

# Stability of democracies – A complex systems perspective

**K Wiesner<sup>1,2</sup>, A Birdi<sup>3</sup>, T Eliassi-Rad<sup>4</sup>, H Farrell<sup>5</sup>, D Garcia<sup>2,6</sup>,  
S Lewandowsky<sup>7</sup>, P Palacios<sup>8</sup>, D Ross<sup>9,10,11</sup>, D Sornette<sup>12</sup> and  
K Thébault<sup>13</sup>**

<sup>1</sup> School of Mathematics, University of Bristol, U.K.

<sup>2</sup> Complexity Science Hub Vienna, Austria

<sup>3</sup> Department of Economics, University of Bristol, U.K.

<sup>4</sup> Network Science Institute, College of Computer and Information Science, Northeastern University, Boston, U.S.

<sup>5</sup> Department of Political Science, George Washington University, Washington, D.C., U.S.

<sup>6</sup> Section for Science of Complex Systems, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria

<sup>7</sup> School of Experimental Psychology, University of Bristol, U.K.

<sup>8</sup> Department of Philosophy, University of Salzburg, Austria

<sup>9</sup> School of Sociology, Philosophy, Criminology, Government, and Politics, University College Cork, Ireland

<sup>10</sup> School of Economics, University of Cape Town, South Africa

<sup>11</sup> Center for Economic Analysis of Risk, Georgia State University, U.S.

<sup>12</sup> Department of Management, Technology and Economics, ETH Zurich, Switzerland

<sup>13</sup> Department of Philosophy, University of Bristol, U.K.

E-mail: [k.wiesner@bristol.ac.uk](mailto:k.wiesner@bristol.ac.uk)

**Abstract.** The idea that democracy is under threat, after being largely dormant for at least 40 years, is looming increasingly large in public discourse. Complex systems theory offers a range of powerful new tools to analyse the stability of social institutions in general, and democracy in particular. What makes a democracy stable? And which processes potentially lead to instability of a democratic system? This paper offers a complex systems perspective on this question, informed by areas of the mathematical, natural, and social sciences. We explain the meaning of the term ‘stability’ in different disciplines and discuss how laws, rules, and regulations, but also norms, conventions, and expectations are decisive for the stability of a social institution such as democracy.

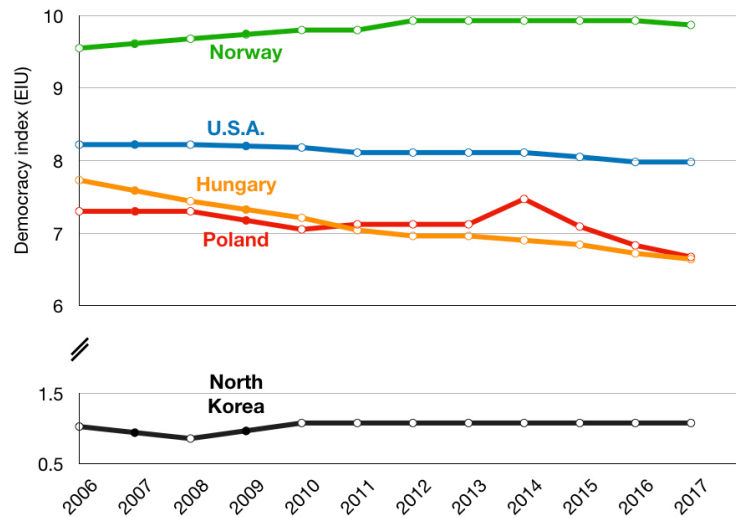
## 1. Introduction

### 1.1. Democratic backsliding

According to the media search engine *Factiva*, there were around 1,500 articles published in the global media during the 40 years between 1970 and 2010 that were concerned with a “threat to democracy.” In the period since then, in just over 8 years, more than 1,700 articles appeared on that topic, with the first 5 months of 2018 ( $N = 309$ ) already topping the total for all of 2017 ( $N = 304$ ). The idea that democracy is under threat, after being largely dormant for at least 40 years, is looming increasingly large in public discourse. When *The Economist* analysed the democracy score of 167 countries in early 2018, more than half the countries saw their democratic-health index decline compared to the preceding decade.‡ How real is this threat? How stable are today’s western democracies, some of which have survived for centuries?

Whilst the last two centuries have featured a wider variety of ‘democratic transitions’, where non-democratic countries become democratic, with some notable exceptions the opposite phenomenon has been less frequent. We agree with Waldner & Lust (2018) who remark: “we lack theories to explain backsliding, though we have long engaged in a perhaps interminable debate about the causes of democratic transitions, democratic breakdowns, authoritarian resilience, and democratic consolidation”. As a consequence there is a lack of work on the question under which circumstances instability of democracy might arise. The literature on democratic transitions studies the circumstances under which regimes fall and which contribute to non-democracies become democracies. But the circumstances under which democracies become non-democracies, which is to say, under which democracies destabilize to the extent until they cannot be described as democracies any more, are much less studied, and mostly only in terms of single-case studies of particular situations, informed by very little in the way of theory. This reflects the world we (in the west) have lived in over the relatively recent past of the last fifty to sixty years. Times have changed. Citizens of democracies are less and less content with their institutions; they are more and more willing to jettison institutions and norms that have traditionally been regarded as central components of democracy; and they are increasingly attracted to alternative, even autocratic regime forms (Survey 2014). Furthermore, recent events in Europe such as in Poland and Hungary, where elected officials have begun to put pressure on critical media and to undermine key institutions such as independent courts, illustrate the need to rethink the assumption of democracies as stable institutions. In the political sciences there is now a term for the opposite of democratic transition: ‘democratic backsliding’. The composition of a list of examples of democratic backsliding is an inherently controversial and politically loaded issue. An example of a plausible methodology for inclusion on the list is a reduction in a ‘democracy index’ (Unit 2017, Freedom House 2018) (see also (Munck 2009)) such that a country falls precipitously from a previously high position. The Economist Intelligence Unit (EIU) defines a ‘democracy index’ based on five categories: electoral process and pluralism, civil liberties, the functioning of government, political participation, and political culture and a range of indicators within these categories (Unit 2017). In Fig. 1 we show the EIU democracy index for Norway, the U.S.A., Hungary and North Korea for the years 2006 to 2017. The index for Hungary shows a clear downward trend over the entire range, that for the U.S.A. shows a slight but monotonic fall from above 8 to just below 8.

‡ <https://www.economist.com/graphic-detail/2018/01/31/democracy-continues-its-disturbing-retreat>



**Figure 1.** Democracy index for selected countries, as measured by the Economist Intelligence Unit for 2006 - 2017. The EIU democracy index takes values between 0 (worst) and 10 (best). Shown are data for Norway (green), United States of America (blue), Hungary (orange), Poland (red) and North Korea (black), the values for 2007 and 2009 are linear interpolations as no indices were reported for these two years.

## 1.2. Complex systems theory

The theory of complex systems is, at the least, a collection of concepts and methods which bridges disciplines and brings into focus processes which act across time and length scales. For some popular accounts of complex systems, see (Mitchell 2009, Holland 1992, Gell-Mann 1995 *a*). For more rigorous treatments, see (Ladyman et al. 2013, Strevens 2005, Gell-Mann 1995 *b*, Holovatch et al. 2017). There are many notions of complex system in the literature. Arguably, a minimal kind of a complex system is a collection of many elements with many repeated interactions between them, which exhibits self-organization, that is, a system which sustains patterns without being controlled by a central or an external element (Ladyman et al. 2013). Living complex systems exhibit additional features such as hierarchical organization and adaptive behaviour (Ladyman & Wiesner ). Examples of social complex systems with astounding levels of hierarchical organisation and adaptive behaviour are the eusocial societies of bees, termites, ants, and wasps. These have been studied extensively and their behavioural dynamics are understood to a great extent (Gordon 2010, Hölldobler & Wilson 2009, Seeley 2010). What is much less understood and has much less theoretical foundation is the adaptive behaviour of complex social groups of humans (Page 2010).

In our discussion, we focus on two features of complex social systems in general, and democratic systems in particular: feedback and stability, and their mutual relationship. It is generally known that negative feedback stabilises a system. A simple example is the thermostat which monitors the temperature in a room and regulates the heating up or down when the room temperature has surpassed a certain threshold. The dynamic of supply and demand in standard economics is another example. An equilibrium price is reached and maintained due to the negative feedback between price of a product and demand for it – at least in theory. Positive feedback, on the other hand is a destabilising phenomenon (‘positive’ in the sense of ‘enhancing’, not in the sense of ‘desirable’). An example of positive feedback is the melting of the polar icecaps upon atmospheric warming. Smaller surface areas of ice will reduce the albedo effect of the ice on the atmospheric temperature and speed up the melting process, an effect we observe at the moment (Chen et al. 2006). The recruiting process in honey bee societies is another example of positive feedback. Scouter bees advertising a new nest site will recruit other bees to the same site. The more bees are recruited the more will advertise the site and again recruit others to it until, eventually, the number of scouter

bees for one site will pass a threshold and the site will be moved to by the entire hive. For examples of positive feedback as a mechanism of financial bubbles, see the discussion by Sornette & Cauwels (2015).

As complex systems science reaches increasing maturity as part of the natural sciences, it is now in the position to offer tools to tackle some of the big open questions relating to social systems such as the robustness of social institutions, in general, and of democracy, in particular. With this perspective article we hope to bring attention to these and other arising opportunities for complex systems theory to address timely and important questions in political science.

## 2. Definitions of stability

### 2.1. (Modal) robustness vs (temporal) stability

The term ‘stability’ has different meanings in different fields and can be ambiguous even within a field. It is typically used in a sense that is synonymous with, for example, ‘robustness’, ‘resilience’, ‘viability’ or ‘adaptability’. In this section we consider the meaning of notions of stability across different disciplinary contexts in order to provide a framework for applying such notions to social institutions in general, and democracy in particular.

The natural place to start our discussion is at the most abstract and general level that is given by the philosophical discussion of stability in the context of scientific models. There we find a useful distinction that can be made between ‘robustness’ and ‘stability’. Robustness is understood as the insensitivity or independence of the behaviour of the system to changes in possible microscopic realisation (Weisberg 2006). Crucially, this is a *modal* notion, in that it refers to different possible states of the world. Robustness in this sense is clearly very important for model building in science since it allows for abstraction and the re-use of models for different target systems (e.g. magnets and fluids) or even different domains (e.g. physical, biological and social systems). An example of modal robustness is found in machine learning. We can define a machine learning problem as follows: ‘A computer program is said to learn from experience  $E$  w.r.t. some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience  $E$ .’ (Mitchell 1997). Generally speaking, a robust learning algorithm is one whose performance  $P$  is not altered substantially when slight modifications are made to its experience  $E$ . Being robust in terms of experience  $E$  is important because machine learning cares about generalization – i.e., performance on new experience. A prototypical example of robustness from the physical sciences is the phenomenon of continuous phase transitions. Phase transitions are a macroscopic phenomenon almost completely independent of fine-grained microscopic details. In fact, one can demonstrate that systems with physically distinct micro-structure, such as fluids and ferromagnets, have the same qualitative and quantitative macroscopic behavior when they undergo a phase transition (Goldenfeld 1993). In contrast, stability is typically used to refer to persistence over time or resistance to change or perturbation. We discuss various examples of stability in the following sections. Thus, at this abstract level, a stable system need not be robust nor a robust system stable.

### 2.2. Stability in nonlinear dynamics and complex networks

Stability in the sense of persistence over time or resistance to perturbation is part of the general mathematical framework of dynamical systems theory. In this context equilibrium points can be defined in terms of a general stationarity property of particular solutions to a system of (possibly non-linear) differential equations. Stability can then be formally defined via the Lyapunov function: if the time rate of change of this function following small perturbations from the equilibrium point is always negative or zero then the equilibrium state is stable (Haddad & Chellaboina 2011). The equilibrium state is the ‘attractor’ of the system. There may be more than one attractor for a given set of equations and, therefore, stability here, also termed ‘resilience’ or ‘engineering stability’, is related to the local response of the system to small perturbations. The systems in question might or might not be systems of static variables. The concepts equally apply to systems of dynamic variables such as flows.

Resilience to perturbation is also the stability concept relevant to the study of complex networks<sup>§</sup>. Stability in complex networks refers to the invariance under perturbations of the network topology and / or of the dynamical processes on the network. These perturbations can

<sup>§</sup> For an introduction to complex networks see, for example Newman (2003)

be random or targeted, can involve removal of nodes and links, or can include changes to the dynamical processes themselves. From a complex network perspective the significant questions are then (for example): under what kind of perturbations does the network break apart into isolated components? How persistent are the communities of nodes to perturbations? How susceptible is a network dynamic to cascading failures? An example in case is the resilience of public transport networks (Berche et al. 2009).

### 2.3. Ecological stability and adaptation

The notion of ‘engineering stability’ can be further generalised to a notion of ‘ecological stability’ (Walker et al. 2004, Kovalenko & Sornette 2013). Ecological stability describes the fact that a system state will return to its initial equilibrium point as long as the perturbations do not kick the system out of the ‘basin of attraction’ of the equilibrium point. This means that the system is stable with respect to non-local finite size perturbations, limited in their size by the size of the basin of attraction.

Although very general, and certainly applicable to some biological and social systems over some time scales, this notion of the ‘ecological stability’ of a dynamical system is still not general enough to effectively encompass the full sense of stability needed. In particular, the dynamical system analogy is limited by the fact that it fails to account for the possibility of adaptation of system constituents under the influence of external stressors (Walker et al. 2004, Kovalenko & Sornette 2013). This requires taking into account the possibility for the system to adapt its constituents so as to enhance its stability, for example by ‘topological transformations’ of the system’s state space. The system may even transform itself into a completely different structure via its capacity to evolve, as described by the theory of complex adaptive systems (Holland 1975, Kauffman 1993). Stability in the sense of adaptation is related to the concept of ‘viability’ as defined in (Aubin 1991, Deffuant & Gilbert 2011).

### 2.4. Stability in statistical mechanics

After closer inspection of the principles underlying robustness and stability there are good reasons to expect the two concepts to be related. That is, the physical principles that underlie the robustness of a system (i.e. insensitivity to modal variations) can be connected to those that underlie the stability of a system (i.e. insensitivity to temporal variations). An example of this is the stability of equilibria in statistical mechanics. Although this is a vexed and controversial issue, recent work on the foundations of Boltzmannian statistical mechanics (Werndl & Frigg 2015) points to a new formal definition of equilibrium based upon the existence of a macro-state in which the system spends a majority of the time in the long run. Crucially for our purposes, such a definition immediately implies that, given a microscopic perturbation that takes the system away from equilibrium, the dynamics will, in general, lead the system back to equilibrium. Thus, the equilibrium state in this sense is a stable state. It is also a robust state in the sense that the equilibrium state is the one with the largest number of fine-grained micro-states or microscopic configurations associated with it. Another sense in which a statistical mechanical equilibrium is robust is that this state is insensitive to modal variation of initial conditions: different *possible* initial conditions will lead back to the same stable equilibrium macro state. A further significant feature of the principles of statistical mechanical equilibria is that they have a natural relationship with the notion of equilibrium found in thermodynamics, where the equilibrium state is that in which the thermodynamical functions of state are approximately invariant. Although such a connection is satisfying from a physical science perspective, it suggests a simple and obvious objection to the viability of this notion of Boltzmannian statistical mechanical equilibrium for the analysis of the stability of social systems. Since social systems are typically wildly out of thermal equilibrium the sense in which we should understand them as stable surely must be disjoint with thermodynamic equilibrium. Moreover, when we are talking about living systems, of which social systems are of course a subset, any reference to ‘the long run’ is surely misplaced: as Maynard Keynes famously remarked, in the long run we are all dead! To understand the stability of social systems we clearly need to move beyond a statistical mechanical or thermodynamical modelling framework. This does not prevent us from using tools of statistical mechanics in the modelling of social dynamics, as we will see in the examples in Section 2.6.

### 2.5. Stability in economics

The final, most difficult and, for our purposes, most significant set of concepts of stability that we should consider are those in economics. By analogy with the physical sciences, the notions of stability found in economics are closely related to the notions of equilibrium. The modern concept of economic equilibrium is a unique or optimal solution to a system of equations in a model. Examples are resource allocation in (Walrasian) welfare economics and vectors of best response strategies in (non-cooperative) game theory. However, as a framework for studying the stability of real economic systems, such an approach is limited because the stability in question in these concepts is essentially modal rather than temporal (these stable states of equilibrium are more like robustness concepts than stability concepts). The history of the concept of equilibrium in economics is complicated but we can distinguish a more intuitive equilibrium concept that pre-dates the modern one. This older version is the equalisation of, for example, supply and demand in a market for trade in goods. Equilibrium in this older sense does arguably identify states that persist upon perturbation and is therefore dynamical rather than modal. Polemicists and policy-makers often make arguments that slide between and conflate these different concepts of equilibrium. ||

Opposed to these theoretical notions of stability, and at least as important, is some kind of overarching empirical stability. How the theoretical notions of equilibrium map onto real systems and to empirical notions of stability is entirely a function of which member of a family of models is under consideration and of the relative tractability (numerical or otherwise) of alternatives. Historically, that mapping was very tenuous or non-existent because tractability constraints were so binding. Over the last 25 to 30 years there has been a sea change in economics mainly because these tractability constraints have been relaxed with more and more computational power becoming available. Economics has undergone a change from a largely theory dominated to a data dominated discipline. This has allowed the focus to be shifted to broader empirical notions of stability (which are arguably closer to older more intuitive conceptions of supply and demand equilibria) which include the absence of large fluctuations in financial markets, low unemployment and inflation, increasing GDP. Thus, the relevant notions of economic stability most relevant prove to be dynamical and wedded to social and political institutions, rather than modal.

Different conceptions of equilibrium, as considered above, can be thought of as nested within more general systems. This may help to better understand the relationship between equilibria and empirical stability. Consider a system S1 (say, a labour market) nested inside a system S2 (say, a macro-economy as guided by a central bank), with models M1 and M2 respectively. If E1 and E2 are equilibria, they will be solutions, respectively, to systems of equations that constitute M1 and M2. E1 will tend to identify a relatively stable state of S1 just in case the state variables that characterize M2 correspond to processes in S2 that evolve on a measurably longer time scale than the state variables in M1 that correspond to processes in S1. Thus one typically cannot assess whether the equilibrium of a model corresponds to an empirically stable configuration of the world without understanding the general ecology of modelled systems in which the model under assessment (i.e., M1) is embedded. Times of profound social and political change might be regarded as circumstances in which the rate of evolution of processes designated by state variables in the highest-level models speeds up, which would then undermine correspondences between equilibria and stability in the lower-level, nested systems. A complication is that S1 and S2 are often entangled rather than constituting a hierarchy. (That is, there are feedback loops between the processes that constitute them.) This can imply conditions in which the fact that a set of state variable values is an equilibrium provides no predictive leverage for assessing the stability of the corresponding state of the world.

An instructive case is that of the 2008 financial crisis which, according to a common charge, had not been foreseen by economists. Macroeconomists running the popular dynamic stochastic

|| International trade economics provides an example of this. A statement such as “universal free trade is the optimally efficient equilibrium” is using ‘equilibrium’ in the modern sense; that is, it is a solution to a set of equations of a model of international trade. It effectively says that in an unregulated world every trade barrier would be ‘arbitraged away’ (effectively disappear through traders who exploit these opportunities). In this sense, the free trade equilibrium is a unique solution in the model. But this isn’t why the economist’s point is supposed to convince anyone to politically support free trade. They’re supposed to be convinced because otherwise there’ll be waste from excess supply or demand somewhere. This latter argument is based on the dynamical sense of equilibrium rather than the modal, ‘modern’ sense of equilibrium. The confusion lies in the fact that unless all countries are small (meaning none can individually influence prices in any sector) there are multiple equilibria in the dynamical sense, and most of them don’t incorporate free trade. So one is only persuaded by “free trade is an equilibrium” if one slides between the two equilibrium concepts. (Of course one might favour free trade on the basis of some other argument.)

general equilibrium (DSGE) models largely believed the economy was stable and that there was little risk of a crisis. Nonetheless, financial economists had predicted the crisis in mortgage markets, and had published models of the mechanisms by which this crisis could be transmitted into the extra-financial economy (Holmstrom & Tirole 1997). The problem was that the financial and DSGE models didn't inform one another. Many macroeconomists think that the charge that they mis-predicted stability is unfair to them. This is because their consultant role, as researchers sharing the concerns of central bankers, was to study the instruments of monetary policy that central banks control. The specific stability they were therefore interested in was around output and investment not financial asset values. And neither output trends in the OECD (which were smoothly rising) nor the trend line in the ratio of investment to output (smoothly falling) were disturbed, except quite briefly, by the 2008 crisis or the subsequent Euro crisis. (There were exceptions in some individual rich countries, namely Ireland, Greece, and Spain.) So DSGE modellers can argue that 'the economy', in the sense that professionally matters to them, was in fact stable just as they said. Note that macroeconomists are divided over the persuasiveness of this defence. This reflects controversy within macroeconomics over whether unemployment is a real variable in the models central banks should use. Roughly, New Keynesians think it is and 'New Classicals' think it isn't. Clearly, the demand for labour was not stable in 2007.

One determinant of the stability of modern democratic orders is the extent to which institutions, especially civil service bureaucracies and agencies, maintain legality and procedural legitimacy against populist and authoritarian threats. It is unlikely that any such unified notion of democratic stability could in principle map onto any refined equilibrium concept from economics. Nonetheless one can consider how institutions, such as legal ones, might support the empirical stability of democratic systems. In current work in progress, Don Ross applies Wynn Stirling's conditional game theory (Stirling 2012, Stirling 2016) to model the capacities of bureaucracies to protect rule of law against populist authoritarian politicians. This protective capacity in Ross' model is a function of the extent to which officials choose strategies and actions under the influence of shared professional norms that they are confident will be maintained by their colleagues. This formal model is intended to provide measures by which the relative strength of this source of stability could be estimated empirically. We will revisit the role of norms in Section 3.3. Next, we discuss some examples of models for complex social systems which use concepts and tools from statistical mechanics and nonlinear dynamics.

## 2.6. Examples of social system modelling

The last twenty years have seen an upswell in a new type of mathematical models of human society, which has been inspired by analogies between social behaviour and the statistical properties of condensed matter physics. Work in this tradition has come to be known as sociophysics (Galam 2012). Although the models developed within this discipline are highly idealized, they have provided insight into the occurrence of extreme political events including democratic backsliding. An important example of sociophysics models of opinion dynamics was developed by Galam (2002) to analyse the causes of minority opinion spreading. The model uses real space renormalization group techniques borrowed from physics to drive the dynamics of opinion formation in a large population of agents divided into small groups. Briefly the model is set up as follows. Each agent is represented by a state variable with two possible states representing two different opinions. Dynamics are determined by the type of agents. *Floater*s, for example, are agents who change their mind according to the opinion of the majority in their vicinity. *Contrarians* oppose the opinion held by the majority, and *inflexibles* never change their mind. The system has two attracting states, each of which is characterised by every agent sharing the same opinion. Which of these attractors the system ends up in is a function of the initial proportion in the population sharing the respective opinion. In the case of all agents being floaters and with no bias in the local majority rule (renormalization group transformation), the threshold is fifty percent. However, the incorporation of heterogeneous agents such as contrarians and inflexibles can shift the threshold from 50 % down to 10 %, thus allowing for the opinion of a tiny minority to propagate and dominate the entire population. These results are important because they suggest a highly non-linear effect of contrarians and inflexibles on opinion formation, which can eventually lead to dictatorial outcomes. The model was used, for example, to explain the French hoax regarding 9/11 (Galam 2003) and to predict the highly improbable political vote outcome in the French referendum on the European constitution (Galam 2002). Empirical studies have confirmed that minorities can have considerable influence, in particular when they are perceived as consistent and competent (Maass & Clark 1984).

Dynamical systems techniques also offer a range of tools to analyse opinion formation. An example is modeling ‘hidden’ opinion formation processes as a saddle-node bifurcation – a change in the nature of the attracting state in a dynamical system from stable to unstable or vice versa (Brummitt et al. 2015). Such an approach, taking into account the perception of other people’s will to rebel and the availability of non-local information about revolutions in other countries, was used to model the Arab Spring rising (Brummitt et al. 2015). We will revisit hidden opinion formation when we discuss the influence of social media on political discourse in Section 3.3. Another example is the modelling of *preference falsification* – misrepresenting one’s wants under social pressure. Preference falsification is relevant for understanding the formation of hidden mass support for anti-democratic sentiments which only becomes visible when political equilibrium is already destabilising. The difference between publicly disclosed and private political preferences was used to explain, for example, why some revolutions were hard to predict (Kuran 1989) and why the polls before Brexit and Trump’s election were apparently misleading (Wu et al. 2018). The recent election of Donald Trump may represent a similar case, as revealed by research into the discrepancy between private and public attitudes towards immigration (Bursztyn et al. 2017).

Complex networks are frequently used to model social dynamics. An example is the user dynamics in online social media. An equilibrium point is considered to be a steady number of active online users of an online social media platform. Assuming a function of cost and benefits for a user being active, the complex network tool of k-core decomposition finds a (Nash) equilibrium between cost and benefit which maximises the number of active users (Garcia et al. 2013). This is reminiscent of economic equilibrium of supply and demand. Despite the apparent simplicity of this approach, calibrating the microdynamics of social behaviour can be a non-trivial task. For example, individual decisions can depend non-monotonically on the social environment (Garcia et al. 2017), leading to highly connected users to trigger cascades of others leaving, an example of an unexpected collective response to changes in the environment of an online community.

Social adaptation can also be modelled with the tools of dynamical systems. While some civilisations persist for centuries, others suddenly disappear over a very short time span (Scheffer & Westley 2007). The Anasazi are an example of this. It is believed that their reliance on established rules, sophisticated norms and traditions created rigidity that slowed down behaviour change and hindered adaptation to new environmental conditions. This phenomenon of a slowing down adaptation ending in catastrophic failure of the system has been modelled as a so-called ‘cusp catastrophe’ – a dynamical system moving on a state-space manifold across points of infinite gradient (Scheffer et al. 2003). A long perspective in the system dynamics of a society informed Turchin’s (2010) prediction of an increase of political instability, as opposed to more optimistic contemporary predictions. This prediction was enabled by the methods of Cliodynamics, i.e. the quantitative analysis of History. In particular Structural-Demographic Theory models society as a dynamical system in which inequality and elite overproduction endogenously generate the kind of instability that we currently observe in democratic countries (Turchin 2016). This approach has shed light on historical stages of political violence such as the French revolution and the American civil war (Turchin 2012), laying the ground for future work on democratic backsliding.

### 3. Stability of democracy

#### 3.1. Destabilisation of social institutions through feedback loops

The above discussion suggests that scholars of democracy need to move away from arguments based on static equilibrium, to more dynamic frameworks, which are better suited to understanding how stable the equilibria are to perturbations. Some perturbations may cause temporary instability. Others may set in train self-reinforcing changes that may have long term consequences, up to and including the transition from democracy to non-democracy. Such a dynamic framework provides a better approach for understanding democratic backsliding than much of the existing work, which (as summarized in (Waldner & Lust 2018)) often focuses on the consequences of static factors for what are inevitably dynamic outcomes. We will now work through some specific ways in which perturbations may affect the political process and lead to politically self-reinforcing consequences.

In the U.S., real income growth has largely stagnated since the 1960s for people at or below the median income. Most of the increase in disposable income has instead benefited top income earners. To illustrate, the total share of income earned by the top 1% of the income distribution has doubled from 10% in 1979 to around 20% in 2013. In 2013 the top 1% of families made more than 25 times as much as the remaining 99% of the population (Sommeiller et al. 2016).



The consequences of inequality are manifold and have been thoroughly examined (see for example (Piketty & Saez 2014, Wilkinson & Pickett 2009)). In a nutshell, greater inequality is associated with poorer health outcomes and social problems, ranging from lower life expectancy, obesity, deteriorated mental health, drug use, lower educational performance, to teenage births and violence to name a few (Wilkinson & Pickett 2009). Particularly relevant in this context is the finding that inequality is also associated with political polarization. For example, U.S. senators from states with high levels of income inequality have been found to be more polarized than other senators (Garand 2010).

As Page & Gilens (2017) have argued, “the health of democracy and the extent of economic inequality have tended to rise and fall together.” The most plausible pathway through which economic inequality can affect the quality of democracy and vice versa involves institutions. Democracy presupposes a basic equality of influence. As economic inequality increases, so too do differences of influence over institutions. Those who have substantial financial resources are better able than those without to influence institutional change.

These dynamics are likely to become especially visible after a financial shock. Consider the following plausible dynamic: under a situation of moderate economic inequality, high influence actors have reshaped institutions to provide them with asymmetrically high economic benefits. They are then able to use some of these benefits to further reshape institutions in ways that advantage them, setting them on a long term trajectory towards high economic inequality, and institutions that reinforce that inequality. This trajectory may appear invisible to most voters, who pay little attention to politics, and in particular to the complex regulatory questions where it is easiest to create self-reinforcing trajectories. However, when there is an exogenous shock requiring painful adjustments, previously invisible institutional protections may become visible, as those who are highly influential shape the adjustments so that the costs fall disproportionately lightly on them, and disproportionately heavily on others. This is likely to generate substantial grievances among a majority that sees itself falling ever further behind. Moreover, the relationship between high inequality and institutional change may have long term consequences by corroding the relationship between voters’ choices and institutional outcomes by, for example, discouraging less well off people from voting, or excluding large groups, and may consequently lead to effective or actual non-democratic rule.

Of course, neither of these tendencies are inevitable. As Page & Gilens (2017) point out, economic shocks and high levels of economic inequality may generate democratic counterpressures too, by giving unhappy citizens greater incentives to mobilize around their shared interests against disproportionately powerful elites. Yet scientific attention to these dynamics - and to understanding when they are simply self-reinforcing, and when they may set in motion countervailing forces - is warranted. Piketty (2015) claimed that

the study of inequality dynamics and institutional change are intimately related. The development of stable institutions and the construction of a legitimate and centralized government are closely linked to the way different societies are able to address the issue of social inequality in a peaceful and orderly manner. ... Institutions do not arise out of harmonious societies populated by representative agents; they arise out of unequal societies and out of conflict.

Recent work by Farrell & Shalizi (2015) explains how some degree of diversity and partisan disagreement in a democracy is healthy and even necessary, if democracy is to harness different perspectives to understand and address joint problems. However, too much diversity may lead to democratic breakdown, as actors forsake the minimal agreement on interests that they need to work together and the minimal agreement on facts that they need to identify and agree on common problems. Waldner & Lust (2018) argue that one crucial causal factor explaining the survival of democracy is whether there is a sufficient coalition that supports democratic institutions. This coalition can be undermined by radicalization and polarization. Radicalization happens when political elites try to reshape politics to secure them permanent advantage by “bend[ing] the rules, ignor[ing] norms, and pursu[ing] strategies that would have seemed off limits.” Polarization involves a breakdown of common faith, so that members of one partisan coalition “ignore potential threats to democracy,” based on the belief that having their opponents come to power would be worse.

These dynamics can become self-reinforcing when political actors have incentives to spread disinformation and increase radicalization and polarization for short term goals. The U.S. has

less heavily regulated media than other advanced industrialised democracies where this has in some cases at least limited this process of polarisation. It is notable that “fake news” has been a problem long before the arrival of online social media. In the U.S., talk radio and Fox News not only catered to a conservative constituency hungry for information and perspectives that validated its beliefs - they helped create that constituency through a feedback loop fed by commercial imperatives between media and audiences (Benkler et al. 2018). In fact, they and their listeners and viewers created an entire mythological universe, in which Barack Obama was collaborating with internationalists to take away guns from conservatives. Most recently, Breitbart.com, prisonplanet.com and other websites have employed a similar approach to build up audience share.

Similarly, partisan competition and the desire to support or frustrate policy goals may lead to feedback loops between media and political actors. The debate on climate change is an example of this. Climate change became a banner of partisan identity thanks to the combination of conservative media and pseudo-scientific think tanks such as the Heartland Institute. Fox News played a key role in publicizing - and indeed organizing - the “Tea Party” (Skocpol & Williamson 2016), despite the misgivings of its CEO Roger Ailes (Sherman 2014).

Both processes may reinforce each other, and both may be reinforced by the newer incentive systems of social media, which have created an algorithmically moderated information market that is characteristically biased towards user “engagement” which in practice may frequently mean shocking or salacious content. Small scale investigations suggest, for example, that YouTube recommendation algorithms tend to lead viewers quickly towards conspiratorial videos (Bessi et al. 2016). Together, social inequality and feedback loops between media, audiences and political actors may weaken democracy through increasing polarization among citizens, and encouraging politicians to bend rules without fearing punishment from their supporters.

### 3.2. Democracy as public good

It is instructive in this context to consider democracy as a ‘public good’, something that is both useful for everyone and that is in principle available to and can be used by everyone. Ideally, the collective use of democracy ensures the allocation of a Nation’s resources to the common social good. However, it is well recognised that public goods are susceptible to the “tragedy of the commons”, another example of a positive feedback loop leading ultimately to collapse. It is often illustrated by the example of cattle herders sharing a common parcel of land on which they are each entitled to let their cows graze. If unregulated, each herder will put more than their allotted number of cattle on the common to receive additional benefits. This will ultimately lead to overgrazing, a harm shared by the whole group. While all herders made the individually rational economic decision, the commons becomes depleted or even destroyed, to the detriment of all. A similar and therefore rather pessimistic prediction has been made for the democratic process of voting. Alexander Tyler (a Scottish history professor at The University of Edinburgh) wrote in 1787 on “The Fall of The Athenian Republic” some 2,000 years prior:

A democracy is always temporary in nature; it simply cannot exist as a permanent form of government. A democracy will continue to exist up until the time that voters discover that they can vote themselves generous gifts from the public treasury. From that moment on, the majority always votes for the candidates who promise the most benefits from the public treasury, with the result that every democracy will finally collapse due to loose fiscal policy, (which is) always followed by a dictatorship.

Reminiscent of the tragedy of the commons, Tyler predicts that a majority of voters exploit their democratic power to “vote themselves generous gifts”, the economic structure collapses and crises follow. Elinor Ostrom and colleagues have studied many cases susceptible to the tragedy of the commons and have outlined an approach, called “collective actions”, which identifies eight design principles for stable local common pool resource management:

- 1-Clearly defined boundaries (effective exclusion of external un-entitled parties);
- 2-Rules regarding the appropriation and provision of common resources that are adapted to local conditions;
- 3-Collective-choice arrangements that allow most resource appropriators to participate in the decision-making process;
- 4-Effective monitoring by monitors who are part of or accountable to the appropriators;
- 5-A scale of graduated sanctions for resource appropriators who violate community rules;

- 6-Mechanisms of conflict resolution that are cheap and of easy access;
- 7-Self-determination of the community recognized by higher-level authorities;
- 8-In the case of larger common-pool resources, organization in the form of multiple layers of nested enterprises, with small local CPRs (common pool resources) at the base level.

It is not obvious, a priori, how to apply these design principles to the democratic process as it is implemented in most modern democracies. But these principles can help in rethinking the voting act because they are designed specifically to interrupt positive feedback loops leading to detrimental outcomes for the common good. One might argue that Switzerland’s direct democracy is close to implementing some of these eight design principles, for instance by having clear boundaries for the political voting process (local, cantonal, and federal votes – principle 1), by regularly putting decisions directly to the voters (principle 3), and by the existence of a clear self-determination and empowerment of communities who see the direct consequences of their votes at the local as well as federal level (principle 7).

### 3.3. Destabilisation of social institutions through erosion of social norms

Democracy is not merely a codification of laws, rules, and regulations. Much of democracy relies on norms, conventions, and expectations of people’s behaviour. Numerous psychological processes thus contribute to the stability of a democracy or threats to it. To shed light on some of these, it is worth revisiting the role of online social media. Much has been written about the impact of social media on public discourse and the multiple threats it presents to democracy (Persily 2017). Two of those impacts have received relatively little attention to date and deserve to be highlighted.

First, there is a plethora of evidence to suggest that the strength with which people hold an opinion is proportionate to the extent to which they believe it to be shared by others. We are unlikely to give credence to a claim that the Earth is flat because we know that everybody believes it to be round. But what if this signal is distorted? We have already seen how extreme views can move into the mainstream when they are legitimized by actual or presumed majority endorsement. There is a related process, uniquely facilitated by social media, that serves to entrench extreme opinions and make them resilient to change. To illustrate, a recent study by Leviston et al. (2013) on people’s attitudes about climate change, showed that only a small minority of people, between 5% and 7% in their Australian sample, denied that climate change was happening. However, those minority respondents thought that their opinion was shared by between 43% and 49% of the population. The massive discrepancy between actual and conjectured prevalence of an opinion (around 40% in this case) is known as the false consensus effect (Krueger & Zeiger 1993). When people believe that their opinion is widely shared, they are particularly resistant to belief revision (Leviston et al. 2013), are less likely to compromise, and more likely to insist that their own views prevail (Miller 1993). The fact that *any* opinion, no matter how absurd, will be shared by at least *some* of the more than one billion Facebook users worldwide (Synnott et al. 2017), creates an opportunity for the emergence of a false consensus effect around any fringe opinion because the social signal is distorted by global interconnectivity.

Second, people’s information diet on social media is now at least partly determined by algorithms (Pariser 2011). It is known that people’s personality attributes can be inferred from their pattern of behaviour on social media. Youyou et al. (2015) showed that a computer algorithm could infer people’s personality on the basis of just 10 Facebook likes more accurately than human work colleagues. This success rate increased with the number of likes, and the program outperformed people’s spouses—the best available human judges—when it had access to 300 likes. When such highly personal data are made accessible to political operatives, as demonstrably happened during the recent Brexit referendum in the U.K. and during the 2016 U.S. presidential election, this opens the door to “micro-targeting” of messages that exploit people’s unique vulnerabilities. It has been claimed that this enabled the small company Cambridge Analytica to depress turnout among potential Clinton voters and to discover hidden Trump voters (Persily 2017). The effects of micro-targeting are not fully understood, and may indeed not be especially significant using current technologies (Benkler et al. 2018). What must be noted, however, is that the micro-targeting of messages may be at odds with the democratic fundamentals. The foundational idea of a democracy is that it provides a public market place of ideas, however imperfect, where competing positions are discussed and decided upon. We suggest that this entails a normative imperative to provide the opportunity that opponents can rebut each other’s arguments. This possibility for engagement and debate is destroyed when messages are disseminated in secret, targeting individuals based on their personal vulnerabilities to persuasion, without their knowledge and without the opponent being

able to rebut any of those arguments. These normative and ethical aspects of microtargeting are only just beginning to be debated (Zuiderveen Borgesius et al. 2018). These impacts of social media on public discourse show how democracies can be vulnerable in ways against which institutional structures and historical traditions offer little protection.

#### 4. Concluding remarks

In this paper, we discussed some of the many mechanisms resulting in institutional instability, ranging from social inequality and financial shocks to disconnected information flow on modern social media. We have illustrated how many of these mechanisms are interconnected, meaning that their temporal and / or spatial dynamics are not separable but influence each other to a significant degree. We have focused in particular on the role of positive feedback (positive in the sense of ‘enhancing’, not in the sense of ‘desirable’) which leads to mutual enhancement. In addition, many of the mechanisms relevant for the democratic process are multiply interconnected. This interconnectedness might seem an insurmountable hindrance for analysing, modelling and, ultimately, understanding these processes for which we, not least, require the knowledge and tools from many different disciplines of the natural and social sciences. But successful cross-disciplinary research of interconnected phenomena is possible. To mention a few examples relevant in this context, Hsiang et al. (2013) showed the strong influence that deviations of climate variables from their mean have on the occurrence of human conflict and institutional change; Sornette & Cauwels (2014) used tools from the statistical mechanical theory of phase transitions to identify the many financial bubbles which were flourishing over the pre-2008 period as a warning sign for the great crash to come; and an example in practice is the SARS epidemic in 2003 which was successfully contained by the WHO in collaboration with officials using tools beyond epidemiology, informed by complex network theory, web-based alert systems, and air transportation control¶.

We conclude that the scientific and quantitative analysis of the question of stability of democracy is possible and, indeed, necessary. It requires a concerted effort across the mathematical, natural and social sciences. In particular, the insights and tools from complexity science deserve more attention in this endeavour.

#### Acknowledgments

This paper arose out of discussions during the workshop “How stable are democracies? — Complex systems perspectives on modern society”, held in Bristol in January 2018, funded through the U.K. Engineering and Physical Sciences Research Council grant EP/E501214/1. DG acknowledges funding from the Vienna Science and Technology Fund through Vienna Research Group Grant “Emotional Well-Being in the Digital Society” VRG16-005.

**Data access statement:** This study did not involve any underlying data.

- Aubin J P 1991 *Viability theory. Systems & control: Foundations & applications* Vol. 10 Birkhäuser, Boston.
- Benkler Y, Faris R & Roberts H 2018 *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* Oxford University Press.
- Berche B, von Ferber C, Holovatch T & Holovatch Y 2009 *The European Physical Journal B* **71**(1), 125–137.
- Bessi A, Zollo F, Del Vicario M, Puliga M, Scala A, Caldarelli G, Uzzi B & Quattrocioni W 2016 *PloS one* **11**(8), e0159641.
- Brummitt C D, Barnett G & D’Souza R M 2015 *Journal of The Royal Society Interface* **12**(112), 20150712.
- Bursztyn L, Egorov G & Fiorin S 2017 From extreme to mainstream: How social norms unravel Technical Report 23415 National Bureau of Economic Research Cambridge, MA.  
**URL:** <http://www.nber.org/papers/w23415>
- Chen J, Wilson C & Tapley B 2006 *Science* **313**(5795), 1958–1960.
- Deffuant G & Gilbert N 2011 *Viability and resilience of complex systems: concepts, methods and case studies from ecology and society* Springer Science & Business Media.
- Farrell H & Shalizi C R 2015 in D Allen & J Light, eds, ‘From Voice to Influence: Understanding Citizenship in a Digital Age’ University of Chicago Press pp. 211–231.
- Freedom House 2018 Nations in transit 2018 Technical report.  
**URL:** <http://www.freedomhouse.org>
- Galam S 2002 *The European Physical Journal B* **25**(4), 403–406.
- Galam S 2003 *Physica A: Statistical Mechanics and its Applications* **320**, 571–580.

¶ <https://www.nytimes.com/2013/03/15/opinion/global/how-sars-was-contained.html>

- Galam S 2012 *Sociophysics: a physicist's modeling of psycho-political phenomena* Springer Science & Business Media.
- Garand J C 2010 *The Journal of Politics* **72**, 1109–1128.
- Garcia D, Mavrodiev P, Casati D & Schweitzer F 2017 *Policy & Internet* **9**(3), 343–364.
- Garcia D, Mavrodiev P & Schweitzer F 2013 in 'Proceedings of the first ACM conference on Online social networks' ACM pp. 39–50.
- Gell-Mann M 1995a *The Quark and the Jaguar: Adventures in the Simple and the Complex* St. Martin's Griffin.
- Gell-Mann M 1995b *Complexity* **1**(1), 16–19.
- Goldenfeld N 1993 *Lectures on phase transitions and the renormalization group* Addison-Wesley.
- Gordon D M 2010 *Ant encounters: interaction networks and colony behavior* Princeton University Press.
- Haddad W M & Chellaboina V 2011 *Nonlinear dynamical systems and control: a Lyapunov-based approach* Princeton University Press.
- Holland J H 1975 *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence* University of Michigan Press Ann Arbor.
- Holland J H 1992 *Daedalus* **121**(1), 17–30.
- Hölldobler B & Wilson E O 2009 *The superorganism: the beauty, elegance, and strangeness of insect societies* W.W. Norton & Company.
- Holmstrom B & Tirole J 1997 *The Quarterly Journal of Economics* **112**(3), 663–691.
- Holovatch Y, Kenna R & Thurner S 2017 *European Journal of Physics* **38**(2), 023002.
- Hsiang S M, Burke M & Miguel E 2013 *Science* **341**(6151), 1235367.
- Kauffman S A 1993 *The origin of order* Oxford University Press New York.
- Kovalenko T & Sornette D 2013 *Davos, Global Risk Forum (GRF) Davos* **1**(1), 7–33.
- Krueger J & Zeiger J S 1993 *Journal of Personality and Social Psychology* **65**, 670–680.
- Kuran T 1989 *Public choice* **61**(1), 41–74.
- Ladyman J, Lambert J & Wiesner K 2013 *European Journal for Philosophy of Science* **3**(1), 33–67.
- Ladyman J & Wiesner K Princeton University Press. Forthcoming.
- Leviston Z, Walker I & Morwinski S 2013 *Nature Climate Change* **3**(4), 334–337.
- Maass A & Clark R D 1984 *Psychological Bulletin* **95**, 428–450.
- Miller C T 1993 *Personality and Social Psychology Bulletin* **19**, 389–398.
- Mitchell M 2009 *Complexity: A Guided Tour* Oxford University Press, U.S.A.
- Mitchell T M 1997 *Machine Learning* 1 edn McGraw-Hill Series in Computer Science.
- Munck G L 2009 *Measuring Democracy: A Bridge between Scholarship and Politics* Johns Hopkins University Press.
- Newman M E 2003 *SIAM review* **45**(2), 167–256.
- Page B I & Gilens M 2017 *Democracy in America?: What has gone wrong and what we can do about it* University of Chicago Press.
- Page S E 2010 *Diversity and complexity* Princeton University Press.
- Pariser E 2011 *The filter bubble: What the internet is hiding from you* Penguin Press New York.
- Persily N 2017 *Journal of Democracy* **28**, 63–76.
- Piketty T 2015 *Journal of Economic Perspectives* **29**(1), 67–88.
- Piketty T & Saez E 2014 *Science* **344**(6186), 838–843.
- Scheffer M, Westley F & Brock W 2003 *Ecosystems* **6**(5), 493–502.
- Scheffer M & Westley F R 2007 *Ecology and Society* **12**(2).
- Seeley T D 2010 *Honeybee democracy* Princeton University Press.
- Sherman G 2014 *The Loudest Voice in the Room: How the Brilliant, Bombastic Roger Ailes Built Fox News—and Divided a Country* Random House Incorporated.
- Skocpol T & Williamson V 2016 *The Tea Party and the Remaking of Republican Conservatism* Oxford University Press.
- Sommeiller E, Price M & Wazeter E 2016 Income inequality in the US by state, metropolitan area, and county Technical report Economic Policy Institute.  
**URL:** <https://www.epi.org/publication/the-new-gilded-age-income-inequality-in-the-u-s-by-state-metropolitan-area-and-county/>
- Sornette D & Cauwels P 2014 *Risks* **2**(2), 103–131.
- Sornette D & Cauwels P 2015 *Review of Behavioral Economics* **2**(3), 279–305.
- Stirling W C 2012 *Theory of conditional games* Cambridge University Press.
- Stirling W C 2016 *Theory of social choice on networks* Cambridge University Press.
- Strevens M 2005 *Philosophy of Science* **72**(4), 531–556.
- Survey W V 2014 World values survey: All rounds Technical report.  
**URL:** <http://www.worldvaluessurvey.org>
- Synnott J, Coulias A & Ioannou M 2017 *Computers in Human Behavior* **71**, 70–78.
- Turchin P 2010 *Nature* **463**(7281), 608.
- Turchin P 2012 *Journal of Peace Research* **49**(4), 577–591.
- Turchin P 2016 *Ages of Discord* Beresta Books, Chaplin, CT.
- Unit T E I 2017 Democracy index 2017 Technical report.  
**URL:** <http://www.eiu.com>
- Waldner D & Lust E 2018 *Annual Review of Political Science* **21**, 93–113.
- Walker B, Holling C S, Carpenter S & Kinzig A 2004 *Ecology and society* **9**(2).
- Weisberg M 2006 *Philosophy of Science* **73**(5), 730–742.
- Wernrdl C & Frigg R 2015 *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **49**, 19–31.
- Wilkinson R & Pickett K 2009 *The spirit level: why more equal societies almost always do better* Allen Lane London.
- Wu K, Wheatley S & Sornette D 2018 Brexit vs trump: quantifying Pound and Peso market efficiency with a

natural experiment Technical report ETH Zurich. preprint.

**URL:** <http://ssrn.com/abstract=2940173>

Youyou W, Kosinski M & Stillwell D 2015 *Proceedings of the National Academy of Sciences* **112**(4), 1036–1040.

Zuiderveen Borgesius F J, Moller J, Kruikemeier S, Fathaigh R Ó, Irion K, Dobber T, Bodo B & de Vreese C 2018 *Utrecht Law Review* **14**, 82–96.