Tournament Remix

Kyrylo Krocha

November 20, 2025

Contents

1 Introduction

Iterated Prisoner's Dilemma (IPD) is a famous mathematical model which can be used to describe various behavioral phenomena and has applications in such areas as biology [1], psychology [2], social studies [3], as well as other areas. This model has been widely researched as a framework for studying the evolution of cooperation and competition among rational agents.

In the classical Prisoner's Dilemma, two players simultaneously choose between two possible actions: *cooperate (C)* or *defect (D)*, without knowing the other's choice. The outcome of the interaction is determined by a payoff matrix, which assigns rewards based on the combination of decisions made by both players:

	Cooperate (C)	Defect (D)
Cooperate (C)	(R, R)	(S, T)
Defect (D)	(T, S)	(P, P)

Table 1: Payoff matrix for the Prisoner's Dilemma

When both players choose to cooperate, they each receive the reward payoff R; if one defects while the other cooperates, the defector obtains the temptation payoff T while the cooperator receives the sucker's payoff S; and when both players defect, they both receive the punishment payoff P. The values of payoffs in this work are set to $R = 3$, $P = 1$, $T = 5$, and $S = 0$, which corresponds to the parameters used in the original Axelrod tournaments [4] and have been widely adopted in subsequent research.

Unlike the single-round Prisoner's Dilemma, the iterated version allows the same participants to interact repeatedly, enabling the development of complex strategic behaviors. Over the years, the study of such behaviors has become a central topic in the analysis of cooperation and competition, leading to the development of numerous strategies and extensive theoretical and computational investigations [5, 6, 7].

The Axelrod tournaments, in particular, played a key role in demonstrating how cooperation can emerge and persist even in competitive environments. In these computational experiments, organized by Robert Axelrod in the early 1980s, researchers submitted strategies that competed in repeated Prisoner's Dilemma games. Each strategy played against every other strategy, as well as against a copy of itself, and rankings were based on the average payoff across all matches. The simple Tit for Tat strategy, proposed by Anatol Rapoport, achieved the best overall performance, showing that cooperation, when combined with reciprocity and forgiveness, can outperform more exploitative approaches [4]. A second tournament with more sophisticated opponents confirmed the robustness of this result, highlighting that cooperative behavior can be evolutionarily stable even among self-interested agents. These findings established the Axelrod tournament as a benchmark for studying cooperation and strategic interaction.

While traditional tournaments rank strategies based solely on their average score, this approach may not fully capture the complexity of strategic interactions in the IPD. A strategy can achieve a high average score by performing consistently well overall, even if it loses in most pairwise encounters. As a result, the average-score metric emphasizes efficiency but may overlook dominance relationships between individual competitors. To address these limitations, alternative evaluation methods for determining the winner of a computer tournament are implemented and compared. Namely, the following methods are considered:

- **Win-based tournament:** ranks strategies according to the number of head-to-head victories, focusing on direct competitive success rather than aggregate performance. This measure highlights strategies that are more dominant in individual matchups, even if their overall scores are slightly lower.

- **Schulze method:** ranks strategies based on strength of their wins, where the strength of each victory is defined as the difference in points in their head-to-head match. These differences form a weighted directed graph of pairwise outcomes, where strategy may rank above another even if it loses directly, as long as it can reach it through a chain of strong victories (e.g., A beats C, and C beats B). The algorithm computes the strongest such paths, producing a ranking that reflects not only who wins directly, but also how convincingly each strategy can dominate others through sequences of weighted victories.

By implementing and comparing these different evaluation frameworks: classic (average-score), win-based, and Schulze-style, this project aims to uncover how alternative ranking methods influence the identification of dominant strategies in the IPD setting. Through these comparisons, we can gain deeper insight into the structural dynamics of competition and the mechanisms that sustain cooperation across different strategic environments.

2 Data collection

The main tool for running tournaments and collecting data is Axelrod-Python [8] - an open-source research framework designed specifically for the study of the IPD. It provides a comprehensive collection of tools for simulating tournaments and analyzing the behavior of a wide variety of strategies. The library also includes hundreds of predefined strategies, ranging from classic rule-based approaches such as *Tit for Tat* and *Grim Trigger* to more sophisticated stochastic, memory-based, and machine learning–driven algorithms.

2.1 Tournament parameters

Using the Axelrod-Python library, we can create tournaments using a predefined set of strategies. Each tournament consists of pairwise matches, and each match is played for a fixed number of rounds. When noise is included, matches may also terminate early with a specified probability.

The parameters for running the tournaments remain consistent throughout this work and are summarized in the table below. These parameters correspond to those used in the first Axelrod tournament [4].

Parameter	Value
Number of rounds per match	200
Number of repetitions per opponent pair	5
Noise probability	0.0
Match payoff values	$T = 5, R = 3, P = 1, S = 0$

Table 2: Tournament Parameters

2.2 Data pipeline and source code

After the tournament is executed, all results are saved in a CSV file, which is then processed by a separate analysis program. This program is responsible for computing key performance metrics, applying alternative ranking methods such as the win-based and Schulze formats, and generating statistical and graphical summaries of the outcomes.

The full source code used in this project, including tournament setup, data processing, and analysis scripts, can be found at this GitHub repository.

3 Alternative evaluation methods

3.1 Win-based method

The win-based method evaluates tournament performance by counting the number of head-to-head victories each strategy achieves against others. This method focuses on direct competitiveness - how often a strategy performs better than its opponent in individual matches. Such an evaluation can reveal dominance relations that average-based rankings might obscure, especially in cases where a strategy scores highly overall but consistently loses to certain opponents. By emphasizing pairwise outcomes, the win-based method provides an alternative perspective on strategic strength and robustness within the tournament.

Although this ranking approach is not directly implemented in the Axelrod-Python library, the tournament summary it produces contains all the necessary data to compute the number of wins for each strategy.

3.2 Schulze method

The Schulze method is a ranking algorithm originally developed by Markus Schulze [9] for preferential voting systems, designed to determine the most representative or dominant option among multiple candidates. In the context of the IPD, it can be adapted to rank strategies based on the strength of their pairwise victories. Rather than considering only the number of wins, the Schulze method analyzes the paths of dominance - indirect chains of victories that reveal which strategies consistently outperform others, even through intermediaries.

As a Condorcet completion method [10], it ensures that if a majority prefers one option over another (e.g., in most matches strategy A wins over B), that preference is respected whenever possible. The Schulze method resolves cyclic ties - situations where no clear majority winner exists, by considering the strength of indirect victories, resulting in a stable and transitive ranking that better reflects the underlying dominance structure among strategies.

The code for implementing Schulze method uses the results provided by Axelrod-Python library to calculate a pairwise winning matrix, which contains differences in scores between each pair of strategies. The algorithm for evaluating the best strategies is presented below:

Algorithm 1 Schulze method for strategy ranking

1: **Step 1: Build pairwise comparison matrix**
2: **for all** pairs (A, B) **do**
3:     Compute $d(A, B)$ – number of matches where A scores higher than B
4: **end for**
5: **Step 2: Initialize path strengths**
6: **for all** pairs (A, B) **do**
7:     **if** $A \neq B$ and $d(A, B) > d(B, A)$ **then**
8:         $p(A, B) \leftarrow d(A, B)$
9:     **else**
10:         $p(A, B) \leftarrow 0$
11:     **end if**
12: **end for**
13: **Step 3: Compute strongest paths**
14: **for all** strategies i **do**
15:     **for all** distinct (A, B) where $A \neq i$ and $B \neq i$ **do**
16:         $p(A, B) \leftarrow \max(p(A, B), \min(p(A, i), p(i, B)))$
17:     **end for**
18: **end for**
19: **Step 4: Determine final ranking**
20: Strategy A is ranked higher than B if $p(A, B) > p(B, A)$

---

4 Remixing original tournament

4.1 Baseline: Classic tournament

To analyze the effectiveness of different tournament evaluation methods, we first conducted the recreation of the first Axelrod's tournament, which served as a baseline for subsequent comparisons. This classic format is based on the average score obtained by each strategy over multiple pairwise interactions. Each strategy plays against every other strategy, including itself, over a fixed number of rounds, and the final ranking is determined by the mean payoff per match.

By beginning with this traditional format, we establish a well-understood reference against which alternative evaluation criteria, such as win-based ranking and the Schulze method, can be directly compared. Axelrod's first tournament included 14 strategies (plus a random "strategy", which cooperates with probability 0.5) and they are listed below[4]:

Name	Author
Tit For Tat	Anatol Rapoport
Tideman and Chieruzzi	T. Nicolaus Tideman and Paula Chieruzzi
Nydegger	Rudy Nydegger
Grofman	Bernard Grofman
Shubik	Martin Shubik
Stein and Rapoport	Stein and Anatol Rapoport
Grudger	James W. Friedman
Davis	Morton Davis
Graaskamp	Jim Graaskamp
FirstByDowning	Leslie Downing
Feld	Scott Feld
Joss	Johann Joss
Tullock	Gordon Tullock
(Name withheld)	Unknown
Random	Unknown

Table 3: Strategies from Axelrod's first tournament. Information about the behavior of these strategies can be found at <u>this link</u>.

After running the tournament, the summary of results was obtained, part of which is listed below:

Rank	Name	Median_score	Cooperation_rating	Wins
1	First by Stein and Rapoport	2.583571	0.602929	10
2	First by Grofman	2.5275	0.848357	1
3	First by Shubik	2.508214	0.613429	3
4	Tit For Tat	2.4925	0.717357	0
5	First by Nydegger	2.461071	0.996	0
6	First by Tideman and Chieruzzi	2.458929	0.597286	11
7	Grudger	2.401071	0.537929	5
8	First by Davis	2.399643	0.547571	4
9	First by Graaskamp	2.262857	0.565571	4
10	First by Downing	2.2375	0.362571	7
11	First by Feld	1.902143	0.290357	10
12	First by Tullock	1.833571	0.311357	8
13	First by Joss	1.774286	0.2895	11
14	First by Anonymous	1.641071	0.505	3
15	Random: 0.5	1.620714	0.502714	3

Table 4: Summary of the first tournament

As can be seen from a table above, *Tit for Tat* didn't win in this recreation, which is due to unavailability of the exact conditions of the original tournament, which is why in modern recreations *Tit for Tat* rarely wins, although still scores highly. It can also be observed that the best-performing strategies in the classic tournament are not necessarily those that achieve the highest number of wins. For example, *First by Joss* records 11 wins but is ranked only 12th overall. This observation is further supported by the Spearman's rank correlation between the strategy's position in the classic ranking and its number

of wins, which in this setup equals approximately 0.2428. Such a low correlation indicates that the alternative evaluation formats proposed in this work are likely to produce substantially different rankings compared to the original tournament.

4.2   Adding new formats

After running the tournament, new methods of determining the winner had been applied using the summary obtained above. These methods were integrated into the post-processing stage of the analysis pipeline, allowing comparison of their outcomes with the traditional average-score ranking. Different rankings of strategies can be seen below:

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

Table 5: All rankings

The results can also be summarized visually using a comparative ranking graph (where a lower rank indicates better performance):
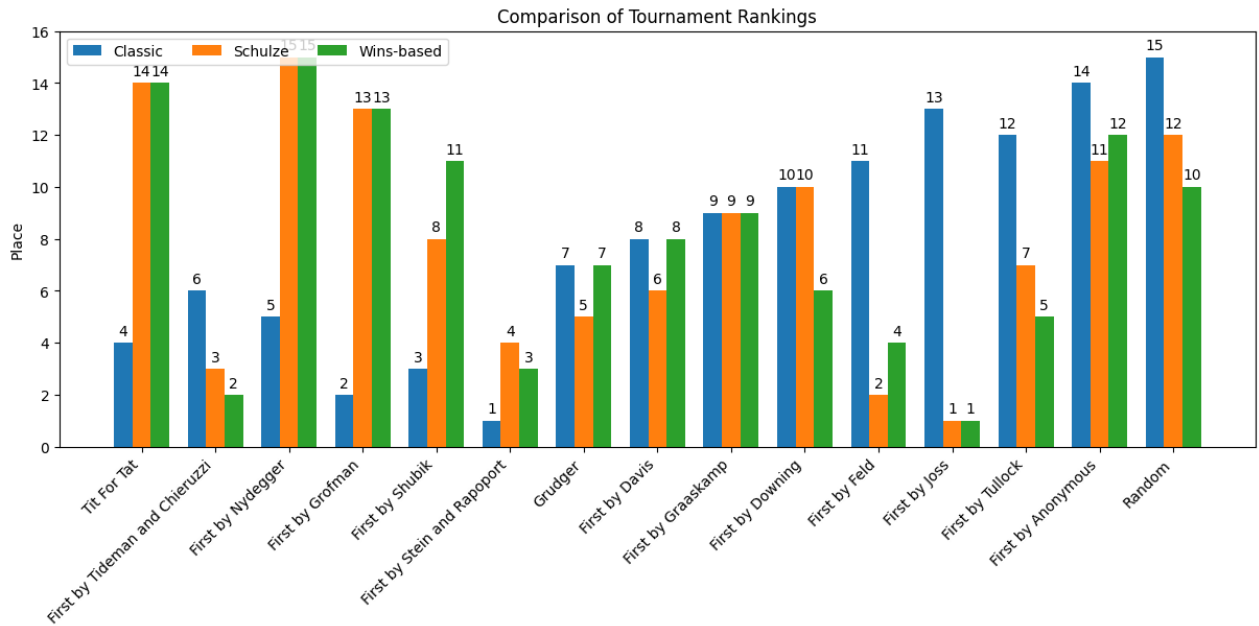
Figure 1: Rankings across different methods

The Spearman's correlation between the classic ranking and the win-based method is relatively low (approximately –0.2357), as is the correlation between the classic ranking and the Schulze method (–0.1428).

However, the correlation between the win-based and Schulze methods is notably high (0.9142), as illustrated in Figure 2.
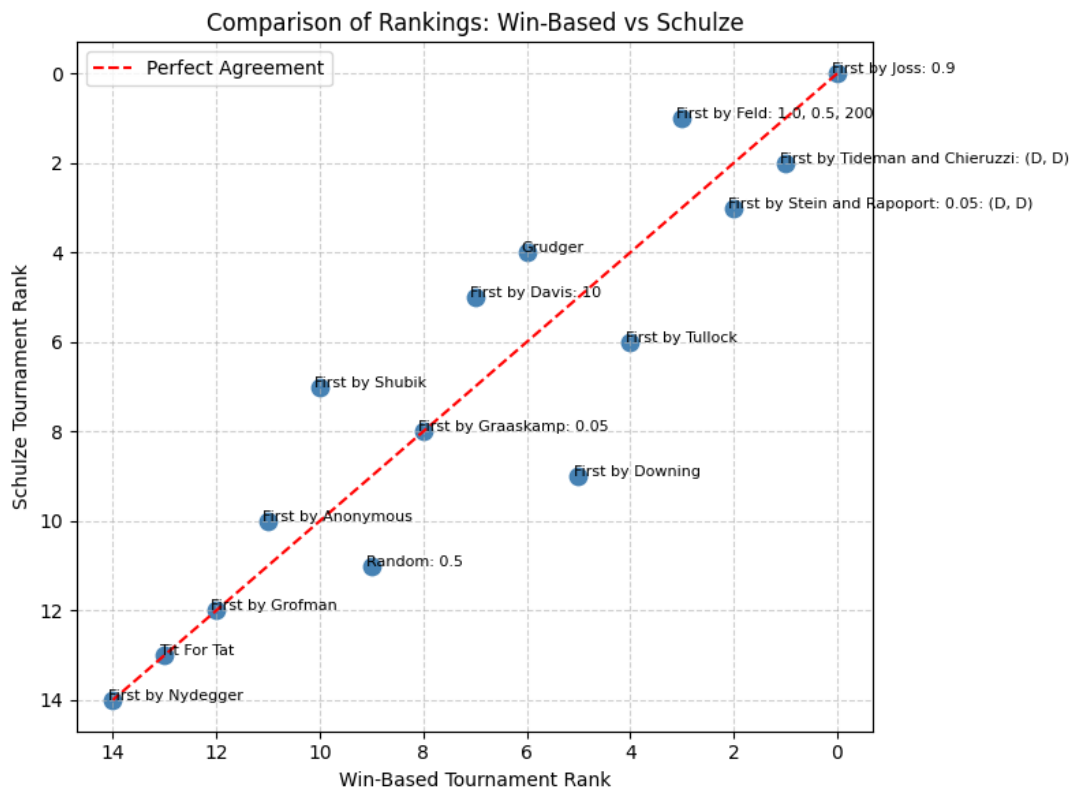

Figure 2: Comparison of Schulze and win-based

This strong positive correlation can be attributed to the conceptual similarity between the two approaches - both evaluate strategies based on the outcomes of pairwise interactions. The primary difference is that the Schulze method additionally considers the strength of indirect victories, enabling it to resolve cyclic dominance relationships and produce a more stable overall ranking. Due to this methodological proximity, the subsequent analysis focuses primarily on the Schulze method, as it provides a richer and more general framework for assessing strategy dominance and robustness.

As illustrated in Figure 3, five out of fifteen strategies experienced a decrease in their ranking positions when transitioning from the classic tournament format to the Schulze method. This indicates that while these strategies achieved high average scores overall, they were less dominant in direct pairwise comparisons. Conversely, eight strategies improved their positions under the Schulze ranking, suggesting that their relative strength in head-to-head matchups was underestimated by the traditional average-score approach.
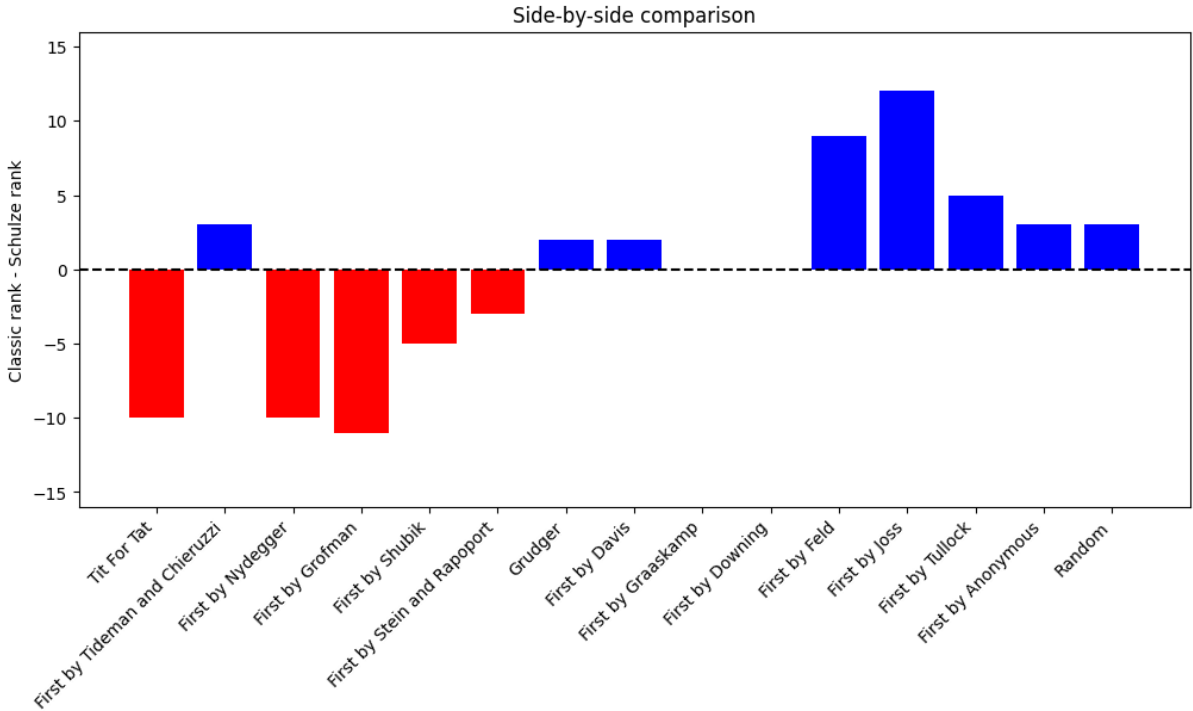

Figure 3: Classic to Schulze difference

The most significant improvement in ranking is demonstrated by First by Joss, which moved twelve positions upward under the Schulze evaluation. First by Joss is a pretty aggressive strategy - it begins by cooperating but introduces random defections with a small probability, even when the opponent has been cooperative. This behavior allows it to exploit overly trusting strategies while still maintaining some level of cooperation against reciprocal ones. Its sharp rise in the Schulze ranking suggests that it performs particularly well in direct confrontations because of its aggressive nature. This outcome is especially noteworthy, as it contrasts with Axelrod's original findings [4], which emphasized that successful strategies tend to be nice - avoiding unprovoked defections and fostering mutual cooperation.

Such results suggest that the new tournament environment created by the Schulze method is inherently more provocative, favoring strategies that adopt more aggressive or exploitative behavior. To evaluate this hypothesis, a set of correlations was computed,

	Classic Rank	Win-Based Rank	Schulze Rank
Median_score	-1.0	0.235	0.142
Cooperation_rating	-0.835	0.635	0.553
Initial_C_rate	-0.548	-0.132	-0.334
CC_rate	-0.903	0.389	0.321
DD_rate	0.714	-0.671	-0.632
CD_rate	0.374	0.560	0.725
DC_rate	0.575	-0.253	-0.142
CC_to_C_rate	-0.846	0.336	0.223
DD_to_C_rate	0.165	0.617	0.726

Table 6: Correlations of rankings with strategies features

as presented in Table 6, in order to quantify how the distribution of ranks shifts with respect to strategy cooperation levels and defection tendencies.

It can be seen from the Table 6 that the correlation with cooperation rating, CC rate, and CC-to-C rate is high and negative for the classic ranking, indicating that strategies inclined toward cooperation tend to achieve lower (i.e., better) ranks. In contrast, for both the win-based and Schulze methods, these features show medium positive correlations with rankings, while the DD rate is negatively correlated, implying that the best-performing strategies defect more frequently. It is also noteworthy that, for the Schulze and win-based methods, the DD-to-C rate - a measure of how often strategies cooperate following mutual defection - exhibits a moderately high positive correlation with ranking. This suggests that, under the new formats, the most successful strategies are less forgiving.

The results of the correlation analysis therefore support the hypothesis that the new tournament formats favor more aggressive and less cooperative behaviors. The best-performing strategies tend to be less forgiving, more provocative, and less inclined toward sustained cooperation. This indicates that the new tournament environments differ fundamentally from the original Axelrod tournaments and may provide better tools for exploring behavioral phenomena that classical tournament structures cannot fully capture.

5   Bigger tournament

To validate the conclusions drawn in the previous section, a larger tournament was conducted using an expanded set of strategies. Increasing the diversity of strategic behaviors allows for a more precise and comprehensive evaluation of the tournament formats. The inclusion of a wider range of strategies helps to reduce bias caused by specific interactions in smaller groups and provides a more robust basis for comparing the effectiveness and behavioral implications of the classic, win-based, and Schulze ranking methods.

5.1   Choosing the strategies

To conduct the larger tournament, the first step was to select the strategies to be included. The Axelrod-Python library provides convenient classification tools that allow filtering strategies based on their behavioral and computational properties. This ensures that only

valid and compatible strategies are used in the experiment.

In this work, several restrictions were applied. Only strategies that do not manipulate the internal state of opponents, do not inspect or alter source code, and do not require excessive computational time were included. These criteria guarantee that all selected strategies operate fairly and can complete their executions within reasonable time limits. Applying these filters yields a large and diverse set of valid strategies that consists of 224 strategies. Full list of strategies in CSV format can be found at this link. The other parameters for running the tournament did not change and can be found in Table 2.

5.2 Running the tournament

After running the tournament, the collected data was analyzed using all three ranking methods discussed earlier - Classic (average-score based), win-based, and Schulze. In the larger-scale tournament, the correlations between these ranking methods changed notably. For instance, there is now a moderate negative correlation (-0.5462) between the Classic and win-based approaches, suggesting that strategies achieving higher average scores tend to lose more frequently in direct pairwise encounters. The correlation between the Classic and Schulze rankings is close to zero (0.0638), indicating that average score has almost no influence on performance within the Schulze method.

The most striking difference, however, appears in the correlation between the Schulze and win-based rankings: while previously these two methods were very closely aligned, the current correlation has decreased to a moderate level (0.5363). This shift implies that as the number and diversity of strategies increase, the two methods begin to capture distinct aspects of strategic dominance. Consequently, both new formats should be included in further analysis, as they yield meaningfully different insights into strategic performance.

In the Table 7 we can see the top five performers by Schulze ranking and their main features, and Figure 4 shows the most changes in ranks by strategies.

Schulze_Rank	Classic_Rank	Win_Based_Rank	Name	Median_score
0	213	0	Aggravater	1.868946
1	223	1	Defector	1.796143
2	178	25	SolutionB5	2.020022
3	144	44	Suspicious Tit For Tat	2.208274
4	208	26	UsuallyDefects	1.877556

Table 7: Schulze top rankings

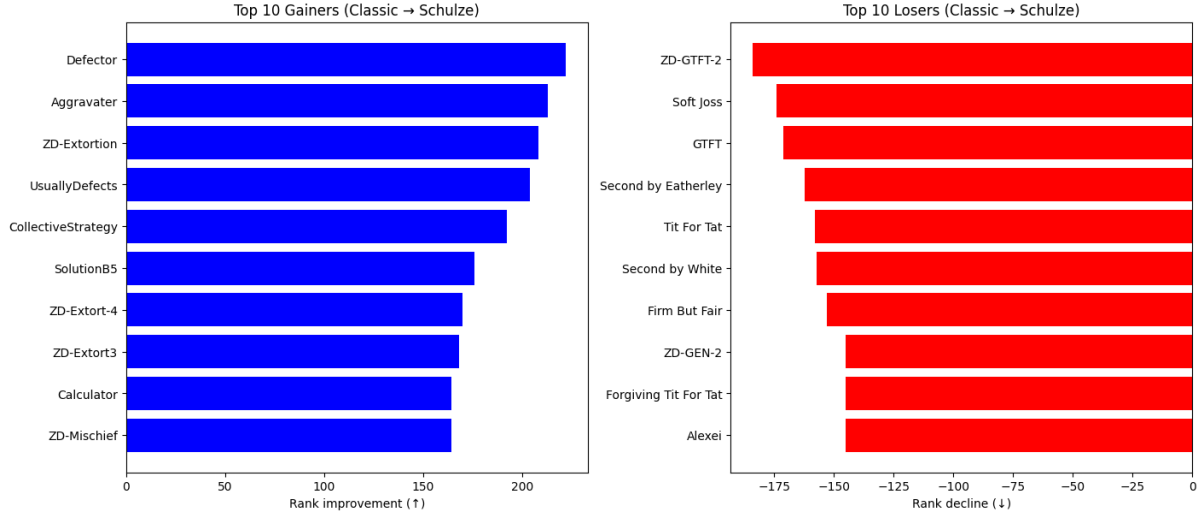| Win_Based_Rank	Rank	Schulze_Rank	Name	Median_score
0	213	0	Aggravater	1.868946
1	223	1	Defector	1.796143
2	197	5	CollectiveStrategy	1.897399
3	216	52	ZD-Mischief: 0.1, 0.0, 1	1.864305
4	214	6	ZD-Extortion: 0.2, 0.1, 1	1.865516

Table 8: Win-Based top rankings

Figure 4: Classic to Schulze most change in rank

The same characteristics and graph are shown for win-based method by Table 8 and Figure 5.
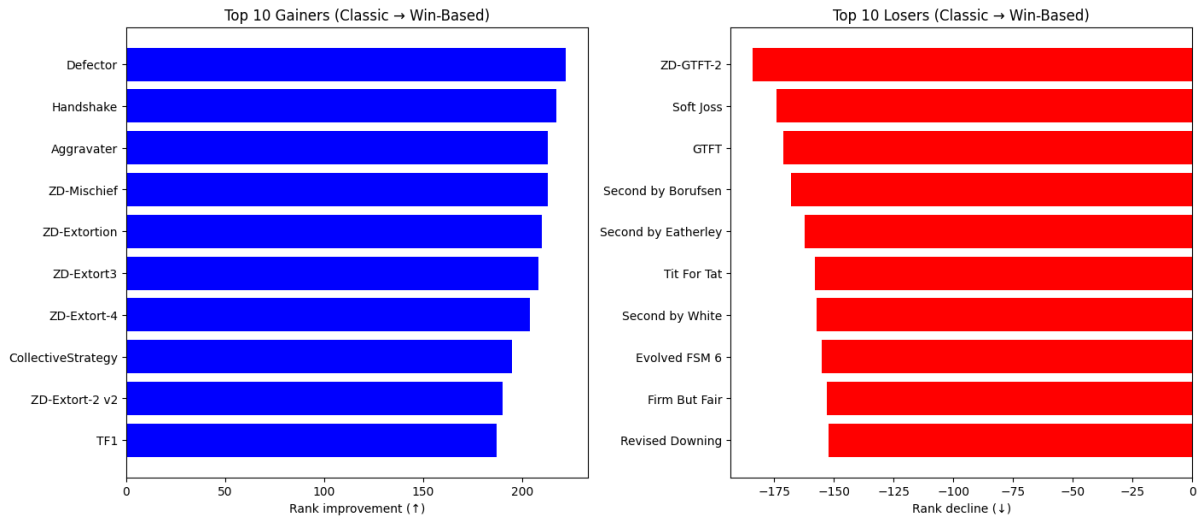

Figure 5: Classic to Win-Based most change in rank

Overall, when transitioning from the Classic to the Schulze ranking, 47.77% of strategies improved their positions, while 51.79% experienced a decline. In contrast, under the win-based method, exactly 50% of strategies improved their rank. It is also worth to note that for Schulze method only approximately 8% of strategies improved their position by

	Classic Rank	Win-Based Rank	Schulze Rank
Median_score	-1	0.546	-0.063
Cooperation_rating	-0.454	0.837	0.629
Initial_C_rate	-0.495	0.317	0.005
CC_rate	-0.727	0.843	0.381
DD_rate	0.387	-0.765	-0.650
CD_rate	0.510	0.078	0.607
DC_rate	0.402	-0.716	-0.427
CC_to_C_rate	-0.505	0.494	0.075
DD_to_C_rate	0.199	0.076	0.375

Table 9: Correlation table

more than half of their rank, and 10% declined by more than half of their rank. In contrast, for win-based method these values are 22% and 24% respectively. Figure 6 shows the distribution of losses/gains in ranks for both methods.
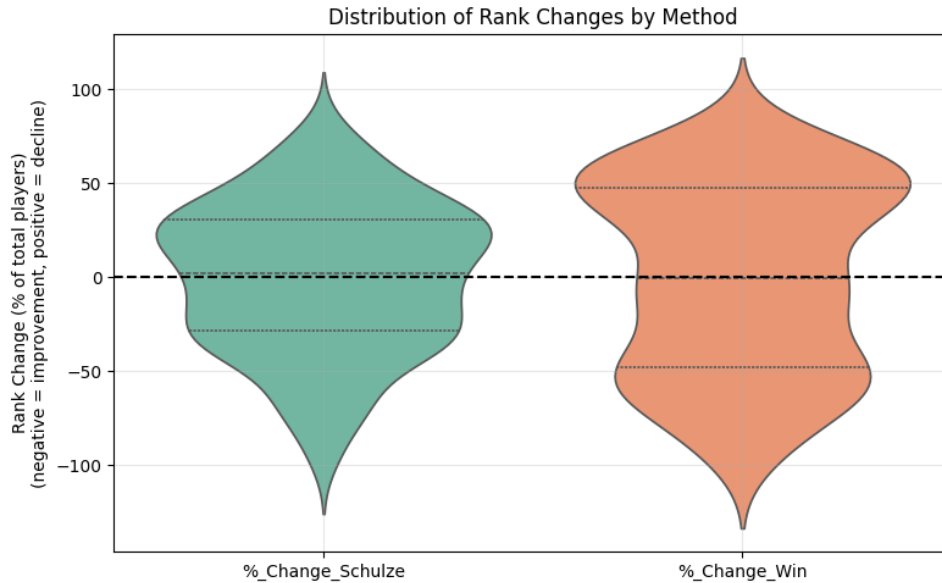

Figure 6: Distribution of rank changes

It is also notable that the top two positions remain identical in both the win-based and Schulze rankings - Aggravater and Defector. Both strategies also appear within the top 10 most improved ranks across both comparisons. This consistency can be attributed to their highly defect-oriented behavior, which allows them to dominate pairwise matchups and remain largely unbeatable in head-to-head encounters.

The correlation analysis has been performed on both Schulze and win-based methods, the table containing the summary of this analysis is shown in Table 9.

From Table 9 we can see that correlation with cooperation rating for classic ranking method dropped significantly in the new tournament, compared with the recreation of first Axelrod's tournament, showing that with bigger pool of strategies cooperation doesn't affect average score as significantly. For Schulze and win-based rankings that correlation dropped slightly, but still remains moderately high, confirming the conclusions from last section. Overall, the correlation table doesn't differ much from the table in section 4,

12

which suggest that even on the larger scale, the environment created by two new methods of ranking still remains hostile and unforgiving and favors "meaner" strategies.

To further confirm these conclusions, it has been decided to fit the multivariate linear regression model for each of the ranking methods. For regression coefficients to be more easily understood and explained, we will use the normalized rank instead of simple rank in our model. The normalized rank takes on values in range from 0 to 1 and is calculated by following formula[5]:

$$R_n = R/N,$$

where R – the absolute rank of the strategy, and N is the total number of strategies in the tournament. As can be seen from formula above, $R_n = 0$ corresponds to the best performing strategy, and $R_n = 1$ corresponds to worst-performing one.

After fitting the linear regression model to the data obtained from the tournament, and omitting some features to avoid multicollinearity, we obtain the model shown in Table 10.

	Classic ranking		Schulze ranking		Win-Based ranking	
	Coefficient $R^2_{adj} = 0.952$	**p-value**	**Coefficient** $R^2_{adj} = 0.493$	**p-value**	**Coefficient** $R^2_{adj} = 0.707$	**p-value**
constant	2.7543	0.000	1.8976	0.000	0.3499	0.017
Median score	-1.0241	0.000	-0.7074	0.000	-0.1406	0.048
Initial C rate	0.0079	0.555	0.0198	0.651	-0.0303	0.363
CC rate	0.3462	0.000	1.0052	0.000	1.4131	0.000
DD rate	0.1422	0.033	-0.5800	0.008	-0.0145	0.930
CC_to_C_rate	-0.0323	0.259	-0.1644	0.079	-0.3072	0.000
DD_to_C_rate	0.0003	0.988	0.1095	0.099	0.1022	0.043

Table 10: Regression coefficients across different tournament environments

The regression model we got is not surprising and largely confirms the conclusions from above – for instance it confirms the negative effect of high CC rate on strategy performance in both Schulze and win-based tournaments, implying once again that new formats tend to favour aggressive behavior. The relatively high adjusted R^2 values, especially for the Classic and win-based models, indicate that the selected variables explain a significant proportion of variation in normalized ranks. However, the lower R^2 for the Schulze ranking suggests that the Schulze method may have more complex, non-linear dominance relationships that cannot be fully described by linear dependence on basic behavioral features.

5.3   Section summary

In this section, a larger tournament was conducted using 224 strategies from the Axelrod-Python library and three different ranking methods - Classic (placement based on average score), win-based (number of wins in pairwise matchups), and Schulze (winners determined by beatpaths). This experiment was designed to validate the conclusions drawn from the smaller-scale tournament involving only 15 strategies.

Overall, the main observations identified in the smaller 15-strategy tournament – such as the reduced importance of average score and the rise of more unforgiving strategies, were confirmed on a larger scale, clarifying how the alternative ranking formats reshape the competitive landscape. In particular, the Schulze ranking highlights several notable patterns:

1. Winning strategies are no longer "nice" - they exhibit more suspicious and provocative behavior;

2. Average score is no longer a significant factor influencing strategy performance;

3. Top-performing strategies tend to be highly unforgiving and less cooperative.

These findings reinforce the hypothesis that the new ranking formats fundamentally reshape the competitive dynamics of the IPD. In the next section, the obtained results are discussed in the broader context of strategy evolution and cooperation theory, drawing final conclusions from the conducted experiments.

6   Conclusion

The study presented in this work explored the effects of alternative tournament ranking methodologies in the Iterated Prisoner's Dilemma, with the goal of providing a more comprehensive understanding of how different evaluation frameworks influence the identification of dominant strategies. Building upon the foundation of Robert Axelrod's tournaments, this research sought to extend the classic setup by introducing and comparing three ranking systems: the traditional average-score (Classic) method, a win-based method emphasizing direct pairwise success, and the Schulze method, adapted from preferential voting theory to resolve cyclic dominance relationships among strategies.

The first stage of the analysis involved replicating the conditions of Axelrod's original tournament, consisting of 15 well-known strategies. Comparison of results across the three formats revealed substantial differences in ranking outcomes. Notably, the correlation between the Classic and alternative methods was weak, indicating that strategies performing well on the basis of average score did not necessarily excel in direct confrontations. The Schulze and win-based rankings, in particular, tended to favor less cooperative, more provocative strategies - in contrast to the cooperative and forgiving nature of the top performers in Axelrod's experiments. This shift in dominance patterns demonstrated that the evaluation metric itself can fundamentally alter which behaviors are rewarded within the competitive environment.

The most significant transformation was observed in the performance of the *First by Joss* strategy, which rose dramatically in the Schulze ranking compared to its placement under the Classic approach. Its aggressive behavior - characterized by spontaneous defections even against cooperative opponents - proved advantageous in head-to-head contests. This finding challenged the long-standing notion that "nice" strategies always prevail, suggesting that the Schulze-style environment rewards caution and opportunism over unconditional trust. These results laid the foundation for the hypothesis that the new tournament structures create a more antagonistic and less cooperative ecosystem.

To validate these insights, the study extended the experiment to a much larger tournament involving 224 strategies. The expanded scope introduced greater behavioral diversity, allowing for a more statistically robust examination of patterns. Correlation analysis across ranking methods again confirmed the earlier observations: the Classic ranking maintained a weak negative correlation with the new methods, while the Schulze and win-based rankings showed moderate agreement. This divergence emphasized that the alternative evaluation systems capture distinct dimensions of performance - efficiency in the classic sense versus dominance and resilience in direct interactions.

Regression analysis further illuminated the structural differences among the ranking methods. In the classic tournament, performance was strongly associated with high median score and consistent cooperation, echoing the cooperative ethos of Axelrod's findings. However, for both the win-based and Schulze rankings, the influence of cooperative indicators diminished significantly. Instead, metrics reflecting unprovoked defections and low forgiveness rates gained importance, confirming that in these new competitive frameworks, success favors strategies that are assertive, less forgiving, and strategically cautious.

The broader implications of these findings extend beyond the numerical rankings themselves. They illustrate how seemingly minor methodological changes - such as redefining what constitutes a "win" - can reshape the entire strategic landscape of the Iterated Prisoner's Dilemma. In essence, the payoff structure of social interaction models

is not solely determined by the game's matrix, but also by the criteria used to evaluate success. By rewarding dominance over cooperation, the new ranking systems simulate environments where mistrust, competition, and provocation become stable strategies - conditions that may better mirror certain real-world social and economic systems.

Future work could further investigate how these ranking methods interact with strategy populations in dynamic or noisy environments. Incorporating incomplete information into the model would enable simulation of more realistic social ecosystems, bridging the gap between theoretical prediction and empirical behavior. Additionally, hybrid ranking mechanisms that balance dominance and cooperation metrics could provide more nuanced perspectives on the evolution of trust in competitive societies.

In conclusion, the research demonstrates that the structure of the tournament - not only the strategies within it - fundamentally shapes the dynamics of competition and cooperation. By extending the classical framework of the Iterated Prisoner's Dilemma through win-based and Schulze methods, this project reveals how alternative evaluation paradigms can lead to qualitatively different equilibria. The results underscore the importance of critically re-examining traditional metrics of success in strategic modeling and highlight the potential of multi-format tournaments as a tool for exploring complex adaptive behaviors in social and computational systems.

References

[1] Kalliopi Kastampolidou, Maria Nefeli Nikiforos, and Theodore Andronikos. A brief survey of the prisoners' dilemma game and its potential use in biology. *GeNeDis 2018: Computational Biology and Bioinformatics*, pages 315–322, 2020.

[2] Christophe Boone, Bert De Brabander, and Arjen Van Witteloostuijn. The impact of personality on behavior in five prisoner's dilemma games. *Journal of Economic Psychology*, 20(3):343–377, 1999.

[3] Philip Pettit. The prisoner's dilemma and social theory: An overview of some issues. *Politics*, 20(1):1–11, 1985.

[4] Robert Axelrod. Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, 24(1):3–25, 1980.

[5] Nikoleta E. Glynatsi, Vincent Knight, and Marc Harper. Properties of winning iterated prisoner's dilemma strategies. *PLOS Computational Biology*, 20(12):1–19, 12 2024.

[6] David B Fogel. Evolving behaviors in the iterated prisoner's dilemma. *Evolutionary Computation*, 1(1):77–97, 1993.

[7] Ethan Akin. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems*, pages 77–107, 2016.

[8] Vince Knight, Owen Campbell, Marc Harper, VSN Reddy Janga, James Campbell, Karol M. Langner, Sourav Singh, eric-s s, Nikoleta, Thomas Campbell, Jason Young, Julie Rymer, Geraint Palmer, Kristian Glass, edouardArgenson, Martin Jones, Cameron Davidson-Pilon, alajara, Ranjini Das, Marios Zoulias, Aaron Kratz, Timothy Standen, Paul Slavin, Adam Pohl, andy boot, WonkySpecs, Jochen Müller, Georgios Koutsovoulos, Areeb Ahmed, and Tomáš Ehrlich. Axelrod-python/axelrod: v3.0.0, June 2017.

[9] Markus Schulze. The schulze method of voting. *arXiv preprint arXiv:1804.02973*, 2018.

[10] Markus Schulze. A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Social choice and Welfare*, 36(2):267–303, 2011.