockout stage.

Results

The first simulation of the knockout tournament using the objective criterion of maximizing scores across all interactions resulted in no winner (essentially a draw happened at every stage). Details of this tournament is provided in Appendix 6.0.

```
...  Round 1:
     First by Tullock vs. First by Graaskamp: 532 vs. 527
     Tit For 2 Tats vs. Random: 701 vs. 1176
     First by Shubik vs. First by Feld: 1017 vs. 1102
     First by Davis vs. First by Downing: 300 vs. 325
     First by Anonymous vs. Grudger: 117 vs. 767
     First by Nydegger vs. First by Stein and Rapoport: 18 vs. 18
     First by Grofman vs. First by Tideman and Chieruzzi: 5472 vs. 5472
     First by Joss vs. Tit For Tat: 58 vs. 53
     Moving on to next round: ['First by Tullock', 'Random', 'First by Feld', 'First by Downing', 'Grudger', 'First
     Round 2:
     First by Tullock vs. Random: 936 vs. 826
     First by Feld vs. First by Downing: 1362 vs. 642
     Grudger vs. First by Joss: 29 vs. 29
     Moving on to next round: ['First by Tullock', 'First by Feld']
     Round 3:
     First by Tullock vs. First by Feld: 232 vs. 232
     Moving on to next round: []
     Draws: [First by Nydegger, First by Stein and Rapoport: 0.05: (D, D), First by Grofman, First by Tideman and Ch
```

Appendix 6.0: Knockout format with score maximization criterion (winner: no result)

Next, a competitive knockout tournament was simulated with the no. of wins as the performance criterion shown in Appendix 7.0. Ironically, the strategies which had poor performance in the first simulated Axelrod's tournament excelled in this tournament - as was the case for the winner: Joss. The Joss strategy was 13th out of 16 when considering score maximization which suggests the strategy was built to win individual interactions over anything lese. Notice that both TFT strategies did not make it past the first round as they were shown to win 0 encounters in the first tournament.

```
...  Round 1:
     First by Grofman vs. First by Joss
     First by Stein and Rapoport vs. First by Shubik
     Tit For 2 Tats vs. First by Tideman and Chieruzzi
     First by Feld vs. Tit For Tat
     First by Downing vs. Random
     First by Anonymous vs. First by Graaskamp
     First by Nydegger vs. First by Davis
     First by Tullock vs. Grudger
     Moving on to next round: ['First by Joss', 'First by Feld', 'First by Downing', 'First by Graaskamp']
     Round 2:
     First by Joss vs. First by Feld
     First by Downing vs. First by Graaskamp
     Moving on to next round: ['First by Joss', 'First by Graaskamp']
     Round 3:
     First by Joss vs. First by Graaskamp
     Moving on to next round: ['First by Joss']
     Draws: [First by Stein and Rapoport: 0.05: (D, D), First by Shubik, Tit For 2 Tats, First by Tideman and Chieruzzi: (D
     Winner: First by Joss
```

Appendix 7.0: Knockout format with no. of wins criterion (winner: Joss)

Finally, a mixed approach of consisted of

- A round robin stage to initially determine strategies which qualify to play in the next stage.
- A subsequent knockout stage consisting of many rounds to decide an overall winner.

The result for this approach is also disappointing as no strategy made it to the second round (as highlighted in Appendix 8.0). Interestingly, these top 8 strategies will always result in a draw when played against one another since they are considered "nice" strategies.

```
       run_tournament(top_participants)
[253]  ✓ 0.0s

...  Round 1:
     First by Nydegger vs. Tit For Tat: 1881 vs. 1881
     First by Tideman and Chieruzzi vs. First by Shubik: 162 vs. 162
     Tit For 2 Tats vs. First by Grofman: 1317 vs. 1317
     Grudger vs. First by Stein and Rapoport: 4023 vs. 4023
     Moving on to next round: []
     Draws: [First by Nydegger, Tit For Tat, First by Tideman and Chieruzzi: (D, D), First by Shubik, Tit For 2 Tats,
```

Appendix 8.0: Round robin + Knockout 2-stage format (winner: no result)

Discussion

The results revealed an admittedly apparent flaw in using a knockout format approach – that is playing IPD can (and often does) results in a draw. The knockout format is more conducive of circumstances where a single winner needs to be decided. The answer to why knockout format was rarely discussed for Axelrod's tournament may have been obvious. IPD is not a zero-sum game and there is not always a "winner" in the traditional sense. This contradicts with the entire format of a knockout tournament.

Furthermore, Axelrod (1980) is extremely careful in defining a strategy as simply effective rather than labels such as the "best" or "most effective" as to not suggest there can only be one. Instead, he presents the group of strategies considered co-operative to perform well in his original tournament setup.

These simulations for a knockout tournament format were not necessarily expected to propose a rational approach to competing under IPD. Rather, it was the consequence of personal curiosity regarding such a dominating preference of using round-robin formats.

The knockout tournament format is seldom even acknowledged in studies discussing Axelrod's tournament or in IPD. The few studies which do acknowledge it as an option, such as Rapoport et. al. (2015) and Schimit et. al. (2015), simply present the notion that round robin is better. Whilst this may be true, it gives the reader very little reason as to why knockout tournaments are inherently incompatible with the discussion of effective strategies in an IPD context.

Ultimately, these results (or rather lack thereof) opens a discussion of the importance in both tournament setup and objective criterion when evaluating a strategy's performance. Often, it is easy to overlook the lack of discussion surrounding these dependencies or accept an oversimplified rationalisation of the selected methodology. Studies may have proven to acknowledge the interdependency of the pool of competing strategies with the effectiveness of a given strategy, but they often fail to realize the influence tournament configuration and objective criterion also possess. Therefore, when attempting to find effective strategies in IPD, it can be argued there is significant merit in investigating a rational tournament format and the relevant criterion for success.

Conclusion

This study does not intend to discredit or discourage the exploration of strategy within IPD - but rather, it encourages studies based on Axelrod's tournament to avoid simply disputing which strategy is best. Instead of proclaiming "Tit-for-Tat is not the best strategy" as Ferguson's (2024) article is titled, a much more rational motivation would be to uncover what environments is TFT considered effective and ask what reasons allowed TFT to win Axelrod's tournament in the first place? As the results of this study show, the reason is not as simple as the pool of competing strategies – although that is undeniably a major factor in influencing performance (Hoffman, 2020).

Moreover, Axelrod is intentionally vague in the justification of both his choice of tournament format (i.e. why total points accumulated/round robin was chosen) and in the criteria for evaluating performance (Rapoport et. al., 2015). In other words, it was likely he did not arbitrarily select these configurations with no insight but perhaps he deemed it insignificant in the overall discussion surrounding effective strategies for playing IPD. However, results from this study proves otherwise.

Tournament configurations, the criterion for success and the pool of competing strategies are just a few factors identified to greatly influence the environment in which a strategy is evaluated in. It is this environment (and the factors pertaining to it) which determines how effectively a strategy performs under IPD. Furthermore, it is also this environment that explains the conflicting results of TFT's effectiveness from the perspective of many different studies.

Thus, the concept of attempting to "beat" TFT is called into question. If strategy performance is conducive of environmental factors such as tournament configurations, criterion for success and pool of competing strategies, how can consistency or objectivity be ensured in this case? The answer may be that it is very difficult or impossible to do so. Therefore, studies of the Axelrod tournament should do their best to acknowledge these intricate contingencies when attempting to contest TFT's performance.

This study ultimately conveys the idea that it may be more beneficial to analyze the environments in which specific strategies flourish and understand effectiveness from that perspective, rather than attempting to find a supposed "best" strategy.

References

Axelrod, R., 1980. Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, 24(1): 3-25.

Axelrod, R., 1980. More effective choice in the prisoner's dilemma. *Journal of Conflict Resolution, 24* (3), 379-403

Axelrod, R. and Hamilton, W.D., 1981. The evolution of cooperation. science, 211(4489), pp.1390-1396.

Bendor, J., Kramer, R.M. and Stout, S., 1991. When in doubt... Cooperation in a noisy prisoner's dilemma. Journal of conflict resolution, 35(4), pp.691-719.

Diekmann, A. and Mitter, P. eds., 2012. *Paradoxical effects of social behavior: Essays in honor of Anatol Rapoport*. Springer Science & Business Media.

Dyer, D.W., 2004. Opponent modelling and strategy evolution in the iterated prisoner's dilemma. *Master's thesis, School of Computer Science and Software Engineering, The Univ. of Western Australia*.

Ferguson, S., 2024. Prisoner's Dilemma Revisited - Tit-for-Tat is not the best strategy. Medium.
https://medium.com/@metaform3d/prisoners-dilemma-revisited-bfdaa0e02c80

Glynatsi, N.E. and Knight, V.A., 2020. A meta analysis of tournaments and an evaluation of performance in the Iterated Prisoner's Dilemma. arXiv preprint arXiv:2001.05911.

Hoffmann, R., 2000. Twenty years on: The evolution of cooperation revisited. *Journal of Artificial Societies and Social Simulation*, *3*(2), pp.1390-1396.

Kaznatcheev, A. 2015. Short history of iterated prisoner's dilemma tournaments. Axelrod's tournaments and TIT-for-TAT. Theory, Evolution and Games Group.
https://egtheory.wordpress.com/2015/03/02/ipd/

Knight, V., 2015. Welcome to the documentation for the Axelrod Python library. Axelrod.
https://axelrod.readthedocs.io/en/stable/index.html

Kopelman, S., 2020. Tit for tat and beyond: The legendary work of Anatol Rapoport. Negotiation and Conflict Management Research, 13(1), pp.60-84.

Milinski, M., 1987. Tit for tat in sticklebacks and the evolution of cooperation. *nature*, *325*(6103), pp.433-435.

Nowak, M. and Sigmund, K., 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature, 364(6432), pp.56-58.

Player, N., 2024. The Morality and Practicality of Tit for Tat. *The Philosophy, Politics, and Economics Review*.

Rapoport, A., Seale, D.A. and Colman, A.M., 2015. Is tit-for-tat the answer? On the conclusions drawn from Axelrod's tournaments. *PloS one*, *10*(7), p.e0134128.

Schimit, P.H.T., Santos, B.O. and Soares, C.A., 2015. Evolution of cooperation in Axelrod tournament using cellular automata. *Physica A: Statistical Mechanics and its Applications*, *437*, pp.204-217.

Shah, R., 2021. Robert Axelrod: The Prisoner's Dilemma Simulation. Shortform. https://www.shortform.com/blog/robert-axelrod/

StackExchange, 2015. Axelrod's Prisoner's Dilemma source code. Math StackExchange. https://math.stackexchange.com/questions/465054/axelrods-prisoners-dilemma-source-code

Vincent, F.M. and Fryer, D.E., 2021. Top Score in Axelrod Tournament. *arXiv preprint arXiv:2104.03347*.