Suspicion: Recognising and evaluating the effectiveness of extortion in the Iterated Prisoner's Dilemma

Vincent A. Knight Nikoleta E. Glynatsi

June 24, 2018

Abstract

The Iterated Prisoner's Dilemma is a model for rational and evolutionary interactive behaviour. It has applications both in the study of human social behaviour as well as in biology. It is used to understand when and how a rational individual might accept an immediate cost to their own utility for the direct benefit of another.

Much attention has been given to a class of strategies called Zero Determinant strategies. It has been theoretically shown that these strategies can "extort" any player.

In this work, an approach to identify if observed strategies are playing in an extortionate way is described. Furthermore, experimental analysis of a large tournament with 204 strategies is considered. In this setting the most highly performing strategies do not play in an extortionate way against each other but do against lower performing strategies. This suggests that whilst the theory of Zero Determinant strategies indicates that memory is not of fundamental importance to the evolution of cooperative behaviour, this is incomplete.

1 Introduction

Agent based game theoretic models have become a stalwart of the underpinning mathematics of interactive behaviours. One of the major pieces of work in this area is the pair of original computer tournaments run by Robert Axelrod [2, 3]. These tournaments pitted submitted computer strategies against each other in plays of the Iterated Prisoner's Dilemma. A common game where agents can choose to pay a slight cost to their immediate utility in the hope of building a reputation. This has been used in economic and evolutionary game theory to understand the evolution of cooperative behaviour.

Recently, a class of strategies was described in [16] that can provably extort any given opponent. In [8, 12] some questions have already been asked about the true effectiveness of these strategies in an evolutionary setting. Here another question is asked: is it possible to recognise this extortionate behaviour? A mathematical procedure for suspicion is presented: in the same way that the continued actions of an extortionate individual might raise suspicion.

This work makes use of the Axelrod Python library [13, 11] with a large number of Prisoner Dilemma strategies available to give an extensive numerical example of the ideas presented. The approach is presented in Section 2. All of the code and data discussed in Section 3 is open sourced, archived and written according to best scientific principles [21]. The data archive can be found at [10].

2 Recognising Extortion

In [16], given a match between 2 memory-one strategies, the concept of Zero Determinant (ZD) strategies is introduced. The main result of that paper shows that given two memory one players $p, q \in \mathbb{R}^4$ a linear relationship between the players' scores could be forced by one of the players.

Using the notation of [16], assuming the utilities for player p are given by $S_x = (R, S, T, P)$ and for player q by $S_y = (R, T, S, P)$ and that the stationary scores of each player is given by S_X and S_Y respectively. The main result of [16] is that if

$$\tilde{p} = \alpha S_x + \beta S_y + \gamma \tag{1}$$

or

$$\tilde{q} = \alpha S_x + \beta S_y + \gamma \tag{2}$$

where $\tilde{p} = (1 - p_1, 1 - p_2, p_3, p_4)$ and $\tilde{q} = (1 - q_1, 1 - q_2, q_3, q_4)$ then:

$$\alpha S_X + \beta S_Y + \gamma = 0 \tag{3}$$

Here, the reverse problem is considered: given a $p \in \mathbb{R}^4$ how does one identify α, β, γ if they exist.

To do this, equation (1) is expressed linear algebraically as:

$$Mx = \tilde{p} \qquad x = (\alpha, \beta, \gamma) \tag{4}$$

with $M \in \mathbb{R}^{4\times 3}$ given by:

$$M = \begin{bmatrix} R & R & 1 \\ S & T & 1 \\ T & S & 1 \\ P & P & 1 \end{bmatrix} \tag{5}$$

Note that in general, equation (4) will not necessarily have a solution. From the Rouché-Capelli theorem if there is a solution it is unique as $\text{rank}(M) = 3$ which is the dimension of the variable x. Furthermore, removing a single row of M would ensure that the corresponding matrix is invertible. This corresponds to the fact that a ZD strategy is defined by only 3 of its values.

As an example, consider the known ZD strategy $p = (8/9, 1/2, 1/3, 0)$ from [18] which is referred to as `Extort-2`. In the standard case of $(R, S, T, P) = (3, 0, 5, 1)$ the inverse of $M_{(4)}$ (removing the last row of M) is given by:

$$M_{(4)}^{-1} = \begin{bmatrix} 1 & -\frac{3}{5} & -\frac{2}{5} \\ 1 & -\frac{2}{5} & -\frac{3}{5} \\ -5 & 3 & 3 \end{bmatrix} \tag{6}$$

The $x = (\alpha, \beta, \gamma)$ that corresponds to $\tilde{p}$ is given by:

$$x = M_{(4)}^{-1}\tilde{p}_{(4)} = \begin{bmatrix} \frac{1}{18} & -\frac{1}{9} & \frac{1}{18} \end{bmatrix} \tag{7}$$

Using (4) gives that these values lead to the correct value for $p_4 = 0$ which confirms that p is a ZD strategy.

This approach could in fact be used to confirm that a given strategy is acting in a ZD manner even if it is not a memory one strategy. However, in practice, if a closed form for p is not known, then due to measurement and/or numerical error this would not work.

Thus, an approach based on least squares [6] is proposed. This approach finds the best fitting $\bar{x} = (\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ which minimises:

$$\text{SSError} = \|Mx - \tilde{p}\|_2^2 = \sum_{i=1}^{4} \left((M\bar{x})_i - \tilde{p}_i\right)^2 \tag{8}$$

Note that SSError, which is the square of the Frobenius norm [6], becomes a measure of how close a strategy is to being a ZD strategy. In [16] a particular type of ZD is defined, if:

$$P(\alpha + \beta) + \gamma = 0 \tag{9}$$

then the player can ensure they get a score χ times larger than the opponent. This extortion coefficient is given by:

$$\chi = \frac{-\beta}{\alpha} \tag{10}$$

Thus, if (9) holds and $\chi > 1$ a player is said to extort their opponent. Recalling (5), equation (9) corresponds to the last row of M, thus a ZD player p with low enough SSError extorts their opponent if and only if $p_4 = 0$ and the estimated $\chi > 1$. Using the method of least squares, the strategy can be considered ZD for a given threshold of SSError and χ can be approximated. Figure 1 illustrates this for potential strategies: for all of these we see that for the strategy to be extortionate p_2 and p_3 seem to have a linear relationship: for example if $p_1 = 1$ and $p_4 = 0$ then all strategies where $p_2 = p_3$ are extortionate. Suspicion of extortion then corresponds to a threshold on SSError.

By observing interactions (human or otherwise), the transition rates of given players, their memory one representation can be approximated and this approach can be used to recognise extortionate behaviour. The notion of comparing theoretic and actual plays of the IPD is not novel, see for example [17]. Immediately it is noted that if the environment is noisy [22] then no strategy can be considered to be extortionate as $p_4 > 0$.

In the next section, this idea will be illustrated by observing the interactions that take place in a computer based tournament of the IPD.

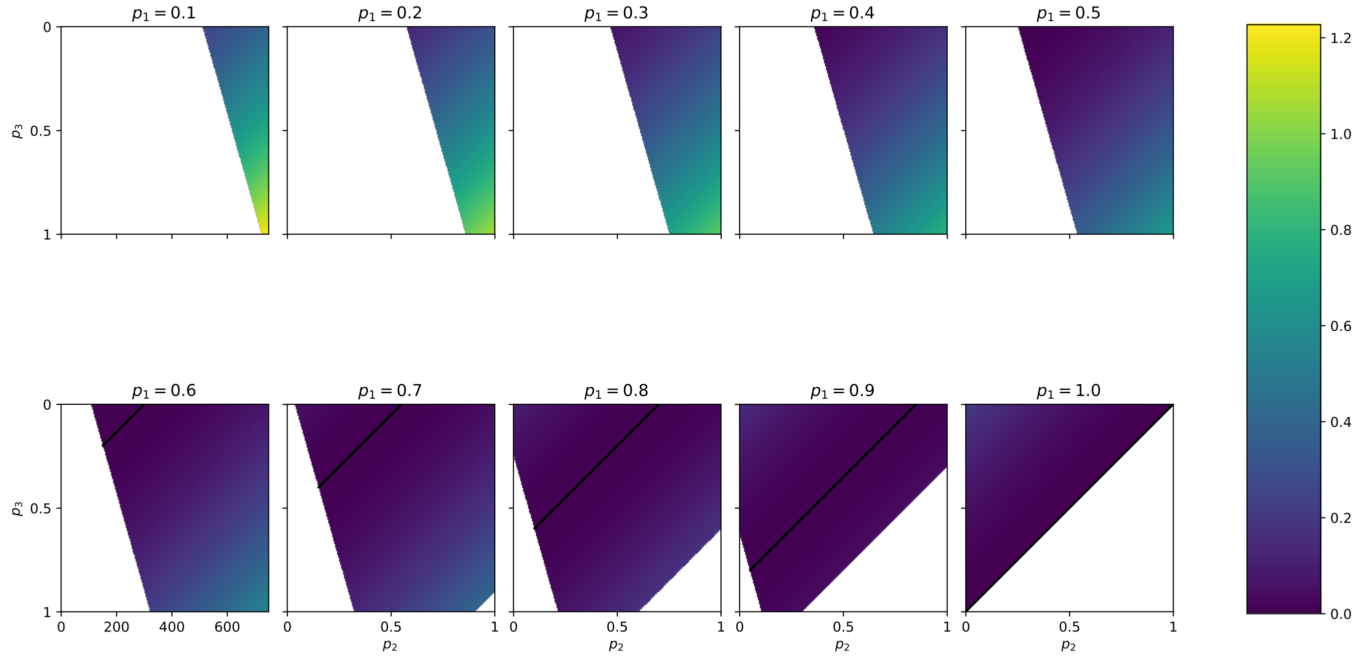

Figure 1: Strategies of the form: $p = (p_1, p_2, p_3, 0)$. The solid line shows the values for which SSError $< 10^{-6}$. Only values for which $\chi > 1$ are displayed, strategies outside of this region can never be extortionate.

3 Numerical experiments

In [18] results from a tournament with 19 strategies, was presented with specific consideration given to ZD strategies. This tournament is reproduced here using the Axelrod-Python project [11]. To obtain a good measure of the corresponding transition rates for each strategy all matches have been run for 2000 turns and every match has been repeated 60 times. All of this interaction data is available at [10].

Using the pair wise interactions the transition rates p, q can be measured and the steady state probabilities inferred and compared to the actual probabilities of each state. This is done numerically by computing the singular eigenvector of the matrix A [19]:

$$A = \begin{bmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_2 & p_2(1-q_2) & (1-p_2)q_2 & (1-p_2)(1-q_2) \\ p_3q_3 & p_3(1-q_3) & (1-p_3)q_3 & (1-p_3)(1-q_3) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{bmatrix}$$

Figure 2 shows a regression line fitted to every pairwise interaction with a reported SSError value (pairwise interactions with missing states were omitted). This serves to validate the approach: a part from some edge cases the relationship is consistent.

Figure 3 shows the SSError values for all the strategies in the tournament, as reported in [18] the extortionate strategy (which has an expected SSError approximately 0) gains a large number of wins.

Here, the work of [18] is extended by investigating a tournament with 204 strategies.

The results of this analysis are shown in Figure 4. The top ranking strategies by number of wins seem to be extortionate (but not against all strategies) and it can be seen that a small sub group of strategies achieve mutual defection. All the top ranking strategies according to score achieve mutual cooperation and do not extort each other, however they **do** exhibit extortionate behaviour towards a number of the lower ranking strategies.

4 Conclusion

This work defines an approach to measure whether or not a player is playing a strategy that corresponds to an extortionate strategy as defined in [16]: a mathematical model for suspicion. Indeed Figure 1 classifies all extortionate strategies. This is done through a linear algebraic approach for approximating the solution of a linear system. Using this, a large number of pairwise interactions is simulated and in fact very few strategies are found to act extortionately.

The work of [16], whilst showing that a clever approach to taking advantage of another memory one strategy exists: this is incomplete. Whilst the elegance of this result is very attractive, just as the simplicity of the victory of Tit For Tat in Axelrod's original tournaments was, it is incomplete. Extortionate strategies achieve a high number of wins

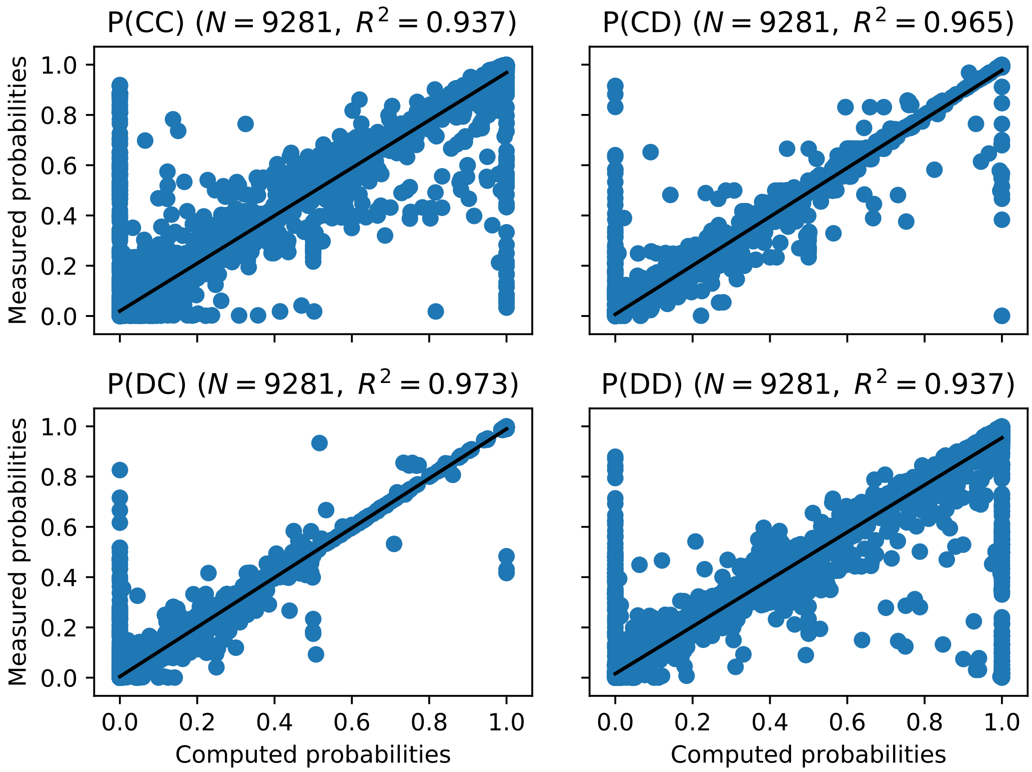

Figure 2: The relationship between the steady state probabilities inferred from the measured transitions and the actual steady state probabilities. A linear regression line is included validating the approach.

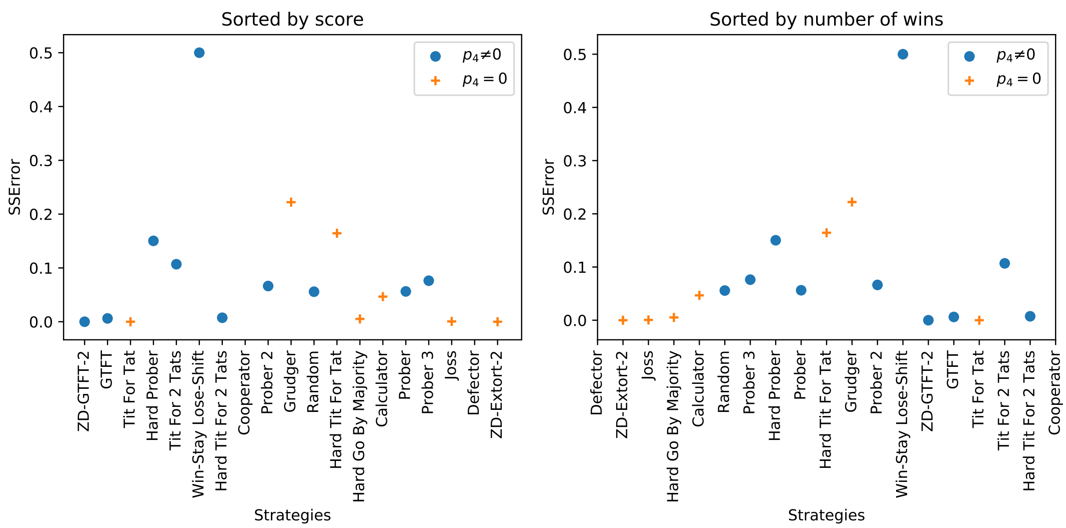

Figure 3: SSError for the strategies of [18], ordered both by number of wins and overall score. Cooperator and Defector are omitted as they do not visit all the states.

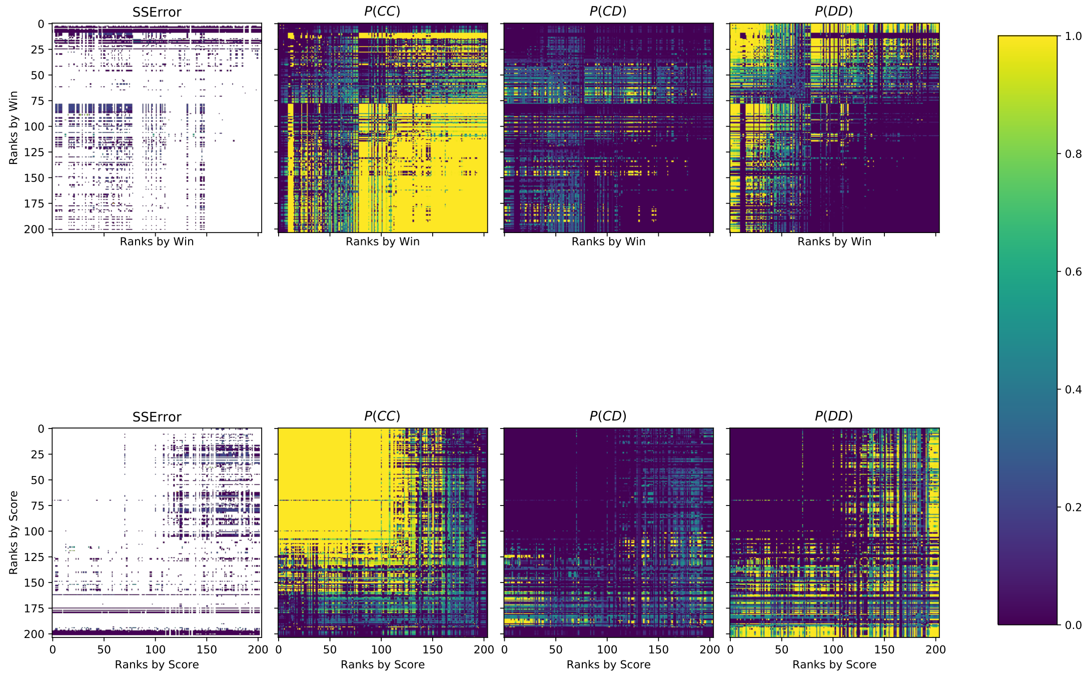

Figure 4: SSError for the strategies for the full tournament. Only strategy interactions for which $p_4 = 0$ and $\chi > 1$ are displayed.

but they do not achieve a high score which corresponds to the fitness landscape in an evolutionary sense. From the large number of interactions a payoff matrix P can be measured where P_{ij} denotes the score (using standard values of $(R, S, T, P) = (3, 0, 5, 1)$) of the ith strategy against the jth strategy. Using this, the replicator equation describes the evolution of the system based on a population density fitness function:

$$\frac{dx}{dt} = x(P - x^T P x) \tag{11}$$

Equation (11) is solved numerically through an integration technique described in [15] and Figure 5 shows the evolution of the distribution of the system: the various strategies are ranked by scores. It is clear to see that only the high ranking strategies survive the evolutionary process (in fact, only 18 have a final distribution greater than 10^{-2}). This confirms the findings of [12] in which sophisticated strategies resist evolutionary invasion of shorter memory strategies. Recalling Figure 4 this demonstrates that:

- Cooperation emerges through the evolutionary process: the high scoring strategies do not exhibit extortionate behaviour towards each other.
- Extortionate strategies do not survive the evolutionary process.

This work can be used to classify plays of the IPD: data can be collected from actual interactions (in lab or in the field). Furthermore, this allows for a classification method similar to the notion of fingerprinting presented in [1]. Trained strategies can potentially be classified as extortionate or not or it could be possible to even constraint the reinforcement learning approaches that are becoming prevalent in the literature. It is worth noting that, as described in [7], the top ranking strategies in the full tournament are obtained using reinforcement learning techniques, thus suspicion of extortionate behaviour could in fact be an evolutionary trait.

Acknowledgements

The following open source software libraries were used in this research:

- The Axelrod library (IPD strategies and tournaments) [11, 13].
- The matplotlib library (visualisation) [5].
- The pandas, dask and NumPy libraries (data manipulation) [20, 4, 14].
- The SciPy library (numerical integration of replicator equation) [9].

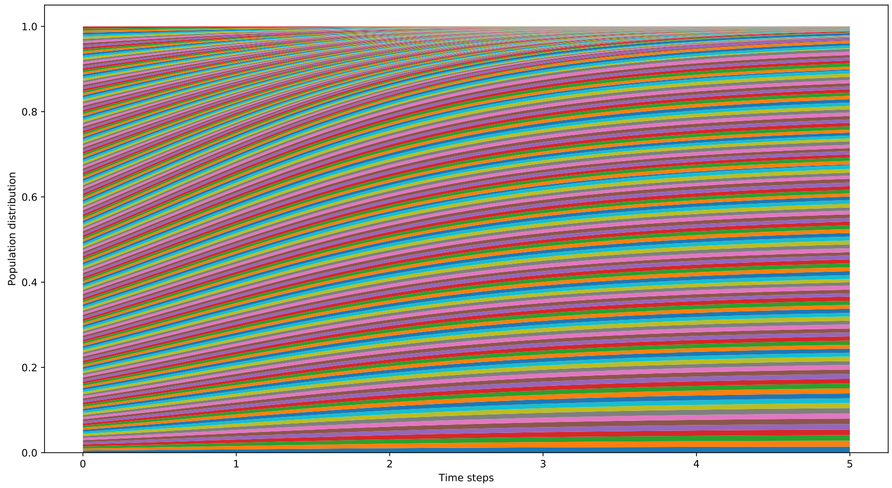

Figure 5: Numerical simulation of the replicator equation (11): strategies are ordered by score, only the strategies with a high score survive the evolutionary process.

References

[1] D. Ashlock and E.-Y. Kim. "Fingerprinting: Visualization and Automatic Analysis of Prisoner's Dilemma Strategies". In: *IEEE Transactions on Evolutionary Computation* 12.5 (Oct. 2008), pp. 647–659. DOI: 10.1109/tevc.2008.920675.

[2] Robert Axelrod. "Effective Choice in the Prisoner's Dilemma". In: *Journal of Conflict Resolution* 24.1 (Mar. 1980), pp. 3–25. DOI: 10.1177/002200278002400101.

[3] Robert Axelrod. "More Effective Choice in the Prisoner's Dilemma". In: *Journal of Conflict Resolution* 24.3 (Sept. 1980), pp. 379–403. DOI: 10.1177/002200278002400301.

[4] Dask Development Team. *Dask: Library for dynamic task scheduling.* 2016. URL: http://dask.pydata.org.

[5] Michael Droettboom et al. *Matplotlib/Matplotlib V2.2.2.* 2018. DOI: 10.5281/zenodo.1202077.

[6] Gene H. Golub. *Matrix Computations.* J. Hopkins Uni. Press, Jan. 7, 2013. ISBN: 1421407949. URL: https://www.ebook.de/de/product/20241149/gene_h_golub_matrix_computations.html.

[7] Marc Harper et al. "Reinforcement learning produces dominant strategies for the Iterated Prisoner's Dilemma". In: *PLOS ONE* 12.12 (Dec. 2017). Ed. by Yong Deng, e0188046. DOI: 10.1371/journal.pone.0188046.

[8] C. Hilbe, M. A. Nowak, and K. Sigmund. "Evolution of extortion in Iterated Prisoner's Dilemma games". In: *Proceedings of the National Academy of Sciences* 110.17 (Apr. 2013), pp. 6913–6918. DOI: 10.1073/pnas.1214834110.

[9] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open source scientific tools for Python.* [Online; accessed ¡today¿]. 2001–. URL: http://www.scipy.org/.

[10] Vincent Knight. *Raw data for: "Suspicion: Recognising and evaluating the effectiveness of extortion in the Iterated Prisoner's Dilemma".* June 2018. DOI: 10.5281/zenodo.1297075. URL: https://doi.org/10.5281/zenodo.1297075.

[11] Vincent Knight et al. "An Open Framework for the Reproducible Study of the Iterated Prisoner's Dilemma". In: *Journal of Open Research Software* 4 (Aug. 2016). DOI: 10.5334/jors.125.

[12] Vincent Knight et al. "Evolution Reinforces Cooperation with the Emergence of Self-Recognition Mechanisms: an empirical study of the Moran process for the iterated Prisoner's dilemma". In: 1707.

[13] Vince Knight et al. *Axelrod-Python/Axelrod: V4.2.0.* 2018. DOI: 10.5281/zenodo.1252994.

[14] Travis E. Oliphant. *Guide to NumPy: 2nd Edition.* CreateSpace Independent Publishing Platform, 2015. ISBN: 9781517300074. URL: https://www.amazon.com/Guide-NumPy-Travis-Oliphant-PhD/dp/151730007X?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=151730007X.

[15] Linda Petzold. “Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations”. In: *SIAM Journal on Scientific and Statistical Computing* 4.1 (Mar. 1983), pp. 136–148. DOI: 10.1137/0904010.

[16] W. H. Press and F. J. Dyson. “Iterated Prisoner’s Dilemma contains strategies that dominate any evolutionary opponent”. In: *Proceedings of the National Academy of Sciences* 109.26 (May 2012), pp. 10409–10413. DOI: 10.1073/pnas.1206569109.

[17] David G. Rand and Martin A. Nowak. “Human cooperation”. In: *Trends in Cognitive Sciences* 17.8 (Aug. 2013), pp. 413–425. DOI: 10.1016/j.tics.2013.06.003.

[18] A. J. Stewart and J. B. Plotkin. “Extortion and cooperation in the Prisoner’s Dilemma”. In: *Proceedings of the National Academy of Sciences* 109.26 (June 2012), pp. 10134–10135. DOI: 10.1073/pnas.1208087109.

[19] William J. Stewart. *Probability, Markov Chains, Queues, and Simulation.* Princeton Univers. Press, July 11, 2009. 760 pp. ISBN: 0691140626. URL: https://www.ebook.de/de/product/8052317/william_j_stewart_probability_markov_chains_queues_and_simulation.html.

[20] Data Structures et al. “PROC. OF THE 9th PYTHON IN SCIENCE CONF. (SCIPY 2010)”. In: 2010.

[21] Greg Wilson et al. “Best Practices for Scientific Computing”. In: *PLoS Biology* 12.1 (Jan. 2014). Ed. by Jonathan A. Eisen, e1001745. DOI: 10.1371/journal.pbio.1001745.

[22] Jianzhong Wu and Robert Axelrod. “How to Cope with Noise in the Iterated Prisoner’s Dilemma”. In: *Journal of Conflict Resolution* 39.1 (Mar. 1995), pp. 183–189. DOI: 10.1177/0022002795039001008.