Predicting regime shifts in social systems modelled with agent-based methods

Manfred Füllsack[1] · **Simon Plakolb[1]** · **Georg Jäger[1]**

Abstract

We investigate the possibility to apply a method of calculus analytics developed for predicting critical transitions in complex systems to social systems modelled with agent-based methods (ABMs). We introduce this method on the example of an equation-based modelled system and subsequently test it—to our knowledge for the first time—on ABMs. Our experiments show that the method may have wide applicability in the analysis of social systems. The method can help to approximate abrupt and thus unpredictable regime shifts, even though it may be constrained by stochastics and require a bit more experimentation in selecting suitable variables for making it work in ABMs.

Keywords Critical transitions · Regime shifts · Prediction · Critical Slowing Down · Agent-based modelling

Introduction

Like other complex dynamical systems too, social systems can experience abrupt transitions from one equilibrium state to another [1]. The fall of the socialist societies in the late 1980s [2], the sudden emergence of chaotic riots in London 2011 [3], the Arab spring in 2010/11 [4], the outbreak of workers protests [5] or the transition of financial markets from a balanced to a crisis condition in 2008 [6] are examples of such sudden state changes, which all have been subjected to analytical modelling. Since many of these changes concern shifts from socially preferable to adverse

✉ Manfred Füllsack
manfred.fuellsack@uni-graz.at

Simon Plakolb
simon.plakolb@uni-graz.at

Georg Jäger
georg.jaeger@uni-graz.at

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

🖄 Springer

or detrimental states, the purpose of modelling often is to find ways to predict the tipping point in these critical transitions, that is, the point where dynamics incontrovertibly shift.

Theory suggests that critical transitions are caused by reinforcing feedbacks generating alternate attraction states, between which the system may shift abruptly when its equilibria change from stable to unstable or back. Such equilibria changes are called bifurcations and analysed in the framework of calculus analytics, for which a mathematical representation of the system in terms of differential equations is needed. Once the differential equations are known, at least for simpler systems, bifurcation diagrams can be derived from the development of the dominant Eigenvalues of the Jacobian matrix, thus giving an idea of where bifurcations can be expected [7–9].

Due to this proximity to calculus, it is common to conduct critical transition analyses in the framework of equation-based modelling (EBM), that is, based on systems being represented in terms of coupled differential equations. Given the theoretic proposition, however, that critical transitions origin in the distributed interactions on a system's component- or micro-level and are only abridged mirrored by the aggregated macro-level representations as obtained from EBMs, this seems somehow at odds. Particularly when considering structured interactions, as they are norm in social systems, EBMs do not seem to capture the relevant details. In these cases, agent-based modelling methods (ABMs) appear to be the appropriate mean.

ABMs, however, are primarily rule-driven and, therefore, less directly affiliated with mathematics. Only rarely they can be directly expressed in terms of differential equations. Predicting regime shifts in an ABM-context thus needs to draw on simulations and statistical means. Recently in this regard, an elaborate apparatus of methods has been introduced, comprised under the term Early Warning Signals ([10–12] that basically consists in the analysis of statistical traces that systems generate when being perturbed in their approach to a tipping point. A core concept in this analysis is Critical Slowing Down, referring to the tendency of systems to show prolonged recovery times from perturbations when approaching a tipping [13, 14]). Since Critical Slowing Down analysis does not depend on a mathematical representations of a system—it just needs a way to test systems systematically for their resilience—, the method is applicable to ABM-generated data [15], although it is still most often applied to EBMs so far.

Alternative to the concept of Early Warning Signals but based on Critical Slowing Down analysis, a team at the University of Michigan [16–18] has suggested a method for drafting the essential parts of a system's equilibria curves, in particular in bi-stable systems showing hysteresis, without mathematical representation. Again, this method has been developed and investigated primarily in the framework of EBMs, which gives reason to try to apply the method to ABM-generated data. As our investigation shows, this application can be constrained by stochasticity to certain extent, which is expectable given the higher structural resolution of social interactions in ABMs. Moreover, it can be slightly hampered by the difficulty to select appropriate variables as central observables, which is a problem in EBMs as well, but it may be somehow aggravated due to the larger complexity that ABMs allow to consider. Despite these limitations, however, we find that the method provides

interesting insights into the behaviour of complex interactions and thus can help to approximate abrupt and thus hard to predict regime shifts in social systems.

In the remainder of this paper, we will first describe and test the method on the example of simple EBM systems prone to critical transitions (Sect. 2). We will then apply it to three different ABMs to indicate not only the possibilities but also the difficulties this method holds for analysis, with the first ABM known to undergo a supercritical pitchfork bifurcation and the other two arbitrarily tuned to show bi-stability in certain ranges of a critical parameter. The first ABM is a two-dimensional version of the famous Ising model of magnetization and is contained in Netlogo's model library (Sect. 3). The second ABM depicts a stylized repeated Public Good Game that determines contribution behaviour of agents from a combination of payoffs and social influence (Sect. 4). And the third ABM models a stylized job market in recession and recovery of economy, which causes unemployed agents to alter skill-updating activities in dependence of their neighbourhood's employment status (Sect. 5). Section 6 discusses the results and insights and concludes with an outlook on needed further research. To keep the method accessible to social scientists, its explanation is deliberately kept somewhat less formal in mathematical terms.

Applying bifurcation prediction to an equation-based model (EBM)

We first demonstrate the method on the example of an EBM, which among others has been suggested in [19] to model economies coupled by trade, or patches of eco-systems coupled by movement of organisms. The modelled system is subject to a combined saddle-node-bifurcation, which causes a particular troublesome form of critical transitions. The system's bi-stability (hysteresis) makes it shift states differently in dependence of the direction from which the critical transitions are approached [9]. The model reads:

$$x' = -x^3 + cx + a. \tag{1}$$

Figure 1a shows a numerical simulation of the system variable x undergoing critical transitions in dependence of whether the parameter a increases (black curve) or decreases (grey curve). The equilibria curves of the system are shown in red. Additionally, to provide one example of Early Warning Signals analysis, Fig. 1b, c shows the autocorrelation with lag 1 (AC-1) of time series of length 100 (after a transient phase) at each simulated equilibrium. As can be seen, autocorrelation clearly increases towards 1 in the wake of a critical transition depending on whether the critical parameter a grows or declines. Simulation parameters are $c = 0.9$, a as the critical parameter varying in 1000 steps between -1 and 1, and a noise term $\sigma x dW$ added with $\sigma = 0.001$.. The equilibria curves (red) show the bi-stable phase of the system and indicate critical transitions at $a = -0.3286$ (B1) and $a = 0.3286$ (B2) (i.e. hysteresis).

In what follows, we assume that the mathematical model is unknown, but the system can be tested for its resilience by systematically perturbing it at different equilibria states. To do this, we generate time series at different values of a
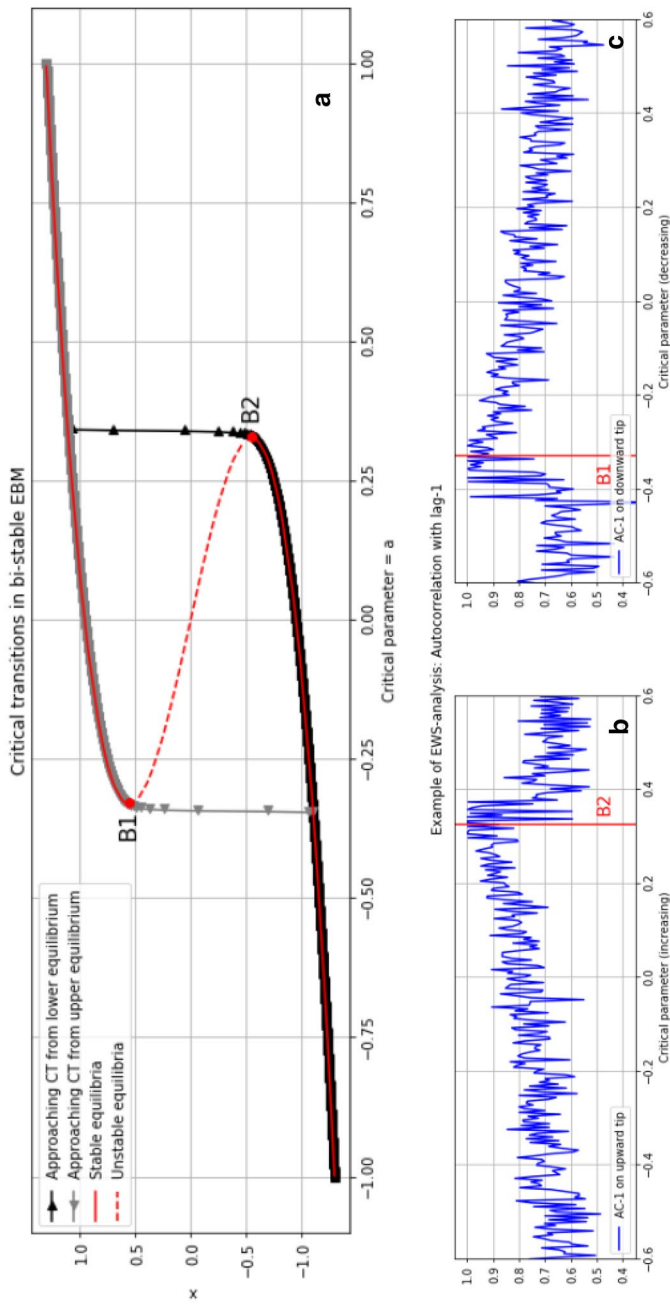
Fig. 1 Showing numerical simulations of the system expressed by Eq. (1), and as an example of Early Warning Signals-analysis, the autocorrelation with lag 1 (AC-1) of the black (1b, *a* increasing) and grey (1c, *a* decreasing) curves in 1a. In both cases, AC-1 clearly increases towards 1 in approaching the tipping points (indicated with red vertical lines)
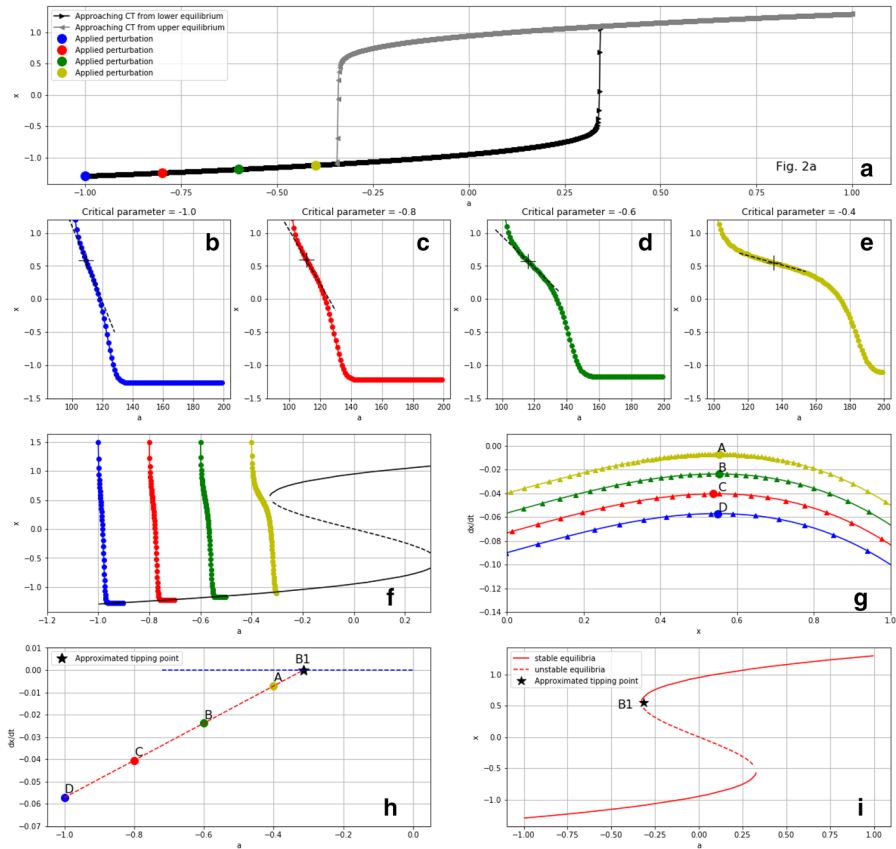
Fig. 2 Showing simulations (**a**), recovery curves from perturbations (**b–f**), their differentiations (**g**), a linearization of the maximum of deflection (**h**) and the prediction of the tipping point (**i**) of the system expressed by Eq. (1). **f** shows the four (averaged) recovery curves of **b–e** in relation to the equilibria curve, projected onto the locations (the a-values on the x-axes) of the perturbations

(blue, red, green and yellow dots in Fig. 2a) and perturb the system in each of these states after a transient time of 100 time steps by raising the system variable x beyond the alternate stable equilibrium (Fig. 2b–e, similar to [17]). Thereby, the values of the critical parameter a are chosen so that they (a) either increase or decrease sequentially, that (b) the resilience of the system still guarantees a return to the original equilibrium, and that (c) the recovery over these different a-values happens with increasing delays caused by Critical Slowing Down. For filtering noise from the system's recover, these perturbations are repeated several times at each value and averaged for further processing. Figures 2b–e shows averages over 20 recovery curves of 4 systematic perturbations at different a-values increasing from -0.7 to -0.4. (Note that the number of perturbations is arbitrary. The method works with a minimum of two perturbations.)

In the case of systems showing combined saddle-node-bifurcations, theory suggests that near-by bifurcation points will exert attraction on the system states recovering from perturbation and thus deflect the recovery curves increasingly when the perturbation approaches the critical transition. This so-called Critical Slowing Down can be seen in Fig. 2f, where the recovery curves of 2b–e are shown in relation to the calculated equilibria curve, projected onto the locations of the perturbations.

In this case, the probably most straight-forward way of the proposed method is to calculate the first order discrete differences of the recovery curves,[1] focusing on the part of their slopes at which the system has not yet returned to equilibrium. This part of the differentiated curves then is approximated with a least square polynomial fit of degree 3 and plotted in relation to the systems variable during recovery (Fig. 2g). Assuming a linear approach to the critical transition and depending on the distinctiveness of the deflection, one can take the maxima of the fitted curves (the bold blue, red, green and yellow points in 2 g, also marked with black crosses in 2b–2e) as correlated indicators of the approaching tipping. These maxima then are plotted against the critical parameter values at which the system was perturbed (Fig. 2h). Fitting the four points with linear regression and extending the resulting line (dotted red) to the zero line (dotted blue) indicates the parameter value at which the slope (dotted black in 2b–2e) of a corresponding recovery curve would be horizontal, or in other words, at which no more recovery would take place. This point hence indicates the bifurcation point B1, in this case at $a = -0.3286$ (Fig. 2h). Averaging the x-axes-values of the maxima in Fig. 2g, additionally indicates the variable value at this bifurcation, so that the point B1, i.e. the left bifurcation point in Fig. 1b, can be approximated as $B1 = (-0.3286, 05508)$, which in this case corresponds perfectly with what can be analytically derived (Fig. 2i).

If this procedure is extended by considering additional points to the left and right of the maxima of the differentiated recovery curves (A, B, C, D in Fig. 3a), the left part of the hysteresis curve as shown in Fig. 1b can be approximated (Fig. 3b).

For deriving bifurcation point B2 in Fig. 1a (i.e. the right part of the equilibria curve), there are two possibilities in this case of a simple and well-understood system. The one, suggested by [17] is to choose further critical parameter values for perturbation past the B1-bifurcation point, that is, in the range between B1 and an assumed B2-point. In this case, the system variable cannot be perturbed beyond the unstable part of the equilibrium curve (dotted red line in 1a), since too large perturbations would drive the system over the unstable part towards the upper stable equilibrium. No return curve would be gained.

In the experiment shown in Fig. 4a, b below, perturbations were applied at the critical parameter values $a = \{-0.1, -0.03, 0.03, 0.1\}$ and the system variable was raised to the value of -0.4 (instead of 1 as in the B1 case). Since in this case, no deflection in the recovery curves can be observed, the minima, instead of maxima, of the differentiated and fitted return curves were considered for linearization to the

---

[1] Alternatively, [16] suggest to differentiate the recovery curves RC by considering $\lambda = \frac{lnx_{t-dt} - lnx_{t+dt}}{2dt}$ for each x_t in $RC_{[dt:-dt]}$, which in our experiments yield similar results as the discrete differences, with the cutback that this differentiation cannot be applied to negative ranges.
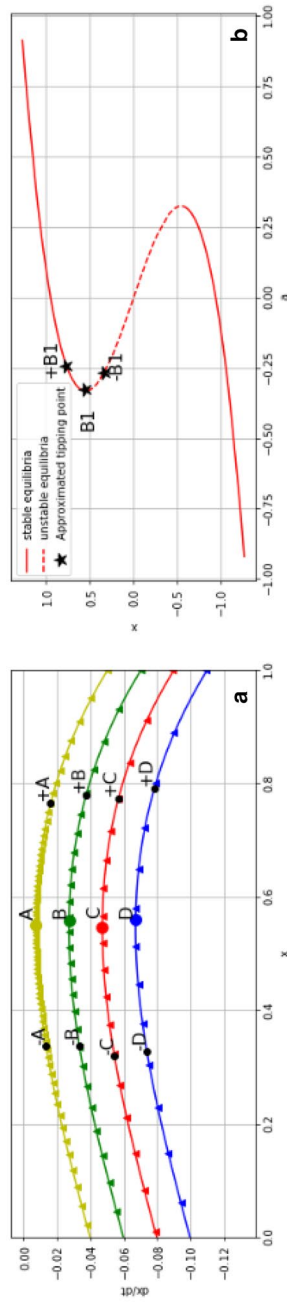
Fig. 3 Showing differentiations of recovery curves with one additional point to the left and the right of the maximum considered (**a**), and corresponding projections to approximate equilibria (**b**)
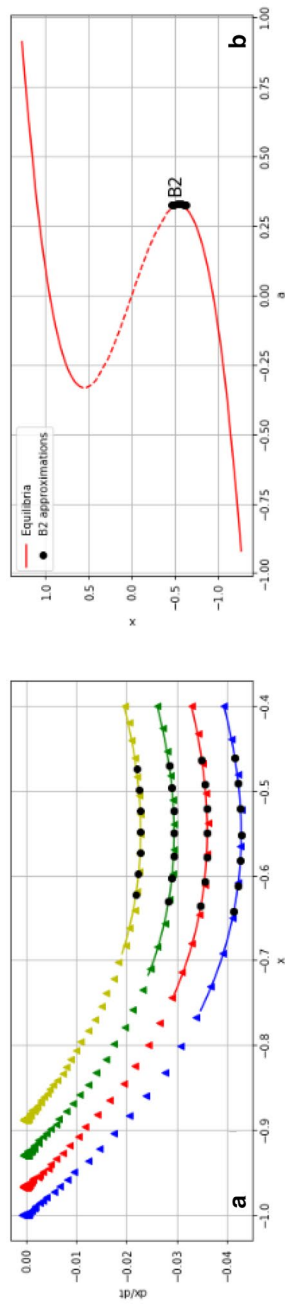
Fig. 4 Showing differentiations of recovery curves from alternate perturbations at parameter values $a = \{-0.1, -0.03, 0.03, 0.1\}$, with several additional points (black dots) to the left and the right of the minima considered (**a**), and corresponding projections to approximate the bifurcation B2 of the system expressed by Eq. (1) (**b**)
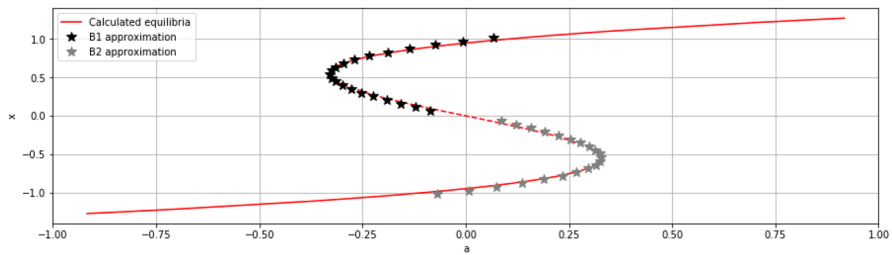
Fig. 5 Showing the combined bifurcation approximations (black and grey stars) on the foreground of the analytically derived equilibrium curve of the system expressed by Eq. (1)

zero line, together with three further points in equal distance to the left and right of the minima (black dots in Fig. 4a), yielding the B2 approximations (black) as shown in Fig. 4b. Note, however, that, as described below, it is not mandatory to consider extrema of the differentiated recovery curves. As long as the selected points are correspondingly alligned over all curves the method yields useable results.

As an alternative to what is suggested in [17], we also applied the procedure symmetrically to what is shown in Fig. 2 by running the system from right to left, that is in the case at hand, by considering critical parameter values in the range from 0.7 to 0.4. For perturbation, we lowered the system variable to − 1 and took the minima of the first part of the fit of the differentiated return curve. Combined with the approximation of B1, as described above, this procures more details of the equilibria curve and hence a more accurate estimate of the bifurcation points, as shown in Fig. 5. It also seems to be better applicable to respective investigations in ABMs (see next section).

Not all systems prone to critical transitions, however, are subject to combined saddle-node-bifurcations. Perturbing systems with sub- or supercritical pitchfork-bifurcations for instance does not necessarily yield clear deflections in the recovery curves, although Critical Slowing Down can be observed in the wake of a tipping point (see Fig. 6a–d). In these cases, when extremes are out of the considered range, it may be more applicable, instead of considering maxima or minima, to iterate over the differentiated recovery curves with a small running window around a value $r.$, taking the averages of the windows and using these terms for the linearized projection to the zero line as described above (and indicated in Figs. 2h and 6f). Figure 6 shows this procedure on the example of the supercritical system $x' = x(a − x^2)$ (2), with a serving as the critical parameter while considering the λ-differentiation as described in footnote 1 (plain discrete differentiation, however, yields very similar results).

Figure 6a–d shows the recovery curves from perturbing the system after a transient time of 100 iterations. Figure 6e shows the λ-differentiations of these curves (analogue colours to 6a–d) and an (arbitrarily chosen) example point at $r = 0.22$ (blue vertical line), chosen as centre of a small window from which the average is plotted in Fig. 6f in relation to the corresponding a-value at which the system was perturbed (using an average over a small window might seem unnecessary in this case, but can improve results enormously in the case of noise in the recoveries, as
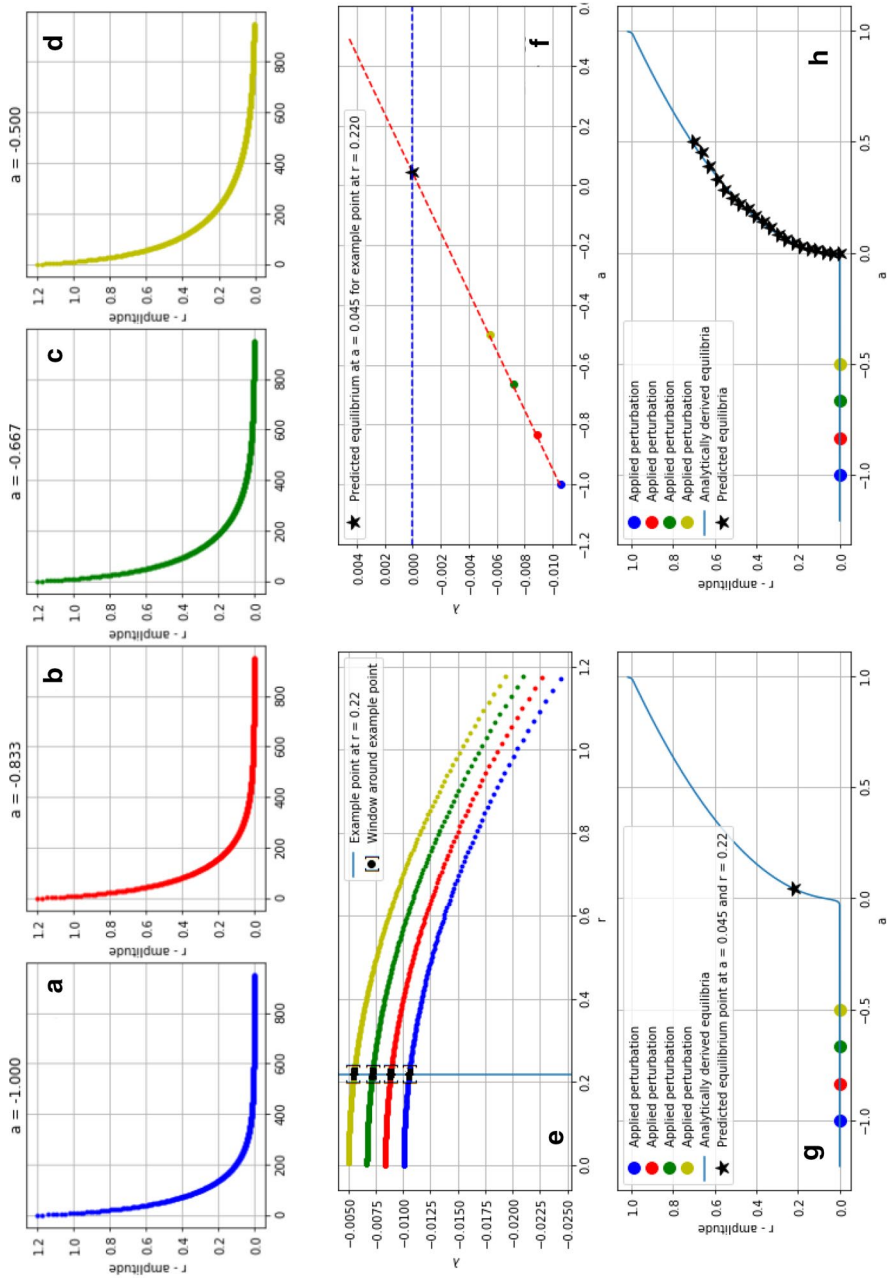
◄**Fig. 6** Showing recovery curves (**a–d**), differentiations (**e**), an example linearization (**f**) and its corresponding equilibrium point approximation (**g**), and the result of iterating several such linearizations with different values of r (**h**) of the system expressed by Eq. (2)

usually is the case in ABMs). As can be seen again from 6f, linearizing the points obtained in this way from the λ-differentiated recovery curves towards the zero line indicates a point at $a = 0.045$, which together with $r = 0.22$ approximates one point on the equilibria curve (black star) as shown in Fig. 6g. Repeating the same procedure for various r-values by iterating over the differentiation curves yields a close approximation of the systems equilibria, as shown in Fig. 6h.

Note that, irrespective of the deployed differentiation, the method is quite sensitive to small disturbances. Figure 7a–h shows the procedure as applied in Fig. 6a–h with a small noise term added to the state variable $[x' = x(a − x^2) + \sigma x dW$. with $\sigma = 0.005$ (3)]. While the recovery curves 7a–d look smooth and not overly different from 6a–d, their differentiations are heavily scattered (7e). Consequently, the critical transition prediction is significantly less accurate (7f). As will be seen in the next sections, this can be a crucial constraint in applying the method to agent-based modelled systems.

In the following three sections, we show results from applying the above-described method to attempts of predicting tipping points in agent-based modelled systems for which no direct mathematical representation is available. All considered ABMs were implemented in Netlogo (https://ccl.northwestern.edu/netlogo/) and actuated with the Python module NL4Py (https://github.com/chathika/NL4Py).

Ising ABM

In a first exploratory experiment, we apply the method to a two-dimensional-version of the Ising model on ferromagnetism [20], which often also is used to model social phenomena like opinion changes [21] or spreading rumours [22]. The used ABM is contained in Netlogo's model library.

The model represents magnetic dipole moments of atomic spins as points on a lattice, which can be in one of two states ($+1$ or $−1$), dependent on an ambient temperature T and an energy, which is defined as the negative of the sum of the products of a spin with each of its four neighbouring spins. Spins are seeking a low energy state, causing them to flip depending on a potential gain in energy E_{diff}, which determines the flipping probability p using the Metropolis algorithm as $p = \frac{E_{\text{diff}}}{t}$. Consequently, as temperature increases, flipping to a higher energy state becomes increasingly likely, but as the energy to be gained by flipping increases, the likelihood of flipping decreases.

As a simplified model of particle behaviour in magnetic metals, the model allows the identification of phase transitions. It is known to exhibit a (symmetric) super-critical pitchfork bifurcation at a critical temperature value of about 2.27. As in the EBM-example in the preceding section, our experiment considers only the upper branch of the pitchfork (Fig. 8b), simulated by running the ABM 100 times for 1000 ticks on a parameter range from $T = 1.42$ to $T = 3.12$, with an initial value of $p = 1$
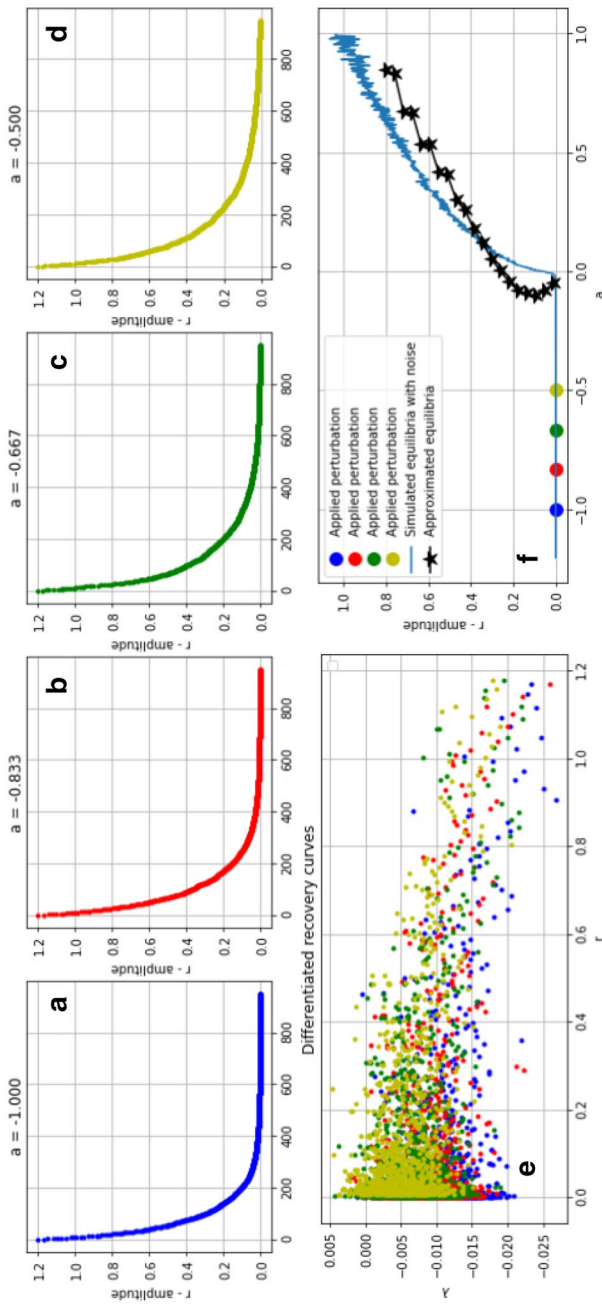
Fig. 7 Showing recovery curves (**a–d**), their differentiations (**e**), and approximation results (black stars) compared to a simulation (blue) of the noise-driven system expressed by Eq. (3)
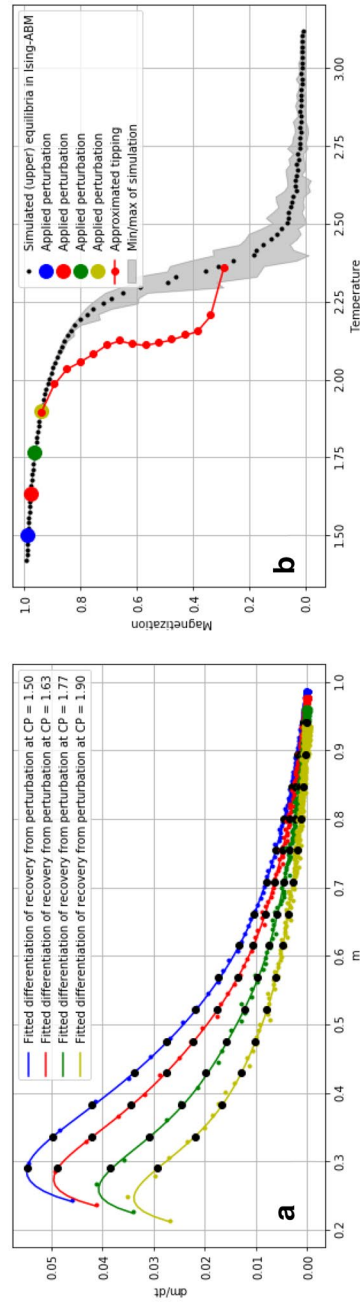
Fig. 8 Showing differentiated recovery curves (**a**, $m =$ recovery of magnetization) from perturbations applied at parameter values indicated by the blue, red, green and yellow dots in (**b**) and approximation results (red dots) on the foreground of simulation results (black) of the Ising ABM as contained in Netlogo's model library

of all spins being in state $+1$. Perturbations were applied 100 times at 4 equally intervaled values from $T = 1.5$ to $T = 1.9$, (that is, in some distance to the assumed critical transition at $T = 2.27$) by making 40% of the particles change their spin's orientation (i.e. from 1 to -1 or vice versa) for one timestep. Altering larger fractions than 40% caused state shifts prior to the theoretically assumed critical transition at $T = 2.27$ and thus was avoided.

The differentiations of the averaged recovery curves are shown in Fig. 7a. To cope with the still quite considerable noise in the curves, we fitted them with a degree-5 polynomial and iterated over 15 values in the interval (0.29, 0.93) to generate a not overly accurate, but still somehow indicative approximation of the Ising model's equilibria, as shown as red dots in Fig. 8b.

Repeated public good game-ABM

The second experiment was performed on an ABM simulating a stylized form of a Repeated Public Good Game, in which agents adjust an individual probability for contributing (a) in respect to their payoffs in relation to payoffs from preceding rounds of the game and (b) in respect to the majority of other agents' respective behaviour. Due to the consideration of individual agent interactions under social influence, this system cannot be represented by an EBM.

A Repeated Public Good Game is a well-known game theoretic formalization of a situation in which cooperation can procure a common good of high social value, but contribution is impeded by the possibility to realize still higher payoffs by free riding [23]. Due to social contagion, this situation when repeated tends to evolve towards a socially sub-optimal Nash-equilibrium of pervasive defection [24] with nobody being willing to cooperate as long as nobody else does [25, 26]. Although experiments show that contributions in Repeated Public Good Games do not decline to absolute zero [26–28], changes in cooperation behaviour can be abrupt [29], in particular when feedback-driven reciprocal entrainment is involved [30].

Our model mimics these feedback-driven shifts in a simple way, which is primarily tuned for generating critical transitions, and not for modelling realistic behaviour (see for model details Appendix A). It considers a population of agents playing a Repeated Public Good Game conditioned on an individual contribution probability, which determines whether agents invest all or nothing of an endowment into a common pool. After each investment round the pool is multiplied with an enhancement factor and evenly distributed among all agents earning them a final payoff dependent on investment.

While iterating the game the agents' contribution probability is updated with a simple influence mechanism making agents compare their current payoff with the one of the preceding rounds of the game. Dependent on payoff-gains or –losses, they adjust their contribution probability in respect to the fraction of other agents contributing or not contributing to the public good (see details in Appendix A). Which one of the counteracting updating dynamics dominates depends on a parameter p that serves as the *critical parameter* in this case, causing contribution to tip abruptly after a sustained decline when starting contribution near 100% and driving p from
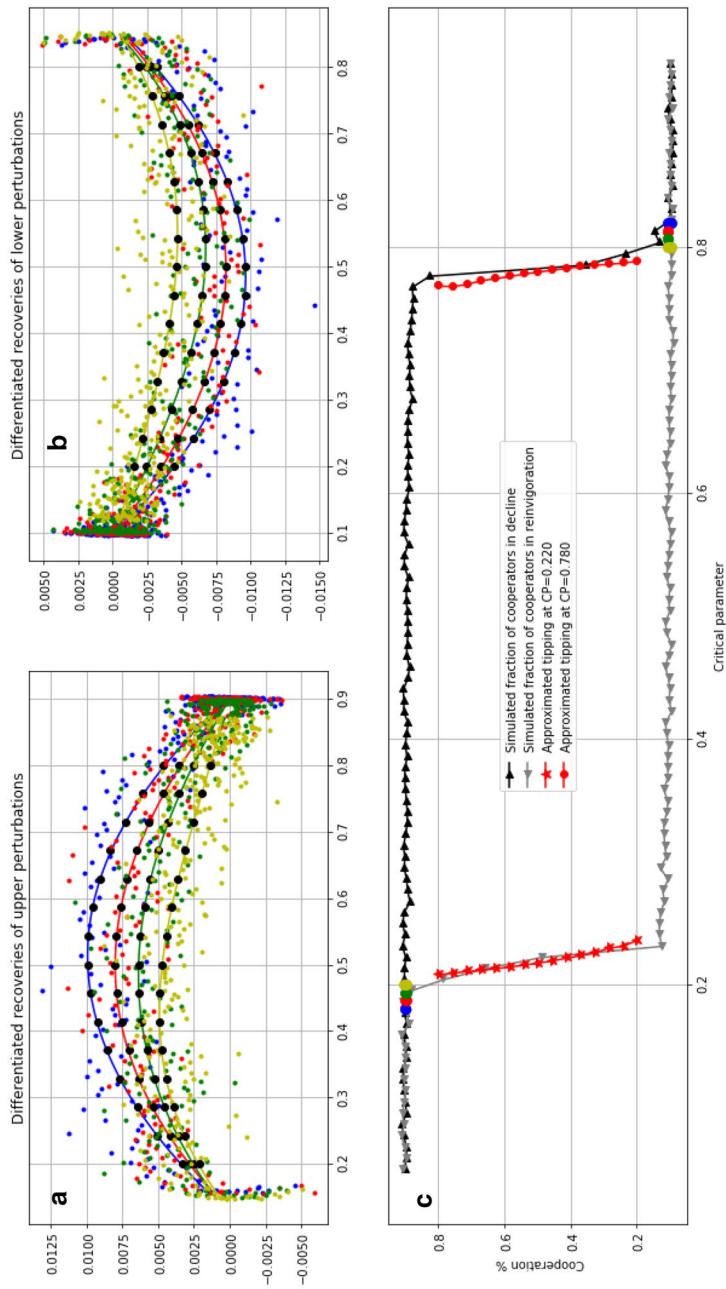
Fig. 9 Showing differentiated recovery curves (**a**, **b**) from perturbations applied at parameter values indicated by the upper (**a**) and the lower (**b**) blue, red, green and yellow dots in (**c**). **c** shows approximation results (red dots) on the foreground of averaged simulation results (black and grey triangles) of the Repeated Public Good Game-ABM (as detailed in Appendix A)

0 to 1 (from left to right in Fig. 9c, black triangles), while generating a symmetric upwards-dynamic (i.e. a backward change from the Nash-equilibrium to the social optimum of near overall cooperation) when starting contribution near 0% and decreasing parameter p from 1 to 0 (from right to left in Fig. 9c, grey triangles).

Applying perturbations for predicting critical transitions holds some subtleties in this case. As in the EBM, the system is run at different values of p over some transient time and then is systematically perturbed for one time step by diminishing (increasing) the cooperation probability of 95% of the agents by the value 0.4. As observable for the recovery curves served the percentage of agents cooperating. This, however, turned out to yield not sufficiently fine-grained curves to provide useful data for differentiation. As a remedy, the update regime of the Netlogo model had to be altered so that in each time step (i.e. Netlogo-tick) several rounds of the Repeated Public Good Game could be considered, which improved granularity, but also caused severe stochastic variations. Perturbations had to be averaged over large numbers of instances (100 in the plots shown in Fig. 9c) to yield usable results.

Even with this, the variance in the differentiated recovery curves (Fig. 9a, b) remains large, suggesting again to fit the data points with a higher degree (3 in this case) polynomial to get clear enough signals for approximating the tipping parameters. The resulting predictions in this case seem plausible to some extent. In Fig. 9c, they are plotted (in red) together with the simulated decline and recovery of the Repeated Public Good Game's cooperation dynamics.

Skill-matching ABM

Our third experimental ABM builds loosely on an observation by [31] noting that the abrupt decline of employment during an economic recession may not be followed by a similarly rapid reinvigoration once the economy recovers, consequently causing hysteresis in the unemployment rate. Our model follows an argument of [32], which as a cause for the delayed upswing suggests that workers may lose skills while being unemployed and thus face difficulties finding new jobs after recession [33]. Or in other words, reestablishing skills costs them more time than losing them. Our model accounts for this aspect with the additional proposition that the loss and the subsequent update of skills are socially mediated. It assumes that employed neighbourhoods raise the probability of updating skills, while neighbourhoods with high unemployment diminish this probability, with neighbourhoods having their own time to form and disintegrate (see details in Appendix B). Or to put it differently, skill-updating together with the situation on the job market is taken to determine neighbourhoods, with neighbourhoods in their turn forming clusters of self-enforcing stability that cause unemployment to persist even so the demand on the job market reaches pre-crises levels (see Fig. 10b).

These dynamics of the model induce a distinct upward tipping of the unemployment rate (black curve in Fig. 10) when the economy E (x-axes, taken as critical parameter) deteriorates beneath a certain level and a delayed recovery when recession is over. The green curve in Fig. 10 shows the parallel dynamics of the agents'
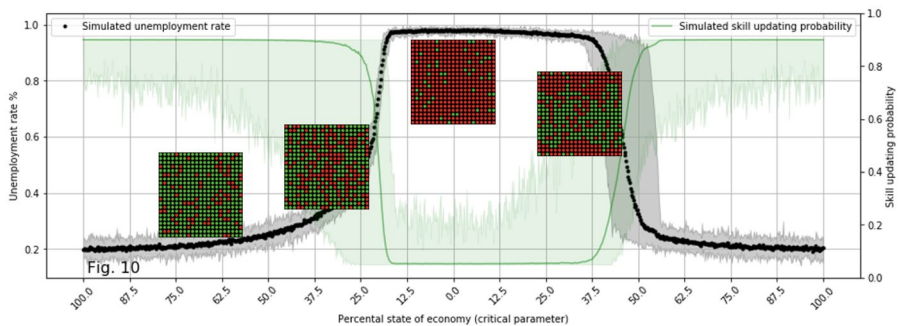
Fig. 10 Showing the simulated unemployment rate (black) with its variance (grey) generated with the skill-matching ABM as detailed in Appendix B. A linearly declining and recovering economic situation (*x*-axes) serves as critical parameter. The green curve shows the corresponding average skill-updating probability of agents with its variance (in light-green). The inserted grid-displays show four example neighbourhoods constraining skill-updating in varying extent, with inserts 2 and 4 depicting conditions at the same percental state of economy, one in the onset of the crisis (2) and the other in the recovery phase (4). Unemployment (red) is clearly higher in the second case, indicating hysteresis

skill-updating probability, impeded or amplified by the development of the neighbourhoods (shown in the inserted 2-*d* displays). Note that the inserts 2 and 4 are obtained at the same economic state value, but show distinctly different neighbourhood clusters.

Perturbations in this case were applied to the unemployment rate by setting the employment status of the agents to "true" for one timestep in the onset of the unemployment upswing (at $E = \{0.6, 0.53, 0.47, 0.4\}$, see lower coloured dots in Fig. 11c) and to "false" after the economic turnaround at the peak level of unemployment (at $E = \{0.05, 0.08, 0.11, 0.15\}$, upper lower coloured dots in Fig. 11c). The resulting recovery curves again had to be averaged over large numbers, and their differentiations again were fitted with a degree 3 polynomial to gain sufficiently clear signals for the critical transition prediction (see 11a, b). Note that the second set of perturbations at the upper level of unemployment (11b) obviously exerts less impact on the system and thus yields significantly shorter recovery curves with less information about the critical transition. The results from linearized zero-projection of the differentiations, shown in red in Fig. 11c, again are not as highly correlated to the dynamics of the state variable as in the case of the EBM in Sect. 1. Still they may be seen as indicative of changing dynamics in the onset of a critical transition and may thus be interpreted as an early warning signal for an imminent regime shift.

Summary and discussion

We investigate the prospects of applying a method for predicting critical transitions in complex dynamic systems, which was developed and deployed so far only on EBMs, to ABMs. This method bases on the fact that the resilience of a system decreases when approaching a critical transition point and the time to recover from small perturbations increases. Assuming a linear increase of recovery time, the
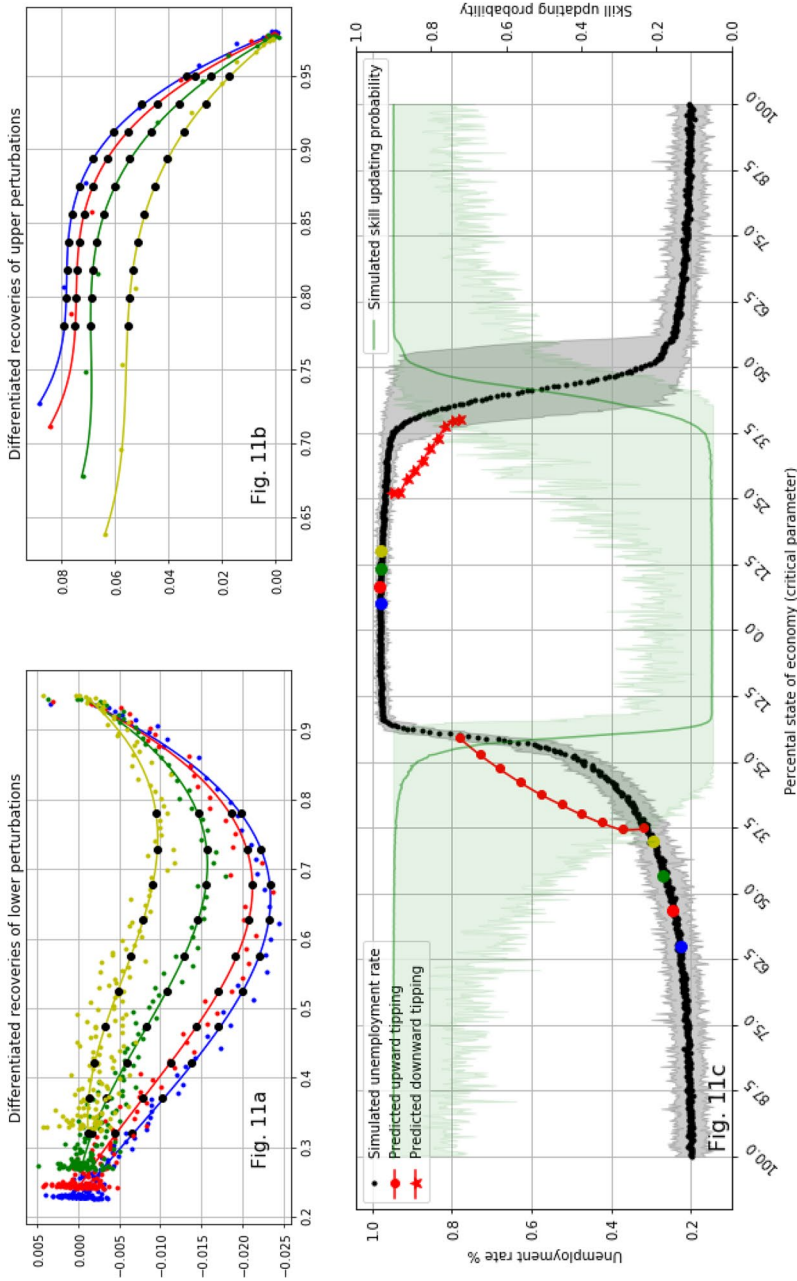
Fig. 11 Showing differentiated recovery curves (**a**, **b**) from perturbations applied at parameter values indicated by the lower (**a**) and the upper (**b**) blue, red, green and yellow dots in (**c**). **c** shows approximation results (red dots) on the foreground of averaged simulation results (black) of the skill-matching ABM as detailed in Appendix B

changes in the slope of the recovery curves can be considered as an indicator of the distance to the tipping point. Linearly projecting these slopes (mathematically the differentiations at certain points of the recovery curves) to the zero line (that is to a horizontal slope and thus a state of no recovery) yields approximations of critical parameter values at which the system shifts its regime from one equilibrium state to another (it bifurcates). This allows predicting critical transitions without exact mathematical representation of the system, which here is done for three different ABMs holding various degrees of complexity and yielding results with different levels of accuracy and predictive power.

The first ABM is a two-dimensional implementation of the well-known Ising model for identifying phase transitions in ferromagnetism. The difficulties in this ABM concern primarily the limited possibilities for perturbing the main state variable, that is, the spin of the particles. The analogous example-EBM with the supercritical pitchfork bifurcation in Sect. 2 (Figs. 6 and 7) suggests perturbing the Ising-system in the range $T > 2.27$, that is, while decreasing the critical parameter temperature from values beyond the critical transition. Since in this range, particles have an equal probability of being either in state $+1$ or -1, switching their spin does not change the state of the system. Perturbations, therefore, had to be applied at the upper equilibrium branch in the range $T < 2.27$ (for details see Sect. 3). Additionally, perturbations had to be restricted to just 40% of the particles for not driving the system beyond the unstable equilibrium. This, together with the strong noise caused by the particles' individual interactions, restricted the results to a rather approximate transition forecast, as shown in Fig. 8b.

The second ABM is arbitrarily modelled to show a combined saddle-node bifurcation on the example of a Repeated Public Good Game with an artificial symmetry that, strictly speaking, makes one of the two sets of perturbations redundant. As a proof of concept, however, this example shows that critical transitions can be approximated quite accurately with the method described, even though recovery curves and their differentiations are overly noisy in this case as well.

The dynamics of the third ABM eventually are explicitly subjected to local neighbourhood interactions and, therefore, are probably the ones most affected by stochastic variations. The system is asymmetric and consequently provides different conditions for perturbations on the lower and on the upper equilibrium. While the perturbation amplitude for the lower set covers nearly the whole range between equilibria, the upper equilibrium allowed only for minor displacements, although the same number of agents was exposed to perturbations in both cases. The usable signals thus, and their differentiations are significantly shorter in the second case. Despite averaging and curve-fitting in the analysis, the critical transition-predictions turned out to be inaccurate but not completely out of line, as shown in Fig. 11c. Given the strong noise from truly local neighbourhood interactions, it still can be argued that the method may help to give an approximate indication of where critical transitions in agent-based modelled systems occur.

Two key problems occur in all three ABMs. One concerns the question about which of the state variables is most suitable for being perturbed. While the Ising model seems to offer switching the particle spins quite naturally, the other two models hold several possibilities in this regard. Instead of increasing or

decreasing the cooperation probability in the Repeated Public Good Game-ABM, one could simply alter the number of cooperators. Analogously, instead of changing the number of unemployed in the skill-matching ABM, one could change the skill-matching probability. Which of the variables is best suited to generate a clear Critical Slowing Down seems to depend very much on the system in question and thus needs to be subjected to experimentation.

The second problem concerns the question of where in the parameter range best to apply the perturbations. Our experiments show that the distinctiveness and clarity of Critical Slowing Down does not necessarily increase continuously with the approach to a critical transition. In respect to this problem, future research could focus on the possibility of combining the statistical method of Early Warning Signals (as mentioned in Sect. 1) with the method described here and to tap its insights for pinpointing prospective parameter ranges at which perturbations can be applied. Regardless of these problems, we find that the method presented has great potential to further the analysis of complex systems and, especially when applied to ABM-generated data, can enhance the understanding of real-world behavior in social systems.

Compliance with ethical standards

Appendix A: Repeated public good game-ABM details

A population of N agents is made to play Repeated Public Good Games with an agent's cooperativeness depending on a contribution probability c_i, which initially is normally distributed either close to 1 or close to zero ($c_i = 1 - s$ or $c_i = 0 + s$. with $s \in \{0.01, 0.04\}$). In each game iteration, agents, dependent on c_i, invest either all or nothing of an endowment E into a common pool, earning them a final payoff

$$\pi_i = E - I_i + f \frac{\sum_{j=1}^{n} I_j}{N} \tag{4}$$

with the initial endowment E, the individual investment I_i, the number of participants N in the game, and an enhancement factor f, expressing the added value from social cooperation.

After each game, c_i is updated in respect to the following symmetric dynamics (5) and (6)

$$\text{if } \pi_{i,t-1} > \pi_{i,t} \text{ then} : c_{i,t+1} = c_{i,t} - g\left(\frac{N|_{I=0}}{N}\right)^3,$$

$$\text{else} : c_{i,t+1} = c_{i,t} \qquad (5)$$

and

$$\text{if } \pi_{i,t-1} < \pi_{i,\,t} \text{ then} : c_{i,t+1} = c_{i,t} + g\left(\frac{N|_{I=1}}{N}\right)^3$$

$$\text{else} : c_{i,t+1} = c_{i,t} \qquad (6)$$

with $N|_{I=0}$ ($N|_{I=1}$) indicating the number of non-contributors (contributors) in the population, and g being an arbitrary scaling factor which governs the dynamics of the development.

The second dynamic (6) is introduced to mimic system behaviour similar to the one of the EBM in Sect. 2. An explanation for it could be that once the majority cooperation is lost and the Repeated Public Good Game is close to its Nash equilibrium, the difference between the actual payoff and the payoff, which could be gained with a majority investing, might cause individuals to be dissatisfied, leading to scattered trial investments and eventually to a revival of cooperation. For considering a transition between these two effects, that is, between dynamic (5) and dynamic (6), we define a critical parameter p that lets dynamic (5) take effect if a random real number $r \in \{0, 1\} < p$, and dynamic (6) otherwise, with p being decremented (incremented) by steps of 0.003 after each 100 rounds of the game.

Parameters are: N = 100, E = 1, $I \in \{0, 1\}$, $f = 1.6$, g = 3.

Appendix B: Skill-matching ABM details

The model assumes economic decline causing deteriorating job supply, with job supply evolving according to $J = \frac{E^3}{(h+E^3)}$ if $1 > E > 0$, with E indicating the percentual state of the economy, varying linearly between 1 (full blown) and 0 (deteriorated). Agents employment status is determined by J and a skill-matching probability s, which deteriorates with the unemployment rate in an agent's neighborhood. The assumption here is that group dynamics rather than individual employment status are the driver of regular skill updating. The number of unemployed agents a_u in the Moore-neighborhood of an agent determines a probability f according to the

sigmoid function $f = \frac{a_u^p}{a_u^p + h^p}$, with $h = 5.7, p = 10$, where f is taken to equilibrate the skill-matching probability s of an agent with $dt = 0.01$ over 200 time steps in relation to the employment condition in its neighborhood. To increase the resolution of the generated data, the described dynamics were executed five times per simulation step (tick).

References

1. Levy, M. (2005). Social phase transitions. *Journal of Economic Behavior & Organization, 57*(1), 71–87.
2. Kuran, T. (1991). Now out of never: the element of surprise in the east european revolution of 1989. *World Polit., 44*(1), 7–48.
3. Davies, T. P., Fry, H. M., Wilson, A. G., & Bishop, S. R. (2013). A mathematical model of the London riots and their policing. *Sci. Rep., 3*, 1303.
4. Leonov, G. A., Kudryashova, E. V., & Kuznetsov, N. V. (2015). Modeling and identification of the Tunisian social system in 2011–2014: bifurcation, revolution, and controlled stabilization. *IFAC-Pap., 48*(11), 725–729.
5. Kim, J.-W., & Hanneman, R. (2009). A computational model of worker protest. *Journal of Artificial Societies and Social Simulation, 14*(3), 1.
6. Diks, C., Hommes, C., & Wang, J. (2019). Critical slowing down as an early warning signal for financial crises? *Empirical Economics*, *57*(4), 1201–1228.
7. Kuehn, C. (2011). A mathematical framework for critical transitions: bifurcations, fast–slow systems and stochastic dynamics. *Physics Nonlinear Phenomena, 240*(12), 1020–1035.
8. Kuehn, C. (2013). A mathematical framework for critical transitions: normal forms, variance and applications. *Journal of Nonlinear Science, 23*(3), 457–510.
9. Sayama, H. (2015). Introduction to the modeling and analysis of complex systems, Print edition. Open SUNY Textbooks.
10. Scheffer, M., et al. (2009). Early-warning signals for critical transitions. *Nature, 461*(7260), 53–59.
11. Lade, S. J., & Gross, T. (2012). Early warning signals for critical transitions: a generalized modeling approach. *PLoS Computational Biology, 8*(2), e1002360.
12. Dai, L., Vorselen, D., Korolev, K. S., & Gore, J. (2012). Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science, 336*(6085), 1175–1177.
13. Van Nes, E. H., & Scheffer, M. (2007). Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *The American Naturalist, 169*(6), 738–747.
14. Dakos, V., & Bascompte, J. (2014). Critical slowing down as early warning for the onset of collapse in mutualistic communities. *Proceedings of the National Academy of Sciences, 111*(49), 17546–17551.
15. Hofer, C., Jäger, G., Füllsack, M. (2018). Critical transitions and early warning signals in repeated cooperation games. *Journal of Dynamics & Games 5*, 223–230, https://www.aimsciences.org/article/doi/10.3934/jdg.2018014. Accessed: 24 Jul 2018.
16. Lim, J., & Epureanu, B. I. (2011). Forecasting a class of bifurcations: theory and experiment. *Physical Review E*, *83*, 016203.
17. D'Souza, K., Epureanu, B. I., & Pascual, M. (2015). Forecasting bifurcations from large perturbation recoveries in feedback ecosystems. *PLoS ONE, 10*(9), e0137779.
18. Ghadami, A., & Epureanu, B. I. (2018). Forecasting critical points and post-critical limit cycles in nonlinear oscillatory systems using pre-critical transient responses. *International Journal of Non-Linear Mechanics, 101*, 146–156.
19. Brummitt, C. D., Barnett, G., & D'Souza, R. M. (2015). Coupled catastrophes: sudden shifts cascade and hop among interdependent systems. *Journal of the Royal Society, Interface, 12*(112), 20150712.
20. McCoy, B. M., & Wu, T. T. (2014). *The two-dimensional ising model, reprint 2014*. Cambridge: Harvard University Press.

21. Grabowski, A., & Kosiński, R. A. (2006). Ising-based model of opinion formation in a complex network of interpersonal interactions. *Phys. Stat. Mech. Its Appl., 361*(2), 651–664.

22. Ostilli, M., Yoneki, E., Leung, I. X. Y., Mendes, J. F. F., Lió, P., & Crowcroft, J. (2010). Statistical mechanics of rumour spreading in network communities. *Procedia Computer Science, 1*(1), 2331–2339.

23. Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31*(1), 169–193.

24. Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences, 36*(1), 48–49.

25. Ledyard, J. O. (1994). Public goods. A survey of experimental research

26. Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review, 90*(4), 980–994.

27. Marwell, G., & Ames, R. E. (1981). Economists free ride, does anyone else?: Experiments on the provision of public goods, IV. *Journal of Public Economics, 15*(3), 295–310.

28. Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140.

29. Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences, 107*(12), 5334–5338.

30. Suri, S., & Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE, 6*(3), e16836.

31. Blanchard O. J., Summers, L. H. (1986). Hysteresis and the European Unemployment Problem. National Bureau of Economic Research, Working Paper 1950, Jun..

32. Hargreaves Heap, S. (1980). Choosing the wrong 'natural' rate: accelerating inflation or decelerating employment and growth? *Journal of Enonomics, 90*(359), 611–620.

33. Dosi, G., Pereira, M. C., Roventini, A., & Virgillito, M. E. (2018). Causes and consequences of hysteresis: aggregate demand, productivity, and employment. *Industry and Corporate Change, 27*(6), 1015–1044.