CRITICAL TRANSITIONS AND EARLY WARNING SIGNALS IN REPEATED COOPERATION GAMES

CHRISTIAN HOFER
Institute of Systems Sciences, Innovation and Sustainability Research
University of Graz
Graz, Austria

GEORG JÄGER* AND MANFRED FÜLSSACK
Institute of Systems Sciences, Innovation and Sustainability Research
University of Graz
Graz, Austria

(Communicated by Gabrielle Demange)

Abstract. Scanning a system’s dynamics for critical transitions, i.e. sudden shifts from one system state to another, with the methodology of Early Warning Signals has been shown to yield promising results in many scientific fields. So far however, such investigations focus on aggregated system dynamics modeled with equation-based methods. In this paper the methodology of Early Warning Signals is applied to critical transitions found in the context of Cooperation Games. Since equation-based methods are not well suited to account for interactions in game theoretic settings, an agent-based model of a repeated Cooperation Game is used to generate data. We find that Early Warning Signals can be detected in agent-based simulations of such systems.

1. Introduction. Anticipating critical transitions, i.e. abrupt changes from one system state to a different system state is an important challenge in many scientific fields, like climatology, ecology, medicine, economics, or sociology.

In order to find statistical traces of impending regime shifts, an elaborate apparatus of methods has been introduced recently, comprised under the term Early Warning Signals (EWSs) [24, 6, 5]. Such methods have been successfully used in the fields of ecology [25], climate change [17, 16], and medicine [18, 20]. Both measured data, as well as data from computer simulations is investigated. Also in the field of social science (agent-based) computer simulations are performed, mainly in the fields of finance and bubble detection [27, 12, 13, 26] and the results are scanned for EWSs.

In this paper, we introduce an agent-based model of a Cooperation Game that is known to exhibit critical transitions from the state of overall cooperation to the state of no cooperation. We then analyze the data for EWSs and see if it is possible to detect the impending critical transition before it occurs.

2010 Mathematics Subject Classification. Primary: 37M10; Secondary: 62-07, 62J10, 62M10.

Key words and phrases. Critical transition, Early Warning Signals, Cooperation Games, repeated games, agent-based models.

* Corresponding author: georg.jaeger@uni-graz.at.
The agent-based approach makes it possible to model the system bottom-up, starting from the rules that govern the decisions of each individual agent. On the other hand, an equation-based top-down model would require an equation, ideally a differential equation that gives insight into the cooperation behavior of all the players involved in the game. Such a set of equations is very difficult to find and substantiating it is challenging, while the set of rules that are needed for the agent-based approach can be found intuitively and can be backed up by empirical evidence, which makes this approach well-suited for game theory.

The paper is organized as follows: Section 1 briefly summarizes aspects of stability analysis in dynamical systems and explains the theoretical background of EWSs. The statistical methods that are deployed in this investigation are introduced. Section 2 describes the model subjected to EWS analysis and details on this analysis. Section 3 presents results of both model and data analysis, while Section 4 discusses the significance of these results and implications for future research.

Systems theory describes dynamical systems as having an eigenbehavior [11] which makes them evolve towards stable states, so called equilibria. Internal interaction dynamics can cause systems to have several such equilibria, spanning basins of attraction. In the reach of these basins dynamics tend to return to equilibrium after (small) perturbations [19]. The edges of these attractor basins (or potential wells) mark tipping points (i.e. unstable equilibria) at which transitions to alternative stable states occur. Depending on the strength of positive feedback effects, these transitions can be abrupt, showing sudden shifts of regime [14], so called critical transitions.

With respect to the suddenness of these shifts, attempts have been made to derive signals for the approach of a tipping point at which critical transitions occur. One such indicator is the time a system needs to return to equilibrium after perturbation. The slowing down of recovery time, so called critical slowing down (CSD), has been found to indicate the loss of a system’s resilience and thus the approach of a transition [6].

CSD is visible in the form of various statistical signals. One of these signals concerns changes in the correlation structure of a time series, caused by an increase in the short-term memory (i.e. the correlation at low lags) of a system prior to a transition and is measured with lag-1 autocorrelation of consecutive observations [24]. Another indicator can be found in the tendency of a system to drift more widely around its stable state when approaching an unstable equilibrium, i.e. a tipping point. This causes the variance in the time series to increase, measured for example by the standard deviation. Furthermore, variance can show asymmetries and more extreme values when a system’s state gets close to the attraction of an alternative equilibrium. This causes skewness and kurtosis to change. All these signals are signs that can be used to predict an imminent regime shift.

2. Methodology.

2.1. The model. The model simulates a simple Cooperation Game that reproduces the empirical finding that without additional mechanisms such as punishment [8] contribution in Cooperation Games declines towards the free riding Nash equilibrium [22] of overall defection [15, 4]. In this model the transition between cooperation and non-cooperation is critical, which makes it ideal for studying EWSs. Agents are set up on a grid with periodic boundary conditions, so that every agent has exactly 8 neighbors, i.e. Moore neighborhood. Each round, all agents get an
endowment of 1 monetary unit and are asked if they want to invest their endowment into a common pool or not. After each agent made its decision, the common pool is multiplied by a synergy factor $s$ ($1 \leq s \leq 2$) and is then evenly distributed among all agents, even those who did not invest into the common pool. The only Nash equilibrium of this system is overall defection, since switching to defection always increases personal payoff, while decreasing the sum of all payoffs. So the state with the highest sum of all payoffs is a state in which everyone cooperates, a state completely opposite the Nash equilibrium. The simulation begins with nearly all agents cooperating and continues until the Nash equilibrium is reached.

The calculation of individual payoffs is straightforward. Agents, who decided to invest in the common pool, get the payoff $P_0$, which can be calculated via

$$P_0 = \frac{A_C \cdot s}{A_C + A_D},$$

with the synergy factor $s$, the number of cooperating agents $A_C$ and the number of defecting agents $A_D$. The defecting agents get the same share from the common pool $P_0$, but they also keep their endowment, increasing their overall payoff by 1 to $P_1$:

$$P_1 = P_0 + 1$$

Whether or not an agent decides to cooperate or defect depends on its personal cooperation probability $c_p$. $c_p$ is defined as the chance of choosing to cooperate ($0 \leq c_p \leq 1$) and is initialized with a value close to 1, e.g. 0.98. This cooperation probability is not constant, but can be influenced by one of two effects.

The first effect, which can either increase or decrease an agent’s cooperation probability, is called conditional cooperation [10]. If all neighbors do cooperate, the cooperation probability is increased by the change rate $c$. For each neighbor that does not cooperate, the cooperation probability is decreased by $8c$, with 8 being the number of neighbors that influence the decision (both direct neighbors and diagonal neighbors on the grid are considered). This value is chosen such that the negative influence of one defector can be compensated by the positive influence of all neighbors in a following turn. The numerical value of $c$ is arbitrary and it serves only as a scaling factor. For high $c$ the Nash equilibrium is reached fast, while a low $c$ leads to a longer time series. However, apart from the scaling, the value of $c$ has no influence on the qualitative behavior of the system.

The second effect, which can only decrease an agent’s cooperation probability, is called reference switching. Reference switching occurs every time the payoff of an agent is smaller than its investment, i.e. the agent experiences a net loss. In this case, the cooperation probability decreases proportional to the experienced loss:

$$c_{p_{i+1}} = c_{p_i} - 0.01 (I - P),$$

with the payoff $P$, the investment $I$, the cooperation probability in round $i$ $c_{p_i}$ and the cooperation probability in round $i+1$ $c_{p_{i+1}}$. That way, agents reduce their cooperation probability by 1% of the difference between investment and payoff.

Although more complex agent behaviors like learning [2, 23], fairness [9, 1] and social preferences [21, 3] would make the model more realistic, these effects are not considered, since then the system would be too complicated to guarantee critical transitions, which are of course essential for scanning for EWSs.

2.2. Data analysis. The EWS analysis that we applied focuses on the number of cooperators at each point in time. For this, a window of investigation was chosen
ending ahead of the breakdown of cooperation so that EWSs can be used to predict imminent collapses. Often in literature EWS analysis is applied to pure data [24, 5]. In the case of a Cooperation Game, however, the development of cooperation shows visible downward trends well before tipping. Since our interest was on the properties of the fluctuations, not on global shifts to lower values, we decided to detrend the data [7]. For this, a polynomial fit was calculated and subtracted from raw data. The resulting residuals were then analyzed for EWSs by calculating lag-1 autocorrelation, standard deviation, skewness and kurtosis for the data in appropriately sized rolling windows. Carefully choosing the size of these windows is essential. If the interval is too large, the signal gets averaged out; too small intervals lead to a random noise on top of the signal due to random fluctuations. We found that an interval of roughly 100 data points is a suitable window size for our system.

3. Results.

3.1. Simulation results. The described model is initialized with an agent population of 1000 agents on a grid, an initial cooperation probability of 0.98 and a synergy factor of 1.2. The simulation then runs until the Nash equilibrium is reached, which takes a few thousand turns with these parameters. In every turn the number of cooperating agents is saved, leading to a time series that can be analyzed for EWSs.

A typical result of such a simulation run is shown in Figure 1. The percentage of cooperators is high in the beginning and slowly decreases due to changes in the cooperation probability of the agents. After roughly 3500 turns a critical transition occurs, bringing the system abruptly to the state of overall defection. The important part of this time series is highlighted in orange. Here the critical transition is imminent, yet there is no abrupt change in the number of cooperators yet. This makes this data ideal for scanning for EWSs.

Figure 1. Time development of the rate of cooperators in a simulated Cooperation Game with a population of 1000 agents. The highlighted fraction of the time series (orange) was considered for EWS analysis.

Figure 2 shows details on the critical transition. In an equation-based model, one would expect the transition to have a shape similar to a Hill function, however, in an agent-based model there is no analytic function that describes the number of
cooperators. Nevertheless, the simulated number of cooperators shows very good agreement with the theoretically expected Hill function.

![Graph showing cooperation in a Cooperation Game.](image)

**Figure 2.** Critical transition of cooperation in a Cooperation Game. Simulated data (orange) is compared to a Hill function (blue) that is known to replicate critical transitions on an aggregated level.

3.2. **Scanning for EWSs.** Figure 3 shows the results of the EWS analysis for pure data (left column) as well as for detrended data (right column). The first panel in the left column shows a detailed view of the considered data that is highlighted in Figure 1 and the fit that is used to calculate the residuals. The first panel in the right column shows the residuals. The second row of panels reports on autocorrelation, the third row on standard deviation, the fourth row on skewness and the fifth row on kurtosis.

EWSs can be detected in both the pure data (left column) as well as in the detrended data (right column) for lag-1 autocorrelation, standard deviation and skewness. Only kurtosis shows no signal. For lag-1 autocorrelation there is more noise for the residuals, yet the sharp increase in autocorrelation shortly before the critical transition is more pronounced. Standard deviation shows a similar increase in both pure and detrended data, only the value of the standard deviation is shifted downwards for the detrended data, which is to be expected, since the polynomial fit was subtracted. The skewness shows qualitatively similar signals in both pure and detrended data, with a visible peak before the transition. Only the kurtosis does not produce an EWS. The main reason for this may be that an increase in kurtosis is a way of detecting flickering, i.e. the occurrence of more extreme states prior to the critical transition. Since the investigated system exhibits no such flickering the kurtosis does not produce an EWS.
4. Conclusion. We report on investigations of the possibility of analyzing the resilience of cooperation in simulated Cooperation Games. The analysis aims at predicting regime shifts with the concept of Early Warning Signals. An agent-based Cooperation Game model was designed to generate rapid regime shifts in response to accumulating positive feedback, and was used as an example case on which the possibilities of EWS analysis were tested. The tests considered two kinds of data: the easily observable bare number of cooperators and the detrended number of cooperators.
cooperators, filtering out slow changes in cooperation, thus focusing on fluctuations. In both cases EWS analysis showed clear indications of an approaching regime shift.

Thus it was possible to show that the concept of EWSs can be applied to system dynamics derived from interaction on component level and is thus apt to capture dynamics in games, here in particular the development of simulated social dilemmas. The EWS concept thus could serve as a powerful tool to predict regime shifts and critical transitions in this context. This result is not specific to the investigated system, but a universal property of such interactions. Considering the overall aim to derive possibilities of countering negative dynamics in the development of cooperation in social dilemmas or at least to derive warning signals, our results substantiate that EWSs in the form of variance or autocorrelation of observables can be used and that such an approach yields promising results in the context of an agent-based simulation of a Cooperation Game.

REFERENCES

[1] G. E. Bolton and A. Ockenfels, Self-centered fairness in games with more than two players, Handbook of Experimental Economics Results, 1 (2008), 531–540.
[2] M. N. Burton-Chellew, H. H. Nax and S. A. West, Payoff-based learning explains the decline in cooperation in public goods games, Proceedings of the Royal Society of London B: Biological Sciences, 282 (2015), 20142678.
[3] G. Charness and M. Rabin, Understanding social preferences with simple tests, The Quarterly Journal of Economics, 117 (2002), 817–869.
[4] A. Chaudhuri, Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature, Experimental Economics, 14 (2011), 47–83.
[5] L. Dai, D. Vorselen, K. S. Korolev and J. Gore, Generic indicators for loss of resilience before a tipping point leading to population collapse, Science, 336 (2012), 1175–1177.
[6] V. Dakos and J. Bascompte, Critical slowing down as early warning for the onset of collapse in mutualistic communities, Proceedings of the National Academy of Sciences, 111 (2014), 17546–17551.
[7] V. Dakos, S. R. Carpenter, W. A. Brock, A. M. Ellison, V. Guttal, A. R. Ives, S. Kéfi, V. Livina, D. A. Seckell, E. H. van Nes and M. Scheffer, Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data, PLOS ONE, 7 (2012), 1–20.
[8] E. Fehr and S. Gächter, Cooperation and punishment in public goods experiments, American Economic Review, 90 (2000), 980–994.
[9] E. Fehr and K. M. Schmidt, Fairness, incentives, and contractual choices, European Economic Review, 44 (2000), 1057–1068.
[10] U. Fischbacher, S. Gächter and E. Fehr, Are people conditionally cooperative? evidence from a public goods experiment, Economics Letters, 71 (2001), 397–404.
[11] H. v. Foerster, Objects: Tokens for (eigen-)behaviors, ASC Cybernetics Forum, 8 (1976), 91–96. Reprinted in: Foerster H. von (1981) Observing systems. Intersystems Publications, Seaside CA: 274–285., Reprinted in: Foerster H. von (2003) Understanding understanding: Essays on cybernetics and cognition. Springer, New York: 261–271.
[12] G. Harras and D. Sornette, How to grow a bubble: A model of myopic adapting agents, Journal of Economic Behavior & Organization, 80 (2011), 137–152.
[13] Z.-Q. Jiang, W.-X. Zhou, D. Sornette, R. Woodard, K. Bastiaensen and P. Cauwels, Bubble diagnosis and prediction of the 2005–2007 and 2008–2009 chinese stock market bubbles, Journal of Economic Behavior & Organization, 74 (2010), 149–162.
[14] C. Kuehn, A mathematical framework for critical transitions: Bifurcations, fast-slow systems and stochastic dynamics, Physica D: Nonlinear Phenomena, 240 (2011), 1020–1035.
[15] J. O. Ledyard, Public goods: A survey of experimental research, 1994.
[16] T. M. Lenton, Early warning of climate tipping points, Nature Clim. Change, 1 (2011), 201–209.
[17] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf and H. J. Schellnhuber, Tipping elements in the earth’s climate system, Proceedings of the National Academy of Sciences, 105 (2008), 1786–1793.
[18] B. Litt, R. Esteller, J. Echauz, M. D’Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter and G. Vachtsevanos, Epileptic seizures may begin hours in advance of clinical onset: A report of five patients, Neuron, 30 (2001), 51–64.

[19] R. M. May, Thresholds and breakpoints in ecosystems with a multiplicity of stable states, Nature, 269 (1977), 471–477.

[20] P. E. McSharry, L. A. Smith and L. Tarassenko, Prediction of epileptic seizures: Are nonlinear methods relevant?, Nat Med, 9 (2003), 241–242.

[21] R. O. Murphy and K. A. Ackermann, Social value orientation: Theoretical and measurement issues in the study of social preferences, Personality and Social Psychology Review, 18 (2014), 13–41.

[22] J. F. Nash et al, Equilibrium points in n-person games, Proceedings of the National Academy of Sciences, 36 (1950), 48–49.

[23] H. H. Nax and M. Perc, Directional learning and the provisioning of public goods, Scientific Reports, 5 (2015), P8010.

[24] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk and G. Sugihara, Early-warning signals for critical transitions, Nature, 461 (2009), 53–59.

[25] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke and B. Walker, Catastrophic shifts in ecosystems, Nature, 413 (2001), 591–596.

[26] D. Sornette, Critical market crashes, Physics Reports, 378 (2003), 1–98.

[27] D. Sornette, Physics and financial economics (1776–2014): Puzzles, ising and agent-based models, Reports on Progress in Physics, 77 (2014), 062001,28pp.

Received November 2017; revised February 2018.

E-mail address: christian.hofer@uni-graz.at
E-mail address: georg.jaeger@uni-graz.at
E-mail address: manfred.fuellsack@uni-graz.at