Michaeli, Moti; Nicolaievsky Spiro, Daniel

Working Paper

The dynamics of revolutions

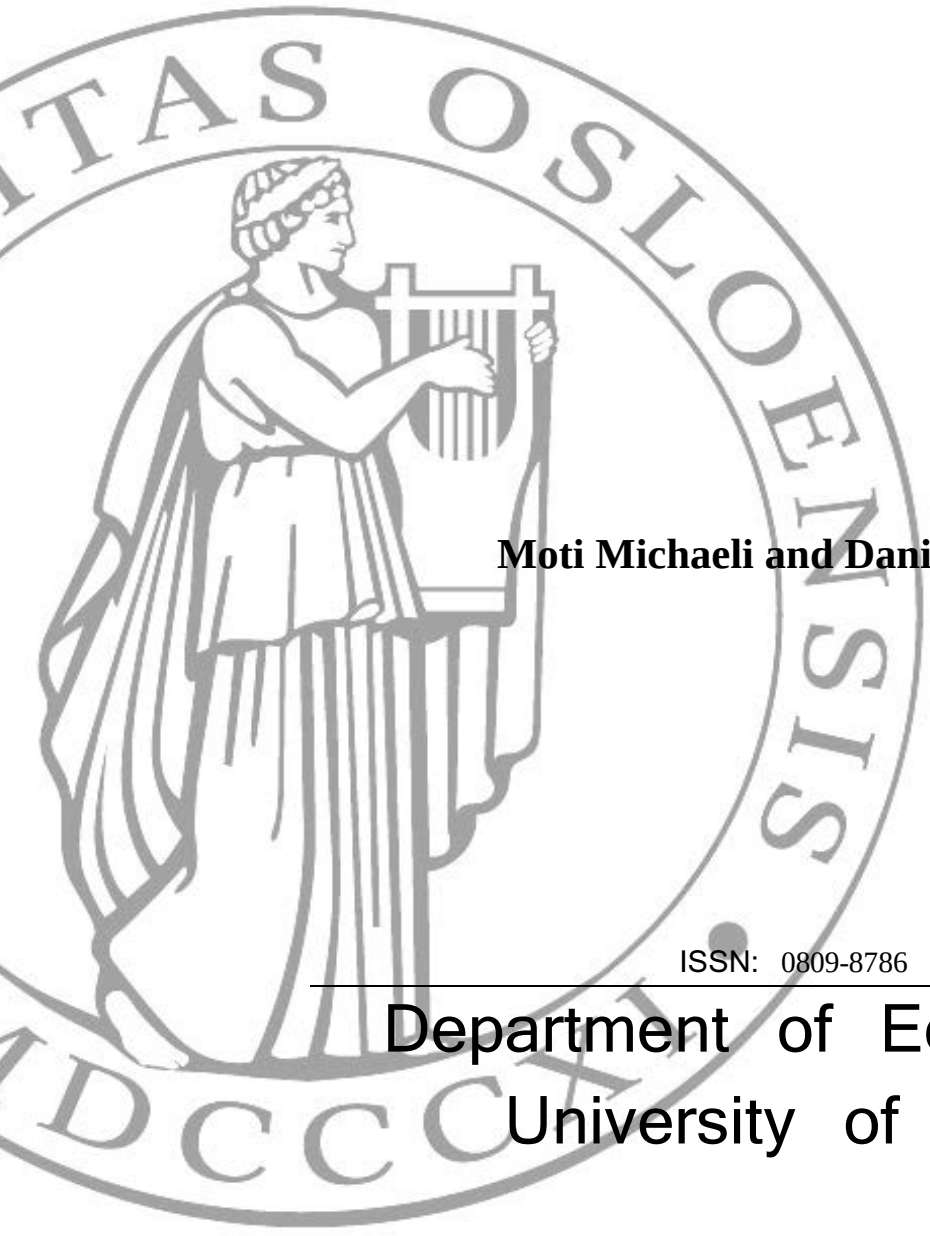
Memorandum, No. 16/2016

Provided in Cooperation with:
Department of Economics, University of Oslo

MEMORANDUM

No 16/2016

The dynamics of revolutions



Moti Michaeli and Daniel Spiro

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 15/16	Geir B. Asheim, Mark Voorneveld and Jörgen W. Weibull *Epistemically robust strategy subsets*
No 14/16	Torbjørn Hanson *Estimating output mix effectiveness: A scenario approach*
No 13/16	Halvor Mehlum and Kalle Moene *Unequal power and the dynamics of rivalry*
No 12/16	Halvor Mehlum *Another model of sales. Price discrimination in a horizontally differentiated duopoly market*
No 11/16	Vladimir W. Krivonozhko, Finn R. Førsund and Andrey V. Lychev *Smoothing the frontier in the DEA models*
No 10/16	Finn R. Førsund *Pollution Modelling and Multiple-Output Production Theory**
No 09/16	Frikk Nesje and Geir B. Asheim *Intergenerational altruism: A solution to the climate problem?**
No 08/16	Michael Hoel *Optimal control theory with applications to resource and environmental economics*
No 07/16	Geir B. Asheim *Sustainable growth*
No 06/16	Arnaldur Sölvi Kristjánsson *Optimal Taxation with Endogenous Return to Capital*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/

The dynamics of revolutions*

Moti Michaeli† Daniel Spiro‡

Abstract

This paper studies the dynamic process of revolutions and mass protests. In a unified framework we explain three classes of revolutions that have been observed historically and earlier models cannot explain: 1) a revolution where the most extreme opponents of the regime protest fiercely and gradually recruit more moderate dissidents; 2) a revolution where moderates, or regime insiders, lead the way and gradually recruit more extreme regime opponents; 3) a revolution where those who dislike the regime the most are gradually pushing the freedom of speech, backed by increased dissent of moderate individuals. These revolutions match the dynamics of many major revolutionary processes, such as the Iranian revolution in 1979, the fall of the communist regimes in eastern Europe in 1989, the Arab Spring in Egypt in 2011, the April Revolution in South Korea in 1960 and the protests on Tiananmen Square in 1989.

Key words: Revolution; Mass protest; Regime; Dissent.
JEL: D74; P26: P5; Z12.

*We wish to thank Sylvain Chassang, Bård Harstad, Anirban Mitra, Kalle Moene, Manuel Oechslin, Paolo Piacquadio, Debraj Ray, Kjetil Storesletten, Sareh Vosooghi, Yikai Wang, and seminar participants at George Washington University, University of Oslo, Tilburg University, King's College, EUI and the ASREC, IMEBESS and NCBEE conferences for valuable comments.

†Department of Economics, European University Institute, Italy and the University of Haifa, Israel. Email: motimich@gmail.com.

‡Corresponding author, Department of Economics, University of Oslo, Norway. daniel.spiro@econ.uio.no, Tel: +47 22855137, Fax: +47 22855035.

1 Introduction

Throughout history, revolutions have led to fast and massive changes in institutional, economic and social environments and, as such, most social-science disciplines have been interested in understanding their causes and dynamics. It is common to divide revolutions against a regime into two classes (Tanter and Midlarsky, 1967). First, *coup detats,* performed by elites or a competing party to the regime.[1] Second, *major revolutions,* driven not by a small group of elites but by popular protest and large social movements. This paper is concerned with the latter, which includes, e.g., the French revolution, the toppling of the Shah in Iran in 1978-79, the collapse of the communist regimes in Eastern Europe and the recent Arab Spring. In particular, we are interested in understanding who will participate in a revolution, which stances these individuals will express and what may spark the revolution.

The workhorse model of revolutions and mass protests – developed by Granovetter (1978) and discussed and applied in a series of papers by Kuran (1989a, 1989b, 1995) – is binary. That is, each individual can either support the regime or protest against it, individuals differ in their propensity for each of these two alternatives and, importantly, the larger the share of individuals that choose an alternative is, the more each individual is inclined to choose so as well. The binary model provides valuable insights on, for instance, thresholds for regime stability. However, it is very limited in its ability to explain different patterns of revolutions. Since an individual in that model only has the choice between complete obedience to the regime and full-blown protest, the binary model is silent about who in society – moderates or extremists, leftists or rightists (or both) – will participate in the revolution at its various stages, how fiercely each individual will protest, and how this will change over the course of the revolution. To see why these aspects are important, consider the following illustrative examples of three distinct classes of mass movements and revolutions (a richer account of these examples is provided later in the paper).

1. A wave-type revolution going from the outside-in, which can be illustrated by the Islamic Revolution against the Shah in Iran in 1978-79 (Razi 1987, Moaddel 1992, Ghamari-Tabrizi 2008). The protests started by religious extremists and gradually less extreme fractions of the population joined, until the Shah's closest support also abandoned him. The central characteristic of this type of revolution is that those who are most critical to the regime initiate the revolution by dissenting fiercely and gradually less critical individuals join the protests but dissent less than the initiators.

2. A wave-type revolution going from the inside-out, which can be illustrated by the Arab Spring in Egypt in 2011. The first protests were performed by moderate liberals and conservatives, while those most critical to Mubaraq's regime – the Muslim

[1] Examples of these are plentiful in both Africa and Latin America and they are typically modeled by assuming the existence of an elite group in society (e.g. Acemoglu and Robinson, 2001).

Brotherhood and the Salafis – were the last to join (BBC 2013). The central characteristic of this type of revolution is that those most critical to the regime are not taking part in the protests initially. Rather, the revolution starts with moderates expressing moderate views and, gradually, less moderate individuals join and express more harsh critique. Another example of this kind of revolution is the fall of the communist regimes in Hungary and Poland in 1989 (Lohmann, 1994).

3. A stretching-out type of revolution, which can be illustrated by the April Revolution in South Korea in 1960. What started as a students' protest against a governmental directive to attend school on a Sunday, turned into a massive riot by several hundred thousand people, led by the students, demanding (and succeeding) to overthrow the regime (Kim 1996). The central characteristic of this type of revolution is that the expressed dissent by all participants, leaders and followers alike, becomes more extreme over time. Another example of this kind of revolutionary process is the evolution of the protests on Tiananmen Square in Beijing in 1989, which eventually failed to lead to a change of regime.

As can be noted from these examples, a revolution will not always start with the most extreme opposition, as implicitly implied by the binary model. Neither is the critique homogenous – it is differentiated amongst protesters and is also changing over time. The purpose of this paper is to present a unified framework whereby all three classes of revolutions and mass protests can be explained. Our theory provides predictions for which of the three classes of revolutions will occur, and hence which individuals will participate in a revolution at various stages and how extreme their stances will be. We also answer questions such as: What are the catalytic events starting a revolution? When will a revolution be one-sided and when will there be critique coming from both ends of the political spectrum? At what stage is a revolution most likely to fail and what can the regime do to achieve it? We also highlight when the revolutionary momentum is mainly driven by new recruits and when by gradual increases of dissent of current participants.

The theory we develop contains all the core components of the standard binary model with a seemingly simple extension. Rather than having a binary choice between obeying the regime or dissenting against it, an individual can choose the *extent* to which she dissents from a continuum. The more an individual dissents, the more she will be sanctioned. What makes her possibly dissent despite this sanctioning is that she has a private bliss point (a political view or an economic interest) from which it is costly for her to deviate. Hence, each individual trades off the sanctioning for disobeying the regime and the cost of deviating from her bliss point. This individual trade off is conceptually the same as the individual trade off relevant for choosing the extent of norm conformity in Michaeli and Spiro (2015). However, the current paper analyzes the full dynamic consequences of the collective behavior and there is a central addition: the strength of the regime (i.e., how heavily it is able to

sanction dissent) is endogenous – it decreases in the extent of dissent in society in terms of the number of dissenters and the views they publicly express. This further implies that staying silent is equivalent to supporting the regime, as this silence does not contribute to weakening the regime. This creates an interaction between the population and the regime, and the behavior of each individual has a positive externality on other individuals, thus also capturing the collective-action problem emphasized in the binary models (Granovetter 1978 and Kuran 1995).

Our static analysis characterizes the conditions for a steady state with a stable regime and the equilibrium extent of dissent in terms of which individuals dissent and what opinions they express. Our dynamic analysis characterizes the kind of shocks and policy changes that can destabilize this steady state – referred to as catalytic events – and how the dissent will change following such events. The evolution of dissent depends, to a large extent, on the sanctioning structure the regime is using. A regime that uses a concave sanctioning system barely differentiates between small and large dissent and essentially requires full obedience by the individual. Then, since those who dislike the regime the most perceive the highest cost of obeying it, it will be these individuals – the extremists – who will be first to dissent and the low marginal punishment once dissenting will push them to express very deviant views. During the course of the revolution the overall sanctioning power of the regime will gradually fall, and less extreme dissent will follow too, which explains the outside-in pattern of the first class of revolutions (Iran 1979). In contrast, a regime that uses a convex sanctioning system ensures that small dissent is not so costly while large dissent is very costly. Hence, under such sanctioning, no one dissents a lot. During a revolution, the sanctioning gets gradually weaker and hence the maximal dissent increases over the course of the revolution – the freedom of speech is pushed further. This explains the inside-out nature of revolutions of categories 2 and 3 (Egypt 2011 and South Korea 1960 respectively).

Who in society will be dissenting (and thereby pushing the revolution) depends on the cost of deviating from one's ideological bliss point. In a society that is characterized by individuals with a convex cost of deviating from their bliss points, individuals will find it easy to deviate slightly from their bliss points, while large deviations will be very costly. Hence, in such societies, individuals with private views close to the regime will tend to obey it, while individuals whose private views are very far from the regime's will dissent more. Consequently, the most extreme types will be the ones dissenting the most and thus leading the way during the revolution. This explains the pattern of the first (Iran) and third (South Korea) classes of revolutions. In contrast, in a society that is characterized by individuals with a concave cost of deviating from their bliss points, individuals will find it very costly to deviate even a little from their bliss points, but deviating more will not change much for them. Hence, if they do deviate, they might as well largely align with the regime, for instance by remaining silent. Those who will find it the hardest to express their true opinions and hence are prone to stay silent are the extremists, because their private

views are sanctioned more than the private views of moderates. Thus, they will be the ones aligning with the regime. This means that the most deviant expressions will be stated by moderates who speak their minds. During the revolutionary process, as the regime's sanctions get weaker, more and more extreme types will find it possible to express their private opinions and thus start dissenting. This explains the pattern of the second class of revolutions (Egypt), whereby the most extreme regime opponents are the last ones to join the revolution and the revolution is initiated by moderate forces or party insiders.

The curvature of the sanctioning has further implications for whether the dissent will be two-sided or one-sided. To help fix ideas, consider a society where the individuals' bliss points are distributed along a left-to-right political scale. Suppose now the regime's policy is biased to the left. In a wave-type revolution going from the outside-in (class 1), the only one dissenting initially are the extremists. With a sufficiently left-biased regime there *are* no types on the left who are extreme vis-à-vis the regime, which implies that dissent is expressed only on the right. Consequently, here a revolution will initially be one-sided. However, at some point, if the revolution evolves sufficiently, the regime becomes so weak that moderate types will dissent as well and hence there may be dissent also on the left. At this point the revolution gains momentum (recruiting supporters on both sides) and the regime is bound to collapse. Hence, the revolution will be fragile initially, when it only recruits on one side, and it is at these early stages that the revolution may fail. In contrast, in a wave-type revolution going from the outside-in (class 2), the only dissenters initially are moderates. So even when the regime is left-biased, there will be dissent both on the left and on the right of the regime. Consequently, a revolution will be two-sided right from the very start with "strange bedfellows" – some of which are criticizing the regime's policy from the left and some of which are criticizing it from the right (this was particularly clearly observed during the Arab Spring in Egypt where the first protesters were both moderate liberals and moderate conservatives). As the regime gets weaker during the revolution, views that are less moderate are expressed as well, and at some point the left side of the political spectrum is exhausted of new recruits and the revolution loses momentum. Hence, in this case it is in the late stages of the revolution, when new recruiters come only from one side of the spectrum, that the revolution may fail. Finally, the third class of revolutions is not a wave-type revolution hence its development is not built upon recruiting new dissenters but instead on the gradual increase in the extent of dissent by moderates and extremists alike. Here the revolution is two-sided right from the start and is not particularly fragile in any of its stages – if the regime does not take action against it, it is bound to succeed.

Most of the previous theoretical literature on major revolutions and mass protests utilize a binary model (see, for instance, Granovetter 1978; Kuran 1989a, 1989b, 1995; Naylor 1989; Angeletos et al. 2007; and Rubin, 2014), from which our model is clearly different.[2]

[2]Note that in Rubin's (2014) paper, the individual has a binary decision to support or not support the regime but the political regime itself can choose a more popular political policy (on a continuum) to avoid social unrest. However, the existence of a political regime is taken as exogenous.

More broadly, however, obeying a regime and conforming to a social norm are theoretically quite similar and in the literature on social norms some non-binary models exist (Bernheim 1994, Kuran and Sandholm 2008, Manski and Mayshar 2003, Michaeli and Spiro 2015) and contain a similar individual trade-off as ours. Bernheim (1994), Manski and Mayshar (2003) and Michaeli and Spiro (2015) are concerned with static equilibria. Kuran and Sandholm (2008) analyze integration between groups and no regime exists in their framework.

Granovetter (1978) and Kuran (1995) offer a dynamic setting in which individuals take the actions of others as given. That kind of analysis, which abstracts from strategic considerations on the individual level, seems adequate for analyzing major revolutions and mass protests whereby, literally, millions of individuals may participate. We therefore adopt this approach in our paper too. Strategic considerations by revolutionary participants are analyzed by Angeletos et al. (2007) and Edmond (2013), but in the limited binary framework. Strategic behavior of revolutionary leaders has been analyzed by Bueno de Mesquita (2010) and Shadmehr (2015a), where the latter is the paper most related to ours, as it endogenizes the policy of those leading the protest. In particular, in Shadmehr's (2015a) paper the protest leader offers an alternative policy to the regime while taking into account the support she will get by the population. The modeling of Shadmehr (2015a) implies that those with extreme views are in all scenarios part of the revolution while those sufficiently close to the regime are never part of the revolution. As we exemplify with the Arab Spring events in Egypt, a broader account of revolutions reveals that in some revolutions the extremists will be the last to join a revolution that is initiated by the individuals closest to the regime. Furthermore, we answer questions (not analyzed by Shadmehr 2015a) such as how the statements will evolve over time, who will join the revolution at what point in time, when a revolutionary movement will contain ideological adversaries and when a revolution is most likely to fail. Our focus is mainly on the dynamics during the revolution, just like in Granovetter (1978) and Kuran (1995).

Our paper is also related to the literature on collective action (see Olson 1971 and Tullock 1971 for early treatments and, e.g., Oliver & Marwell, 1988; Esteban, 2001; Esteban & Ray, 2001 for more recent work). This literature focuses on individual homogeneity, where all agree what would be a collectively good outcome but individuals are disincentivized to take an action themselves. In contrast, we focus on individual heterogeneity of preferences or ideology – aspects that have been identified as important determinants of revolutions (Goldstone, 2001).

The next section outlines the model and presents the main results. Sections 3-5 analyze, each in turn, the three classes of revolutions more in depth and provide more details on the historic examples briefly discussed earlier. Section 6 provides empirical predictions and Section 7 concludes.

2 The model

We start by describing a static version of the model and then add a dynamic structure to it. Society consists of a continuum of atomistic individuals and of a political regime. The regime has a policy $R \in [-1, 1]$ which can be thought of as a point on a left-to-right political scale. Focusing on revolutions and mass protests against a regime we let R be exogenously given (though there are several straight forward options of endogenizing it). Each individual expresses a political opinion (or stance) $s \in \mathbb{R}$, where $s = R$ is equivalent to being silent. The regime sanctions expressions that deviate from its policy, with larger deviations representing harsher critique of the regime, which in turn is sanctioned more heavily. This is represented by the following *punishment* function P

$$P(s, R, K) = K |s - R|^{\beta} \ , \ \beta > 0. \tag{1}$$

The overall severity of punishment (sanctioning), as captured by the scaling variable K in (1), is endogenous and depends on the aggregate dissent in society. Let S denote a distribution of stances in society. The *approval* of the regime, denoted by A, has the following properties.

$$A \in [0, 1] : \text{for each individual, } A(s, S) \geq A\left(s', S\right) \text{ iff } |s - R| \leq \left|s' - R\right| \tag{2}$$

That is, the approval of the regime is decreasing the more dissenting each individual is. The overall severity of punishment, to which we also refer as the strength of the regime, is then given by

$$K = \bar{K} A$$

where $\bar{K}$ is an exogenous parameter capturing the *force* of the regime. K is endogenously capturing the actual strength of the regime, so that the more approving the population is of it, the easier it is for the regime to punish dissenters. One interpretation is that $\bar{K}$ represents the per capita law-enforcement forces actively used by the regime to sanction dissent, with A representing the proportion of them that stay loyal to the regime when asked to use force against dissenting civilians. Then, the more dissent there is, the less likely it is for an individual dissenter to get caught by the regime. β captures the curvature of the sanctioning system, which will be important for the analysis. A regime with a large β (> 1) uses convex sanctioning and hence is tolerant to critique as long as it is not too extreme. A regime with a small β (< 1) uses concave sanctioning whereby it punishes rather heavily even small dissent but does not distinguish much between small and large dissent.

Each individual has a privately preferred political policy or opinion $t \in T \subset \mathbb{R}$, also referred to as the individual's bliss point or type. When expressing a stance s, the individual

bears a cost for deviating from her bliss point:

$$D(s,t) = |s-t|^{\alpha} \ , \ \alpha > 0. \tag{3}$$

D can be interpreted as discomfort from expressing a political opinion not in line with a person's conviction (or, if t reflects one's most preferred economic behavior, as a material cost of deviating from the person's economic interests). The choice problem of an individual with $t \neq R$ is how to trade off the sanctioning when dissenting against the regime and the disutility of deviating from her own privately held opinion. That is, the individual minimizes

$$L(s;t,R,K(S)) = D(s,t) + P(s,R,K(S)). \tag{4}$$

It is immediate from this choice problem that the individual will take a stance somewhere weakly in between t and R. The extent to which the individual feels forced to go towards the regime depends on the regime's strength $K(S)$ and hence indirectly on the stances taken by all individuals in society.

Being interested in revolutionary dynamics, i.e., how a regime's strength interacts dynamically with the behavior of individuals in society, we will now add a simple dynamic structure to the model. These dynamics are standard in games with large populations and in the analysis of revolutions (Granovetter 1978; Young 1993; Kuran 1995; Kaniovski et al. 2000; Young 2015). The stances at period $i+1$ is a mapping from the type space to the stance space $s^*_{i+1} : T \to \mathbb{R}$, such that,

$$s^*_{i+1}(t,R,K(s^*_i)) = \arg\min_{s} \{L(s_{i+1};t,R,K(s^*_i))\}. \tag{5}$$

This dynamic formulation means that, at period $i+1$, the strength of the regime is determined by the regime's approval at period i, hence

$$A_{i+1} = f(A_i)$$

where f describes the dynamics of approval between periods. For example, this could represent the loyalty or determinism of the regime's troops at day $i+1$ of a revolution after observing (and internalizing) the dissent of the population on the previous day. We wish to emphasize that we choose these adaptive dynamics for tractability and for brevity in presenting the results but the specific dynamic modeling does not drive our results.[3] A steady state (which is equivalent to a Nash equilibrium in the static model) is achieved when $s^*_{i+1}(t,R,K(s^*_i)) = s^*_i \ \forall t$, which also yields $f(A_i) = A_i$. We set $A_i = 0$ when $s^*_i(t) = t \ \ \forall t$, in which case we say that no regime exists (this is always a steady state as it implies that

[3] We have also solved a version of the model with forward-looking agents and strategic interaction between the agents. The main results about the three classes of revolutions are the same but it is substantially more complicated to show our further results about regime stability and failed revolutions.

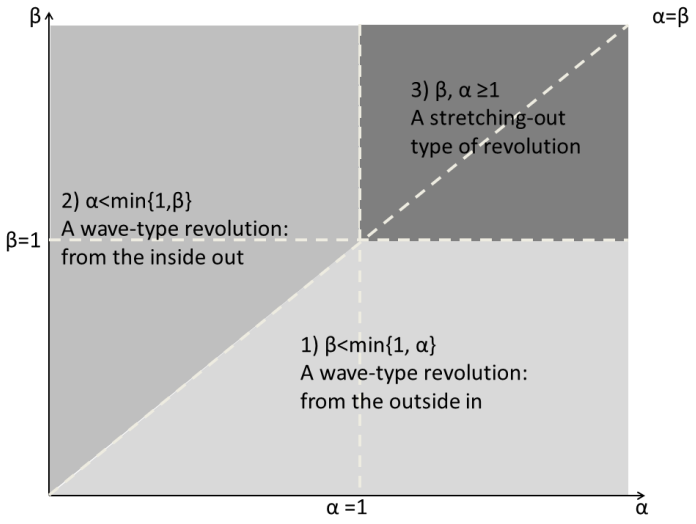


Figure 1: Parameter space of the different classes of revolutions.

$P(s, R, K) = 0$ and so $s^*_{i+1}(t) = t \ \ \forall t$). We will consider a steady state to be stable, with its approval denoted by A_{ss}, if there is convergence back to it following a small perturbation to A_{ss}. Otherwise the steady state is unstable, with its approval denoted by A_{uss}. Our measure for the stability of the regime following a shock to its approval is the distance between the regime's approval at the steady state A_{ss} and the approval in the closest unstable steady state to the left, i.e., $A_{ss} - A_{uss}$, because the zone of convergence to A_{ss} from below is (A_{uss}, A_{ss}). A revolution is defined as a dynamic process where the approval is converging to a new, lower, steady state (i.e., a revolution is not a situation where a small change to a parameter leads to a small change in the steady state approval). A successful revolution is one where $A = 0$ in the new steady state and a failed revolution is one where $A > 0$ in the new steady state. Catalytic events are events that may trigger a revolution. These are exogenous changes or shocks that either imply that a previously stable steady state seizes to exist or decrease the approval to a point where the approval will, endogenously, decrease further.

The main focus of our analysis is on the evolution of participation (i.e., which types dissent) and of statements (i.e., which stances they express) during the revolution. For short, we will refer to individuals with private views far from the regime (large $|t - R|$) as extremists and we will refer to those with private views close to the regime (small $|t - R|$) as moderates. That is, a type's extremeness is always relative to the regime – a liberal democrat under the Taliban regime is an extremist in our definition. The model predicts three classes of revolutions depending on the combination of the parameters β and α, as depicted in Figure 1. The following proposition expresses the main features of these revolutions.[4]

Proposition 1 *There are three exhaustive classes of revolutions:*

[4]For brevity we ignore here the special case of $\alpha = \beta \leq 1$ with its unique technicalities.

1. *(**A wave-type revolution from the outside-in**) If $\beta < \min\{1, \alpha\}$, then initially the dissenters in the revolution are extremists, and later in the revolution more moderate types join but dissent less than the initial extremists.*

2. *(**A wave-type revolution from the inside-out**) If $\alpha < \min\{1, \beta\}$, then initially the most deviant expressions are mild and are made by moderates, and later in the revolution the most deviant expressions are extreme and are made by extremists.*

3. *(**A stretching-out type of revolution**) If $\beta \geq 1$ and $\alpha \geq 1$ (but at least one inequality is strict), then throughout the revolution the most dissenting types are extremists, and during the revolution all types gradually increase their dissent.*

These three classes of revolutions fundamentally differ in how participation and statements evolve during the revolution. The first is a revolution that starts with extremists voicing extreme critique and gradually recruiting individuals who are more moderate (thus resembling a wave of dissent from the outside-in). The second is a revolution in which moderates start voicing their mild critique of the regime, gradually making more and more extreme individuals stand up for their views as well (thus resembling a wave from the inside-out). The third is a revolution in which the composition of participants and their inner ranking (in terms of who dissents more) are constant throughout the revolution, and where everyone constantly increases her dissent (hence the name – stretching-out). This revolution spreads from the inside-out as well, in the sense that the expressed dissent becomes more and more extreme.

In the upcoming three sections we analyze each one of these classes of revolutions separately, provide intuition for the dynamics described in Proposition 1, more results on the stability of regimes and catalytic events initiating a revolution, and predictions on when a revolution is most likely to fail and which measures can help the regime stay in power. Most of the results in the paper, and in particular those stated in Proposition 1, hold with the very general formulation of the approval function A in (2) and in fact can be derived also with more general functional forms for P and D. However, showing some specific further results requires a more explicit functional form of A and of the distribution of types. For analytical tractability we will assume throughout that $t \sim U(-1, 1)$ and that the approval of the regime is linear in the deviations from it. That is,

$$A = \max\left\{0, 1 - \lambda \int_{-1}^{1} |s(t) - R| \, dt\right\}. \tag{6}$$

This is a special case of (2) where $A = 1$ if nobody dissents ($s(t) = R \ \forall t$) and the regime attains its maximum strength $\bar{K}$. We normalize $\lambda = 1$ so that a non-biased regime ($R = 0$) has zero approval precisely when all types speak their minds ($s(t) = t \ \forall t$).[5] This ensures

[5] I.e., $\lambda = 1/\left(2\int_0^1 t dt\right) = 1$.

that for any $R \in [-1,1]$ we get $A = 0$ when all types speak their minds ($s(t) = t \ \forall t$). It further implies that, when a regime is biased ($R \neq 0$), A may equal 0 also without all types speaking their minds (which can be motivated by the fact that, when all speak their minds under a biased regime, dissent is larger than when all speak their minds under a non-biased one). This normalization is mostly without consequence apart from implying that even a central regime loses all of its strength when all speak their minds (which would not be true for $\lambda < 1$). Throughout the paper we relate to $|R|$ as the bias of the regime, and to a regime with $R \neq 0$ as a biased regime.

3 A wave-type revolution: from the outside-in

3.1 Analysis

We start by considering the case where $\beta < \min\{\alpha, 1\}$. This case can be further divided into two subcases: $\beta < \alpha \leq 1$ and $\beta < 1 < \alpha$. While these two cases differ in some details, they are largely the same from the point of view of what we are interested in. Hence, for brevity, we will focus on the subcase $\beta < \alpha \leq 1$.[6]

By differentiating L twice with respect to s it is immediate that when $\beta < \alpha \leq 1$ the second order condition is not fulfilled, implying that an individual will choose either $s(t) = R$ or $s(t) = t$. It is simple to further show that there exists a cutoff distance Δ such that all types closer to the regime than Δ will fully follow the regime ($s(t) = R$) while types further from the regime than Δ will speak their minds ($s(t) = t$), as illustrated for a biased regime in Figure 2. Hence, in this type of society, the regime induces silence by those who largely agree with it. The intuition is easy to understand. The important property of this case is that β is relatively small, which implies that the regime applies a (very) concave punishment whereby even small dissent is heavily punished while more extreme dissent is punished only slightly more. This will induce an individual to either fully follow the regime or, if she does not fully follow it, she may as well dissent quite a lot. Then, since types far from the regime perceive the highest cost of discomfort from following the regime, it will be these types who may dissent – and dissent quite a lot if they do given their extreme views – while types close to the regime will fully conform. The cutoff between those following the regime and those who do not (if they exist), denoted by Δ, is naturally increasing in the strength of the regime K, so that a stronger regime has less dissent.

This result, that extremists speak their minds and moderates are completely supporting the regime, has important implications for the stability of regimes and for the revolutionary dynamics as expressed in the following proposition.

Proposition 2 *When* $\beta < \alpha \leq 1$:

[6] See sections A.2.2 and A.3.1 in the appendix for a treatment of the other subcase ($\beta < 1 < \alpha$).

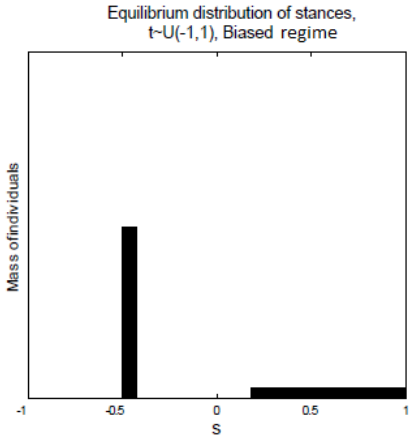


Figure 2: An illustration of an equilibrium distribution of stances for a biased regime for $\beta < \alpha \leq 1$.

1. ***Existence of a stable steady state****: A stable regime exists iff it employs sufficient force, and the more biased its policy is the more force it needs to employ.*

2. ***Catalytic events****: A revolution may start following a shock to the regime's approval or force or following implementation of unpopular policies.*

3. ***Revolutionary participants****:*

   (a) *Initially only the most extreme types participate in the revolution, but over time types who are more moderate join it too.*

   (b) *For any regime with $|R| \neq 0$, the revolution will start only on one side of the political spectrum.*[7]

4. ***Revolutionary statements****: The revolution goes from the outside-in, with initially only very extreme statements, and then, as more people join the revolution, the new statements become more moderate.*

We start by explaining the dynamics of the revolution (parts 3 and 4) since this largely determines what makes a regime stable and which events may initiate a revolution. The revolutionary process follows from the dynamics of the cutoff between those who dissent and those who do not (Δ_i). As explained earlier, when the regime uses (very) concave sanctioning, it induces dissent by extremists but not by moderates. This means that, if a revolution starts, the first ones to dissent are the most extreme types. When these extremists start dissenting, the strength of the regime falls, which makes it possible also for less extreme types to dissent. This way, increasingly moderate types join the revolution and they dissent less than those who started it – the revolution goes from the outside toward

[7] Unless there is a very large shock to the force or approval of the regime.

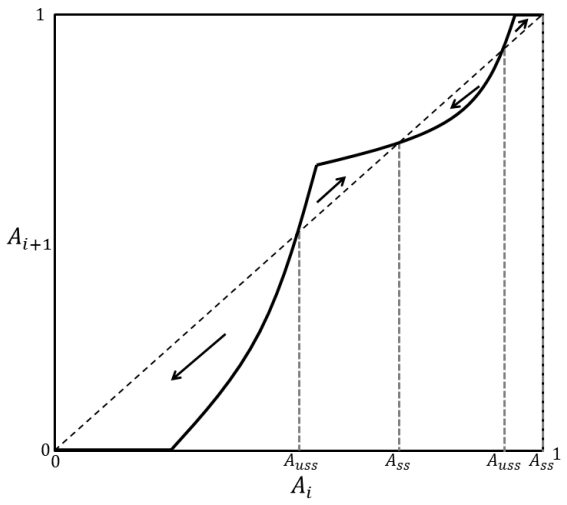


Figure 3: Stylized phase diagram for the case $\beta < \min\{1, \alpha\}$ and biased regime. The full line depicts the equilibrium function $A_{i+1}(A_i)$ and the dashed line depicts the 45-degree line where $A_{i+1} = A_i$. The vertical lines depict the stable (A_{ss}) and unstable (A_{uss}) steady states.

the inside (as summarized in parts 3a and 4 of the proposition). If the regime is, say, left of center, the first dissenters will be on the far right – the revolution starts only on one side (part 3b of the proposition). During this phase, the revolutionary momentum is rather low, since new recruits are only coming from the right, while later, if the regime has gotten sufficiently weak, new recruits might appear also on the left side of the regime. This has important implications for the fragility and success of a revolution as will be explained later.

As a tool to understand the additional results, consider the phase diagram in Figure 3, which depicts a stylized example of the intertemporal-dynamics function $A_{i+1}(A_i)$ for a moderately biased regime ($|R| \in]0, 0.5[$). The higher is current approval (A_i), the higher is the regime's strength (K_i), which implies less dissent thus higher approval in the next period (A_{i+1}). Hence, A_{i+1} is a (weakly) increasing function of A_i as can be seen in the figure. Quite naturally, for any approval level A_i, an increase of the regime's force $\bar{K}$ raises A_{i+1} (through an increase of $K_{i+1} = K(s_i^*)$). This implies that the function $A_{i+1} = f(A_i)$ in Figure 3 shifts up as $\bar{K}$ is increased. For sufficiently small $\bar{K}$, A_{i+1} is always below the 45-degree line, implying no stable regime exists which naturally implies that there is a minimum amount of force a regime has to employ to stay stable. For sufficiently large $\bar{K}$ there may exist one, two, or three (inner) intersections with the 45-degree line. The first intersection from the left is an unstable steady state, the second is stable and the third is unstable. Additionally, as is the case in the figure, there is one stable steady state at $A_i = 0$ (where the regime does not exist by our definition) and there may be one at $A_i = 1$.

As those with views far from the regime are the ones dissenting, a biased regime, with policies far from most of the population's views, will induce more dissent for any given level

of regime strength K_i (biasness shifts down the dynamic approval function in Figure 3). Hence more force ($\bar{K}$) is needed as a compensation for the existence of a stable steady state (part 1 of the proposition).[8] Thus, increasing the bias decreases the approval function (A_{i+1}) for any current level of approval (A_i), so that the unstable steady states move right while the stable steady states move left in the phase diagram. This means that biased regimes are inherently less stable and that an implementation of unpopular policies (increase in the bias) may ultimately be the catalytic event that starts a revolution by making a previously stable steady state seize to exist (as stated in part 2). Other catalytic events are a shock to the regime's force or a shock to the approval, if they throw A_{i+1} outside the convergence zone of the current stable steady state, leading the political system to a zone with convergence downwards.

The further properties of the phase diagram depicted in Figure 3 is that the function $A_{i+1}(A_i)$ is first flat near zero (unless $R = 0$), then rises convexly, kinks downwards and then rises convexly again. The flat initial part is where current approval is so low that the regime will not be able to gain any approval at the next period ($f(A_i) = 0$). To see why a kink exists, consider a left-biased regime (like the one depicted in Figure 2). When approval A_i is low, there will be dissent on both sides of the regime in the next period – Δ_i is small. As A_i increases, the dissent falls on *both* sides of the regime implying A_{i+1} is a steep function of A_i. The kink is the point where A_i induces $\Delta_{i+1} = 1 - |R|$ (in Figure 2, $1 - |R|$ is the distance from the regime to the left edge corner). After this point there is no way to further increase dissent on the left side of the regime. From here onwards an increase in A_i will reduce dissent only on the right side of the regime, as illustrated in Figure 2, which implies that A_{i+1} becomes less steep.

This description of the phase diagram applies to a moderately biased regime ($|R| \in]0, 0.5[$). When the regime is instead very biased ($|R| \in [0.5, 1]$), it is also weaker, as explained above. This has implications for the phase diagram, which will not contain the left convex part hence will not have a kink. This reflects the fact that a very biased regime is so far from the dissenters at the opposite extreme of the political scale that approval goes to zero whenever Δ_i is sufficiently small to imply dissent on both sides. In fact even if dissent is only on one side of the regime but is practiced by sufficiently many individuals on that side, it will be sufficiently strong to break the regime. It also means that if the regime is, say, very biased to the left, and a revolution starts on the far right, the regime will collapse by a revolution that is purely right-wing (i.e., all participants have $t > 0$). The fact that it is enough for the initiators of the revolution to recruit people on their own side in order to successfully topple the regime is a reflection of the weakness of very biased

[8] More precisely, an increase in bias shifts the convex part to the right of the kink downwards and at the same time widens this part outwards in both directions. This has to do with the fact that the biasness affects dissent not through affecting Δ – which is independent of $|R|$ – but through affecting the actual mass of types at distance larger than Δ from the regime, which increases in $|R|$ when the regime is sufficiently biased to induce dissent only the opposite extreme, as can be seen in Figure 2.

regimes. Thus, it does not apply to revolutions against regimes that are only moderately biased (i.e. $|R| \in]0, 0.5[$), where the kink in their phase diagram has important implications for the success and failure of revolutions.

As explained earlier, when the kink exists in the phase diagram there may be up to two stable steady states with $A_{ss} > 0$ (one internal and one where $A_{ss} = 1$), implying three possible revolutionary scenarios. The first scenario occurs when $A_{ss} = 1$ is the unique stable steady state (i.e., the kink is below the 45-degree line so there are no intersections in the phase diagram). Here, a shock that eliminates the stable steady state triggers a successful revolution; as no other stable steady state exists, dissent will start at the opposite end of the political spectrum, with new recruits appearing initially only on that same side but later on (once A goes below the kink) on both sides of the regime, until the regime collapses. The second scenario occurs when there is a unique inner stable steady state, so that the regime starts with approval $A_{ss} < 1$ (reflecting the existence of a pocket of dissenters at the far end of the regime). Here again, a shock that triggers a revolution will collapse the regime, with the revolution progressing in a similar way – first recruiting only on the far side of the regime and then on both sides. The third and last scenario occurs when two stable steady states with $A_{ss} > 0$ exist, one internal and one where $A_{ss} = 1$ (as in Figure 3). Here, a revolution that starts from the second steady state may fail to topple the regime. The revolution will start as before, with the extremists on the far side recruiting less extreme followers on their side of the political scale. But, since recruits are made only on one side of the regime, the momentum of the revolution will be low and the revolution fragile. In particular, the revolution is bound to fail if the shock that starts the revolution eliminates the stable steady state with $A_{ss} = 1$, which is the pre-revolution state, while the internal steady state potentially moves but is not eliminated. In this case, the approval will converge to a new (but still strictly positive) level and the regime will survive. Overall, we get that this kind of revolution, which progresses from the outside in, will be fragile initially but strong at later stages of the revolutionary process as they gain momentum by the faster recruitment on both sides of the political scale.

In order to prevent the success of the revolution, the regime has to either increase its force or implement policies that are more popular (thereby lifting the dynamic approval function). It is further worth noting that a shift of private opinions, say to the right, is equivalent to the regime changing its policies to the left. This is since it is the relative position of the type space vis-à-vis R that matters. This means that all of our results about a change in the regime's policy have an equivalent in an opposite change of private preferences. For instance, what may start a revolution is that the private preferences of the population over time shift away from the regime's policies as depicted in Figure 4. This fact will be used in describing the following historical example.

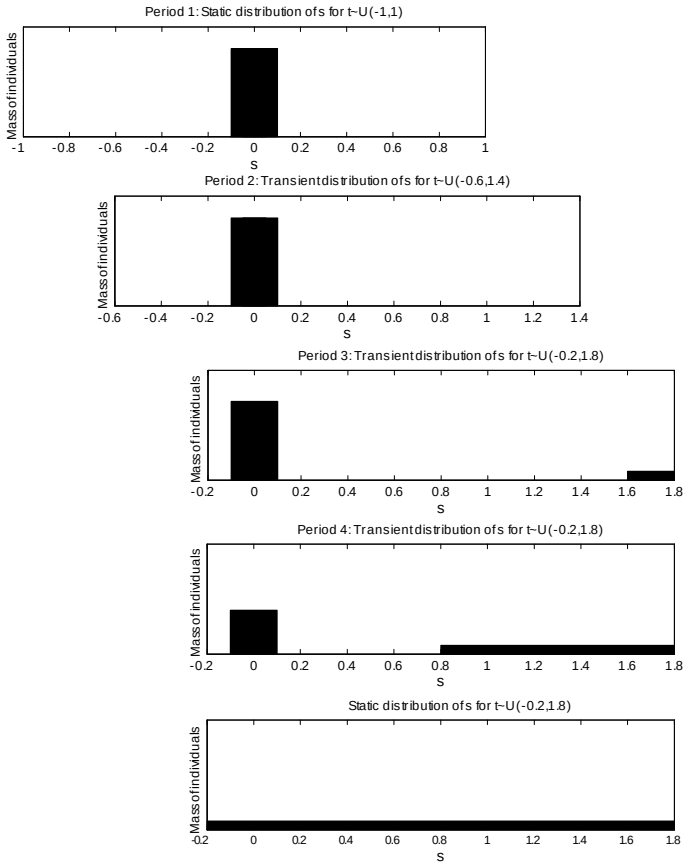


Figure 4: Distribution of stances over time in a stylized case of an outside-in revolution ($\beta < \alpha \leq 1$). $R = -0.8$ and fixed while the distribution of types changes.

3.2 A historic example

The overall pattern of the class of revolutions just described seems to provide a reasonable description for the Iranian Revolution in 1978-79. This revolution began following a gradual increase in the misalignment between the Shah and the increasingly religious sentiments in society (Moaddel, 1992). In line with Figure 4, following this misalignment, the revolution was initiated by the hardest opponents of the regime (i.e., by Khomeini, who held an extremely religious ideology, which he also expressed in public). Then gradually more moderate individuals joined the revolution (Razi 1987, Moaddel 1992, Ghamari-Tabrizi 2008, Shadmehr 2015b). These moderate individuals, while being part of the revolution, advocated less extreme policies and used less extreme slogans than Khomeini and even among Khomeini's closest supporters many were advocating a less religious policy than Khomeini (Ghamari-Tabrizi, 2008). During this process, the weakening of the punishment on dissent came in the form of removal of censorship and increased usage of televised debates, i.e. an acceptance of dissent in general (Milani 1994, p117). In line with Proposition 2 it has been claimed that in order to remain in power, the Shah "either had to crush the growing movement or to relinquish some of his power and strike a deal with the moderate faction of the popular movement. He opted to do neither" (Milani 1994, p116). That is, what could have possibly saved the Shah was either an increase of force ($\bar{K}$ in our model) or the implementation of popular policies (moving R in the religious direction). We will later see a successful execution of the first strategy by the Chinese regime in Tiananmen Square (see Section 5).

4 A wave-type revolution: from the inside-out

4.1 Analysis

We move now to the case where $\alpha < \min\{\beta, 1\}$. This case can be further divided into two subcases: $\alpha < \beta \leq 1$ and $\alpha < 1 < \beta$. While these two cases differ in some details, they are largely the same from the point of view of what we are interested in. Hence, for brevity, we will focus on the subcase $\alpha < \beta \leq 1$.[9]

By differentiating L twice with respect to s it is immediate that when $\alpha < \beta \leq 1$ the second order condition is not fulfilled, implying that an individual will choose one of the corner solutions: either $s(t) = R$ or $s(t) = t$. It is simple to further show that there exists a cutoff distance Δ such that all types further from the regime than Δ will fully follow the regime ($s(t) = R$) while types closer to the regime than Δ will speak their minds ($s(t) = t$), as illustrated for a biased regime in Figure 5. Hence, in this type of society, the regime induces silence by those who dislike it the most while those who largely agree with the regime pose mild critique of it. To understand the intuition behind this result note

[9] See sections A.2.3 and A.3.2 in the appendix for a treatment of the other subcase ($\beta < 1 < \alpha$).

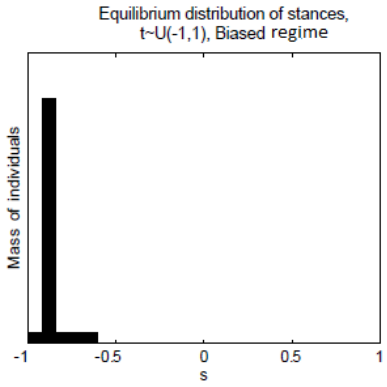


Figure 5: An illustration of an equilibrium distribution of stances for a biased regime for $\alpha < \beta \leq 1$.

that the important property of this case is that α is very small and in particular smaller than β. Consider, for instance, the special case of $\beta = 1$. First note that, as $\alpha < 1$, types will perceive a relatively high cost from even a very small deviation from their private bliss points. Hence, they will either speak their minds or, if this is too difficult given the punishment, be willing to go a long way to please the regime. Then, as $\beta = 1$ implies that speaking one's mind is considerably harder for extremists, they will be the ones submitting to the pressure and following the regime. The cutoff value Δ – between those speaking their minds and those staying silent – is decreasing in K, reflecting that the stronger the regime is, the smaller is the share of the population speaking their minds. This result, that extremists keep silent while the moderates are speaking their minds, has important implications for the stability of regimes and for the revolutionary dynamics as expressed in the following proposition.

Proposition 3 *When $\alpha < \beta \leq 1$:*

1. ***Existence of a stable steady state**: A stable regime exists iff it employs sufficient force, and the more biased its policy is the less force it needs to employ.*

2. ***Catalytic events**: A revolution may start following a shock to the regime's approval or force or following implementation of popular policies.*

3. ***Revolutionary participants**:*

   (a) *Initially only the most moderate types participate in the revolution, but over time types who are more extreme join it too.*

   (b) *For any regime with $|R| \neq 1$ the revolution will be two sided throughout.*

4. ***Revolutionary statements**: The revolution goes from the inside-out, whereby the most dissenting statements become more dissenting over time.*

We start by explaining the dynamics of the revolutions (parts 3 and 4 of the proposition) since this largely determines what makes a regime stable and which events may initiate a revolution. The revolutionary process follows from the dynamics of the cutoff between those who dissent and those who do not (Δ_i). As explained earlier, when the citizens perceive a (very) concave cost of deviating from their blisspoints, it induces dissent by moderates but not by extremists. This means that if a revolution breaks, the first ones to dissent are types with blisspoints close to the regime. When these moderates dissent, the strength of the regime falls, which makes it possible also for more extreme types to dissent. This way, increasingly extreme types join the revolution and they dissent more than those who started it – the revolution goes from the inside toward the outside (as summarized in parts 3a and 4 of the proposition). This further implies that for any regime (except for the most biased regimes, in which $|R| = 1$) the dissent will be two sided right from the start (part 3b of the proposition) – some will be complaining that the regime is too leftist and some that it is too rightist.

To explain the additional results, consider the dynamic equilibrium function $A_{i+1} = f(A_i)$ depicted in Figure 6. It is first flat and then starts to increase concavely. A_{i+1} then kinks upwards at some point (provided that $|R| \neq 0$) and is concave thereafter. The presence of the kink has important implications for the potential revolution. To see why this kink exists, recall that the cutoff between those who speak their minds and those who obey the regime (Δ) is large for a small K. Consider now a left-biased regime (as the one depicted in Figure 5). The flat initial part of $A_{i+1} = f(A_i)$ in the phase diagram (Figure 6) is where current approval is so low that the regime will not be able to gain any approval at the next period ($f(A_i) = 0$). If A_i is a bit larger but still small, this will imply a small K_{i+1} which will lead some types on the far right to obey the regime. Meanwhile, all types on the left side of the regime (and many on the right) speak their minds, as can be seen in Figure 5. At this point, an increase of A_i adds people obeying the regime only on the right side of it. However, at some point, when A_i has increased sufficiently, Δ_{i+1} will be sufficiently small so that regime followers will be added also on the left. At this point A_{i+1} becomes steeper – this is the kink – as an increase in A_i from that moment on adds obedience on both sides of the regime. The reason why $A_{i+1}(1) < 1$ is that, for any finite K_i, there will always exist sufficiently moderate types who will choose to speak their minds.[10] Put together, we get that the phase diagram has the shape depicted in Figure 6.

As before, for any approval level A_i, an increase in the regime's force ($\bar{K}$) raises $A_{i+1} = f(A_i)$. For sufficiently small $\bar{K}$, no intersection with the 45-degree line exists, but for larger $\bar{K}$ the A_{i+1} function intersects the 45-degree line either twice or four times (bar tangency points). The possibility of four intersections is precisely because of the kink. The intuitive reason for why a biased regime can employ less force yet remain stable (part 1 of

[10] To see why this is the case, note that, when $\alpha < \beta$, $D(x)$ is steeper than $KP(x)$ for sufficiently small values of x and any finite K, which implies that for a type sufficiently close to the regime it is more costly to deviate from her bliss point than to speak her mind and bear the sanctioning for doing so.

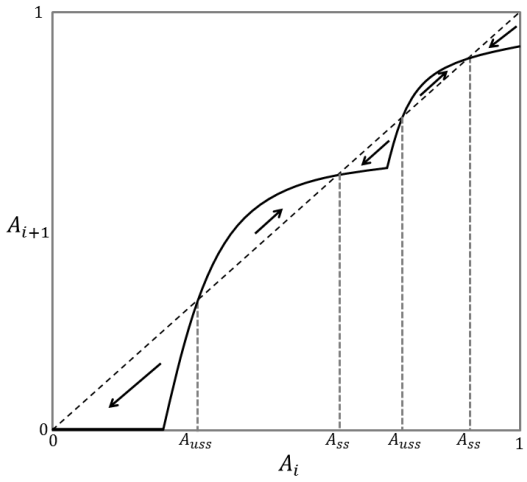


Figure 6: Stylized phase diagram for the case $\alpha < \min\{1, \beta\}$ and biased regime. The full line depicts the equilibrium function $A_{i+1}(A_i)$ and the dashed line depicts the 45-degree line where $A_{i+1} = A_i$. The vertical lines depict the stable (A_{ss}) and unstable (A_{uss}) steady states.

the proposition) is that in the case of $\alpha < \beta \leq 1$ the regime induces obedience from those with private opinions sufficiently far from it. Hence, a biased regime, whose policy is far from many in society, will be stronger than a central regime. In Figure 6 this means that biasness shifts the graph upwards.[11] This shift implies that the stable steady states move rightward while the unstable steady states move leftward in the phase diagram, implying that shocks to the approval can be larger without initiating a revolution – a biased regime is more stable.

If a regime implements popular policies, thus decreasing its bias with respect to the preferences of the population, it lowers A_{i+1} in the phase diagram. This may imply that a stable steady state disappears and a revolution is initiated (part 2 of the proposition). The reason for this is that the ones who largely agree with the regime are the ones dissenting against it by posing mild critique. Roughly speaking, when the regime implements popular policies it aligns with the views of more people thus inducing more people to speak their minds. This increases the number of individuals posing mild critique, which weakens the regime and may ultimately be the catalytic event that starts the revolution. A similar process may be triggered by a temporary shock to the regime's force, and, alternatively, a shock to the approval may lead the political system to a zone with convergence downwards (part 2 of the proposition). The model predicts that if the regime reacts to these events by implementing even more popular policies (reforms), as regimes under threat often naturally do, this will exacerbate its predicament. This is quite surprising but, again, stems from

[11] More precisely, an increase in bias shifts the concave part to the left of the kink upwards and at the same time widens this part outwards in both directions.

the fact that a reform induces more people to speak their minds. It further implies that implementation of *unpopular* policies could help the regime stop the revolution.

The fact that the revolution is initially driven by moderates and that the momentum of the revolution is driven by new recruits has important implications for the fragility of the revolutionary process at different stages. To see this, consider a left-biased regime which starts at the rightmost stable steady state in Figure 6. Suppose now that the force of the regime ($\bar{K}$) decreases for some reason so that the rightmost steady state disappears. Then, since the right part of the curve is concave, the momentum will increase initially. During this phase the revolution recruits new individuals on both sides of the regime. However, eventually no new recruits can be added from the left, as all leftists already speak their minds. This is where the dynamics reach the kink in the phase diagram, after which the momentum decreases since the new recruits come from one side only. At this point the revolution may fail if there still exists an intermediate stable steady state (i.e., if the middle region of the curve intersects the 45-degree line). In other words, the regime holds on but with less approval than it previously had (as in the leftmost A_{ss} in Figure 6). Hence, unlike the revolution described in the previous section, now the revolution is most fragile toward the end, when its momentum is dependent on recruits from one side only.

The revolutionary dynamics following a shift in public sentiments are illustrated in Figure 7. As mentioned earlier, this is equivalent to a shift in the regime's policies. The left hand side (Case 1) illustrates that if the private sentiments in society shift right and the regime is left-biased then this will only strengthen the regime. On the other hand, if the population's opinions shift to the left, as illustrated in Case 2, a revolution following the pattern in parts 3 and 4 of the proposition will commence. The first thing that happens is an increase of dissent by moderates on the left. This will weaken the regime's strength thus making it possible also for less moderate people to speak up against the regime, which increases dissent also on the right. The regime then weakens further. This way dissent will increase on both sides, but since the regime is left-biased, at a certain point new recruits will appear only on the right – what started as a leftist revolution, following a leftward movement of public sentiments, ends up being a rightist revolution, whereby the center of expressed opinions is eventually to the right of the regime that collapsed.

4.2 A historic example

The revolutionary pattern just described provides a theoretically consistent explanation for an important class of mass revolutions that were previously unexplained by formal theory. For example, it seems to provide a reasonable description for the protest movements that led to the collapse of some of the communist regimes in Eastern Europe in 1989-90 and to the recent Arab Spring revolution in Egypt. In Eastern Europe, the trigger was probably not a shift in public sentiments toward the communist regimes but instead a movement of the regimes themselves in the direction of the growing liberal sentiments in society (offering

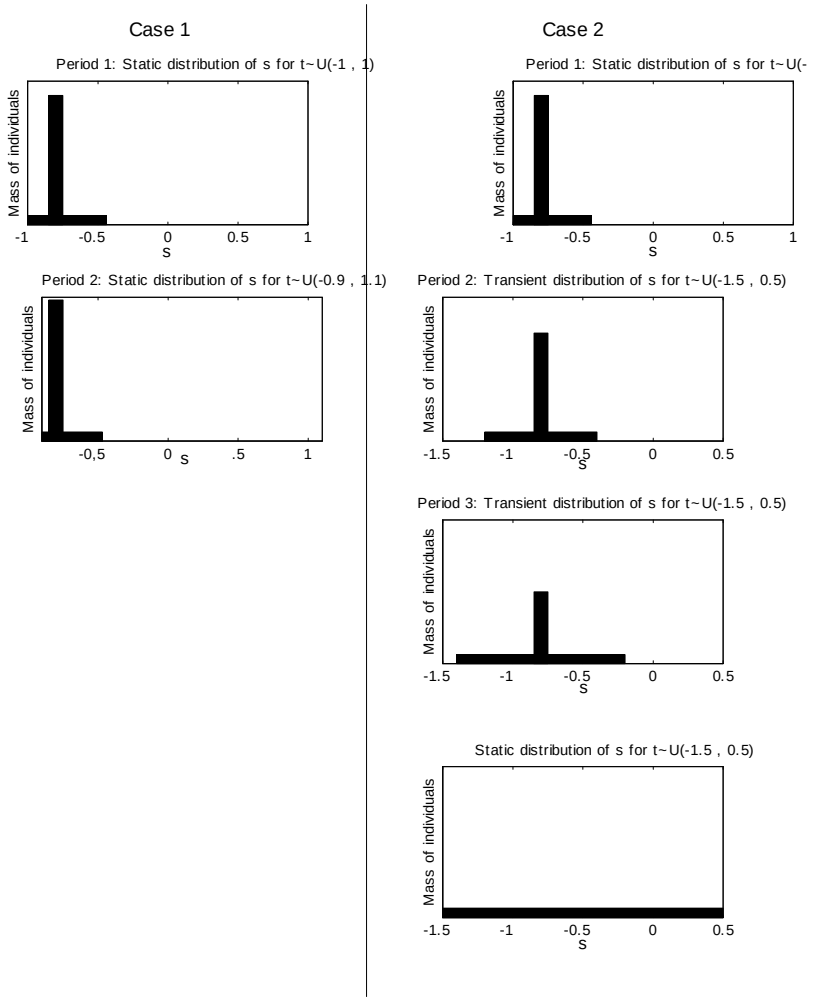


Figure 7: Distribution of stances over time in two stylized cases of inside-out revolutions ($\alpha < \beta \leq 1$). In both cases $R = -0.8$ and fixed while the distribution of types changes.

liberal reforms, most notably the Glasnost). The initial protesters were not very extreme. For instance, Hungarian communist party leader Karoly Grosz stated that "the party was shattered not by its opponent but – paradoxically – from within" (Przeworski 1991:56). Furthermore, in Poland and Hungary, moderate dissidents instigated liberal reforms and made demands for free elections (Pfaff, 2006). Similarly, as was reported about Egypt, the most extreme factions (i.e., the Muslim Brotherhood and the Salafi movement), were hardly present in the protests initially.[12] In the case of Egypt though, the trigger seems to have been a shift in the sentiments of society in the direction of the (relatively liberal) regime, as will be explained below.

Beyond the inside-out progress, another important feature of this kind of revolution is that the undermining of the regime is initiated by individuals with moderate views from *both* sides of the political spectrum (unless the regime is so biased that on one of its sides there are no individuals). This implies that regimes may be undermined by truly "strange bedfellows", in the sense that they are pulling the public opinion in two different directions. This was a clear pattern in the Arab Spring revolution in Egypt. The protesters on the Tahrir Square consisted of some who suggested that Mubaraq was not sufficiently liberal and of others who said he was not sufficiently conservative. While the spark may have been a shift in private opinions towards more liberalism (a leftward movement of the opinion axis when moving from right hand schedule 1 to right hand schedule 2 in Figure 7), the later elections showed that in fact Egyptian society as a whole was even more conservative than Mubaraq's regime (in line with the description in Figure 7, where the average opinion after the shift is to the right of $R = -0.8$, which represents Mubaraq's regime in that figure). This way, as predicted by the model, what started as mainly a leftist (liberal) revolution ended up being a rightist (conservative) revolution instead.

5 A stretching-out type revolution

5.1 Analysis

The final case is when $\alpha > 1$ and $\beta \geq 1$.[13] This case shares the inside-out progress of statements during the revolution with the case of the previous section, while sharing the leading role of extremists in the revolution with the class of revolutions described in Section 3.

An important feature of this case is that $\beta > 1$ represents a regime that is tolerant to small dissent while punishing harshly larger dissent. By differentiating (4) twice it is

[12] For instance, a BBC news profile on the Muslim Brotherhood reports that initially "(t)he group's traditional slogans were not seen in Cairo's Tahrir Square. But as the protests grew and the government began to offer concessions, including a promise by Mr Mubarak not to seek re-election in September 2011, Egypt's largest opposition force took a more assertive role". See http://www.bbc.com/news/world-middle-east-12313405.

[13] When α equals exactly one, some small technicalities need to be kept in mind. The results with respect to what will be presented are however the same, so we will simply ignore this case in our analysis.

immediate that the second order condition holds so that each type has an inner solution.[14] This means that each type compromises between fully obeying the regime and speaking her mind. This is intuitive since when the regime is tolerant toward small dissent, the citizens do not have an incentive to keep silent. At the same time, when D is convex, the citizens are lax about small deviations from their bliss point and hence do not mind compromising a little. Furthermore, extremists dissent more than moderates since the convexity of D makes large deviations from one's bliss point very costly.

This result, that extremists are compromising yet dissent more than the moderates, has important implications for the stability of regimes and for the revolutionary dynamics as expressed in the following proposition.

Proposition 4 *When $\alpha > 1$, $\beta \geq 1$:*

1. ***Existence of a stable steady state**: A stable regime exists iff it employs sufficient force, and the more biased its policy is the more force it needs to employ.*

2. ***Catalytic events**: A revolution may start following a shock to the regime's approval or force or following implementation of unpopular policies.*

3. ***Revolutionary statements**:*

   (a) *The revolution goes from the inside-out, whereby the most dissenting statements become more dissenting over time.*

   (b) *For any regime with $|R| \neq 1$ the revolution will be two-sided.*

4. ***Revolutionary participants**: At all time periods during a revolution with the most extreme types dissenting the most.*

Again, we start by explaining the dynamics of the revolutions (parts 3 and 4 of the proposition) since this largely determines what makes a regime stable and which events may initiate a revolution. Part 3 of the proposition says that, once the revolution starts, dissent becomes more and more fierce over time – the revolution goes from the inside out, as depicted for a stylized example in Figure 8. The intuition for this is that a convex punishment ($\beta > 1$) implies a relatively heavy sanctioning on extreme dissent. Hence, very extreme dissent will be absent initially. During the process of a revolution, as the approval and hence also the strength of the regime fall, all types are induced to dissent more, and in particular it becomes possible to express views that are more extreme than was possible before. This further weakens the regime, causing more dissent and so on. As expressed in the fourth part, the ones who are dissenting the most are those with private views far from the regime – in a sense they are pushing the freedom of speech. This is an important difference between this class of revolutions and the wave-type revolution that goes from

[14] In case $\beta = 1$ types close to the regime fully obey it while types far have an inner solution.

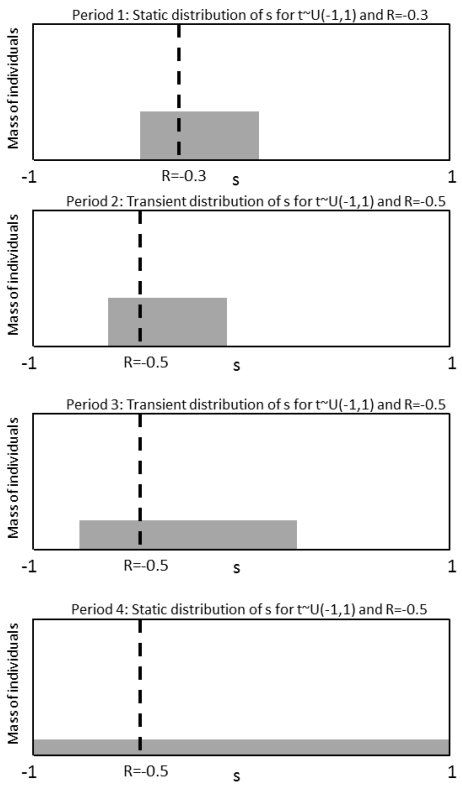


Figure 8: Distribution of stances over time in a stylized case of a stretching-out revolution ($\alpha > 1$, $\beta \geq 1$). The regime starts at $R = -0.3$ and after the first period moves to $R = -0.5$ (which triggers the revolution) and stays there, while the distribution of types gradually changes. The diagram depicts the case of $\alpha = \beta$ for ease of exposition

the inside-out described in Section 4. While in the wave-type revolution the moderates are those pushing the freedom of speech and the most extreme types remain silent for a very long time and are the last to join, here the extremists are the ones constantly pushing the freedom of speech, backed-up from behind by the moderates.

The dynamics of approval are depicted for a stylized example with $\beta > 1$ in Figure 9. As in the previous phase diagrams, the dynamic equilibrium function $A_{i+1} = f(A_i)$ is first flat at zero (unless $R = 0$) and then increases. As can be seen in the figure, $A_{i+1}(1) < 1$, reflecting that there cannot be full obedience in equilibrium. This is since $\beta > 1$ while if β exactly equals 1 and $\bar{K}$ is large there can be full obedience. Depending on the values of $\bar{K}$ and R, the intertemporal dynamics function $f(A_i)$ has either no intersections with the 45 degree line, or one tangency point, or two intersections, but never more than two. Considering the case of two intersections as in the figure, the fact that $A_{i+1}(1) < 1$ implies that the rightmost intersection is stable while the leftmost is unstable and the meeting point at zero is stable too.

The first part of the proposition is thus intuitive: more force (i.e., larger $\bar{K}$) increases

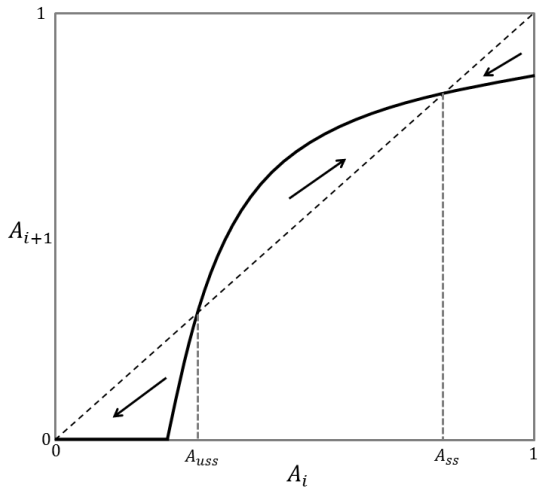


Figure 9: Stylized phase diagram for the case $\alpha > 1$ and $\beta \geq 1$ when the regims is biased. The full line depicts the equilibrium function $A_{i+1}(A_i)$ and the dashed line depicts 45-degree line where $A_{i+1} = A_i$. The vertical lines depict the stable (A_{ss}) and unstable (A_{uss}) steady states.

the public approval (A_{i+1}) of the regime since it decreases dissent for any level of previous approval (A_i); this shifts the function A_{i+1} up in the phase diagram, thus enabling the existence of a steady state. An increase in the bias of the regime's policy, on the other hand, decreases approval. To see this, compare two cases, one where the regime is at $R = 0$ and one where it is at $R = -1$. The former has a mass of types on its left with private opinions at distances between zero and one and this image is mirrored on the right. Switching from $R = 0$ to $R = -1$ is like replacing the mass of types on the left with a mass of types on the right, but now with opinions at distances of one to two. That is, we replace individuals who are moderately critical to the regime in private with individuals who are much more critical. Since types who are more extreme dissent more, we get that approval is decreasing in the bias of R. Hence, biasness shifts the A_{i+1} function down, implying that larger bias of the regime has to be compensated for by the employment of more force for a steady state to exist. Furthermore, by lowering the approval function, biasness reduces the distance between the unstable and stable steady states, hence increases the vulnerability of the regime to shocks to its approval or force – biased regimes are less stable.

It is worth noting that in this class of revolutions the regime will always eventually collapse (because there is no other stable steady state to the left of the initial state besides the one at $A = 0$).[15] Unlike the two wave-type classes of revolutions, here the revolution never loses its momentum since it is the gradual shift of statements that drives it instead of

[15] $R = 0$ with $\alpha > 2$ is a special case where there is no stable steady state at 0 because $f(A_i)$ starts with an infinite slope. In this case, $f(A_i)$ is concave throughout with exactly one stable steady state with $A \in]0, 1[$, corresponding to a regime that cannot collapse.

recruitment of new protesters. Hence, once a revolution has started it will always succeed, unless the regime reacts on time by either increasing its force (e.g., by recruiting more troops) or implementing popular policies to pease the population.

5.2 A historic example

In practical terms, the main feature that distinguishes the stretching-out class of revolutions from the wave-type revolutions that go from the inside-out is that while in the latter the groups that express the strongest criticism of the regime are constantly changing, with increasingly extreme groups voicing their increasingly extreme demands (as in the Egyptian Arab Spring), here during the whole revolution it is the same group of dissenters that expresses the strongest criticism, while constantly increasing its dissent against the regime and pulling other fractions of society to join forces.

As a historic example, consider the April Revolution in South Korea, 1960. We provide here a summary of the gradual escalation of events in this students-led revolution (taken from Kim, 1996); from the minor protest against governmental intervention to the mass demand of resignation of the president. The protest movement began on February 28th 1960, two weeks before the planned (rigged) national elections, when students from Taegu marched into the streets in protest against a governmental directive to attend school on a Sunday (in order to discourage them from showing up at mass rallies held by the opposition). On election day, March 15, angry citizens of the city of Masan, whose names had been removed from the voter registration roster, marched into the city hall asking that voting slips be given to them. About a month after the elections, on 18-19th of April, students from almost all of the major universities in Seoul protested in front of the principal government buildings in demand for new, fair elections. A week later, university professors marched into the streets, loudly demanding the resignation of the president. In the evening, several hundred thousand people rioted in attempt to overthrow the regime. President Rhee indeed resigned the following morning, 27th of April. This revolution seems to have been triggered by implementation of unpopular policies – President Rhee was trying to become an autocrat in a society with strong liberal sentiments, and he tried doing so while having under his control a mostly liberal enforcement system ($\beta > 1$).

A similar chain of events characterized the students-led protests on Tiananmen Square in Beijing in 1989 (for a detailed account see Zhao 2001), with the important difference that in the case of Beijing the protests did not develop into a successful revolution that would end with a change of the regime. The former politician, Hu Yaobang, who was popular among students, passed away on April 15 1989 and this led a large number of students to mourn his death on that day (Pan, 2008). Two days later, a commemoration (which was considered more dissenting than individual mourning) was organized. This organization quickly evolved into a declaration of demands for political reform and thereafter, on April 18, to a sit-in where students demanded to meet with the leadership of the political party.

On April 21, students began organizing themselves formally into unions and some workers into a federation, writing texts challenging the regime (Walder and Xiaoxia, 1993) and on April 22 serious rioting broke out in several places. Quiet prevailed for a few days but then, on April 27, the Autonomous Student Union staged a march to the square breaking through police lines after which the leaders of the union, Wang Dan and Wu'erkaixi, called for more radical measures to regain momentum. This led to hunger strikes and also to the expressed support for the strikes by others who did not themselves strike. On May 17–18, around a million Beijing residents, including low ranked representatives of the regime such as party officials and police officers, demonstrated in solidarity of the hunger strikers. This was a sign of the decreased approval of the regime, as predicted by our model, with another sign of the regime's weakening being the increasingly open and positive reports about the protests in the media. All this time, the soft approach of the regime – of showing sympathy toward the demonstrators and looking for a dialogue with them – as advocated by Zhao Ziyang, the General Secretary of the Communist Party, was giving the tone. This approach of containment of the moderately deviant expressions is represented in the model by the convex sanctioning ($\beta > 1$), which, if not interrupted, is predicted to have led to the eventual collapse of the regime. However, division within the regime with regards to the appropriate reaction to the demonstrations constantly intensified, with the hard-liners rallying behind Premier Li Peng who increasingly gained power. On May 17, a leadership meeting was called at Chairman Deng Xiaoping's residence, where Zhao Ziyang's concessions-based strategy was thoroughly criticized and it was decided to declare martial law. In terms of our model, this was a decision to substantially increases $\bar{K}$. The implementation of the martial law that led to the heavy-handed crackdown on the protests on June 4 eventually stopped the mass protests. Our model predicts that without the extensive use of force, which followed the change in the balance of power within the leadership and the adoption of the hard line approach, the regime would not have been able to stop the revolution. However, the model further implies that the use of force lifted the approval function back to where a *new* inner stable steady state appeared, to which approval converged, and where this new point appeared to the left of the initial point. In real life terms, this means an end to the revolution, with the same regime in power but with less approval.

One interesting detail to note about this attempted revolution is that it happened even though, allegedly contrary to the model's prediction, the regime had been implementing popular policies throughout the Eighties. These reforms were initiated by Chairman Deng Xiaoping, the successor of Mao Zedong, and were indeed generally well received by the public. However, judging from the demands of the protesters – who called for major political reforms such as the implementation of democracy, greater accountability, freedom of the press and freedom of speech – it seems that the shift of views among the people was faster and more far-reaching, thus increasing the bias of the regime to the point where the riots broke.

6 Further results and empirical predictions

In this section we discuss some results that apply generally to the whole model and formulate them in the form of testable predictions. The inter-society variability of each of the two main parameters of the model, β and α, should in principle enable one to test these parameter-based predictions.

Starting with β, the curvature of the sanctioning system, the previous propositions show that it is only for sufficiently small β that the evolution of statements during the revolution starts with expression of extreme views and continues with expression of more moderate views (revolution going from the outside-in).

- **Prediction 1**: Holding all else fixed, the stricter the regime is (i.e. the lower is β), the more likely it is that the dynamics of the revolution will be from the outside in – the most dissenting expressions will appear right from the start.

This prediction holds also if α is heterogeneous between individuals – the smaller is β, the more likely it is that it is smaller than any given α and hence, conditioned that $\beta \leq 1$, that it will induce extremists to be the first speaking their minds against the regime. In principle, the sanctioning structure of the regime is observable. One way for obtaining a proxy for β is to look at past protests and how the regime sanctioned deviant expressions – if only extreme dissent was punished then β is large, and if the regime punished all dissent, large and small, roughly the same, then β is small. One can then look at how the current revolution evolves.

As for α, the curvature of disutility from bliss point deviations, it is harder to observe. Hence we need to identify two observable variables that depend on α and that the model predicts should be related in a certain way. One such prediction relates to policy changes. According to the previous propositions, implementation of popular policies (or reforms initiated by the regime) should trigger a revolution if and only if $\alpha \leq \{1, \beta\}$, in which case the revolution will start with moderate forces that will gradually recruit the more extremists. This leads to the following prediction.

- **Prediction 2:** Holding all else fixed, there is a positive correlation between implementation of popular policies and revolutions that are initiated by regime supporters and spread like a wave from the inside out.

In principle it is observable whether a policy change is in the interest of most of the population or not. It is also observable whether a revolution starts afterwards, who were the initial protesters, and which views they expressed.

The third prediction is a mirror image of prediction 2 and has to do with the effectiveness of the regime's response to the revolution. A very common reaction of regimes that see an escalation in dissent and their approval deteriorating is to offer reforms – popular policies

that are meant to pease the population and consequently cool down the angry civilians. Our model predicts however that this measure will not be effective against the second class of revolutions.

- **Prediction 3:** Revolutions that are initiated by regime supporters and spread like a wave from the inside out will not subside following implementation of popular policies.

Finally, a general point – about the timing of regime reactions, martial law and rule by decree – is revealed when looking at the phase diagrams in Figures 3, 6 and 9. To illustrate this, suppose there is a shock to the regime's force ($\bar{K}$) which lowers the function $A_{i+1}(A_i)$ sufficiently so that the old A_{ss} is now in a zone of downward convergence (a revolution is initiated). Suppose further the regime intends to use increased force to cool down the protests. Then, if a few time periods have already passed since the revolution started, it will not be sufficient to simply restore the old level of force in order to achieve upward convergence of approval. This is since the approval may have deteriorated to a level A_i where also the old dynamic function $A_{i+1}(A_i)$ implies downward convergence. The model thus predicts that what the regime has to do in this case is to overshoot its force, as is often done indeed by the implementation of martial laws or rule by decree. What is further interesting is that it is sufficient that these extraordinary measures be temporary – once the approval has sufficiently grown, the martial laws can be abandoned without leading back to downward convergence of the approval. This explains why harsh temporary measures often tend to work, like was illustrated in the mass protests in Beijing. The same logic applies to other measures that shift the approval function, such as the implementation of new policies (unpopular ones in the case of $\alpha < \min\{1, \beta\}$ and popular ones otherwise) – the regime has to temporarily overshoot with these policies but can then restore the old policies once the protests have calmed down. Moreover, the timing of the intervention of the regime is crucial, as expressed in the following prediction.

- **Prediction 4:** Holding all else fixed, the later the regime implements a given increase in force, the more likely the revolution is to succeed.

The prediction follows directly from the phase diagrams and stems from the fact that the longer the regime waits with reacting to a revolution or protests, the lower the approval level (A_i) will be at the time of reaction, and hence a given increase of $\bar{K}$ is less likely to ensure that the approval function $A_{i+1}(A_i)$ rises sufficiently to imply upward convergence.

7 Conclusions

This paper classifies popular revolutions and mass protests into three classes: wave-type revolutions going from the outside in, wave-type revolutions going from the inside out and stretching-out type of revolutions. This classification is shown to be exhaustive in our

model. It spans the parameter space of α and β, the parameters that capture the curvature of the two costs affecting the individual choice of stance at the time of revolution: the cost of deviating from her privately held opinion and the cost of dissenting against the regime. Each class of revolutions has its own unique set of attributes, characterizing who in society – moderates or extremists – initiate the revolution, how it progresses to other parts of society, which views are expressed by participants at all stages, how the regime may unknowingly trigger the revolution and what it can or cannot do to stop the revolution at different stages. It would be presumptuous on our side to actually claim that all real life revolutions follow one of these patterns. Our analysis abstracts from important real life factors such as intervention of outside forces, conflicts within different revolutionary groups about the targets of the revolution and the appropriate means to achieve them, changes in the regime's leadership during the revolution (as was shown to be crucial in the case of the protests on Tiananmen Square) and heterogeneity in the private costs (α). However, as the historic examples provided in the paper demonstrate, our model *is* able to capture many important aspects of real revolutions that cannot be captured with the existing models and accordingly we provide testable predictions on the progress of revolutions, the catalytic events leading to them and the effective and ineffective responses of regimes. In particular, these predictions do not require a homogenous α in society. We believe that our framework could serve as a new workhorse for the study of revolutions, where further questions could be answered and more parameters could be endogenized.

A Analytical derivations and proofs

A.1 Some auxiliary results

Using equation (6) with $\lambda = 1$ we have

$$A_{i+1} = \begin{cases} 1 - \Psi\left(s_i; R, A_i\right) \text{ when } 1 - \Psi\left(s_i; R, A_i\right) \geq 0 \\ 0 \text{ when } 1 - \Psi\left(s_i; R, A_i\right) < 0 \end{cases} \text{ where} \tag{7}$$

$$\Psi\left(s_i; R, A_i\right) \equiv \int_{-1}^{1} \left|s_i^*\left(t\right) - R\right| dt. \tag{8}$$

A.2 Individual stances

The individual minimizes the loss function given by (4), (1) and (3) when $K > 0$. Using the implicit function theorem we get the following derivatives of $s^*(t)$ in inner solutions:

$$\frac{ds^*}{dt} = \frac{D''\left(t - s^*\right)}{P''\left(s^*\right) + D''\left(t - s^*\right)} \tag{9}$$

Let t_l and t_h denote the left and the right edges of distribution of types, and let

$$\Delta \equiv K^{\frac{1}{\alpha - \beta}}.$$

A.2.1 Case (1): $\max\{\alpha, \beta\} \leq 1$

The second-order condition of the loss function is positive when $\alpha < \beta \leq 1$ or $\beta < \alpha \leq 1$, which implies that any inner extreme point is a maximum. The corner solutions are then either $L(s = R) = |t - R|^{\alpha}$ or $L(s = t) = K|t - R|^{\beta}$. When $\beta < \alpha$ this implies that $L(s = R) < L(s = t)$ iff $|t - R| < \Delta$, and so $s^*(t) = t$ iff $|t - R| \geq \Delta$, and $s^*(t) = R$ iff $|t - R| < \Delta$. When $\alpha < \beta$ the converse holds,[16] with $s^*(t) = t$ iff $|t - R| \leq \Delta$, and $s^*(t) = R$ iff $|t - R| > \Delta$.

A.2.2 Case (2): $\beta < 1 < \alpha$

We perform the proof for $t \geq R$. The opposite case is similar. We will prove that if $t_h - t_l > 2\Delta$, then types close enough to the regime fully conform, while types far from the regime choose an inner solution and $|s^*(t) - R|$ is increasing for them. Along the way we will also show that for a sufficiently narrow range of types, the distribution is degenerate at R.

We will first show that the only relevant corner solution is $s^* = R$. In order to find the global minimum for a type t, we first need to investigate the behavior of $L(s, t)$ at $s = t$ and $s = R$.

$$L'(s, t) = -\alpha(t - s)^{\alpha - 1} + \beta K(s - R)^{\beta - 1}$$

Hence $\lim_{s \to R} L'(s, t) = \infty$ and $L'(t, t) = \beta K(t - R)^{\beta - 1} > 0$. Therefore $s = R$ may be a solution to the minimization problem while $s = t$ is not. The candidate solution $s = R$ will now be compared to potential local minima in the range $]R, t[$. In inner solutions $L'(s, t) = 0$ and hence we get

$$\begin{aligned} \alpha(t - s)^{\alpha - 1} &= \beta K(s - R)^{\beta - 1} \\ &\Rightarrow (t - s)^{\alpha - 1}(s - R)^{1 - \beta} = \beta K/\alpha. \end{aligned} \tag{10}$$

Define

$$\Phi(s) \equiv (t - s)^{\alpha - 1}(s - R)^{1 - \beta}.$$

For the existence of an inner min point for a given t it is necessary that $\Phi(s) = \beta K/\alpha$ for some $s \in]R, t[$. Note that as $t \to R$ both $(t - s)^{\alpha - 1}$ and $(s - R)^{1 - \beta}$ approach zero implying $\Phi(s) < \beta K/\alpha$ for all $s \in]R, t[$. Hence types with sufficiently small $|t - R|$ do not have an inner local min point and they choose $s^* = R$.

For sufficiently large $|t - R|$ it may be that $\Phi(s) = \beta K/\alpha$ for some $s \in]R, t[$ which we investigate next. Note that, for given t, $\Phi(s)$ is strictly positive in $]R, t[$, and that $\Phi(s, t) = 0$ at both edges of the range (i.e. at $s = R$ and at $s = t$). This means that $\Phi(s)$ has at least one local maximum in $]R, t[$. We now proceed to check whether this local maximum is unique:

$$\Phi'(s) = (t - s)^{\alpha - 2}(s - R)^{-\beta}[(1 - \beta)(t - s) - (\alpha - 1)(s - R)]$$

Since $(t - s)^{\alpha - 2}(s - R)^{-\beta}$ is strictly positive in $]R, t[$, and $[(1 - \beta)(t - s) - (\alpha - 1)(s - R)]$ is linear in s, positive at $s = R$ and negative at $s = t$, $\Phi'(s) = 0$ exactly at one point at

[16] Since then $\frac{1}{\alpha - \beta} < 0$, hence when solving for $K^{\frac{1}{\alpha - \beta}}$ the inequality flips direction.

this range (i.e. a unique local maximum of $\Phi(s)$ in $]R,t[$). From the continuity of $\Phi(s)$ we get that if the value of $\Phi(s)$ at this local maximum is greater than $\beta K/\alpha$, then $L(s,t)$ has exactly two extrema in the range $]R,t[$. From the positive values of $L'(s,t)$ at the edges of this range we finally conclude that the first extremum (where $\Phi(s)$ is rising) is a maximum point of $L(s,t)$, and the second extremum (where $\Phi(s)$ is falling) is a minimum point of $L(s,t)$. The global minimum of $L(s,t)$ is therefore either this local minimum (i.e. an inner solution), or $s=R$ (i.e. a corner solution). If however the value of $\Phi(s)$ at its local maximum point is smaller than $\beta K/\alpha$, then there is no local extremum to $L(s,t)$ in the range $]R,t[$, and therefore $s=R$ is the solution to the minimization problem.

Next we show that if $t_h - t_l > 2\Delta$ then there exists a type who is far enough from the regime to choose the inner solution. First, note that the distance from the regime to the type who is the most remote from it is larger than Δ when $t_h - t_l > 2\Delta$. Suppose this type is $t_{h.}$. Then, comparing only the two corner solutions this type can choose, we get

$$L(R,t_h) - L(t_h,t_h) = |t_h - R|^{\alpha} - K\,|t_h - R|^{\beta},$$

which is strictly positive when $|t_h - R| > \Delta = K^{\frac{1}{\alpha-\beta}}$ and $\beta < \alpha$. This implies that t_h does not choose the corner solution of R, hence must choose an inner solution.

Now we show that if there exists any type t_0 who chooses the inner solution then all types with $t > t_0$ have an inner solution. We also show that types close enough to the regime fully conform, and that in the range of inner solutions $|s^*(t) - R|$ is increasing in t. First note that $\Phi(s)$ is increasing in t, so if there exists a local minimum of $L(s,t_0)$ for some t_0, then there exists a local minimum of $L(s,t)$ for $t > t_0$ too. Also note that for all $s \in]R,t[$ $\Phi(s)$ is increasing in t and that $\lim_{t\to\infty} \Phi(s,t) = \infty > \beta K/\alpha$, implying an inner local min point exists for a broad enough range of types. Second, if there is an inner solution to the minimization problem for some t_0 then there is also an inner solution to the minimization problem for $t > t_0$. To see this let $\Delta L \equiv L(R,t) - L(\tilde{s},t)$, where $\tilde{s}$ is the stance at which $L(s,t)$ gets the local minimum. Type t prefers the inner solution to the corner solution if and only if ΔL is positive. Thus we need to show that ΔL is increasing in t and so if ΔL is positive for t_0 then it is positive for $t_1 > t_0$ too.

$$\Delta L = (t-R)^{\alpha} - (t-\tilde{s})^{\alpha} + K\,(\tilde{s}-R)^{\beta}$$

Differentiating ΔL with respect to t yields

$$\Delta L'_t = \alpha\,(t-R)^{\alpha-1} - \left[\alpha\,(t-\tilde{s})^{\alpha-1}\left(1 - \frac{d\tilde{s}}{dt}\right) + \beta K\,(\tilde{s}-R)^{\beta-1}\,\frac{d\tilde{s}}{dt}\right].$$

Using the first order condition (10)

$$\begin{aligned}\Delta L'_t &= \alpha\,(t-R)^{\alpha-1} - \left[\alpha\,(t-\tilde{s})^{\alpha-1}\left(1 - \frac{d\tilde{s}}{dt}\right) + \alpha\,(t-\tilde{s})^{\alpha-1}\,\frac{d\tilde{s}}{dt}\right] \\ &= \alpha\,(t-R)^{\alpha-1} - \alpha\,(t-\tilde{s})^{\alpha-1} > 0\end{aligned}$$

Differentiating once more

$$\Delta L''_t = \alpha\,(\alpha-1)\left[(t-R)^{\alpha-2} - (1 - d\tilde{s}/dt)\,(t-\tilde{s})^{\alpha-2}\right].$$

By equation (9) we have that $\frac{d\tilde{s}}{dt} > 1$ in an inner solution when P is concave, and so $\Delta L''_t > 0$. Hence ΔL is strictly increasing and strictly convex, implying that for a broad enough range of types (in particular larger than 2Δ, as shown above), types sufficiently far from the regime have an inner solution where $\frac{ds^*}{dt} > 1$, and so $|s^*(t) - R|$ is increasing in t at the range of inner solutions.

A.2.3 Case (3): $\alpha < 1 < \beta$

We perform the analysis for $t \geq R$. The opposite case is similar. We will first show that the only relevant corner solution is $s^* = t$, then that types close to the regime choose this corner solution. In order to find the global minimum we first need to investigate the behavior of $L(s,t)$ near the corner solutions.

$$L'(s,t) = -\alpha (t-s)^{\alpha-1} + \beta K (s-R)^{\beta-1}$$

Hence $L'(R,t) < 0$ and $L'(t,t) < 0$ since $\alpha < 1$. Therefore $s = t$ may be a solution to the minimization problem while $s = R$ is not. The candidate solution $s = t$ will now be compared to potential local minima in the range $[R,t]$. In inner solutions $L'(s,t) = 0$ and hence we get

$$\begin{aligned} \alpha (t-s)^{\alpha-1} &= \beta K (s-R)^{\beta-1} \\ &\Rightarrow \quad (t-s)^{\alpha-1} (s-R)^{1-\beta} = K\beta/\alpha \end{aligned} \tag{11}$$

Define

$$\Phi(s) \equiv (t-s)^{\alpha-1} (s-R)^{1-\beta}.$$

For the existence of an inner min point it is necessary that $\Phi(s) = \beta K/\alpha$ for some $s \in]R,t[$. Since $\alpha < 1$ and $\beta > 1$ follows that $\Phi = \beta K/\alpha$ for all s when t is sufficiently small and K is finite. Hence, sufficiently small t do not have an inner local min point which implies $s^* = t$ is the global optimum for these types. Notice that $\Phi(s)$ is strictly positive in $]R,t[$, and that $\Phi(s) \to \infty$ at both edges of the range (i.e. at $s = R$ and at $s = t$). This means that $\Phi(s)$ has at least one local minimum in $]R,t[$. We now proceed to check whether this local minimum is unique:

$$\Phi'(s) = (t-s)^{\alpha-2} (s-R)^{-\beta} [(1-\beta)(t-s) - (\alpha-1)(s-R)].$$

Since $(t-s)^{\alpha-2} (s-R)^{-\beta}$ is strictly positive in $]R,t[$, and $[(1-\beta)(t-s) - (\alpha-1)(s-R)]$ is linear in s, negative at $s = R$ and positive at $s = t$, $\Phi'(s) = 0$ exactly at one point at this range (i.e. a unique local minimum of $\Phi(s)$ in $]R,t[$).

From the continuity of $\Phi(s)$ we get that if the value of $\Phi(s)$ at this local minimum is smaller than $\beta K/\alpha$, then $L(s,t)$ has exactly two extrema in the range $]R,t[$. From the negative values of $L'(s,t)$ at the edges of this range we finally conclude that the first extremum (where $\Phi(s)$ is falling) is a minimum point of $L(s,t)$, and the second extremum (where $\Phi(s)$ is rising) is a maximum point of $L(s,t)$. The global minimum of $L(s,t)$ is therefore either this local minimum (i.e. an inner solution), or $s = t$ (i.e. a corner solution). If however the value of $\Phi(s)$ at its local minimum point is larger than $\beta K/\alpha$, then there is no local extremum to $L(s,t)$ in the range $]R,t[$, and therefore $s = t$ is the solution to the minimization problem.

Next we show that if $t_h - t_l > 2\Delta$ then there exists a type who is far enough from the

regime to choose the inner solution. First, note that the distance from the regime to the type who is the most remote from it is larger than Δ. Suppose this type is $t_{h.}$. Then, comparing only the two corner solutions this type can choose, we get

$$L(R,t_h) - L(t_h,t_h) = |t_h - R|^{\alpha} - K|t_h - R|^{\beta},$$

which is strictly negative when $|t_h - R| > \Delta = K^{\frac{1}{\alpha-\beta}}$ and $\alpha < \beta$. This implies that t_h does not choose the corner solution of $t = t_h$, hence must choose an inner solution.

We now show that if there exists any type t_0 who chooses the inner solution, then all types with $t > t_0$ have an inner solution too. We also show that in the range of inner solutions $s^*(t)$ is decreasing in t. First notice that $\Phi(s)$ is decreasing in t, so if there exists a local minimum of $L(s,t_0)$ for some t_0, then there exists a local minimum of $L(s,t)$ for $t > t_0$ too. Also note that $\Phi(s)$ is decreasing in t with $\lim_{t\to\infty} \Phi(s) = 0 < \beta K/\alpha$ (for $s \in]R,t[$), implying that an inner local minimum exists for a sufficiently large t. Second, if there is an inner solution to the minimization problem for some t_0, then there is also an inner solution to the minimization problem for $t > t_0$. To see this let $\Delta L \equiv L(t,t) - L(\tilde{s},t)$, where $\tilde{s}$ is the stance at which $L(s,t)$ gets the local minimum. Type t prefers the inner solution to the corner solution if and only if ΔL is positive. Thus we need to show that ΔL is increasing in t and so if ΔL is positive for t_0 it is positive for $t > t_0$ too.

$$\Delta L = K(t-R)^{\beta} - \left[(t-\tilde{s})^{\alpha} + K(\tilde{s}-R)^{\beta}\right].$$

Differentiating ΔL with respect to t yields

$$\Delta L'_t = K\beta(t-R)^{\beta-1} - \left[\alpha(t-\tilde{s})^{\alpha-1}\left(1 - \frac{d\tilde{s}}{dt}\right) + \beta K(\tilde{s}-R)^{\beta-1}\frac{d\tilde{s}}{dt}\right].$$

Using the first order condition

$$\begin{aligned}\Delta L'_t &= K\beta(t-R)^{\beta-1} - \left[\beta K(\tilde{s}-R)^{\beta-1}\left(1 - \frac{d\tilde{s}}{dt}\right) + \beta K(\tilde{s}-R)^{\beta-1}\frac{d\tilde{s}}{dt}\right] \\ &= K\beta(t-R)^{\beta-1} - \beta K(\tilde{s}-R)^{\beta-1} > 0 \text{ when } \beta > 1.\end{aligned}$$

Differentiating once more

$$\Delta L''_t = K\beta(\beta-1)\left[(t-R)^{\beta-2} - \beta K\frac{d\tilde{s}}{dt}(\tilde{s}-R)^{\beta-1}\right].$$

By equation (9) we have that $\frac{d\tilde{s}}{dt} < 0$ in an inner solution when D is concave, and so $\Delta L''_t > 0$. Hence ΔL is strictly increasing and strictly convex, implying that for a broad enough range of types, types sufficiently far from R have an inner solution. Moreover, at this subrange of types, $\frac{ds^*}{dt} < 0$ by (9) when D is concave (the denominator is positive in inner solutions by the second order condition). This implies that $s^*(t)$ is decreasing in the subrange of types with inner solutions.

A.2.4 Case (4): $\min\{\alpha, \beta\} \geq 1$

The minimization problem of type t is symmetric around R, so we will present the first- and second-order conditions for an inner solution only for $t \geq R$.

$$-\alpha (t-s)^{\alpha-1} + \beta K (s-R)^{\beta-1} = 0 \quad (12)$$

$$(\alpha-1)\alpha (t-s)^{\alpha-2} + (\beta-1)\beta K (s-R)^{\beta-2} > 0 \quad (13)$$

We perform the proof first for $\alpha, \beta > 1$, and then for the special cases of $1 = \beta < \alpha$ and $1 = \alpha < \beta$.

$\alpha, \beta > 1$: That every t has a unique inner solution can be easily verified using equations (12) and (13). Moreover, by applying the implicit function theorem to equation (12), we get that $ds^*/dt > 0$, hence $|s^*(t) - R|$ is increasing in the distance to the regime.

$1 = \beta < \alpha$: It is easy to verify that types sufficiently close to the regime choose $s^*(t) = R$ (this is true for any $K > 0$) and types sufficiently far from it have a unique inner solution. For the subrange where all follow the regime we have $ds^*/dt = 0$. For the subrange with inner solutions using $\beta = 1$ and $\alpha > 1$ in equation (9) implies that $ds^*/dt = 1$ and hence $|s^*(t) - R|$ is increasing in the distance to the regime.

$1 = \alpha < \beta$: Solving for the range $t > R$ and then using symmetry around R, it is easy to verify that types sufficiently close to the regime choose $s^*(t) = t$, while types sufficiently far from the regime choose the same inner solution s, s.t. $P'(|s-R|) = 1$ $(= D')$. It thus follows that $|s^*(t) - R|$ is first increasing in the distance from the regime and then it stays constant.

A.3 Proof of Proposition 1

A.3.1 Part 1

We start by showing that initially – i.e., in the steady state – the most dissenting types are extremists (i.e., $\max |s^*(t) - R|$ is achieved for $t = \arg\max_t |t - R|$). For $\alpha \leq 1$ this follows immediately from Section A.2.1. If instead $\alpha > 1$, we know from Section A.2.2 that if the range of types is not sufficiently broad, then $s^*(t) = R$ for everyone hence the claim trivially holds. Otherwise, if the range of types is sufficiently broad so that types sufficiently far from R have an inner solution, Section A.2.2 further tells us that $s^*(t)$ is increasing in the subrange of types with inner solutions, implying that $\max |s^*(t) - R|$ is achieved for $t = \arg\max_t |t - R|$.

To see that, as the revolution evolves, more moderate types join, note first that during the revolution K decreases. Sections A.2.1 and A.2.2 tell us that, when $\beta < \alpha$, types sufficiently close to the regime (moderates) support the regime. Consider now the cutoff type at time i, who supports the regime ($s^*(t) = R$) but is indifferent between R and some $s \neq R$ ($s = t$ in the case of $\alpha \leq 1$ and some inner solution in the case of $\alpha > 1$). This means that, for this type, the difference between the two alternative solutions in terms of regime sanctioning P exactly cancels out with the difference between the two alternative solutions in terms of the discomfort D. At time $i+1$ the regime becomes weaker, hence the difference between the two alternative solutions in terms of regime sanctioning P must become smaller than the difference between the two alternative solutions in terms of the discomfort D, implying that this type will stop supporting the regime and instead join the revolution.

A.3.2 Part 2

We start by showing that initially – i.e., in the steady state – the most dissenting types are moderates (i.e., $\max |s^*(t) - R|$ is achieved for $t \neq \arg\max_t |t - R|$). If $\beta \leq 1$, we know from Section A.2.1 that there exists a distance from the regime, $\Delta = K^{\frac{1}{\alpha-\beta}}$, such that a type at that distance chooses $s^*(t) = t$ and hence has $|t - R| = \Delta$, while any type further away from R has $|t - R| = 0$. Given that, in a steady state with a regime, Δ must be smaller than $\max_t |t - R|$ (as otherwise $s^*(t) = t$ for everyone hence the regime does not exist), this immediately implies that $\max |s^*(t) - R| = \Delta$ is achieved for $t = R \pm \Delta \neq \arg\max_t |t - R|$. Alternatively, if $\beta > 1$, we know from Section A.2.3 that if the range of types is not sufficiently broad, then $s^*(t) = t$ for everyone hence a regime does not exist. If a regime exists it therefore must be that types sufficiently far from R have an inner solution. Moreover, Section A.2.3 further tells us that $s^*(t)$ is decreasing in the subrange of types with inner solutions, implying that $\max |s^*(t) - R|$ is achieved for $t \neq \arg\max_t |t - R|$.

To see that, as the revolution evolves, more extreme types (compared to $\arg\max_t |s^*(t) - R|$ at the steady state) dissent the most, note first that during the revolution K decreases. This implies that the most dissenting type at time $i+1$ (who, at this point in time, chooses $s^*(t) = t$) must have had a different solution at time i ($s_i^*(t) = R$ if $\beta \leq 1$, or an inner solution if $\beta > 1$), implying that she is further away from the regime (= a more extreme type) than the type who was most dissenting at time i (who herself is more extreme than the one most dissenting at time $i-1$ and so on until we reach the steady state).

A.3.3 Part 3

That initially – i.e., in the steady state – the most dissenting types are extremists (i.e., $\max |s^*(t) - R|$ is achieved for $t = \arg\max_t |t - R|$), follows immediately from Section A.2.4, where we show that $|s^*(t) - R|$ is increasing in the distance to the regime. During the revolution K decreases, making any type with an inner solution choose a new stance further away from the regime. In the special case where $1 = \beta < \alpha$ and we start with a steady state were all follow the regime, the revolution will be triggered by someone stopping to follow it, where the analysis in Section A.2.4 implies that these will be the types furthest away from the regime, and they will have inner solutions, hence, again, will gradually choose solutions further and further away from the regime.

A.4 A wave type revolution: from the outside in $\beta < \alpha \leq 1$

A.4.1 The phase diagram

We start by analyzing the behavior of A_{i+1} as a function of A_i, as depicted graphically in the phase diagram (Figure 3). As will be proved below, the phase diagram contains at most four parts, corresponding to the following cases (described from left to right in the diagram):

1. A sufficiently small A_i, which produces $A_{i+1} = 0$, indicating the case where $s_{i+1}(t) = t$ $\forall t$, and the phase diagram is flat.

2. A bit larger A_i, for which types far from the regime on both sides of it choose $s_{i+1}(t) = t$, while for the rest $s_{i+1}(t) = R$.

3. An even larger A_i, for which only types far from the regime on the far side of it choose $s_{i+1}(t) = t$, while for the rest $s_{i+1}(t) = R$.

4. A sufficiently large A_i, which produces $A_{i+1} = 1$, reflecting the case where $s_{i+1}(t) = R$ $\forall t$, and the phase diagram is flat.

We now prove that this is indeed the shape of the phase diagram. The analytical properties of $A_{i+1} = f(A_i)$ and of the individuals' behavior are summarized in the following lemma.

Lemma 1 *Suppose $\beta < \alpha \leq 1$. Then:*

1. *$A_{i+1} = f(A_i)$ is continuous and increasing in A_i.*
2. *There exists an $\varepsilon \geq 0$ such that $A_{i+1} = f(A_i) = 0$ for all $A_i \leq \varepsilon$. $\varepsilon = 0$ iff $|R| = 0$.*
3. *When $R = 0$ then $f(A_i)$ is convex for $A_i > 0$.*
4. *When $R \neq 0$ then for $A_i > \varepsilon$, $f(A_i)$ is convex initially. If $R \in [-1, -1/2[$, it stays convex throughout. Otherwise, if $R \in [-1/2, 0]$, then at the A_i corresponding to $\Delta = 1 + R$ the slope of $f(A_i)$ discontinuously decreases and $f(A_i)$ is convex thereafter until either $f(A_i)$ or A_i reaches 1.*
5. *Holding all else fixed, $f(A_i)$ is weakly decreasing in $|R|$.*
6. *Holding all else fixed, $f(A_i)$ is weakly increasing in $\bar{K}$.*
7. *The unstable steady states (A_{uss}) are increasing in $|R|$ while the stable steady states (A_{ss}) are (weakly) decreasing in $|R|$.*
8. *There exists a $\bar{K}_{c1}$ such that a stable steady state with a regime and $A_{ss} > 0$ exists iff $\bar{K} > \bar{K}_{c1}$.*
9. *$\bar{K}_{c1}$ is increasing in $|R|$.*

Proof. *From Section A.2.1 we know that (for sufficiently large K) there is a cutoff distance Δ between regime conformers (within the cutoff) and those speaking their minds (beyond the cutoff) s.t. $\Delta \equiv K^{\frac{1}{\alpha-\beta}} = \left(\bar{K}A\right)^{\frac{1}{\alpha-\beta}}$. Suppose, without loss of generality, that $R \leq 0$. If $\Delta \leq 1 - |R|$ (which is the distance from the regime to the closest edge of the type distribution), we have by equation (8)*

$$\begin{aligned} \Psi(s_i^*; R, A_i) &= \int_{-1}^{R-\Delta_i} (R - t)\, dt + \int_{R+\Delta_i}^{1} (t - R)\, dt \\ &= ... = R^2 - \Delta_i^2 + 1 \end{aligned}$$

while if $\Delta > 1 - |R|$ we have

$$\begin{aligned} \Psi(s_i^*; R, A_i) &= \int_{R+\Delta_i}^{1} (t - R)\, dt = ... \\ &= \frac{1}{2} - R - \frac{1}{2}\Delta_i^2 + \frac{1}{2}R^2. \end{aligned}$$

Hence we get

$$\Psi\left(s_i^*; R, A_i\right) = \begin{cases} R^2 - \Delta_i^2 + 1 \text{ when } 0 \leq \Delta_i \leq 1 + R \\ \frac{1}{2} - R - \frac{1}{2}\Delta_i^2 + \frac{1}{2}R^2 \text{ when } 1 + R < \Delta_i < 1 - R \\ 0 \text{ when } \Delta_i \geq 1 - R \end{cases}.$$

Noting that $A_{i+1} = 0$ *by construction whenever* $\Psi\left(s_i^*; R, A_i\right) \geq 1$*, we start by checking whether this inequality may hold in the first region of* $\Psi\left(s_i^*; R, A_i\right)$*.*

$$\begin{aligned} 1 &\leq R^2 - \Delta_i^2 + 1 \\ &\Leftrightarrow \Delta_i \leq -R. \end{aligned}$$

If $R \in [-1, -1/2]$*, this inequality holds throughout the first region (i.e. for any* $0 \leq \Delta_i \leq 1 + R$*), which means that* $\Psi\left(s_i^*; R, A_i\right) \geq 1$ *may hold also for some* Δ_i *in the middle region. Checking when this happens we get*

$$\begin{aligned} \frac{1}{2} - R - \frac{1}{2}\Delta_i^2 + \frac{1}{2}R^2 &= 1 \Rightarrow ... \Rightarrow \\ \Delta_i &= \sqrt{\left(R - \left(1 + \sqrt{2}\right)\right)\left(R - \left(1 - \sqrt{2}\right)\right)}, \end{aligned}$$

which does fall within the range $1 + R < \Delta_i < 1 - R$ *for* $R \in [-1, -1/2]$*. Thus, in this case where* $R \in [-1, -1/2[$ *we get*

$$A_{i+1} \equiv f\left(R, A_i\right) = \begin{cases} 0 \text{ when } \Delta_i \leq -R \\ 1 - \left(\frac{1}{2} - R - \frac{1}{2}\Delta_i^2 + \frac{1}{2}R^2\right) \text{ when } -R < \Delta_i < 1 - R \\ 1 \text{ when } 1 - R < \Delta_i \end{cases}$$

Otherwise, for $R \in [-1/2, 0]$*,* $\Psi\left(s_i^*; R, A_i\right) \geq 1$ *may hold only in the first region, and we get*

$$A_{i+1} \equiv f\left(R, A_i\right) = \begin{cases} 0 \text{ when } 0 \leq \Delta_i \leq -R \\ 1 - \left(R^2 - \Delta_i^2 + 1\right) \text{ when } -R < \Delta_i \leq 1 + R \\ 1 - \left(\frac{1}{2} - R - \frac{1}{2}\Delta_i^2 + \frac{1}{2}R^2\right) \text{ when } 1 + R < \Delta_i < 1 - R \\ 1 \text{ when } 1 - R \leq \Delta_i \end{cases}. \tag{14}$$

These four regions correspond to the four schematically described above. As the three-regions phase diagram for $R \in [-1, -1/2[$ *can be seen as a degenerate version of the four-regions phase diagram for* $R \in [-1/2, 0]$*, we will continue the analysis only for the latter case. Recalling that*

$$\Delta_i = \left(\bar{K} A_i\right)^{\frac{1}{\alpha - \beta}}, \tag{15}$$

and noting that this expression is monotonically increasing in A_i *for* $\beta < \alpha$*, we get that* $A_{i+1} = 0$ *for any* $A_i \leq \varepsilon \equiv \frac{(-R)^{\alpha - \beta}}{\bar{K}}$*, where* $\varepsilon \geq 0$ *and* $\varepsilon = 0$ *iff* $|R| = 0$*. As Figure 3 shows*

and will now be proved, the two middle regions are convex. Using (14) and (15)

$$\frac{df}{dA_i} = \begin{cases} \frac{2}{\alpha-\beta}\Delta_i^2 A_i^{-1} \text{ when } \Delta_i \leq 1+R \\ \frac{1}{\alpha-\beta}\Delta_i^2 A_i^{-1} \text{ when } 1+R < \Delta_i < 1-R \end{cases} > 0$$

$$\frac{d^2 f}{dA_i^2} = \begin{cases} \frac{2}{\alpha-\beta}\frac{\Delta_i^2}{A_i^2}\left(\frac{2}{\alpha-\beta}-1\right) \text{ when } \Delta_i \leq 1+R \\ \frac{1}{\alpha-\beta}\frac{\Delta_i^2}{A_i^2}\left(\frac{2}{\alpha-\beta}-1\right) \text{ when } 1+R < \Delta_i < 1-R \end{cases} > 0$$

since $\alpha - \beta \in (0,1)$. Thus, for $R \in [-1/2, 0]$ the function f has a kink at $\Delta_i = 1+R$ with a lower slope after the kink. These properties imply that the phase-diagram is flat at zero, convexly increasing, then has a downward kink and is convexly increasing after. This proves parts (1)-(4). There are at most two stable steady states, one at $A_i = 1$ and one interior. Since $A_{i+1} = f(A_i)$ is flat at zero it means that the first intersection is unstable, the next is stable, next unstable and next stable. Using (14) and (15)

$$\frac{df}{dR} = \begin{cases} -2R \text{ when } -R < \Delta_i \leq 1+R \\ 1-R \text{ when } 1+R < \Delta_i < 1-R \\ 0 \text{ otherwise} \end{cases} \geq 0$$

since $R \leq 0$, proving part (5). Furthermore,

$$\frac{df}{d\bar{K}} = \begin{cases} \frac{2}{\alpha-\beta}\Delta_i^2/\bar{K} \text{ when } -R < \Delta_i \leq 1+R \\ \frac{1}{\alpha-\beta}\Delta_i^2/\bar{K} \text{ when } 1+R < \Delta_i < 1-R \\ 0 \text{ otherwise} \end{cases} \geq 0,$$

proving part (6). These results imply that the unstable steady states (A_{uss}) are increasing in $|R|$ and decreasing in $\bar{K}$. The stable steady states (A_{ss}) are (weakly) decreasing in $|R|$ and (weakly) increasing in $\bar{K}$. This proves part (7).

Since it was shown that the phase diagram $A_{i+1} = f(A_i)$ starts below the 45-degree line, if follows that a stable steady state exists if A_{i+1} crosses the 45-degree line at least once. For this to happen, one of the following conditions should hold:

1. *The kink is above the 45 degree line: $A_{i+1}|_{\Delta_i=1+R} \geq A_i|_{\Delta_i=1+R} \Leftrightarrow \{$Using (14) and (15)$\} \Leftrightarrow 1+2R \geq (1+R)^{\alpha-\beta}/\bar{K}$. As the RHS is positive, this inequality can hold only if*

$$R \in [-1/2, 0] \text{ and } \bar{K} \geq \frac{(1+R)^{\alpha-\beta}}{1+2R}$$

2. *$A_{i+1}(1) = 1$ (i.e., $1-R \leq \Delta_i$ when $A_i = 1$)$\Leftrightarrow \{$Using (15)$\} \Leftrightarrow 1-R \leq \bar{K}^{\frac{1}{\alpha-\beta}} \Leftrightarrow$*

$$\bar{K} \geq (1-R)^{\alpha-\beta}$$

Denote the smallest $\bar{K}$ fulfilling one of these conditions by $\bar{K}_{c1}$. Thus follows part (8), and it can be verified that $\bar{K}_{c1}$ is increasing in $|R|$ (proving part (9)).

■

A.4.2 Proof of Proposition 2

Part (1): follows from parts (8) and (9) of Lemma 1.

Part (2): We first remind that a revolution is defined as a dynamic process where the approval is converging to a new, lower steady state. From this definition it directly follows that a negative shock to the approval of a regime in a stable steady state (with approval A_{ss}), such that the size of the shock is larger than $|A_{ss} - A_{uss}|$ (where A_{uss} is the approval in the closest unstable steady state to the left), would result in a revolution. A negative shock to the force ($\bar{K}$) of the regime reduces A_{i+1} (part (6) of Lemma 1), and in particular if the shock is such that $\bar{K}$ goes below $\bar{K}_{c1}$, A converges to zero and the regime completely falls (part (8) of Lemma 1). Finally, implementation of unpopular policies means that $|R|$ increases, and as a result the approval of the regime decreases (part (5) of Lemma 1), and in particular a revolution would start if the approval decreases sufficiently to eliminate the pre-existing stable steady state.

Part (3): (a) follows directly from part (1) of Proposition 1. (b) follows from the facts that (i) before the revolution everyone fully supports the regime at least on one side of it (as A_{ss} can only be in the third or fourth region of equation (14) – see Figure 3) and (ii) Δ_i starts above $1 + R$ (where $s(t)$ might be different than R only on one side of the regime).

Part (4): follows from part (1) of Proposition 1 and from the fact that dissenters speak their minds ($s(t) = t$).

A.5 A wave-type revolution: from the inside-out $\alpha < \beta \leq 1$

A.5.1 The phase diagram

We start by analyzing the behavior of A_{i+1} as a function of A_i, as depicted graphically in the phase diagram (Figure 6). As will be proven below, the phase diagram contains at most three parts, corresponding to the following cases (described from left to right in the diagram):

1. A sufficiently small A_i, which produces $A_{i+1} = 0$, indicating the case where $s_{i+1}(t) = t$ $\forall t$, and the phase diagram is flat.

2. A smaller A_i, for which types far from the regime on one side of it choose $s_{i+1}(t) = R$, while for the rest $s_{i+1}(t) = t$.

3. A sufficiently large A_i, for which types far from the regime on both sides of it choose $s_{i+1}(t) = R$, while for the rest $s_{i+1}(t) = t$.

We now prove that this is indeed the shape of the phase diagram. The analytical properties of $A_{i+1} = f(A_i)$ and of the individuals' behavior are summarized in the following lemma.

Lemma 2 *Suppose $\alpha < \beta \leq 1$. Then:*

1. *$A_{i+1} = f(A_i)$ is continuous and increasing in A_i.*

2. *There exists an $\varepsilon > 0$ such that $A_{i+1} = f(A_i) = 0$ for all $A_i \leq \varepsilon$.*

3. *When $R = 0$ then $f(A_i)$ is concave for $A_i > \varepsilon$.*

4. *When $R \neq 0$ then for $A_i > \varepsilon$, $f(A_i)$ is concave initially. At the A_i implied by $\Delta_i = 1 - |R|$ the slope of $f(A_i)$ discontinuously increases and $f(A_i)$ is concave thereafter until A_i reaches 1.*

5. *Holding all else fixed, $f(A_i)$ is weakly increasing in $|R|$.*

6. *Holding all else fixed, $f(A_i)$ is weakly increasing in $\bar{K}$.*

7. *The unstable steady states (A_{uss}) are weakly decreasing in $|R|$ while the stable steady states (A_{ss}) are weakly increasing in $|R|$.*

8. *$f(1) < 1$.*

9. *There exists a $\bar{K}_{c2}$ such that a stable steady state with a regime and $A_{ss} > 0$ exists iff $\bar{K} > \bar{K}_{c2}$.*

10. *$\bar{K}_{c2}$ is weakly decreasing in $|R|$.*

Proof. *From Section A.2.1 we know that (for sufficiently large K) there is a cutoff distance Δ between regime conformers ($|t-R| > \Delta$) and those speaking their minds ($|t-R| \leq \Delta$) such that $\Delta \equiv K^{\frac{1}{\alpha-\beta}} = \left(\bar{K}A\right)^{\frac{1}{\alpha-\beta}}$. Suppose, without loss of generality, that $R \leq 0$. If $\Delta \leq 1-|R|$ (which is the distance from the regime to the closest edge of the type distribution), we have by equation (8)*

$$\begin{aligned}\Psi(s_i^*; R, A_i) &= \int_{R-\Delta_i}^{R} (R-\tau)\, d\tau + \int_{R}^{R+\Delta_i} (\tau - R)\, d\tau \\ &= \Delta_i^2\end{aligned}$$

while if $\Delta > 1-|R|$ we have

$$\begin{aligned}\Psi(s_i^*; R, A_i) &= \int_{-1}^{R} (R-\tau)\, d\tau + \int_{R}^{R+\Delta_i} (\tau - R)\, d\tau \\ &= \frac{1}{2}(1+R)^2 + \frac{1}{2}\Delta_i^2\end{aligned}$$

Hence we get

$$\Psi(s_i^*; R, A_i) = \begin{cases} \Delta_i^2 \text{ when } 0 \leq \Delta_i \leq 1+R \\ \frac{1}{2}(1+R)^2 + \frac{1}{2}\Delta_i^2 \text{ when } 1+R < \Delta_i < 1-R \\ 1+R^2 \text{ when } 1-R \leq \Delta_i \end{cases}$$

noting that $\Psi(s_i^*; R, A_i)$ might equal 1 only in the middle range (unless $R = 0$), and in particular when

$$\begin{aligned} 1 &= \frac{1}{2}(1+R)^2 + \frac{1}{2}\Delta_i^2 \\ &\Leftrightarrow \; 1 - R^2 - 2R = \Delta_i^2 \end{aligned}$$

we get by (7) that

$$A_{i+1} \equiv f(R, A_i) = \begin{cases} 1 - \Delta_i^2 \text{ when } 0 \leq \Delta_i \leq 1+R \\ 1 - \left(\frac{1}{2}(1+R)^2 + \frac{1}{2}\Delta_i^2\right) \text{ when } 1+R < \Delta_i < \sqrt{1-R^2-2R} \\ 0 \text{ when } \sqrt{1-R^2-2R} \leq \Delta_i \end{cases}. \tag{16}$$

These three regions correspond to the three schematically described above. Recalling that

$$\Delta_i = \left(\bar{K} A_i\right)^{\frac{1}{\alpha-\beta}}, \tag{17}$$

and noting that this expression is monotonically deceasing in A_i for $\alpha < \beta$, we get that $A_{i+1} = 0$ for any $A_i \leq \varepsilon \equiv \frac{\left(\sqrt{1-R^2-2R}\right)^{\alpha-\beta}}{\bar{K}}$, where $\varepsilon > 0$. As Figure 6 shows and will now be proved, the two regions in which $A_{i+1} \neq 0$ are concave. Using (16) and (17) we get

$$\begin{aligned}
\frac{df}{dA_i} &= \begin{cases} -\frac{2}{\alpha-\beta}\Delta_i^2 A_i^{-1} \text{ when } \Delta_i \leq 1+R \\ -\frac{1}{\alpha-\beta}\Delta_i^2 A_i^{-1} \text{ when } 1+R < \Delta_i < \sqrt{1-R^2-2R} \end{cases} > 0 \\
\frac{d^2 f}{dA_i^2} &= \begin{cases} -\frac{2}{\alpha-\beta}\frac{\Delta_i^2}{A_i^2}\left(\frac{2}{\alpha-\beta}-1\right) \text{ when } \Delta_i \leq 1+R \\ -\frac{1}{\alpha-\beta}\frac{\Delta_i^2}{A_i^2}\left(\frac{2}{\alpha-\beta}-1\right) \text{ when } 1+R < \Delta_i < \sqrt{1-R^2-2R} \end{cases} < 0
\end{aligned}$$

since $\alpha-\beta \in (-1,0)$. Thus, the function f has a kink at $\Delta_i = 1+R$ with a bigger slope after the kink (note that small values of Δ correspond to high approval and large values correspond to low approval). These properties imply that the phase-diagram is first flat, then concavely increasing, then has an upward kink and is concavely increasing thereafter. This proves parts (1)-(4). There are at most two (interior) stable steady states. Since $A_{i+1} = f(A_i)$ is flat at zero, it means that the first intersection is unstable, the next is stable, next unstable and next stable.

$$\frac{df}{dR} = \begin{cases} -1-R \text{ when } 1+R < \Delta_i < \sqrt{1-R^2-2R} \\ 0 \text{ otherwise} \end{cases} \leq 0$$

since $R \geq -1$, proving part (5). Furthermore,

$$\frac{df}{d\bar{K}} = \begin{cases} -\frac{2}{\alpha-\beta}\Delta_i^2/\bar{K} \text{ when } -R < \Delta_i \leq 1+R \\ -\frac{1}{\alpha-\beta}\Delta_i^2/\bar{K} \text{ when } 1+R < \Delta_i < \sqrt{1-R^2-2R} \\ 0 \text{ otherwise} \end{cases} \geq 0,$$

proving part (6). These results imply that the unstable steady states (A_{uss}) are decreasing in $|R|$ and in $\bar{K}$. The stable steady states (A_{ss}) are increasing in $|R|$ and in $\bar{K}$. This proves part (7). When $A_i = 1$ we get by (17) that Δ_i is strictly positive, hence, by (16), $A_{i+1} < 1$, which proves part (8). This further implies, together with the fact that the phase diagram $A_{i+1} = f(A_i)$ starts below the 45 degree line, that a necessary and sufficient condition for the existence of a stable steady state is that f crosses (and not just touches) the 45-degree line. Now, note that for

$$\bar{K} = \frac{(1+R)^{\alpha-\beta}}{1-(1+R)^2} \tag{18}$$

we get that the kink is exactly on the 45-degree line, because this yield

$$1-(1+R)^2 = (1+R)^{\alpha-\beta}/\bar{K} \Rightarrow \{\text{using (16) and (17)}\} \Rightarrow A_{i+1}|_{\Delta_i=1+R} = A_i|_{\Delta_i=1+R},$$

in which case a stable steady state exists. Next, part (6) implies that f is weakly increasing in K, so that if for a certain K^* a stable steady state exists, then a stable steady state exists for any $K > K^*$. Denote the smallest $\bar{K}$ for which f touches the 45-degree line (given by (18)) by $\bar{K}_{c2}$. Thus follows part (9), and part (10) follows from the fact that f increases in

$\bar{K}$ and $|R|$ (by parts (5) and (6)). ■

A.5.2 Proof of Proposition 3

Part (1) follows from Lemma 2 parts (9) and (10).

Part (2): We first remind that a revolution is defined as a dynamic process where the approval is converging to a new, lower steady state. From this definition it directly follows that a negative shock to the approval of a regime in a stable steady state (with approval A_{ss}), such that the size of the shock is larger than $|A_{ss} - A_{uss}|$ (where A_{uss} is the approval in the closest unstable steady state to the left), would result in a revolution. A negative shock to the force of the regime reduces A_{i+1} (part (6) of Lemma 2), and in particular if the shock is such that $\bar{K}$ goes below $\bar{K}_{c2}$, then A_i converges to zero and the regime completely falls (part (9) of Lemma 2). Finally, implementation of popular policies means that $|R|$ decreases, and as a result the approval of the regime decreases as well (part (5) of Lemma 2), and in particular a revolution would start if the approval decreases sufficiently to eliminate the pre-existing stable steady state.

Part (3): (a) follows directly from part (2) of Proposition 1. (b) follows from the fact that dissent at time i comes from people within the cutoff Δ_i, and for any R s.t. $|R| \neq 1$ this implies dissent on both sides of the regime.

Part (4): follows from the facts that (i) dissent at time i comes from people within the cutoff Δ_i (see part (2) of Proposition 1), (ii) Δ_i increases as A_i decreases during the revolution, and (iii) dissenters speak their minds ($s(t) = t$).

A.6 A stretching-out type revolution $\alpha > 1$, $\beta \geq 1$

When $\alpha > 1$, $\beta \geq 1$, every type $t > R$ has a unique inner solution $s^*(t) \in]R, t[$ and every type $t < R$ has a unique inner solution $s^*(t) \in]t, R[$, with this solution being determined by equation (12) (see Section A.2.4). Substituting variables to $\sigma \equiv |s^*(t) - R|$ and $\tau \equiv |t - R|$ yields

$$\begin{aligned} K_i \beta \sigma^{\beta-1} &= \alpha (\tau - \sigma)^{\alpha-1} \\ &\Leftrightarrow \quad \tau = \sigma + \left(\frac{K_i \beta}{\alpha}\right)^{\frac{1}{\alpha-1}} \sigma^{\frac{\beta-1}{\alpha-1}}. \end{aligned} \tag{19}$$

We turn now to calculating $\Psi(s_i^*; R, A_i)$. To do that, we first remind that $\Psi(s_i^*; R, A_i)$ is the sum of deviations from R (i.e. the sum of $\sigma(t)$ $\forall t$). Hence, it equals the area under the graph of $\sigma(t)$. Now, since σ is an implicit function of t (and of τ), it is difficult to compute the integral of $\sigma(\tau)$ (= the area under $\sigma(t)$). Instead, it is easier to compute it using the explicit expression of $\tau(\sigma)$ in (19). Noting that, at each side of R, σ is monotonous in t, we can substitute the calculation of the area under $\sigma(\tau)$ for positive τ with a calculation of the area above $\tau(\sigma)$ and below a horizontal line at the value $1 - R$ (which is max τ), and the calculation of the area under $\sigma(\tau)$ for negative τ with a calculation of the area below $\tau(\sigma)$ and above a horizontal line at the value $-(1 + R)$ (which is min τ).[17] Finally, using the

[17] To see this it is easiest to draw a generic increasing function $\sigma(\tau)$ between 0 and $1 + R$ and note, by turning the drawing 90 degrees, that the area it creates is the same as the area given by $1 + R - \tau(\sigma)$ with boundaries $\sigma(0)$ and $\sigma(1 + R)$.

symmetry of $\sigma(\tau)$ around 0 we can substitute $\int_{-(1+R)}^{0}\sigma(\tau)\,d\tau$ with $\int_{0}^{1+R}\sigma(\tau)\,d\tau$ to get

$$\begin{aligned}
\Psi(s_i^*;R,A_i) &= \int_0^{1+R}\sigma(\tau)\,d\tau+\int_0^{1-R}\sigma(\tau)\,d\tau \\
&= \int_0^{\check{\sigma}\equiv\sigma(1+R)}[(1+R)-\tau(\sigma)]\,d\sigma+\int_0^{\hat{\sigma}\equiv\sigma(1-R)}[(1-R)-\tau(\sigma)]\,d\sigma \\
&= \int_0^{\check{\sigma}\equiv\sigma(1+R)}\left[(1+R)-\sigma-\left(\frac{K_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\sigma^{\frac{\beta-1}{\alpha-1}}\right]d\sigma \\
&\quad+\int_0^{\hat{\sigma}\equiv\sigma(1-R)}\left[(1-R)-\sigma-\left(\frac{K_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\sigma^{\frac{\beta-1}{\alpha-1}}\right]d\sigma \\
&= (1+R)\,\check{\sigma}-\frac{\check{\sigma}^2}{2}+(1-R)\,\hat{\sigma}-\frac{\hat{\sigma}^2}{2}-\left(\frac{K_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\frac{\check{\sigma}^{\frac{\beta-1}{\alpha-1}+1}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}+1}}{\frac{\beta-1}{\alpha-1}+1} \qquad (20)
\end{aligned}$$

The analytical properties of $A_{i+1}=f(A_i)$ and of the individuals' behavior are summarized in the following lemma.

Lemma 3 *Suppose $\alpha>1$, $\beta\geq 1$. Then:*

1. *$A_{i+1}=f(A_i)$ is continuous and increasing in A_i.*
2. *There exists an $\varepsilon\geq 0$ such that $A_{i+1}=f(A_i)=0$ for all $A_i\leq\varepsilon$. $\varepsilon=0$ iff $|R|=0$.*
3. *For $A_i>\varepsilon$, $f(A_i)$ is first convex then concave, or convex throughout, or concave throughout.*
4. *Holding all else fixed, $f(A_i)$ is decreasing in $|R|$.*
5. *Holding all else fixed, $f(A_i)$ is increasing in $\bar{K}$.*
6. *$f(1)<1$.*
7. *There exists a $\bar{K}_{c3}$ such that a stable steady state with a regime and $A_{ss}>0$ exists iff $\bar{K}>\bar{K}_{c3}$.*
8. *$\bar{K}_{c3}$ is increasing in $|R|$.*
9. *There are at most two steady states with $A>0$, where the first is unstable and the second is stable.*
10. *The unstable steady states (A_{uss}) are increasing in $|R|$ while the stable steady states (A_{ss}) are (weakly) decreasing in $|R|$.*

Proof. *To see that part (1) holds, recall that by construction (6) $A=\max\{0,1-\Psi(s_i^*;R,A_i)\}$*

and note that

$$\begin{aligned}\frac{d\Psi\left(\sigma_i;R,A_i\right)}{dA_i} &= \left(1+R-\check{\sigma}\right)\frac{d\check{\sigma}}{dA_i}+\left(1-R-\hat{\sigma}\right)\frac{d\hat{\sigma}}{dA_i}-\frac{1}{\alpha-1}A_i^{\frac{1}{\alpha-1}-1}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\frac{\check{\sigma}^{\frac{\beta-1}{\alpha-1}+1}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}+1}}{\frac{\beta-1}{\alpha-1}+1}\\ &\quad-\left(\frac{\bar{K}A_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\left(\check{\sigma}^{\frac{\beta-1}{\alpha-1}}\frac{d\check{\sigma}}{dA_i}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}}\frac{d\hat{\sigma}}{dA_i}\right)\\ &= \left(1+R-\check{\sigma}-\left(\frac{\bar{K}A_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\check{\sigma}^{\frac{\beta-1}{\alpha-1}}\right)\frac{d\check{\sigma}}{dA_i}+\left(1-R-\hat{\sigma}-\left(\frac{\bar{K}A_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\hat{\sigma}^{\frac{\beta-1}{\alpha-1}}\right)\frac{d\hat{\sigma}}{dA_i}\\ &\quad-\frac{1}{\alpha-1}A_i^{\frac{1}{\alpha-1}-1}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\frac{\check{\sigma}^{\frac{\beta-1}{\alpha-1}+1}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}+1}}{\frac{\beta-1}{\alpha-1}+1}.\end{aligned}$$

Using $\check{\sigma}$ *and* $\hat{\sigma}$ *in the FOC in (10) we get*

$$\alpha\left(1+R-\check{\sigma}\right)^{\alpha-1} = \bar{K}A_i\beta\check{\sigma}^{\beta-1} \quad (21)$$

$$\alpha\left(1-R-\hat{\sigma}\right)^{\alpha-1} = \bar{K}A_i\beta\hat{\sigma}^{\beta-1}. \quad (22)$$

Using these in the previous expression for $\frac{d\Psi(\sigma_i;R,A_i)}{dA_i}$ *we get that*

$$\frac{d\Psi\left(\sigma_i;R,A_i\right)}{dA_i}=-\frac{1}{\alpha-1}A_i^{\frac{1}{\alpha-1}-1}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\frac{\check{\sigma}^{\frac{\beta-1}{\alpha-1}+1}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}+1}}{\frac{\beta-1}{\alpha-1}+1}<0,$$

hence A_{i+1} *is increasing in* A_i *(continuity follows trivially from the definition of* A_{i+1} *in (6) and the expression of* $\Psi\left(\sigma_i;R,A_i\right)$*). When* $A_i\to 0$ *also* $K_i\to 0$ *hence* $\sigma\left(\tau\right)\to\tau$ *for all types. For* $K_i=0$ *we have* $\sigma\left(\tau\right)=\tau$ *and* $\Psi\left(s_i^*;R,A_i\right)=\frac{(1-R)^2+(1+R)^2}{2}\geq 1$*, with equality only for* $R=0$*. From (6) and (8) it thus follows that* $\exists\varepsilon\geq 0$ *such that* $A_{i+1}=f\left(A_i\right)=0$ *for any* $A_i\leq\varepsilon$*, where* $\varepsilon=0$ *iff* $|R|=0$*. This proves part (2). To prove part (3) we differentiate* $\Psi\left(\sigma_i;R,A_i\right)$ *one more time:*

$$\begin{aligned}\frac{d^2\Psi\left(\sigma_i;R,A_i\right)}{dA_i^2} = &-\frac{1}{\alpha-1}\left(\frac{1}{\alpha-1}-1\right)A_i^{\frac{1}{\alpha-1}-2}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\frac{\check{\sigma}^{\frac{\beta-1}{\alpha-1}+1}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}+1}}{\frac{\beta-1}{\alpha-1}+1} \quad (23)\\ &-\frac{1}{\alpha-1}A_i^{\frac{1}{\alpha-1}-1}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\left(\check{\sigma}^{\frac{\beta-1}{\alpha-1}}\frac{d\check{\sigma}}{dA_i}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}}\frac{d\hat{\sigma}}{dA_i}\right).\end{aligned}$$

Note that $\frac{d\check{\sigma}}{dA_i}$ *and* $\frac{d\hat{\sigma}}{dA_i}$ *are both negative.*[18] *This implies that* $\frac{d^2\Psi(\sigma_i;R,A_i)}{dA_i^2}>0$ *when* $\alpha\geq 2$*, hence* A_{i+1} *is concave.*

[18] This is true since K increases in A_i which in turn makes everyone, including types 1 and -1, choose a solution closer to R.

We now investigate the case $1<\alpha<2$. Revisiting equation (19) we can write

$$\begin{aligned} H &= \sigma+\left(\frac{K_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\sigma^{\frac{\beta-1}{\alpha-1}}-\tau=0 \\ \Rightarrow \quad \frac{d\sigma}{dA_i} &= -\frac{\frac{dH}{dA_i}}{\frac{dH}{d\sigma}}=-\frac{\frac{1}{\alpha-1}A_i^{\frac{1}{\alpha-1}-1}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\sigma^{\frac{\beta-1}{\alpha-1}}}{1+\frac{\beta-1}{\alpha-1}A_i^{\frac{1}{\alpha-1}}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\sigma^{\frac{\beta-1}{\alpha-1}-1}} \\ &= \{using\ (19)\}=-\frac{\frac{1}{\alpha-1}A_i^{-1}(\tau-\sigma)}{1+\frac{\beta-1}{\alpha-1}(\tau-\sigma)\sigma^{-1}}. \end{aligned} \tag{24}$$

Rewriting (23)

$$\frac{d^2\Psi(\sigma_i;R,A_i)}{dA_i^2}=-\frac{1}{\alpha-1}A_i^{\frac{1}{\alpha-1}-2}\left(\frac{\bar{K}\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\left[\left(\frac{1}{\alpha-1}-1\right)\frac{\check{\sigma}^{\frac{\beta-1}{\alpha-1}+1}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}+1}}{\frac{\beta-1}{\alpha-1}+1}+A_i\left(\check{\sigma}^{\frac{\beta-1}{\alpha-1}}\frac{d\check{\sigma}}{dA_i}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}}\frac{d\hat{\sigma}}{dA_i}\right)\right]$$

Using the FOC $\left(\frac{K_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\sigma^{\frac{\beta-1}{\alpha-1}}=\tau-\sigma$ *and (24) we get*

$$\begin{aligned} \frac{d^2\Psi(\sigma_i;R,A_i)}{dA_i^2} &= -\frac{1}{\alpha-1}A_i^{-2}\left[\sigma(\tau-\sigma)\left(\frac{1}{\alpha-1}-1\right)\frac{1}{\frac{\beta-1}{\alpha-1}+1}-(\tau-\sigma)\frac{\frac{1}{\alpha-1}(\tau-\sigma)}{1+\frac{\beta-1}{\alpha-1}\frac{\tau-\sigma}{\sigma}}\right]\Bigg|_{\tau=1+R} \\ &\quad -\frac{1}{\alpha-1}A_i^{-2}\left[\sigma(\tau-\sigma)\left(\frac{1}{\alpha-1}-1\right)\frac{1}{\frac{\beta-1}{\alpha-1}+1}-(\tau-\sigma)\frac{\frac{1}{\alpha-1}(\tau-\sigma)}{1+\frac{\beta-1}{\alpha-1}\frac{\tau-\sigma}{\sigma}}\right]\Bigg|_{\tau=1-R}. \end{aligned} \tag{25}$$

Note that

$$\begin{aligned} &\sigma(\tau-\sigma)\left(\frac{1}{\alpha-1}-1\right)\frac{1}{\frac{\beta-1}{\alpha-1}+1}-(\tau-\sigma)\frac{\frac{1}{\alpha-1}(\tau-\sigma)}{1+\frac{\beta-1}{\alpha-1}\frac{\tau-\sigma}{\sigma}} \\ = \ &\sigma(\tau-\sigma)\left[\frac{2-\alpha}{\alpha-1}\frac{\alpha-1}{\beta+\alpha-2}-\frac{\frac{1}{\alpha-1}(\tau-\sigma)}{\sigma+\frac{\beta-1}{\alpha-1}(\tau-\sigma)}\right] \\ = \ &\sigma(\tau-\sigma)\left[\frac{2-\alpha}{\beta+\alpha-2}-\frac{\frac{\tau-\sigma}{\tau}}{(\alpha-1)\frac{\sigma}{\tau}+(\beta-1)\frac{\tau-\sigma}{\tau}}\right], \end{aligned}$$

where $\frac{2-\alpha}{\beta+\alpha-2}>0$ *for* $1\leq\alpha<2$ *and* $\frac{\frac{\tau-\sigma}{\tau}}{(\alpha-1)\frac{\sigma}{\tau}+(\beta-1)\frac{\tau-\sigma}{\tau}}$ *is positive and increasing in the relative step that type t takes toward the regime,* $\frac{\tau-\sigma}{\tau}\in]0,1[$. *Moreover, for any τ and any α s.t. $1\leq\alpha<2$, the expression in the squared brackets goes from positive to negative as the relative step* $\frac{\tau-\sigma}{\tau}$ *grows from 0 to 1. It can further be verified that* $\frac{\tau-\sigma}{\tau}$ *increases in A_i (because an increase in A_i implies that the regime is stronger and so one needs to accommodate more to R). Returning now to (25) and noting that* $\frac{d^2\Psi(\sigma_i;R,A_i)}{dA_i^2}$ *has the opposite sign of the squared brackets, we get that, as A_i increases,* $\frac{d^2\Psi(\sigma_i;R,A_i)}{dA_i^2}$ *either keeps its sign or changes sign once, from negative to positive. Finally, since* $A_{i+1}=\max\{0,1-\Psi_i(\sigma_i;R,A_i)\}$, *we get that* $A_{i+1}(A_i)$ *is first convex then concave, or convex throughout, or concave throughout, which proves part (3).*

Differentiating equation (20) w.r.t. R and then using (21) and (22) yields

$$\frac{d\Psi\left(s_i^*;R,A_i\right)}{dR}=\check{\sigma}-\hat{\sigma}\leq 0$$

(by the monotonicity of $\sigma(\tau)$), implying that A_{i+1} decreases in $|R|$, which proves part (4). Next, differentiating equation (20) by $\bar{K}$ and then using (21) and (22) yields

$$\frac{d\Psi\left(s_i^*;R,A_i\right)}{d\bar{K}}=-\frac{1}{\alpha-1}\bar{K}^{\frac{1}{\alpha-1}-1}\left(\frac{A_i\beta}{\alpha}\right)^{\frac{1}{\alpha-1}}\frac{\check{\sigma}^{\frac{\beta-1}{\alpha-1}+1}+\hat{\sigma}^{\frac{\beta-1}{\alpha-1}+1}}{\frac{\beta-1}{\alpha-1}+1}<0,$$

hence $f(A_i)$ is increasing in $\bar{K}$, which proves part (5). Part (6) follows from the fact that all types always have inner solutions (for finite $\bar{K}$) to the optimization problem, hence $A_{i+1}=f(1)$ never reaches 1. This further implies, together with the fact that the phase diagram $A_{i+1}=f(A_i)$ starts below the 45 degree line, that a necessary and sufficient condition for the existence of a stable steady state is that this diagram crosses (and not just touches) the 45-degree line. Now, fix α,β and R, and set $\bar{K}$ to be sufficiently large such that for $\max\tau=1-R$ and $A_i=1/2$, the value of σ which solves equation (19) is smaller than $1/2$. The strict monotonicity of $\sigma(\tau)$ implies then that the total sum of deviations from the regime ($\Psi\left(s_i^*;R,A_i\right)$) will be smaller than $1\cdot 1/2$, and so $A_{i+1}>1-1/2=1/2=A_i$. In other words, at $A_i=1/2$ the phase diagram is above the 45-degree line, and together with parts (2) and (6) we get that (for $R\neq 0$) the phase diagram crosses the 45-degree line at least twice, and one of these crossing points must be a stable steady state.[19] *Furthermore, this happens for finite $\bar{K}$. Together with this result, part (5) implies that $f(A_i)$ is increasing in $\bar{K}$, so that if for a certain K^* a stable steady state exists, then a stable steady state exists for any $K>K^*$. Denote the smallest $\bar{K}$ for which the diagram touches the 45-degree line by $\bar{K}_{c3}$. Thus follows part (7), and part (8) follows from the fact that $f(A_i)$ decreases in $|R|$ and decreases in $\bar{K}$ (by part (4) and (5)). Given that the phase diagram starts and ends below the 45-degree line (except for one special case – see previous footnote), it cannot cross this line if it is convex throughout, which (by part (3)) implies that, for $A_i>\varepsilon$, it must be either concave throughout or first convex and then concave. In both cases this leads to at most two crossing points of the 45-degree line, the first from below (hence unstable) and the second from above (hence stable). This proves part (9). Increasing $|R|$ reduces A_{i+1} (by part (4)), and so the new crossing points, if they still exist, lie in the range that previously was above the 45-degree line, $]A_{uss},A_{ss}[$, which means that A_{uss} increases while A_{ss} decreases. This proves part (10).*[20] ■

A.6.1 Proof of Proposition 4

Part (1) follows from Lemma 3 parts (7) and (8).

Part (2): We first remind that a revolution is defined as a dynamic process where the approval is converging to a new, lower steady state. From this definition it directly follows that a negative shock to the approval of a regime in a stable steady state (with approval

[19] If $R=0$ and $A_{i+1}=f(1/2)>1/2$, the phase diagram may have only one crossing point in case it starts above the 45-degree line, but since it starts above the 45-degree line and ends below it, this unique crossing-point must be a stable steady state.

[20] In the special case where $R=0$ and the phase diagram starts above the 45-degree line and has only one crossing point (which was shown to be a stable steady state), a decrease of $A_{i+1}=f(A_i)$ results as well in a decrease of A_{ss}.

A_{ss}), such that the size of the shock is larger than $|A_{ss} - A_{uss}|$ (when A_{uss} exists), would result in a revolution. A negative shock to the force of the regime reduces A_{i+1} (part (5) of Lemma 3), and in particular if the shock is such that $\bar{K}$ goes below $\bar{K}_{c3}$ A converges to zero over time and the regime completely falls (part (7) of Lemma 3). Finally, implementation of unpopular policies means that $|R|$ increases, and as a result the approval function $(f)$of the regime decreases (part (4) of Lemma 3), and in particular a revolution would start if the approval decreases sufficiently to eliminate the pre-existing stable steady state.

Part (3): (a) follows directly from part (3) of Proposition 1. (b) follows from the fact that nobody in society fully follows the regime.

Part (4): The fact that the whole population participates in the revolution follows from the fact that nobody in society fully follows the regime, and the fact that the most extreme types dissent the most follows from part (3) of Proposition 1.

References

[1] Acemoglu, D., & Robinson, A.J., (2001). "A Theory of Political Transitions." *American Economic Review*, 91(4): 938-63.

[2] Angeletos, G. M., Hellwig, C., & Pavan, A. (2007). "Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks," *Econometrica*, 75(3), 711-756.

[3] BBC (2013): http://www.bbc.com/news/world-middle-east-12313405.

[4] Bernheim, D.B., (1994), "A Theory of Conformity", *Journal of Political Economy,* Vol. 102, No. 5, pp. 841-877.

[5] Brinton, C. (1938). "The Anatomy of Revolution". New York, NY, US: W W Norton & Co.

[6] Bueno De Mesquita, E. (2010). "Regime change and revolutionary entrepreneurs". *American Political Science Review*, 104(03), 446-466.

[7] Edmond, C. (2013), "Information Manipulation, Coordination, and Regime Change", *Review of Economic Studies,* Vol. 80, pp.1422–1458

[8] Esteban, J. (2001). "Collective action and the group size paradox." *American Political Science Association* Vol. 95, No. 03, pp. 663-672.

[9] Esteban, J., & Ray, D. (2001). "Social decision rules are not immune to conflict". *Economics of Governance*, 2(1), 59-67.

[10] Ghamari-Tabrizi, Behrouz. 2008. Islam and Dissent in Postrevolutionary Iran. New York: I.B. Tauris.

[11] Goldstone, J. A. (2001). "Toward a fourth generation of revolutionary theory". *Annual Review of Political Science*, 4, 139-187.

[12] Granovetter, M., (1978), "Threshold Models of Collective Behavior", *The American Journal of Sociology,* Vol. 83, No. 6, pp. 1420-1443.

[13] Kaniovski, Y. M., Kryazhimskii, A. V., & Young, H. P. (2000). "Adaptive dynamics in games played by heterogeneous populations". *Games and Economic Behavior*, 31(1), 50-96.

[14] Kim, Q. Y. (1996). "From protest to change of regime: the 4–19 Revolt and the fall of the Rhee regime in South Korea". *Social Forces*, 74(4), 1179-1208.

[15] Kuran, T. (1989a). "Sparks and prairie fires: A theory of unanticipated political revolution", *Public Choice*, 61(1), 41-74.

[16] Kuran, T., (1989b), "Now out of Never, The element of surprise in the east European revolution of 1989", *World Politics*, Vol 44, No 1 pp. 7-48.

[17] Kuran, T., (1995), "The Inevitability of Future Revolutionary Surprises," *The American Journal of Sociology*, Vol. 100, No. 6, pp. 1528-1551.

[18] Kuran, T., & Sandholm, W. H. (2008). "Cultural integration and its discontents". *The Review of Economic Studies*, 75(1), 201-228.

[19] Lohmann, S. (1994). "The dynamics of informational cascades". *World politics*, 47(1), 42-101.

[20] Manski, C.F.,Mayshar, J. (2003)"Private Incentives and Social Interactions: Fertility Puzzles in Israel," *Journal of the European Economic Association*, Vol. 1, No.1, pp. 181-211.

[21] Michaeli, M. & Spiro, D., (2015), "Norm conformity across societies," *J. of Public Economics*, Vol. 132, pp. 51-65.

[22] Milani, M. M. (1988). The making of Iran's Islamic revolution: from monarchy to Islamic republic. Boulder, CO: Westview Press.

[23] Moaddel, M. (1992). "Ideology as episodic discourse: the case of the Iranian revolution". American Sociological Review, 353-379.

[24] Naylor, R. (1989). "Strikes, free riders, and social customs". *The Quarterly Journal of Economics*, 104(4), 771-785.

[25] Oliver, P. E., & Marwell, G. (1988). "The Paradox of Group Size in Collective Action: A Theory of the Critical Mass". *II. American Sociological Review*, Vol. 53, No. 1, pp.1-8.

[26] Olson, M., (1971), *The Logic of Collective Action: Public Groups and the Theory of Groups*. Cambridge and London: Harvard University Press.

[27] Pan, P. P. (2008). Out of Mao's shadow: the struggle for the soul of a new China. Simon and Schuster.

[28] Pfaff, S. (2006). *Exit-voice Dynamics and the Collapse of East Germany: the Crisis of Leninism and the Revolution of 1989*. Duke university Press.

[29] Przeworski, A. (1991). *Democracy and the market: Political and economic reforms in Eastern Europe and Latin America*. Cambridge University Press.

[30] Razi, G. H. (1987). "The Nexus of Legitimacy and Performance: The Lessons of the Iranian Revolution". *Comparative Politics*, 453-469.

[31] Rubin, J. (2014). "Centralized institutions and cascades". *Journal of Comparative Economics.*Vol 42, Iss 2, pp. 340–357

[32] Shadmehr, M. (2015a). "Extremism in Revolutionary Movements". *Games and Economic Behavior*, forthcoming.

[33] Shadmehr, M. (2015b). "Ideology and the Iranian Revolution". mimeo, University of Miami.

[34] Tanter, R., & Midlarsky, M. (1967). A theory of revolution. Journal of Conflict Resolution, 11(3), 264-280.

[35] Tullock, G. (1971). "The paradox of revolution". *Public Choice*, 11(1), 89-99.

[36] Walder, A. G., & Xiaoxia, G. (1993). Workers in the Tiananmen protests: the politics of the Beijing Workers' Autonomous Federation. The Australian Journal of Chinese Affairs, 1-29.

[37] Young, H. P. (1993). The evolution of conventions. Econometrica: Journal of the Econometric Society, 57-84.

[38] Young, H. P. (2015). "The evolutions of social norms", *Annu. Rev. Econ.* 2015. 7:359–87

[39] Zhao, D. (2001). The power of Tiananmen: State-society relations and the 1989 Beijing student movement. University of Chicago Press.